

ANÁLISE DE COMPONENTES INDEPENDENTES APLICADA À
IDENTIFICAÇÃO DE REGIÕES LESIONADAS EM MAMOGRAMAS

Luciana Leite

TESE SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS
PROGRAMAS DE PÓS-GRADUAÇÃO DE ENGENHARIA DA UNIVERSIDADE
FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS
NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM
ENGENHARIA ELÉTRICA.

Aprovada por:

Prof. Eduardo Antônio Barros da Silva, Ph.D.

Prof. Wagner Coelho de Albuquerque Pereira, D.Sc.

Prof. Luiz Pereira Calôba, Dr.Ing.

Prof. Homero Schiabel, D.Sc.

RIO DE JANEIRO, RJ - BRASIL
ABRIL DE 2005

LEITE, LUCIANA

Análise De Componentes Independentes
Aplicada À Identificação De Regiões
Lesionadas Em Mamogramas [Rio de
Janeiro] 2005

XIV, 155 p. 29,7 cm (COPPE/UFRJ, M.Sc.,
Engenharia Elétrica, 2005)

Tese - Universidade Federal do Rio de
Janeiro, COPPE

1. Análise de Componentes Independentes
2. Processamento de Imagens
3. Mamografias

I. COPPE/UFRJ II. Título (série)

Dedicatória

À minha querida mãe, mãe de sangue e de alma, minha referência de força e de coragem, meu baluarte. À mulher mais guerreira que já conheci sobre a Terra, cujo caráter é meu exemplo. Meu colo, minha palmatória, minha amiga, meu espelho ferino, minha veia exposta, minha eterna companheira. À sua contrariedade muda e apoio irrestrito. Para que nunca se esqueça de meu amor sobejo e inexorável.

Agradecimentos

À minha tia-madrinha-mãe, abnegação cega, doação completa, presença certa.

Ao meu padrinho, por seu incentivo e crédito, quando tudo era apenas um sonho.

Ao meu orientador, Eduardo, brilhante em inteligência e retidão, que me enche de orgulho por sua orientação e, sobretudo, sua amizade.

Ao meu orientador, Wagner, por seu irretocável empenho em construir um laboratório de excelência num cenário de recursos escassos e objeções políticas.

À sempre amável Dra. Carolina, radiologista do INCa, sem a qual esse trabalho seria infundado e inconcebível. Por ter me permitido desfrutar de seus ensinamentos de Medicina e de vida. Por sua admirável lição de entrega ao ser humano e de esperança em ver um Sistema de Saúde fundado em respeito e dignidade.

Ao meu amigo André Victor, pelo apoio ao projeto e prontidão sempre recheada com idéias notáveis, base de sua incansável dedicação à pesquisa. À minha amiga Fabiana, por seu trabalho hercúleo em carinhosamente preparar o banco de imagens em conjunto com a Dra. Carolina.

Aos amigos do LUS e do LPS, pelos momentos de partilha na exaustiva caminhada de escrever uma tese.

À minha amiga Rosane, sempre perto para colar os caquinhos do meu frágil coração.

A todos os meus amigos, pelo carinho apesar de minha ausência.

Aos meus companheiros da Ultraflux, por tornar meus dias mais leves com sua alegria, em especial aos meus diretores, Guilherme e André, por sua paciência e compreensão em me dividir com este projeto.

Aos professores da banca examinadora, Homero e Calôba, pelo interesse e colaboração.

Aos funcionários da COPPE-UFRJ, em especial Solange e Amanda, por sua competência e simpatia.

Ao INCa, pelas imagens cedidas e ao CNPq, pelo apoio financeiro ao projeto.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

ANÁLISE DE COMPONENTES INDEPENDENTES APLICADA À
IDENTIFICAÇÃO DE REGIÕES LESIONADAS EM MAMOGRAMAS

Luciana Leite

Abril/2005

Orientadores: Eduardo Antônio Barros da Silva

Wagner Coelho de Albuquerque Pereira

Programa: Engenharia Elétrica

No Brasil, o câncer de mama é a maior causa de óbitos por câncer na população feminina. A tese propõe um novo método de discriminação entre tecidos normais e lesionados em mamogramas, baseados em Análise de Componentes Independentes. Considerando que um mamograma típico consiste em um conjunto de informações heterogêneas de diversas procedências (tecidos, filme, características do mamógrafo, etc.), uma imagem mamográfica pode ser descrita como uma combinação linear de um conjunto de imagens-base estatisticamente independentes entre si. Os coeficientes dessa mistura são os parâmetros de entrada usados por uma rede neural com treinamento supervisionado em *backpropagation* para classificar a região como normal ou lesionada. O método foi testado em um conjunto de imagens do INCa, e obteve uma taxa de reconhecimento de cerca de 86 % das regiões de interesse selecionadas com o auxílio de um radiologista especializado.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

INDEPENDENT COMPONENT ANALYSIS APPLIED TO THE IDENTIFICATION
OF LESIONED REGIONS IN MAMMOGRAMS

Luciana Leite

April/2005

Advisors: Eduardo Antônio Barros da Silva
Wagner Coelho de Albuquerque Pereira

Department: Electrical Engineering

In Brazil, breast cancer is one of the leading causes of mortality amongst women. We propose a new method for discriminating between normal and abnormal tissue on mammograms, based on Independent Component Analysis – namely ICA. Considering that a typical mammogram consists of a mixture of heterogeneous information from different origins (tissues, film, X-ray machine), a mammographic image can be described as linear combination of an unknown set of statistically independent source-images. The coefficients of the mixture are the input features that a neural network will use to accuse abnormality. The method has been tested on a set of mammograms from INCa, and recognized correctly about 86% regions of interest selected with the aid of a radiologist.

Índice

Ficha Catalográfica	ii
Dedicatória	iii
Agradecimentos	iv
Resumo	v
Abstract	vi
Índice	vii
Lista de Figuras	x
Lista de Tabelas	xii
Lista de Siglas	xv
1. Introdução	1
1.1 - Relevância	1
1.1.1 O câncer no Brasil	1
1.1.2 Mamografia preventiva.....	4
1.1.3 Diagnóstico auxiliado por computador.....	5
1.2 - Objetivo	6
1.3 - Estrutura do trabalho	9
2. Aspectos Biológicos.....	10
2.1 - Fisiologia Celular e Carcinogênese	10
2.2 - Mamografia	11
2.3 - Morfologia da Mama	11
2.4 - Mastopatias.....	13
2.4.1 Mastopatias Benignas	13
2.4.2 Mastopatias Malignas	14
2.4.3 Microcalcificações.....	17
2.4.3.1 Calcificações em processos benignos:.....	18
2.4.3.2 Calcificações em processos malignos.....	19
3. Análise de Componentes Independentes	21
3.1 - Introdução.....	21
3.1.1 Motivação	21
3.1.2 Modelo.....	25
3.1.3 Restrições	27
3.1.3.1 As componentes independentes são estatisticamente independentes entre si.	27
3.1.3.2 As componentes independentes devem apresentar distribuições não gaussianas	31
3.1.3.3 A matriz de mistura é quadrada.....	32
3.1.4 Ambigüidades.....	33
3.1.4.1 Não é possível determinar a variância das componentes independentes.....	33
3.1.4.2 Não é possível determinar a ordem das componentes independentes	33
3.2 - ICA por Maximização da Não Gaussianidade	33
3.2.1 Não- Gaussianidade.....	33
3.2.2 Kurtosis como medida de não-gaussianidade.....	35
3.2.2.1 Gradiente	38

3.2.2.2 Ponto-Fixo (FastICA por kurtosis).....	40
3.2.3 Neg-entropia.....	42
3.2.3.1 Gradiente.....	44
3.2.3.2 Ponto-Fixo (FastICA por neg-entropia).....	45
3.3 - ICA por Maximização da Função de Verossimilhança.....	47
3.3.1.1 Gradiente.....	52
3.3.1.2 Ponto Fixo (FastICA por Maximização da Verossimilhança).....	54
3.4 - Estimação de várias componentes.....	55
3.4.1 Ortogonalização por deflação.....	56
3.4.2 Ortogonalização simétrica.....	56
4. Redes Neurais.....	58
4.1 - Redes em <i>backpropagation</i>	60
4.1.1 Associadores e Classificadores.....	60
4.1.2 Treinamento.....	64
4.2 - Componentes de Atuação Conjunta.....	68
4.2.1 Treinamento.....	69
5. Sistema de diagnóstico.....	72
5.1 - Materiais.....	72
5.2 - Intuição sobre o método.....	77
5.3 - Metodologia.....	80
5.3.1 Pré-processamento.....	80
5.3.1.1 Filtragem morfológica.....	81
5.3.2 Extração de parâmetros.....	84
5.3.2.1 ICA.....	84
5.3.2.2 Outros métodos de extração de parâmetros.....	89
5.3.3 Classificação.....	90
5.3.3.1 Backpropagation.....	90
5.3.3.2 Componentes de atuação conjunta (CAC).....	91
6. Resultados.....	94
6.1 - Extração de parâmetros por ICA.....	94
6.1.1 Classificação por <i>backpropagation</i>	94
6.1.1.1 Translação da imagem.....	97
6.1.2 Classificação por Componentes de Atuação Conjunta.....	98
6.1.2.1 ICA sobre imagens pré-processadas.....	100
6.2 - Extração de parâmetros por Matriz de co-ocorrência.....	101
6.3 - Extração de parâmetros por Curva de Complexidade.....	102
7. Conclusão.....	104
Apêndices.....	107
A Teorema do Limite Central.....	107
B Morfologia Matemática.....	109
B.1 Erosão.....	109
B.2 Dilatação.....	110
B.3 Abertura.....	111
B.4 Fechamento.....	113
B.5 Dilatação condicional.....	115
B.6 Erosão condicional.....	116
B.7 Reconstrução Inferior.....	116

B.8	Reconstrução superior	117
B.9	Abertura por reconstrução	118
B.10	Fechamento por reconstrução	118
B.11	Top-Hat.....	119
C	Componentes Principais	122
D	Matriz de Co-ocorrência.....	125
E	Curva de Complexidade	129
F	Tabelas de Resultados	131
Bibliografia.....		154

Lista de Figuras

Figura 1.1 - Taxa bruta de mortalidade entre 1969 a 2000 e projeções para o ano 2003, em mulheres, para algumas localizações primárias (Brasil).....	2
Figura 1.2 - Representação espacial das taxas brutas de incidência por 100.000 mulheres estimadas para o ano 2005, segundo a Unidade da Federação (neoplasia maligna da mama feminina) [2].	4
Figura 1.3 - ROI (<i>Region of Interest</i>) contendo tecido lesionado descrita como uma combinação linear de imagens-base estatisticamente independentes entre si.	7
Figura 1.4 – Esquema de classificação de tecidos normais e lesionados por ICA.	8
Figura 2.1 – Célula Normal x Célula Cancerosa.	10
Figura 2.2 – Mamógrafo.	11
Figura 2.3 – Anatomia da Mama.	12
Figura 3.1 – Sinais de áudio originais.	21
Figura 3.2 – Sinais observados, obtidos a partir da mistura dos sinais da Figura 3.1. ...	22
Figura 3.3 – Sinais originais estimados a partir da mistura observada na Figura 3.2. ...	23
Figura 3.4 – Imagens-base obtidas a partir da separação de imagens naturais em componentes independentes. Fonte: Hyvärinen A., <i>Independent Component Analysis</i> [31]	24
Figura 3.5: a) Distribuição conjunta das componentes independentes s_1 e s_2 com distribuições uniformes. b) Distribuição da mistura das componentes s_1 e s_2	27
Figura 3.6 - Distribuição conjunta da mistura de componentes distribuídas uniformemente após o branqueamento.....	31
Figura 3.7 – Distribuição conjunta de duas variáveis gaussianas.....	32
Figura 3.8 – a) Distribuição conjunta de 2 variáveis gaussianas independentes.....	35
Figura 3.9 – a) Distribuição conjunta de 2 variáveis platokúrticas independentes	36
Figura 3.10 – a) Distribuição conjunta de 2 variáveis leptokúrticas;	36
Figura 3.11 - Curvas de nível da função $F(q)$	38
Figura 3.12 – Funções $G(y)$	44
Figura 3.13 – Funções de condição de estabilidade.	52
Figura 3.14 – Funções não lineares de aproximação.....	54
Figura 4.1 – Constituição da célula neuronal.	58
Figura 4.2 – Neurônio: elemento de processamento.	59
Figura 4.3 – Esquema de operação.....	60
Figura 4.4 – Classes linearmente separáveis.	61
Figura 4.5 – Rede Neural para discriminação entre as classes A e B linearmente separáveis.	61
Figura 4.6 – Projeção dos elementos das classes na direção de w	62
Figura 4.7 – Classes linearmente não-separáveis.	62
Figura 4.8 – Mapeamento das classes A e B no hipercubo unitário.....	63
Figura 4.9 – Rede Neural de separação entre as classes A e B linearmente não separáveis.	64

Figura 4.10 – Função de ativação do tipo tangente hiperbólica: função monótona crescente limitada.	65
Figura 4.11 – Esquema de treinamento.	65
Figura 4.12 – Rede com 2 camadas.	66
Figura 4.13 – Passo 1.	69
Figura 4.14 – Passo 2.	70
Figura 5.1 - ROIs a) normal contendo parte do músculo peitoral(r186); b) loja de gordura (r130); c) parênquima livre(r15); d) área contendo microcalcificações benignas (r82); e) carcinoma ductal infiltrante; r(128) f) calcificações grosseiras (r11).	77
Figura 5.2 - Histogramas correspondentes a: a), b) ROIs normais apresentadas na Figura 5.1 (b) e (c), respectivamente; c), d) ROIs lesionadas apresentadas na Figura 5.1 (e) e (f), respectivamente.	78
Figura 5.3 - a) ROI original; b) ROI normalizada.	80
Figura 5.4 a) ROIs contendo loja de gordura (r130) b) ROI filtrada pelo <i>top-hat</i>	83
Figura 5.5 a) ROIs contendo parênquima livre (r15) b) ROI filtrada pelo <i>top-hat</i> ..	83
Figura 5.6 a) ROIs contendo calcificações grosseiras (r11) b) ROI filtrada pelo <i>top-hat</i> ..	84
Figura 5.7 a) ROI com CDI (r128) b) ROI filtrada pelo <i>top-hat</i>	84
Figura 5.8 – Diagramação da metodologia empregada para o sistema de diagnóstico por ICA.	88
Figura 5.9 - Curvas de Erro de Aprendizado para o Treinamento e Teste.	91
Figura 5.10 – Diagramação da metodologia empregada para o sistema de diagnóstico por CAC.	92
Figura 6.1 – Organograma opções de processamento aplicadas.	94
Figura 6.2 – Desenho Esquemático: ICA + <i>backpropagation</i>	95
Figura 6.3 – Imagens-base (método: <i>neg-entropia</i>).	96
Figura 6.4 – Desenho Esquemático: translação + ICA + <i>backpropagation</i>	97
Figura 6.5 – ROI transladada.	98
Figura 6.6 – Desenho Esquemático: ICA + CAC.....	99
Figura 6.7 – Desenho Esquemático: Top-Hat + ICA + CAC.....	100
Figura 6.8 – Desenho Esquemático: GLCM + <i>backpropagation</i>	101
Figura 6.9 – Desenho Esquemático: CCLx + <i>backpropagation</i>	103

Lista de Tabelas

Tabela 1.1: Estimativas para o ano 2005 de número de casos novos por câncer, em homens e mulheres, segundo localização primária.	1
Tabela 1.2 – Estadiamento do Câncer de Mama [4].....	3
Tabela 4.1 – Classes A e B.....	63
Tabela 5.1 – Diagnóstico Radiológico e Histológico das mamografias utilizadas	72
Tabela 7.1 – Resumo dos melhores resultados obtidos.....	105
Tabela F.1.....	131
Tabela F.2.....	131
Tabela F.3.....	131
Tabela F.4.....	131
Tabela F.5.....	132
Tabela F.6.....	132
Tabela F.7.....	132
Tabela F.8.....	132
Tabela F.9.....	133
Tabela F.10.....	133
Tabela F.11.....	133
Tabela F.12.....	133
Tabela F.13.....	134
Tabela F.14.....	134
Tabela F.15.....	134
Tabela F.16.....	134
Tabela F.17.....	135
Tabela F.18.....	135
Tabela F.19.....	135
Tabela F.20.....	135
Tabela F.21.....	136
Tabela F.22.....	136
Tabela F.23.....	136
Tabela F.24.....	136
Tabela F.25.....	137
Tabela F.26.....	137
Tabela F.27.....	137
Tabela F.28.....	137
Tabela F.29.....	138
Tabela F.30.....	138
Tabela F.31.....	138
Tabela F.32.....	138
Tabela F.33.....	139

Tabela F.34.....	139
Tabela F.35.....	139
Tabela F.36.....	139
Tabela F.37.....	140
Tabela F.38.....	140
Tabela F.39.....	140
Tabela F.40.....	140
Tabela F.41.....	140
Tabela F.42.....	140
Tabela F.43.....	141
Tabela F.44.....	141
Tabela F.45.....	141
Tabela F.46.....	141
Tabela F.47.....	141
Tabela F.48.....	141
Tabela F.49.....	142
Tabela F.50.....	142
Tabela F.51.....	142
Tabela F.52.....	142
Tabela F.53.....	142
Tabela F.54.....	142
Tabela F.55.....	143
Tabela F.56.....	143
Tabela F.57.....	143
Tabela F.58.....	143
Tabela F.59.....	143
Tabela F.60.....	143
Tabela F.61.....	144
Tabela F.62.....	144
Tabela F.63.....	144
Tabela F.64.....	144
Tabela F.65.....	144
Tabela F.66.....	144
Tabela F.67.....	145
Tabela F.68.....	145
Tabela F.69.....	145
Tabela F.70.....	145
Tabela F.71.....	145
Tabela F.72.....	145
Tabela F.73.....	146
Tabela F.74.....	146

Tabela F.75	146
Tabela F.76	146
Tabela F.77	146
Tabela F.78	146
Tabela F.79	146
Tabela F.80	146
Tabela F.81	147
Tabela F.82	147
Tabela F.83	147
Tabela F.84	147
Tabela F.85	148
Tabela F.86	148
Tabela F.87	148
Tabela F.88	148
Tabela F.89	149
Tabela F.90	149
Tabela F.91	149
Tabela F.92	149
Tabela F.93	150
Tabela F.94	150
Tabela F.95	150
Tabela F.96	150
Tabela F.97	151
Tabela F.98	151
Tabela F.99	151
Tabela F.100	151
Tabela F.101	152
Tabela F.102	152
Tabela F.103	152
Tabela F.104	152
Tabela F.105	153

Lista de Siglas

BSS	<i>Blind Source Separation</i>
CAC	Componentes de Atuação Conjunta
CAD	<i>Computer Aided Diagnosis</i>
CDI	Carcinoma Ductal Infiltrante
CDIS	Carcinoma Ductal <i>in situ</i>
CMM	<i>Centre de Morphologie Mathématique</i>
DCT	<i>Discrete Cosine Transform</i>
DFT	<i>Discrete Fourier Transform</i>
DNA	<i>Deoxyribonucleic Acid</i>
EEG	Eletro encefalograma
FN	Falso Negativo
FP	Falso Positivo
GLCM	<i>Grey-Level Co-Occurrence Matrix</i>
ICA	<i>Independent Component Analysis</i>
INCa	Instituto Nacional do Câncer
ML	<i>Maximum Likelihood</i>
MRF	<i>Markov Random Fields</i>
NN	<i>Neural Network</i>
PCA	<i>Principal Component Analysis</i>
PDF	<i>Power Density Function</i>
ROI	<i>Region of Interest</i>
SONAR	<i>Sound Navigation Ranging</i>
VN	Verdadeiro Negativo
VP	Verdadeiro Positivo

1. Introdução

1.1 - Relevância

1.1.1 O câncer no Brasil

No mundo, o câncer de mama, entre as mulheres, apresenta-se como a segunda neoplasia maligna mais incidente, assim como uma causa relevante de mortes por câncer [1].

No Brasil a realidade é similar. Dos cerca de 467 mil novos casos de câncer esperados para 2005, 49.470 são de câncer de mama, representando um risco estimado de 53 casos a cada 100 mil mulheres, segundo dados fornecidos pelo Instituto Nacional do Câncer, situado no Rio de Janeiro.

Este número representa a ocorrência de um novo caso a cada 11 minutos.

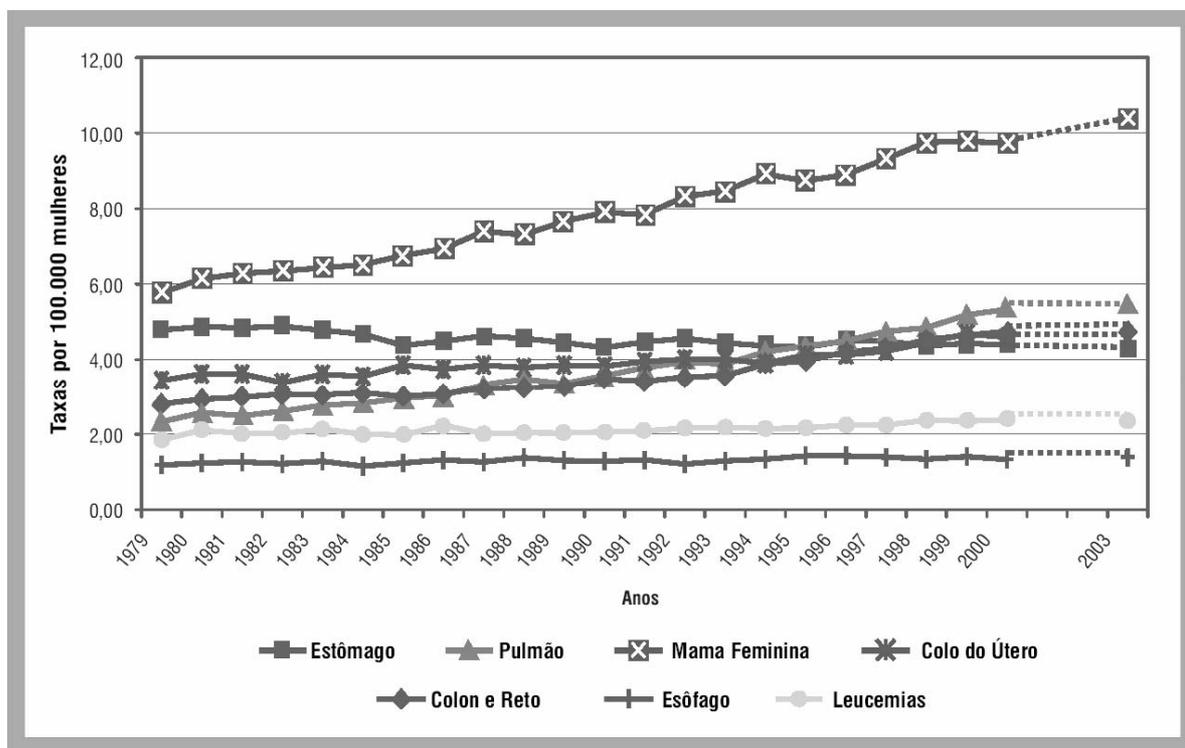
A tabela 1.1 apresenta as estimativas de números de novos casos para o ano de 2005, segundo a sua localização primária.

Tabela 1.1: Estimativas para o ano 2005 de número de casos novos por câncer, em homens e mulheres, segundo localização primária.

Localização Primária Neoplasia Maligna	Estimativa dos Casos Novos		
	Masculino	Feminino	Total
Pele não Melanoma	56.420	56.600	113.020
Mama Feminina	–	49.470	49.470
Traquéia, Brônquio e Pulmão	17.110	8.680	25.790
Estômago	15.170	7.975	23.145
Colo do Útero	–	20.690	20.690
Próstata	46.330	–	46.330
Cólon e Reto	12.410	13.640	26.050
Esôfago	8.140	2.450	10.590
Leucemias	5.115	4.075	9.190
Cavidade Oral	9.985	3.895	13.880
Pele Melanoma	2.755	3.065	5.820
Outras Localizações	56.175	67.290	123.465
Total	229.610	237.830	467.440

No Brasil, ainda, o câncer de mama é o que mais causa mortes entre as mulheres. Em 2002, foram registradas 9.004 mortes decorrentes deste tipo de câncer [2].

A Figura 1.1 revela a evolução da taxa de mortalidade entre mulheres entre o período de 1979 e 2000, e a projeção do INCa para o ano de 2003.



Fontes: MS/FUNASA/CENEPI/Sistema de Informação sobre Mortalidade – SIM

MS/Instituto Nacional do Câncer – INCA

Instituto Brasileiro de Geografia e Estatística/DPE/DEPIS

Figura 1.1 - Taxa bruta de mortalidade entre 1969 a 2000 e projeções para o ano 2003, em mulheres, para algumas localizações primárias (Brasil).

Nas últimas duas décadas observou-se um aumento considerável da taxa de mortalidade por câncer de mama entre mulheres. Esse número passou de 5,77/100.000 em 1979 a 9,74/100.000 em 2000, correspondendo a uma variação percentual relativa de +80,3%.

Apesar de ser considerado como um câncer de relativamente bom prognóstico, se diagnosticado e tratado oportunamente, as taxas de mortalidade por câncer de mama continuam elevadas no Brasil, muito provavelmente porque a doença ainda seja diagnosticada em estádios avançados. Nesses estádios, a doença já evoluiu de forma que o organismo não é capaz de responder ao tratamento, que em geral é mutilante e causa maior sofrimento à mulher. Com base nos dados disponíveis de Registros Hospitalares, 60% dos tumores de mama, em média, são diagnosticados em estádios III ou IV [3]. A

Tabela 1.2 apresenta uma breve descrição das características dos diversos estádios do câncer de mama.

Tabela 1.2 – Estadiamento do Câncer de Mama [4].

Estádio 0	<ul style="list-style-type: none"> ▪ Tumor jovem. Este tipo não se espalha no seio ou fora dele; são chamados câncer <i>in situ</i> ou não-invasivos
Estádio I	<ul style="list-style-type: none"> ▪ O tumor não tem mais que 2 cm de comprimento e não se espalha para fora do seio (tumor jovem)
Estádio II	<ul style="list-style-type: none"> ▪ O tumor não tem mais que 2 cm, mas está espalhado para linfonodos axilares ▪ O tumor está entre 2 e 5 cm, podendo ou não está espalhado para linfonodos axilares. ▪ O tumor está maior que 5 cm, mas não se espalhou para linfonodos axilares.
Estádio III A	<ul style="list-style-type: none"> ▪ Tumor menor que 5 cm e espalhado para os linfonodos axilares. Este pode também estar espalhado para outros linfonodos ▪ Tumor maior que 5 cm espalhado para os linfonodos axilares.
Estádio III B	<ul style="list-style-type: none"> ▪ O câncer espalhou-se para tecidos vizinhos (pele, parede torácica, músculos do tórax e costelas) ▪ O câncer está espalhado pelos linfonodos dentro da parede torácica ao longo do esterno.
Estádio IV	<ul style="list-style-type: none"> ▪ O câncer está espalhado pelas outras partes do corpo, principalmente ossos, pulmões, fígado e cérebro. Pode está espalhado localmente na pele e linfonodos dentro do pescoço, próximo à coluna vertebral.

Investimentos tecnológicos e em recursos humanos no âmbito de um programa estruturado para detecção precoce desta neoplasia, assim como a implementação de um sistema nacional de informações, constituem estratégias importantes no sentido de reverter este cenário.

Nos estados do Centro-Sul, onde os programas de prevenção são mais divulgados, o número de incidências de câncer de mama é maior que nos estados do Norte e Nordeste, que registram, porém, um maior número de óbitos.

O quadro exposto alertou o país para a importância do **controle** e **prevenção** do câncer, institucionalizando programas que abrangem desde ações de promoção à saúde ao fomento a pesquisas voltadas para diagnóstico precoce da doença [3].

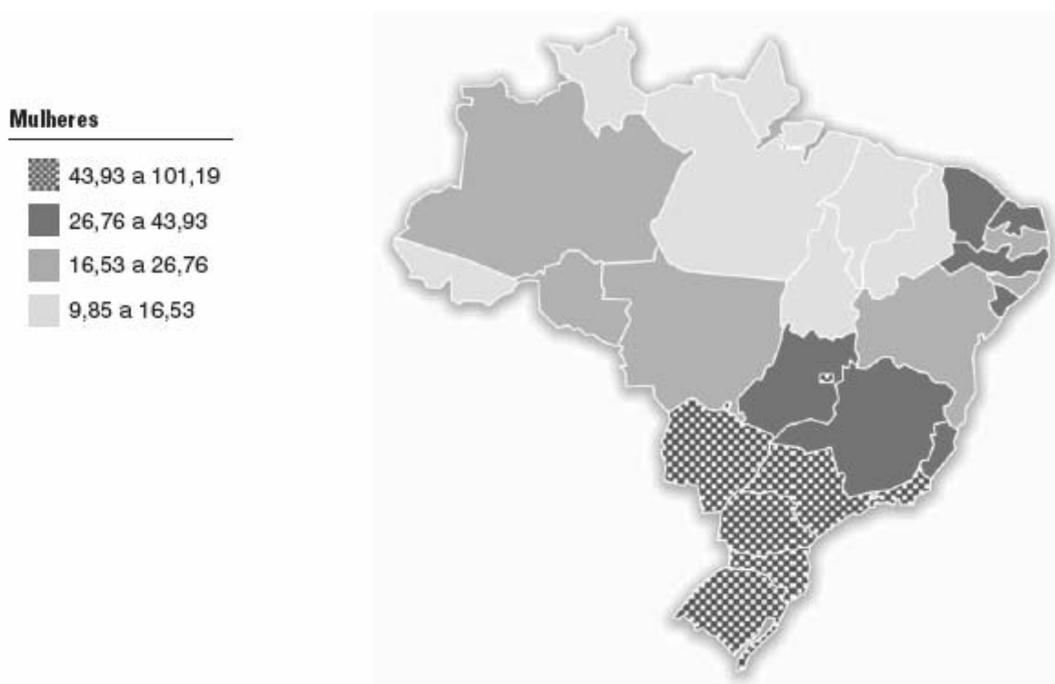


Figura 1.2 - Representação espacial das taxas brutas de incidência por 100.000 mulheres estimadas para o ano 2005, segundo a Unidade da Federação (neoplasia maligna da mama feminina) [2].

1.1.2 Mamografia preventiva

Alguns estudos indicaram que depósitos de cálcio em carcinomas mamários ocorrem dentro das células do tumor e no tecido que o circunda, o que levou a conclusões de que as células cancerígenas sofrem um processo ativo de secreção. Esses depósitos de cálcio, chamados **microcalcificações**, podem não ser apenas resultados da mineralização de tecidos mortos, mas sim, representar um sinal prematuro de malignidade em lesões da mama em mulheres assintomáticas [1].

Atualmente, a mamografia é a técnica mais barata e eficaz na detecção precoce do câncer de mama. Os equipamentos dedicados à mamografia de alta resolução produzem imagens de altíssima qualidade, que permitem detectar e analisar estruturas micrométricas, como as microcalcificações, com um mínimo de radiação - em torno de 0,2 rad por exame - o que não oferece risco para as pacientes.

As calcificações aparecem como regiões mais claras no filme mamográfico, devido à sua maior absorção de fótons de raios-X. Dessa forma, o Colégio Brasileiro de Radiologia, seguindo a diretiva internacional, incentiva a **mamografia** preventiva para

rastreamento de casos de câncer de mama em estágios precoces de sua evolução. Porém, o diagnóstico a partir da análise visual dos mamogramas não é eficiente. Apenas 15% a 30% dos casos mamograficamente suspeitos, não palpáveis, apresentam malignidade quando submetidos à biópsia [2, 5].

Ainda mais graves que os casos de falsos positivos, encontram-se os casos em que as pequenas dimensões dessas estruturas e as limitações dos sistemas fazem com que esses sinais passem despercebidos em uma inspeção visual do mamograma. São os chamados falsos negativos.

A resolução espacial - relação entre o número de *pixels* e a área do tecido analisada - de uma imagem mamográfica é superior à percepção visual do olho humano, [6] capaz de distinguir algumas dezenas de tons de cinza que não são suficientes para classificar uma microcalcificação [7], sobretudo quando a imagem do mamograma apresenta baixo contraste entre as microcalcificações e os tecidos que as circundam. Tecidos densos, comuns em mulheres jovens, podem, dessa forma, mascarar a presença de microcalcificações.

1.1.3 Diagnóstico auxiliado por computador

Esquemas de diagnóstico auxiliado por computador (CAD, de “*Computer-aided diagnosis*”) vêm sendo desenvolvidos com a finalidade de aumentar a probabilidade de detecção [8, 9]. O tema atrai o interesse da autora, que desenvolve e publica pesquisas correlatas desde 1998 [10-14].

O objetivo principal destes esquemas é fornecer aos médicos uma segunda opinião acerca do diagnóstico. Os CADs contornam problemas que surgem da subjetividade de um laudo humano, tais como expectativas, pré-conceitos e cansaço que interferem no diagnóstico do radiologista. Buscam, assim, abaixar as taxas de falsos positivos, ou seja, casos em que a suspeita de malignidade leva a mulher à biópsia - exame altamente invasivo e de maior custo - para um resultado negativo. Vários estudos revelam que algoritmos automáticos de detecção são capazes de proporcionar um aumento de mais de 20% no total de acertos do radiologista e, com isso, consegue-se evitar biópsias desnecessárias. [15]

Um estudo recente publicado pelo Journal of the National Cancer Institute [16] mostra que existe grande quantidade de resultados errados e divergentes em

mamografias realizadas nos Estados Unidos, principalmente se o radiologista for inexperiente. RAJI et al. [17] e NUNES et al. [18] demonstram que a melhoria das imagens através do seu processamento prévio aumenta a capacidade de interpretação tanto visual, realizada pelo radiologista, quanto automática, feita pelo Sistema de Auxílio Diagnóstico (CAD). Dessa forma, procura-se associar a heurística de identificação de padrões reconhecíveis baseada na percepção visual da imagem aos esquemas de CAD para extrair um diagnóstico.

1.2 - Objetivo

O foco do trabalho reside em avaliar a técnica de separação cega de fontes conhecida como ICA, como descritor de anormalidade do tecido mamário. Pretende-se, pois, construir um algoritmo que detecte regiões lesionadas em mamogramas digitalizados. Essas lesões podem ser desde tumores palpáveis ou carcinomas a microcalcificações imperceptíveis a olho nu. A técnica permite encontrar um conjunto de imagens-base em que uma imagem mamográfica genérica pode ser decomposta. Os pesos com que cada imagem-base contribui para formar a imagem original podem, assim, ser usados como parâmetros para avaliação do potencial de anormalidade da região analisada.

Um sistema que diferencie regiões saudáveis e lesionadas em uma mamografia baseia-se, tradicionalmente, na extração de parâmetros quantitativos da imagem que sejam capazes de caracterizar a região de interesse selecionada [19]. Esses parâmetros, normalmente descritores da textura da região, são submetidos a algum método de discriminação, como redes neurais [20, 21], discriminantes de Fisher [22], algoritmos genéticos [23, 24], para fornecer um diagnóstico acerca do potencial de malignidade da região analisada.

Um mamograma típico consiste um conjunto de informações heterogêneas de diversas procedências, como diferentes tipos de tecido (vasos, dutos, pele), a película do filme e características do mamógrafo.

O esquema proposto neste trabalho sugere que uma imagem mamográfica qualquer seja resultado da mistura de um conjunto de imagens-base. Cada mamografia é descrita como uma combinação linear de algumas imagens, ou seja, como uma soma ponderada de imagens que são comuns a todas as mamografias. O biomorfismo da mulher, as características do filme e do aparelho determinam o peso com que cada

imagem-base, também chamada de fonte, contribui para a imagem final. Assim a cada mamografia corresponde um vetor cujos elementos são os coeficientes, ou pesos, de cada fonte na mistura que constitui a imagem mamográfica.

$$\text{Imagem} = a_{i1} \cdot \text{Fonte}_1 + a_{i2} \cdot \text{Fonte}_2 + \dots + a_{in} \cdot \text{Fonte}_n$$

Figura 1.3 - ROI (*Region of Interest*) contendo tecido lesionado descrita como uma combinação linear de imagens-base estatisticamente independentes entre si.

Dessa forma, o vetor de pesos associado a uma região de interesse pode ser um descritor de anormalidade do tecido [19]. No entanto, apenas a imagem mamográfica é observável; tanto as imagens-base como os pesos associados a elas são desconhecidos.

A técnica conhecida como ICA, do inglês *Independent Component Analysis*, é capaz de encontrar fontes que originaram uma determinada mistura, somente a partir de uma mistura observada, supondo independência estatística entre elas. Ou seja, tendo observado um conjunto de dados e supondo que eles tenham sido gerados a partir de uma mistura de outros dados, a técnica é capaz de estimar as fontes que originaram a mistura e os pesos de cada fonte na mistura, a partir de análises estatísticas do sinal observado.

Com o auxílio de um radiologista experiente, foi selecionado um conjunto de regiões de interesse representativo de áreas da mama contendo tecidos saudáveis e lesionados. O conjunto abrange desde tumores palpáveis a microcalcificações imperceptíveis ao olho, e áreas de parênquima, gordura e músculo isentas de qualquer patologia. A partir desse conjunto, a técnica de ICA foi empregada para extrair um conjunto de fontes estatisticamente independentes que representariam as imagens-base da mistura. Conhecidas as imagens-base, a partir de matemática simples, é encontrado um vetor de pesos para cada região de interesse selecionada em cada mamograma a ser analisado.

Em princípio, os vetores-peso são usados como parâmetros de entrada para uma rede neural com treinamento supervisionado em *backpropagation*, que irá classificar a região como normal ou lesionada.

O desempenho do sistema foi avaliado para imagens originais e imagens pré-processadas para ressaltar estruturas que caracterizem o tecido lesionado.

Duas técnicas amplamente utilizadas na literatura para discriminação entre tecidos normais e anormais – a matriz de co-ocorrência [25] e a curva de complexidade [26] - foram aplicadas ao mesmo conjunto de imagens, para comparação dos resultados. Ambas as técnicas estão descritas no apêndice. O diagnóstico por ICA mostrou-se superior às demais técnicas, com 82% de acertos contra 70%, no melhor caso das outras técnicas empregadas.

Na Figura 1.4 está representada uma descrição esquemática do sistema empregado para classificação de amostras do tecido mamário em regiões normais e lesionadas.

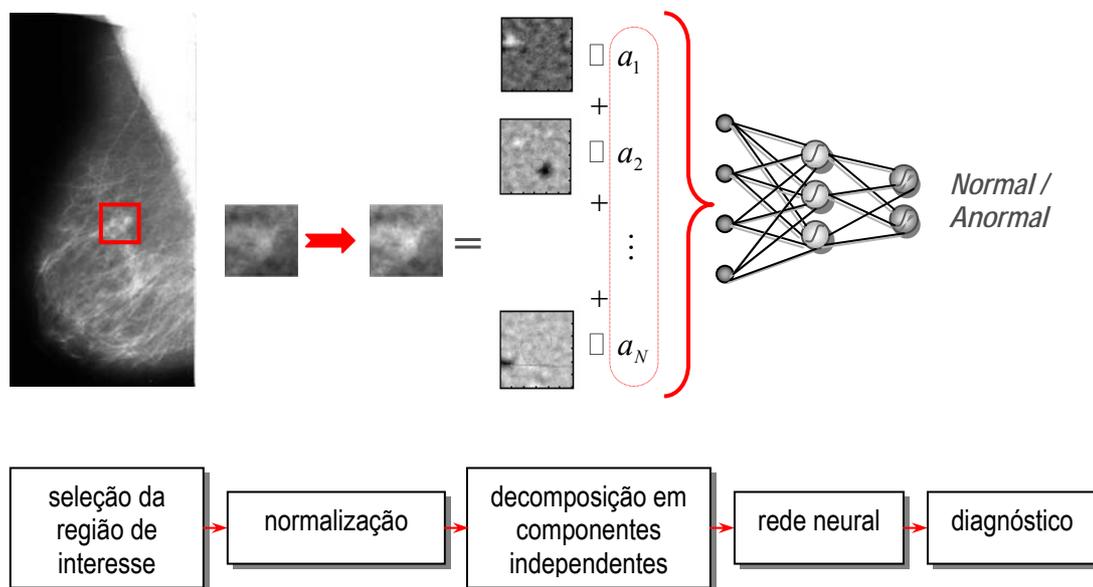


Figura 1.4 – Esquema de classificação de tecidos normais e lesionados por ICA.

Posteriormente, foi testado um novo tipo de rede neural para classificação, com uma primeira camada de neurônios lineares que realizam uma redução da dimensionalidade da entrada, considerando a contribuição de cada componente para a discriminação. Dessa forma, essa camada comprime a informação relevante, eliminando possíveis ruídos que possam atrapalhar o diagnóstico. Alcançam, assim, um resultado de cerca de 86% de acertos.

O método pretende, pois, extrair informações não observáveis na imagem que passem despercebidas às técnicas usuais de descrição de textura, mas que sejam, contudo, relevantes à discriminação entre tecidos saudáveis e lesionados.

A técnica foi empregada, inicialmente, para classificação em tecidos normais ou lesionados. Essa análise preliminar será futuramente ampliada para classificação entre lesões com suspeita de malignidade e regiões seguramente benignas, que podem ser poupadas de biópsia.

1.3 - Estrutura do trabalho

Neste capítulo, foi mostrada a relevância da prevenção do câncer de mama e do papel da mamografia na detecção precoce de lesões na mama. Nesse sentido, as microcalcificações figuram como um elemento patognomônico do câncer e podem auxiliar sua detecção em mulheres assintomáticas. Foram ressaltados os benefícios de esquemas de CAD para o diagnóstico e foi apresentada uma breve descrição do método de classificação de amostras do tecido mamário em áreas normais ou lesionadas por ICA. O **capítulo 2** apresenta uma descrição detalhada dos aspectos biológicos da mama relevantes para o diagnóstico. O **capítulo 3** traz um aprofundamento sobre as duas técnicas mais populares de ICA: a *maximização da não-gaussianidade* e a *maximização da função de verossimilhança*. O **capítulo 4** aborda a teoria sobre redes neurais artificiais supervisionadas, enfatizando o treinamento por *backpropagation*. No **capítulo 5**, a metodologia empregada para o diagnóstico baseado em ICA é detalhada. No **capítulo 6** os resultados são apresentados e discutidos. O **capítulo 7** apresenta a conclusão.

2. Aspectos Biológicos

2.1 - Fisiologia Celular e Carcinogênese

As células que constituem os animais são formadas por três partes: a membrana celular, que é a parte mais externa da célula; o citoplasma, que constitui o corpo da célula; e o núcleo, que contém os cromossomos que por sua vez são compostos de genes. Os genes são arquivos que guardam e fornecem instruções para a organização das estruturas, formas e atividades das células no organismo. Toda a informação genética encontra-se inscrita nos genes, numa "memória química" - o *ácido desoxirribonucléico* (DNA). É através do DNA que os cromossomos passam as informações para o funcionamento da célula [6].

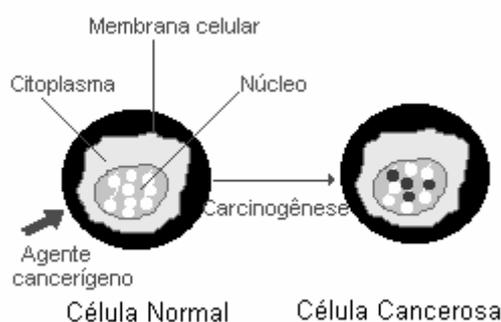


Figura 2.1 – Célula Normal x Célula Cancerosa.

Uma célula normal pode sofrer alterações no DNA dos genes, conforme está esquematizado na Figura 2.1. É o que se chama *mutação genética*. As células cujo material genético foi alterado passam a receber instruções erradas para as suas atividades. As alterações podem ocorrer em genes especiais, denominados *protooncogenes*, que a princípio são inativos em células normais. Quando ativados, os protooncogenes transformam-se em *oncogenes*, resultando em uma reprodução desordenada dessas células e desencadeando o aparecimento de massas celulares denominadas *neoplasias* ou, mais comumente, *tumores*.

Nas neoplasias malignas o crescimento é mais rápido, desordenado e infiltrante; as células não guardam semelhança com as que lhes deram origem e têm capacidade de se desenvolver em outras partes do corpo, fenômeno este denominado *metástase*, que é a característica principal dos tumores malignos [6].

2.2 - Mamografia

O sucesso do exame mamográfico depende da boa qualidade da imagem radiográfica. O médico deve observar a simetria das mamas, a pele, a papila e a aréola. Na apalpação, deve atentar para a situação do parênquima mamário, que varia com a faixa etária, com o ciclo menstrual e na gravidez. Importantes nesse exame são os nódulos e suas caracterizações (número, localização, consistência, sensibilidade, forma, etc.).

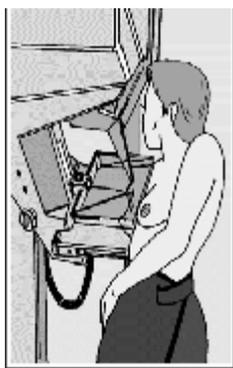


Figura 2.2 – Mamógrafo.

2.3 - Morfologia da Mama

As glândulas mamárias, que têm como principal função a secreção do leite, estão situadas na parede anterior do tórax, conforme se pode verificar na Figura 2.3. Na porção mediana do ápice aparece o *mamilo* ou *papila*, protuberância elástica onde desembocam os ductos mamários. Circundando o mamilo, existe uma membrana pigmentada que constitui a *aréola*, que é permeada de múltiplas e pequenas elevações que são os *tubérculos de Morgagni*. As dimensões da mama variam durante os diferentes períodos de desenvolvimento. São pequenas na puberdade, aumentam no período pós-pubertário, variam de dimensão no período de gravidez e lactação e tornam-se atrofiadas no período pós-menopáusico [6].

As mamas são sempre assimétricas, sendo que a esquerda geralmente é maior. A pele que circunda a região areolar e o mamilo é delgada e móvel. Abaixo da pele estão as fibras musculares lisas que constituem o músculo areolar, formado por fibras radiadas e fibras circulares. Estas últimas aderem ao nível da aréola e estendem-se até a base do mamilo, onde se entrelaçam em torno dos canais excretores. As fibras radiadas

têm direção perpendicular e se originam na derme da aréola e terminam na derme do mamilo. O corpo glandular é circundado por uma cápsula de tecido conjuntivo que emite prolongamentos à derme, denominados *cristas de Duret*.

Todo o restante da mama é preenchido por tecido adiposo, cuja quantidade varia com as características físicas, estado nutricional e idade da mulher. O *parênquima*, ou glândula propriamente dita, é formado por um conjunto de sistemas excretor e secretor. Sua superfície posterior é lisa e plana. A face anterior é irregular e tem forma cônica na região mamilar, onde terminam os canais galactóforos, que são em número de 15 a 20. Cada canal galactóforo divide-se em *lobos* e *lóbulos*, que terminam nos *ácinos*, as menores partes da glândula, responsáveis pela produção do leite durante a lactação, sob estímulos hormonais.

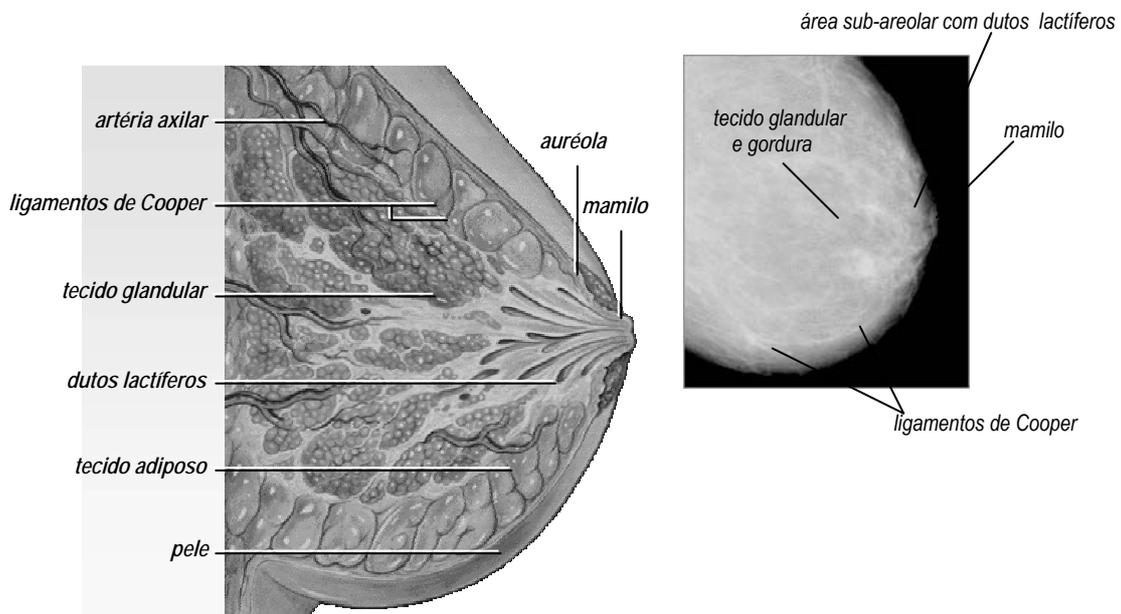


Figura 2.3 – Anatomia da Mama.

O sistema lobular acinar está envolvido pelas estruturas de sustentação que são os *ligamentos de Cooper* e as *cristas de Duret*. Elas prendem o sistema ao tecido celular subcutâneo. Externamente a eles encontram-se o *tecido fibroadiposo* e as *lojas de gordura*.

2.4 - Mastopatias

2.4.1 Mastopatias Benignas

As mastopatias benignas são proliferações ou hiperplasias do tecido glandular mamário, originadas pelas alterações hormonais que levam à perda da relação conjuntivo-glandular. A partir do epitélio, existe uma hiperplasia sólida nos canais galactóforos ou o aparecimento de cavidades císticas cheias de líquidos, margeadas pelo epitélio galactóforo.

Azevedo [27] define as mastopatias benignas localizadas mais comuns como:

- a) *Fibroadenoma*: composto de elementos epiteliais e conjuntivos, ou seja, de tecido fibro-glandular, cuja proliferação se faz harmonicamente. São clinicamente palpáveis, indolores, móveis e não aderentes aos planos adjacentes. A imagem clássica é de um nódulo que geralmente não ultrapassa 4cm. Esse nódulo é unilateral, de contorno irregular, com halo exterior, sem apresentar invasão do tecido glandular ou fibroadiposo adjacente. A presença de calcificações grosseiras múltiplas em seu interior é comum, sendo mais observada nas pacientes pré-menopáusicas, em virtude do estroma esclerosado e hialinizado.
- b) *Cistos*: são coleções líquidas no interior de pequenos ou grandes ductos limitados pelo epitélio ductal, pressionado pela tensão do líquido. Origina-se do acúmulo de líquido dos períodos ovulatórios e menstruais não absorvido após o ciclo menstrual e também da obstrução de ductos ectasiados. Geralmente os cistos simples têm paredes regulares e nítidas (bem delineadas). Na maioria dos casos são bilaterais. Histologicamente estão associados ao tecido fibro-glandular sem interferência ao tecido adiposo. Raramente calcificam, e neste caso, aparecem em torno do nódulo ou em semi-halo. Nas mamas densas não são identificados na mamografia.
- c) *Fibroadenoma gigante*: o aspecto radiológico do fibroadenoma gigante, variante do fibroadenoma pericanalicular, é o de massa densa, de contorno regular (liso), ocupando o corpo glandular parcial ou totalmente. Os sinais radiológicos são muito semelhantes ao do tumor filóide, apresentando, ao contrário deste, uma evolução abrupta. É geralmente único e atinge pacientes na faixa etária entre 10 e 25 anos.
- d) *Tumor filóide*: também chamado de cistossarcoma filóide, é um tumor fibroepitelial que cresce lentamente e se apresenta microscopicamente com áreas císticas alternadas com áreas sólidas em forma de folhas ou lâminas. A imagem radiográfica do tumor filóide é a de massa expansiva e densa, que ocupa um quadrante da mama

ou todo o corpo glandular. Tanto as formas malignas quanto as benignas têm bom prognóstico.

- e) *Lipoma*: são tumores benignos localizados na periferia dos lobos mamários; são enlaçados pelos ligamentos de Cooper e, em geral, apresentam-se encapsulados. A imagem radiográfica é a de uma área lucente, de contorno regular, torneado por fina cápsula. Podem atingir grandes dimensões.
- f) *Fibroadenolipoma*: são massas benignas compostas de tecido gorduroso e tecido glandular, móveis à palpação e menos freqüentes que os lipomas. Na mamografia aparecem como uma massa lucente com até 7 cm de extensão. Geralmente têm forma redonda ou oval e são contornados por fina cápsula. Contêm em seu interior áreas mais densas que correspondem ao tecido glandular.
- g) *Abscesso*: corresponde a um processo inflamatório periductal, com formações pustulosas localizadas na região retroareolar. Apresenta reação no tecido glandular ou no tecido fibro-glandular adjacente. Geralmente estão associados a mastites. Na fase aguda, confunde-se com o carcinoma inflamatório. Na fase crônica apresenta assimetria mamária, com inversão do mamilo e espessamento da pele e da região retroareolar.
- h) *Liponecrose*: a necrose gordurosa é um processo inflamatório não supurativo, geralmente resultante de traumatismos, cirurgia, biópsias, infecções ou terapia por irradiação. Os depósitos de sais de cálcio ao redor do tecido adiposo produzem padrões diversos de calcificações que podem ser redondas ou ovais, de contornos lisos ou regulares, isoladas, esparsas ou agrupadas. A imagem radiográfica é variável e pode ser confundida com carcinoma. Aparecem como áreas de maior densidade relativa, de contorno irregular, áreas de maior densidade com deformação do tecido fibro-glandular adjacente ou imagens simulando espículas.

2.4.2 Mastopatias Malignas

As mastopatias malignas em geral são hormonodependentes. Em geral, têm origem nas células epiteliais dos ductos maiores e sistemas ductais lobares, lobulares e acinares. São raras antes dos 25 anos e a maior incidência está entre mulheres menopáusicas e pós-menopáusicas.

Segundo Azevedo [27], as mastopatias malignas mais comuns são:

- a) *Carcinoma lobular "in situ"*: é uma lesão originada nos ácinos ou porções lobulares da mama. Nesse tipo de lesão os ácinos estão distendidos por proliferação celular

epitelial com padrões histológicos característicos e há associação com microcalcificações agrupadas. É comumente localizado nos quadrantes externos e na região retroareolar. A bilateralidade e multacentralidade são freqüentes. A imagem radiográfica é de pequenas áreas redondas, conglomeradas, sem expressão tumoral, pois não há características infiltrativas. Em muitos casos, apresenta imagem de área de maior densidade relativa de contorno irregular.

- b) *Carcinoma ductal “in situ”*: é o precursor do carcinoma invasivo, o qual não demonstra ao exame microscópico a invasão da membrana basal. Deriva do epitélio do ducto. As células epiteliais proliferadas enchem o lúmen ductal, sofrem alterações necróticas e calcificam. À macroscopia, os debris necróticos no interior do ducto simulam o comedocarcinoma. Apresenta calcificações polimorfas de dimensões variáveis. A bilateralidade e a multacentricidade também são comuns.
- c) *Comedocarcinoma*: tem origem no centro da área tubular sólida do ducto e dos ramos ductais (lóbulos). Histologicamente, os ductos envolvidos estão plenos de massas celulares epiteliais atípicas com acentuado polimorfismo nuclear. O tumor torna-se necrótico no centro das massas e calcificam-se. Observam-se calcificações polimorfas agrupadas, isoladas esparsamente e difusas, associadas ou não, com nódulo dominante.
- d) *Carcinoma papilífero*: a conotação “papilífero” está relacionada com a formação de camadas múltiplas de células epiteliais pleomórficas com pedículo fibrovascular bem definido, em forma de papilas. Existem dois tipos: os intraductais e os intracísticos. O primeiro apresenta-se como uma lesão pequena, não ultrapassando os 3 cm. A imagem radiográfica é de área de maior densidade relativa ou pequeno nódulo dominante, associado ou não a calcificações. No carcinoma intracístico há proliferação dos elementos mesenquimais. Os cistos ocorrem na parede interna do ducto e se devem a hemorragias sucessivas. A lesão é altamente invasiva e a bilateralidade e multacentricidade são comuns.
- e) *Carcinoma cribiforme*: é um tipo de carcinoma ductal “in situ”, variante da forma papilífera. Histologicamente, as micropapilas estão distribuídas em forma de cachos de uva, de onde surge a denominação. Pode estar presente na forma cribiforme pura ou em conjunto com a forma micropapilar
- f) *Carcinoma ductal infiltrante*: é a neoplasia maligna do ducto maior e apresenta massa sólida cuja infiltração ultrapassa a membrana basal e se estende ao tecido fibroglandular adjacente. Em geral ocasiona metástase para as cadeias linfonodais regionais, principalmente os linfonodos axilares e para órgãos distantes. A necrose é

comum assim como a associação com vários tipos de calcificação. A massa infiltrante emite espículas radiantes em forma de estrela. A imagem radiográfica varia bastante, apresentando formas circunscritas parcialmente precisas, formas circunscritas com margens infiltrantes correlacionadas com forma irregular, formas de núcleo denso com radiações espiculadas, massas expansivas densas, e etc. É o tumor maligno mais comum da mama.

- g) *Adenocarcinoma*: é uma variante do carcinoma ductal infiltrante invasivo. A histopatologia mostra grupos celulares variados como mucinosos, papilares e medulares, que podem estar agrupados ou em cordões, em quantidades variadas. A imagem radiográfica mais observada é a de massa densa circunscrita em reação desmoplásica (infiltração) em torno, invadindo as estruturas fibroglandulares adjacentes. Pode ser confundido com massas expansivas benignas.
- h) *Carcinoma lobular invasivo*: as células estão distribuídas no lóbulo ou no ácino em fila indiana. A reação desmoplásica é menor que no ductal infiltrante, sendo freqüente em mamas lipossubstituídas. A imagem radiográfica pode apresentar massa ou nódulo sem grandes infiltrações no tecido fibroglandular adjacente. Por vezes apresenta áreas de maior densidade relativa de contorno irregular. Tem associação com calcificações, na maioria das vezes, polimorfos.
- i) *Carcinoma tubular*: histologicamente apresenta pouco tecido glandular, com células simples em forma de túbulos. Não há evidências de células mioepiteliais e a membrana basal é bem definida. É um tumor pouco invasivo, muito confundido com o adenocarcinoma. Pode apresentar massas de pequenas dimensões, nódulo ou área de maior densidade relativa. Pode ainda apresentar calcificações agrupadas, gerando metástase para os linfonodos axilares.
- j) *Carcinoma medular*: massa de grandes dimensões, circunscritas, com alterações císticas. Pode apresentar-se com ou sem estroma linfóide. As formações císticas apresentam necroses e hemorragias. À microscopia, contém infiltrado linforreticular, composto predominantemente de células plasmáticas e linfócitos, mitoses ativas destas células e modificações nucleares. O padrão de crescimento celular tende a ser túbulo-acinar
- k) *Carcinoma Mucinoso*: tumor palpável, mal definido nas bordas, podendo apresentar massa expansiva de grandes dimensões e limites precisos, simulando lesões císticas. A imagem radiográfica é de nódulo denso, bem definido, que embora comprima o tecido fibroglandular adjacente, não apresenta invasão significativa. No interior das lesões podem ser encontradas calcificações nodulares ou de massa expansiva.

- l) *Doença de Paget*: caracterizada por lesão eczematóide no mamilo em consequência da infiltração da epiderme por carcinoma ductal infiltrante. O eczema mamilar pode produzir infiltração sem massa palpável, associado com microcalcificações esparsas ou agrupadas. Às vezes associa-se com secreção papilar intermitente. A mamografia revela finas calcificações que podem estar disseminadas no corpo glandular, ao longo do trajeto dos ductos maiores associados às alterações mamilares.
- m) *Carcinoma Inflamatório*: apresenta rubor e edema e tem rápida evolução. O mamilo aparece espessado e há, em geral, secreção mamilar sanguinolenta. A imagem radiográfica mostra aumento de volume da mama, espessamento generalizado da pele (generalizado ou difuso), infiltração do tecido celular subcutâneo e ocupação do espaço retroareolar.
- n) *Sarcoma*: são grupos heterogêneos de tumores mesenquimais, apresentando-se tipicamente como uma massa expansiva sólida que cresce invadindo o tecido adiposo e infiltrando o tecido glandular. A imagem radiográfica revela um aumento da densidade relativa de todo o corpo glandular, em razão da infiltração.

2.4.3 Microcalcificações

Calcificações são depósitos de sais de cálcio nos tecidos mamários que podem ter se desenvolvido devido a processos inflamatórios, alterações degenerativas cíclicas, processos tóxicos metabólicos, traumatismos ou mesmo como resultados de processos secretores ativos de células tumorais, além de produtos de mineralização de debris necróticos [27].

Em 1951, Leborgne demonstrou que em 30% dos cânceres de mama pode-se encontrar microcalcificações. Em 1964, treze anos depois, Egan pensava que essas microcalcificações típicas eram tão patognomônicas de câncer, que, ao se fazer a biópsia, se o resultado fosse de lesão benigna, havia grande suspeita de o cirurgião ter selecionado o tecido errado para biópsia ou de ter havido erro do patologista.

Entretanto, dezesseis anos depois, Egan reconsiderou seu achado, dizendo que aqueles sinais não eram tão específicos e que todas as mamografias que mostrassem microcalcificações requereriam estudo histopatológico [28].

Deve-se ressaltar, também, como mostram vários autores (Hermann, 1988; Dershaw, 1989; Ikeda, 1989; Stomper, 1989), que o diagnóstico do *carcinoma ductal in situ* (CDIS) é feito baseado principalmente na presença de microcalcificações.

Bem antes de se difundir o uso da mamografia, o CDIS correspondia a apenas 5% dos novos cânceres detectados, enquanto hoje corresponde a aproximadamente 20% desses tumores e mais de 70% são detectados apenas devido às microcalcificações mamograficamente visíveis. [29]

As calcificações benignas e malignas das mamas tomam a forma das estruturas onde estão contidas e podem ser de origem lobular ou ductal. As patogêneses são diferentes, sendo que as benignas podem ser devidas a secreções calcificadas, hialinização e infarto de tecidos ou ter origem traumática ou inflamatória, enquanto as outras resultam de focos celulares necróticos, de "débris" celulares intraluminais ductais ou acinares ou, ainda, de secreções ou muco nos espaços cribriformes do CDIS não-comedo, ou seja, carcinomas de crescimento lento, que não tendem a se espalhar fora da área onde se originaram [28, 30].

Em Azevedo [27], as calcificações são classificadas em:

2.4.3.1 Calcificações em processos benignos:

- a) *As com centro radioluciente*, encontradas nas paredes de cistos simples, de cistos oleosos (na necrose gordurosa), nas glândulas sebáceas da pele e nas paredes de siliconomas (corpos estranhos representados por glóbulos de silicone líquido industrial injetado nas mamas).
- b) *As vasculares*, que são típicas, representadas por duas linhas paralelas nas paredes das artérias. Neste caso, às vezes há dificuldade em diagnosticá-las quando o processo é inicial. São comuns em pacientes entre 65 e 80 anos e pacientes hipertensos e diabéticos.
- c) *"Em pipoca"*, que são típicas dos *fibroadenomas calcificados*. Neste caso, somente as calcificações pequenas e numerosas, pouco comuns nos fibroadenomas, podem dificultar o diagnóstico diferencial com lesão maligna.
- d) *"Em palito"*, com ou sem centro radioluciente, representando secreções calcificadas no interior ou nas paredes de ductos ectasiados, geralmente orientadas em direção ao mamilo. Como essas calcificações são geralmente grandes, mesmo quando estão localizadas em apenas um segmento da mama, raramente deixam dúvida quanto à sua natureza.
- e) *Pós-cirúrgicas*, que são geralmente grandes e bizarras e dificilmente podem ser confundidas com lesões malignas, especialmente sabendo-se sobre a cirurgia.

- f) *Leite de cálcio em pequenos cistos*, que devem-se ao fato de que, na lactação, o material cálcico do leite fica retido na parede dos grandes ductos e extravasa intersticialmente, causando uma **mastite química**. Elas aparecem como puntiformes na incidência crânio-caudal e semilunares ou curvilíneas na incidência médio-lateral a 90° em relação à primeira. Isto se deve ao fato de que, em perfil, o leite de cálcio no interior do pequeno cisto forma um nível cálcio/líquido e se deposita em sua parede inferior, tomando sua forma. Quando são localizadas e unilaterais, podem ser confundidas, na incidência crânio-caudal, com lesão maligna, mas geralmente a incidência médio-lateral ajuda a análise correta.
- g) *Hiperplasia microcística difusa*, aparecendo predominantemente puntiformes, difusas, bilaterais, representando secreções calcificadas no interior de microcistos.
- h) *Microcalcificações involutivas*, que se encontram em lóbulos atróficos, são finas, puntiformes, monomórficas, difusas ou em agrupamentos.

2.4.3.2 Calcificações em processos malignos

- a) *Carcinoma ductal “in situ”*, ocorrem em conseqüência de secreções estagnadas em fendas nas pequenas dilatações dos ácinos. São calcificações puntiformes (em forma de pequenos pontos, com dimensão menor que 0,3mm) e, às vezes calcificações bizarras e irregulares.
- b) *Carcinoma lobular in situ*, ocorrem em cerca de 10% dos casos e podem apresentar-se de várias formas: agrupadas, esparsas (polimorfas) ou em forma bizarra, semelhante ao carcinoma ductal “in situ”.
- c) *Carcinoma intraductal ou comedocarcinoma*, as células epiteliais da mucosa do ducto que preenchem o lume do ducto central necrosam e calcificam-se. As calcificações dos carcinomas ductais tipo cirrosos são polimorfas e resultam de necroses tumorais, podendo situar-se adjacente ou periféricamente à massa, ou no centro desta. No carcinoma ductal infiltrante tipo tubular, as calcificações encontradas são grosseiras.
- d) *Carcinoma mucinoso*, raramente apresentam calcificações que, quando ocorrem, são grosseiras e bizarras.
- e) *Doença de Paget*, apresentam calcificações de diversas formas, podendo ou não estar associadas à massa tumoral. Às vezes distribuem-se por todo o corpo glandular, apresentando polimorfismo.

As microcalcificações agrupadas, com formas variadas em pacientes de alto risco, são as que engrossam as estatísticas de biópsias de microcalcificações por lesões. Às vezes, um mesmo tipo de microcalcificação, em determinada paciente, sem história pessoal ou familiar de câncer, merece apenas acompanhamento, enquanto em outra, em paciente de alto risco ou com cancerofobia, merece biópsia.

Outras vezes, calcificações maiores que 1,0 mm, deixando de ser microcalcificações, mas estando no interior de hiperdensidades irregulares e de limites imprecisos, necessitam estudo anatomopatológico, pois podem estar associadas a uma neoplasia maligna.

3. Análise de Componentes Independentes

3.1 - Introdução

3.1.1 Motivação

O objetivo da técnica é a análise, ou separação, de fontes estatisticamente independentes a partir de um determinado modelo de mistura das fontes originais.

O exemplo clássico de ICA é a separação da fala, ou, como é popularmente conhecido, *cocktail-party*. Suponha uma sala com três pessoas conversando. Três microfones são posicionados em locais diferentes da sala. Cada um grava uma amostra da conversa ao longo do tempo, que podem ser expressas como $x_1(t)$, $x_2(t)$ e $x_3(t)$. Essas amostras são, portanto, misturas das falas de cada pessoa, que serão denotadas como $s_1(t)$, $s_2(t)$ e $s_3(t)$.

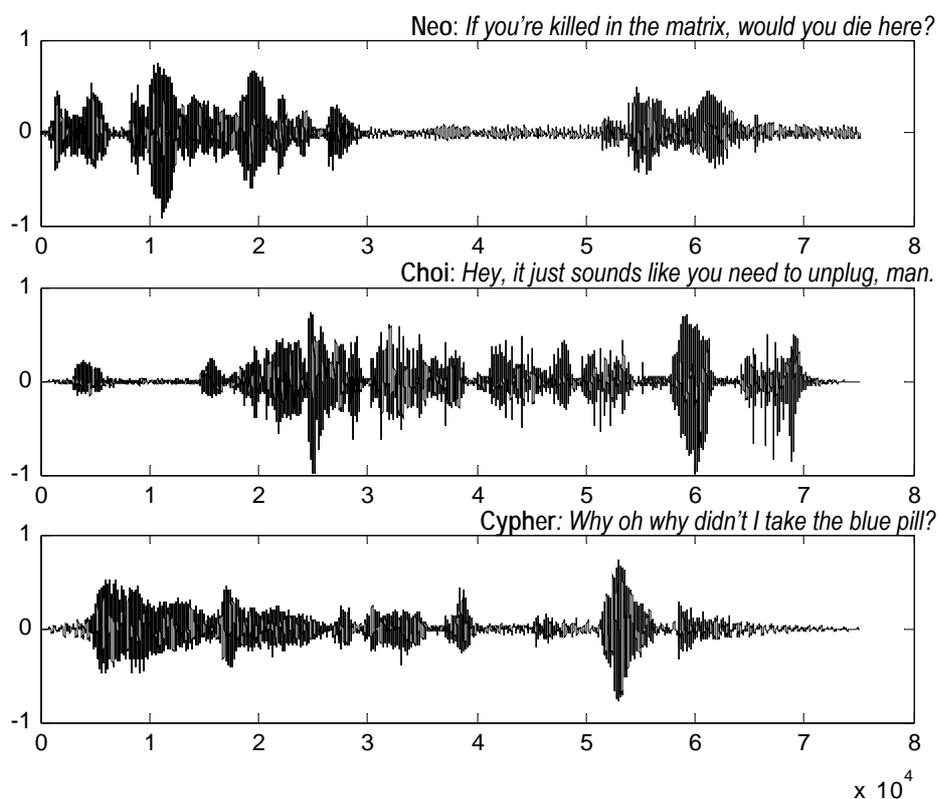


Figura 3.1 – Sinais de áudio originais.

Matematicamente:

$$x_1(t) = a_{11} \cdot s_1(t) + a_{12} \cdot s_2(t) + a_{13} \cdot s_3(t) \quad (3.1)$$

$$x_2(t) = a_{21} \cdot s_1(t) + a_{22} \cdot s_2(t) + a_{23} \cdot s_3(t) \quad (3.2)$$

$$x_3(t) = a_{31} \cdot s_1(t) + a_{32} \cdot s_2(t) + a_{33} \cdot s_3(t) \quad (3.3)$$

Os pesos a_{ij} dependem do ganho do microfone e da distância destes ao interlocutor. Admite-se, por ora, que não haja atrasos ou outros fatores adicionais no modelo de mistura. Se os parâmetros a_{ij} fossem conhecidos, o problema se resumiria na solução de uma equação linear. Porém, nem as fontes, nem os pesos da mistura são conhecidos.

Observe a Figura 3.1. Os arquivos são falas do filme Matrix, gravadas em “.wav”. A Figura 3.2 traz três misturas das falas originais.

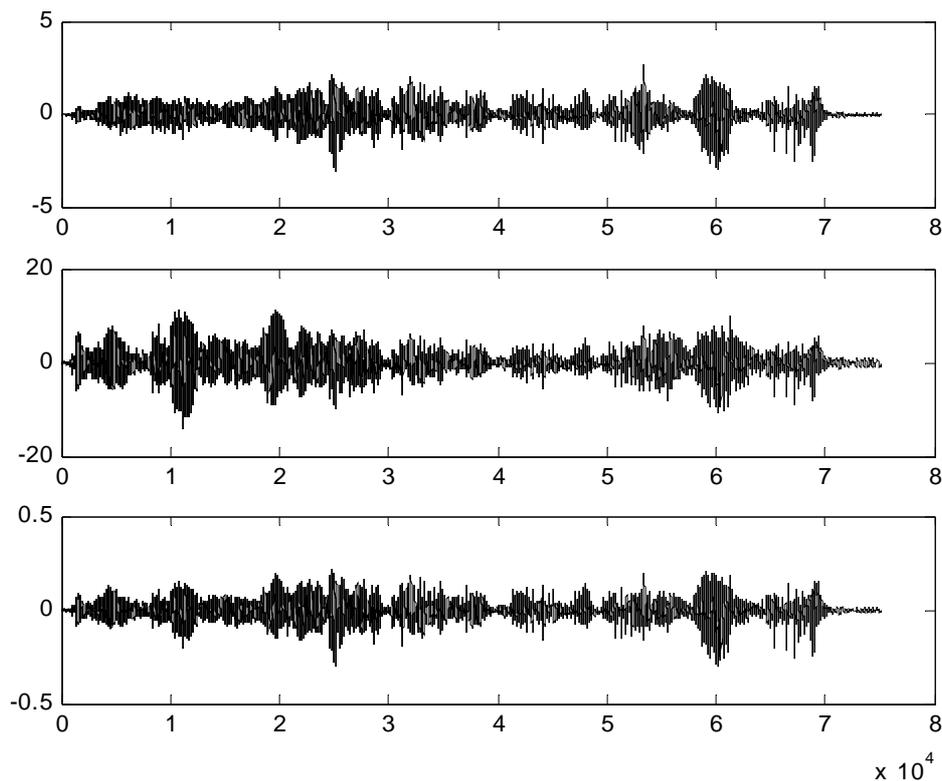


Figura 3.2 – Sinais observados, obtidos a partir da mistura dos sinais da Figura 3.1.

As técnicas de ICA propõem separar as três fontes partindo do simples pressuposto de que elas são estatisticamente independentes entre si.

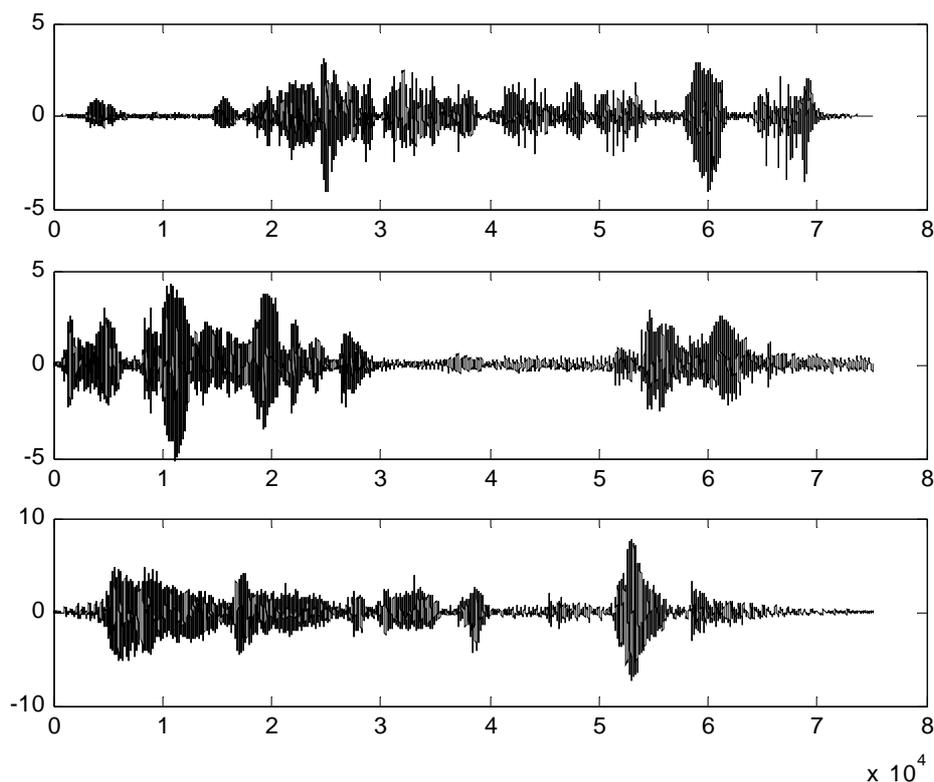


Figura 3.3 – Sinais originais estimados a partir da mistura observada na Figura 3.2.

A Figura 3.3 revela as componentes estimadas. Observe que os sinais 2, 1 e 3 da Figura 3.2 guardam correspondência com os sinais 1,2 3 da Figura 3.1, respectivamente. A ordem das componentes independentes não pode ser estimada. A energia dos sinais também não é a mesma do original, mas o sinal audível é o mesmo.

A técnica de ICA tem uma vasta gama de aplicações. Pode-se citar, como exemplos, dados de econometria, sinais de SONAR, telecomunicações ou sinais da atividade cerebral dadas por um eletro-encefalograma (EEG). Por exemplo, o EEG consiste em registros de potenciais elétricos em locais diferentes do escalpo cerebral. Esses potenciais são gerados a partir de componentes fundamentais da atividade cerebral. Essa situação é similar ao problema de separação da fala, no sentido em que se deseja separar as componentes originais da atividade cerebral, onde apenas são conhecidas as misturas dessas componentes. A ICA pode revelar informações interessantes das misturas ao dar acesso às suas componentes independentes.

A ICA também pode ser usada para extração de parâmetros. Um problema fundamental no processamento de sinais digitais é encontrar uma representação

adequada de uma imagem, sinal de áudio ou outro tipo qualquer de dados, com objetivos diversos como compressão ou retirada de ruído. A representação dos dados é usualmente feita a partir de transformações lineares, como Fourier, DCT, Haar, que apresentam vantagens particulares para cada aplicação. No entanto, teria maior utilidade estimar a transformação linear a partir dos próprios dados, de forma que, em cada caso, a transformada seria idealmente adaptada aos dados sendo processados.

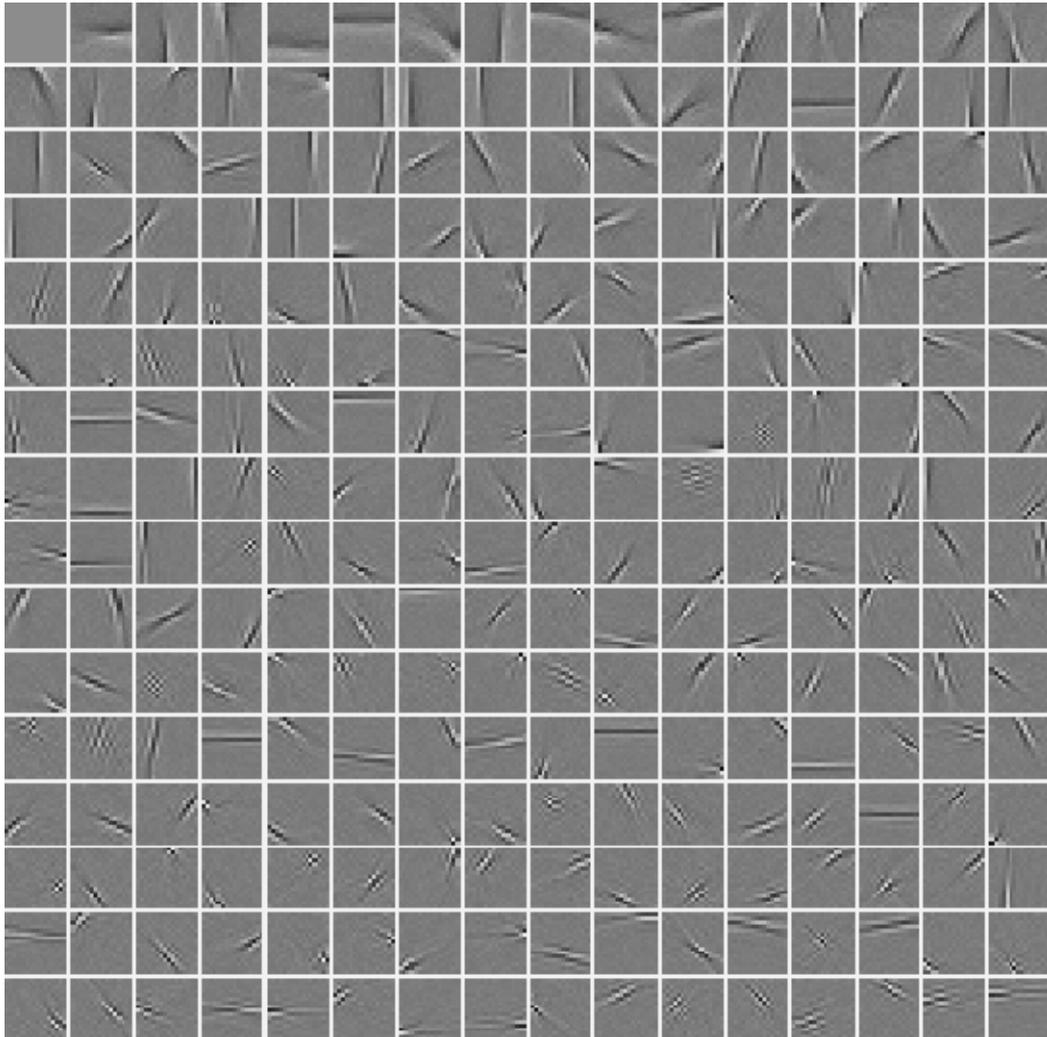


Figura 3.4 – Imagens-base obtidas a partir da separação de imagens naturais em componentes independentes.

Fonte: Hyvärinen A., *Independent Component Analysis* [31]

A Figura 3.4 mostra funções-base obtidas a partir de imagens naturais. Cada imagem do conjunto de treinamento seria uma superposição dessas imagens-base, de forma que os coeficientes da superposição são independentes entre si. Essa é a abordagem que será tratada no presente trabalho. Uma imagem radiográfica da mama

será decomposta em imagens-base, supostamente independentes entre si, a partir de um conjunto de imagens selecionado para o treinamento. Pretende-se mostrar que os coeficientes dessa mistura podem ser um indicativo de anormalidade presente no tecido, e que, portanto, a decomposição em componentes independentes pode ser um método eficaz de diagnóstico de lesões da mama.

3.1.2 Modelo

O modelo de mistura consiste em n dados observáveis x_1, \dots, x_N que resultam de uma combinação linear de N variáveis aleatórias s_1, \dots, s_N estatisticamente independentes entre si:

$$x_i = a_{i1} \cdot s_1 + a_{i2} \cdot s_2 + \dots + a_{iN} s_N, \forall i = 1, \dots, N \quad (3.4)$$

No modelo, apenas as variáveis aleatórias x_i são conhecidas, e a partir delas serão estimados os coeficientes da mistura a_{ij} e as componentes independentes s_j . Em notação matricial [31]:

$$\mathbf{x} = \mathbf{A} \cdot \mathbf{s} \quad (3.5)$$

As componentes independentes são variáveis latentes, o que significa que não são diretamente observadas. Os coeficientes a_{ij} também são desconhecidos. Apenas as variáveis aleatórias x_i são conhecidas e tanto os coeficientes da mistura a_{ij} quanto as componentes independentes s_i devem ser estimadas a partir de x_i .

O modelo de ICA adotado é generativo, o que significa que ele descreve como os dados observados são gerados a partir de um processo de mistura das componentes s_i .

Há casos em que o modelo linear básico é muito simples para descrever adequadamente os dados observados x_i . Nesses casos assume-se que \mathbf{x} é uma função genérica e possivelmente não linear das fontes \mathbf{s} :

$$\mathbf{x} = f(\mathbf{s}) \quad (3.6)$$

Partindo do pressuposto que f é uma função que depende das n componentes de \mathbf{x} , ou seja, existem n sinais observados para n fontes, encontram-se abordagens que estimam as fontes diretamente usando a transformação inversa, ou seja, consistem em buscar um mapeamento $\mathbf{y} = h(\mathbf{x})$ que forneça componentes estatisticamente independentes entre si. Esse modelo é conhecido como *transformação de sinal*, em

contraposição ao modelo generativo, que estima tanto o mapeamento da mistura $f(\cdot)$ quanto as componentes independentes s_i .

A ICA está intimamente relacionada com o método de separação cega de fontes (BSS – *Blind Source Separation*). O método é chamado de cego, pois muito pouco, ou mesmo nada, é conhecido sobre a matriz de mistura, e são feitas poucas considerações acerca das fontes a serem estimadas. ICA é o método mais aplicado para separação cega de fontes.

No caso acima, em que o modelo é não linear, encontrar componentes independentes não é suficiente para resolver o problema de separação de fontes, uma vez que raramente há uma solução única [32]. Isso porque se \mathbf{x} e \mathbf{y} são independentes entre si, qualquer função $f(\mathbf{x})$ e $g(\mathbf{y})$ também serão independentes. Para encontrar as fontes originais é preciso, nesse caso, conhecer algumas informações adicionais de antemão para restringir a solução. Isso vale também quando se sabe que há mais fontes a serem estimadas do que dados observados. [33]

Porém, a abordagem adotada assumirá sempre que o modelo é linear, generativo, e com bases completas, ou seja, há n fontes e n dados observados. Esse modelo básico é também chamado de instantâneo, uma vez que supõe que não haja atrasos na mistura. Ainda, as deduções são feitas supondo sinais em \Re .

A partir de aqui será considerado que as variáveis da mistura têm média igual a zero, sem perda de generalidade. Caso isso não seja verdade na prática, subentende-se que se estará trabalhando com uma variável \mathbf{x} , proveniente da mistura original \mathbf{x}' , cuja média foi retirada:

$$\mathbf{x} = \mathbf{x}' - E\{\mathbf{x}'\} \quad (3.7)$$

As componentes independentes, conseqüentemente, também terão média zero:

$$E\{\mathbf{s}\} = \mathbf{A}^{-1} \cdot E\{\mathbf{x}\} = 0 \quad (3.8)$$

Se for preciso recuperar a média das componentes independentes, basta, portanto, adicionar $\bar{\mathbf{s}} = \mathbf{A}^{-1} \cdot E\{\mathbf{x}'\}$ às componentes encontradas.

A notação adotada assume que todos os vetores são coluna, ou seja, são $N \times 1$. Os vetores são denotados por minúsculas em negrito e as matrizes, por maiúsculas em negrito. Dessa forma, \mathbf{x} é um vetor-coluna, que corresponde a uma observação de N

variáveis aleatórias correspondentes aos N elementos de \mathbf{x} . Em alguns casos, por motivos didáticos, será preferido expressar a mistura em forma matricial. Nesses casos, os dados da mistura serão expressos em uma matriz $\mathbf{X}_{N \times P}$, com $N \times P$ elementos, cujas N linhas correspondem às N variáveis e as colunas a P observações.

3.1.3 Restrições

Adotado o modelo linear, generativo e de bases completas, os algoritmos propostos pretendem encontrar as fontes que originaram um conjunto de dados, apenas a partir da mistura observada, sem conhecimento prévio da matriz de mistura ou das fontes originais. Algumas poucas considerações, porém, precisam ser feitas para garantir a convergência desejada. Todos os cálculos e demonstrações consideram que os sinais são reais.

3.1.3.1 As componentes independentes são estatisticamente independentes entre si.

Por definição, o método proposto para separação de fontes pressupõe que as componentes que originam a mistura sejam estatisticamente independentes entre si.

Variáveis aleatórias são ditas independentes se a informação contida em uma determinada variável não fornece nenhuma informação acerca do valor das demais.

A Figura 3.5 a) revela a distribuição conjunta de duas componentes independentes s_1 e s_2 com distribuições uniformes. A distribuição conjunta s_1 e s_2 é uniforme em um quadrado.

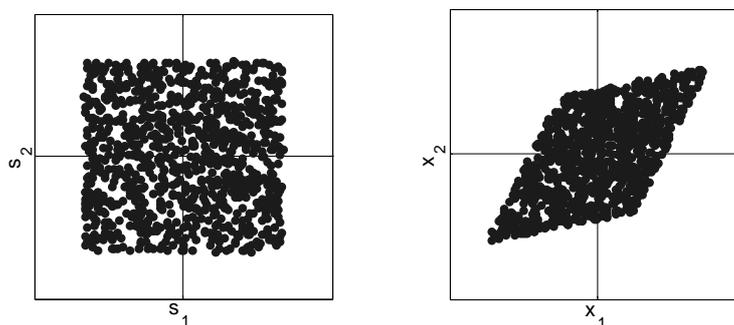


Figura 3.5: a) Distribuição conjunta das componentes independentes s_1 e s_2 com distribuições uniformes. b) Distribuição da mistura das componentes s_1 e s_2

Observe na Figura 3.5 a) que não há como inferir nada sobre a variável s_2 a partir de s_1 . A Figura 3.5 b) mostra a mistura das componentes s_1 e s_2 a partir da matriz

$A = \begin{pmatrix} 5 & 10 \\ 10 & 2 \end{pmatrix}$. Observe que, no caso, se x_1 assume seu valor máximo, praticamente pode-se determinar o valor de x_2 , indicando dependência entre as variáveis.

Matematicamente, a independência estatística é definida em termos das densidades de probabilidade, ou seja, a densidade conjunta pode ser fatorada em um produto das densidades marginais:

$$p(y_1, y_2, \dots, y_n) = p_1(y_1) \cdot p_2(y_2) \cdot \dots \cdot p_n(y_n) \quad (3.9)$$

Para exemplificar a propriedade de independência estatística, considere a densidade de probabilidade conjunta entre a variável aleatória bidimensional $\mathbf{x} = [x_1 \ x_2]^T$ e a variável aleatória y :

$$p_{\mathbf{x},y}(\mathbf{x}, y) = \begin{cases} (x_1 + 3x_2) \cdot y & x_1, x_2 \in [0,1], \text{ e } y \in [0,1] \\ 0, & \text{caso contrário} \end{cases}$$

Verifica-se que o produto das densidades marginais de \mathbf{x} e y é igual à densidade de probabilidade conjunta entre \mathbf{x} e y .

$$p_{\mathbf{x}}(\mathbf{x}) = \int_0^1 (x_1 + 3x_2) \cdot y \cdot dy = \frac{1}{2}x_1 + \frac{3}{2}x_2$$

$$\text{e } p_y(y) = \int_0^1 \int_0^1 (x_1 + 3x_2) \cdot y \cdot dx_1 \cdot dx_2 = 2y$$

Conclui-se então que, como $p_{\mathbf{x}}(\mathbf{x}) \cdot p_y(y) = p_{\mathbf{x},y}(\mathbf{x}, y)$, as variáveis \mathbf{x} e y são estatisticamente independentes.

No entanto, como $p_{\mathbf{x}}(\mathbf{x}) = p_{x_1, x_2}(x_1, x_2) = \frac{1}{2}x_1 + \frac{3}{2}x_2$, então:

$$p_{x_1}(x_1) = \int_0^1 \left(\frac{1}{2}x_1 + \frac{3}{2}x_2\right) \cdot dx_2 = \frac{1}{2}x_1 + \frac{3}{4}$$

e

$$p_{x_2}(x_2) = \int_0^1 \left(\frac{1}{2}x_1 + \frac{3}{2}x_2\right) \cdot dx_1 = 1 + \frac{3}{2}x_2,$$

logo, $p_{x_1}(x_1) \cdot p_{x_2}(x_2) \neq p_{x_1, x_2}(x_1, x_2)$, e portanto, as variáveis x_1 e x_2 não são independentes entre si.

3.1.3.1.1 Descorrelação, branqueamento e independência

Duas variáveis aleatórias são ditas descorrelacionadas se a covariância entre elas for zero:

$$\text{cov}(y_1, y_2) = E\{y_1 \cdot y_2\} - E\{y_1\} \cdot E\{y_2\} = 0 \quad (3.10)$$

Se as variáveis têm média zero, então a covariância é igual à correlação:

$$\text{corr}(y_1, y_2) = E\{y_1 \cdot y_2\} = 0 \quad (3.11)$$

Se duas variáveis aleatórias são independentes, então elas são descorrelacionadas, mas o inverso não é sempre verdade. Por exemplo, suponha duas variáveis aleatórias x e y tais que x é uniforme entre -1 e 1 e $y=x^2$. Portanto, x e y são variáveis estatisticamente dependentes, o que se confirma por:

$$E\{x^2 \cdot y^2\} = E\{x^6\} = \frac{1}{7 \cdot 2^6} \neq \frac{1}{15 \cdot 2^6} = E\{x^2\} \cdot E\{x^4\} = E\{x^2\} \cdot E\{y^2\}$$

Porém, tem-se que:

$$\text{cov}(x, y) = E\{x^3\} - E\{x\} \cdot E\{x^2\} = 0,$$

o que demonstra que x e y são descorrelacionadas, apesar de dependentes entre si.

Se duas variáveis aleatórias são descorrelacionadas e têm variâncias iguais a 1, então elas são chamadas de brancas. Isso incorre em que a matriz de covariância é igual à identidade.

$$E\{\mathbf{z} \cdot \mathbf{z}^T\} = \mathbf{I} \quad (3.12)$$

Variáveis brancas podem ser obtidas a partir de uma transformação linear dos dados:

$$\mathbf{z} = \mathbf{V} \cdot \mathbf{x} \quad (3.13)$$

O branqueamento do vetor aleatório \mathbf{x} pode ser feito pela matriz \mathbf{V} tal que $\mathbf{V} = \mathbf{E} \cdot \mathbf{D}^{-1/2} \cdot (\mathbf{E}^*)^T$, onde \mathbf{E} é a matriz ortogonal cujas colunas são os autovetores

de $E\{\mathbf{x} \cdot \mathbf{x}^T\}$ e \mathbf{D} é a matriz cuja diagonal são os autovalores correspondentes ($\mathbf{D} = \text{diag}(d_1, \dots, d_n)$):

$$E\{\mathbf{x} \cdot \mathbf{x}^T\} = \mathbf{E} \cdot \mathbf{D} \cdot \mathbf{E}^T \quad (3.14)$$

O branqueamento transforma a matriz \mathbf{A} em uma nova matriz $\tilde{\mathbf{A}}$. Logo,

$$\mathbf{z} = \mathbf{V} \cdot \mathbf{A} \cdot \mathbf{s} = \tilde{\mathbf{A}} \cdot \mathbf{s} \quad (3.15)$$

Considere uma transformação ortogonal de \mathbf{z} :

$$\mathbf{y} = \mathbf{U} \cdot \mathbf{z} \quad (3.16)$$

Devido à ortogonalidade de \mathbf{U} , tem-se que:

$$E\{\mathbf{y} \cdot \mathbf{y}^T\} = E\{\mathbf{U} \cdot \mathbf{z} \cdot \mathbf{z}^T \mathbf{U}^T\} = \mathbf{U} \cdot \mathbf{I} \cdot \mathbf{U}^T = \mathbf{I} \quad (3.17)$$

Ou seja, \mathbf{y} também é branco. O branqueamento não é suficiente para que se identifique se as componentes independentes são dadas por \mathbf{y} ou \mathbf{z} . O branqueamento fornece uma transformação ortogonal de \mathbf{s} .

Apesar de não resolver completamente o problema de estimação de fontes, o branqueamento é um pré-processamento útil para o ICA.

Como:

$$E\{\mathbf{z} \cdot \mathbf{z}^T\} = \tilde{\mathbf{A}} \cdot E\{\mathbf{s} \cdot \mathbf{s}^T\} \cdot \tilde{\mathbf{A}}^T = \tilde{\mathbf{A}} \cdot \tilde{\mathbf{A}}^T = \mathbf{I}, \quad (3.18)$$

pode-se concluir que a matriz de mistura que se está procurando é ortogonal.

Ao invés de se estimar n^2 parâmetros da matriz \mathbf{A} , é necessário estimar uma matriz de mistura ortogonal $\tilde{\mathbf{A}}$, que contém apenas $n \cdot (n-1)/2$ graus de liberdade. Em duas dimensões, a matriz de mistura representa uma rotação e apenas o ângulo precisa ser determinado.

A Figura 3.6 mostra a variável \mathbf{x} da Figura 3.5 b) branqueada. Observe que a distribuição de \mathbf{z} é apenas uma rotação da distribuição conjunta de s_1 e s_2 .

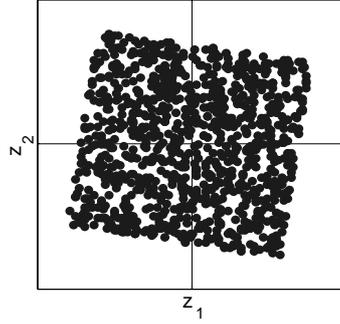


Figura 3.6 - Distribuição conjunta da mistura de componentes distribuídas uniformemente após o branqueamento.

3.1.3.2 As componentes independentes devem apresentar distribuições não gaussianas

A probabilidade conjunta de duas variáveis gaussianas estatisticamente independentes s_1 e s_2 é dada por [34]:

$$p(s_1, s_2) = \frac{1}{2\pi} \exp\left(-\frac{s_1^2 + s_2^2}{2}\right) = \frac{1}{2\pi} \exp\left(-\frac{\|\mathbf{s}\|^2}{2}\right) \quad (3.19)$$

Assumindo que a matriz \mathbf{A} é ortogonal, tem-se que $(\mathbf{A}^*)^T = \mathbf{A}^{-1}$. Essa suposição não invalida a dedução, uma vez que, como exposto na seção 3.1.3, o branqueamento pode ser realizado previamente sem prejuízos ao método. Pode-se, então, expressar a *pdf* das misturas x_1 e x_2 por:

$$p(x_1, x_2) = \frac{1}{2\pi} \exp\left(-\frac{\|\mathbf{A}^T \cdot \mathbf{s}\|^2}{2} |\det \mathbf{A}^T|\right) \quad (3.20)$$

Como \mathbf{A} é ortogonal, tem-se também que $\|\mathbf{A}^T \cdot \mathbf{s}\|^2 = \|\mathbf{s}\|^2$ e que $|\det \mathbf{A}| = 1$.

Portanto,

$$p(x_1, x_2) = \frac{1}{2\pi} \exp\left(-\frac{\|\mathbf{s}\|^2}{2}\right) \quad (3.21)$$

Logo, para variáveis gaussianas, uma matriz de mistura ortogonal não altera a *pdf* conjunta das fontes, ou melhor, a matriz \mathbf{A} sequer aparece na *pdf* da mistura. Isso ocorre porque variáveis gaussianas descorrelacionadas são, necessariamente, independentes. Portanto supor que as componentes são independentes não é nada além de branqueá-las.

Graficamente, pode-se interpretar que distribuição conjunta de s_1 e s_2 é isotrópica, não sendo alterada quando submetida a uma rotação. Portanto, a *pdf* não contém nenhuma informação sobre a matriz \mathbf{A} , o que impede de estimá-la somente a partir das distribuições de x_1 e x_2 .

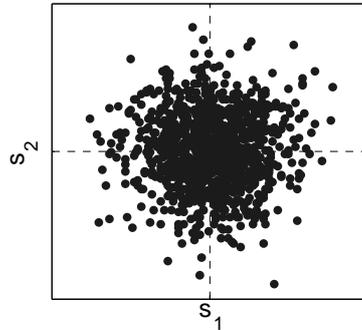


Figura 3.7 – Distribuição conjunta de duas variáveis gaussianas.

Ao tentar-se encontrar as componentes independentes a partir de uma mistura que contenha algumas fontes gaussianas e outras não, as componentes não-gaussianas serão estimadas corretamente. No entanto, não será possível distinguir as gaussianas entre si; estas permanecerão como uma combinação linear qualquer das componentes gaussianas originais.

Apesar dessa restrição, não é necessário ter nenhum outro conhecimento prévio sobre a distribuição das componentes independentes para que se possa estimá-las.

3.1.3.3 A matriz de mistura é quadrada

Os métodos de separação cega de fontes que serão apresentados supõem que a matriz de mistura \mathbf{A} é quadrada, ou seja, que o número de componentes independentes que se deseja estimar é o mesmo que o número de dados observáveis.

No caso de haver mais misturas que fontes é possível omitir misturas redundantes, recaindo-se no caso anterior.

No entanto, se o número de fontes for maior que o número de misturas, encontrar a matriz de mistura não resolve o problema de estimação das componentes independentes, que se torna consideravelmente mais complexo. Os métodos de estimação apresentados na literatura para esse caso ainda são bastante incipientes. [35]

3.1.4 Ambigüidades

3.1.4.1 Não é possível determinar a variância das componentes independentes

Como ambos \mathbf{s} e \mathbf{A} são desconhecidos, pode-se multiplicar um escalar em qualquer das fontes s_i e cancelá-lo dividindo a coluna correspondente \mathbf{a}_i pelo mesmo valor α_i :

$$\mathbf{x} = \sum_i \left(\frac{1}{\alpha_i} \mathbf{a}_i \right) \cdot (s_i \cdot \alpha_i) \quad (3.22)$$

Por conveniência, assume-se sempre que as componentes independentes têm variância unitária, ou seja, $E\{s_i^2\} = 1$. Essa determinação ainda deixa uma ambigüidade no sinal, pois sempre se pode multiplicar a componente por -1 sem mudar o sinal. Essa ambigüidade é irrelevante na maioria das aplicações.

3.1.4.2 Não é possível determinar a ordem das componentes independentes

Seja \mathbf{P} uma matriz de permutação qualquer. Se multiplica-se o vetor-fontes \mathbf{s} pela matriz \mathbf{P} , obter-se-ão as mesmas variáveis s_i , mas em outra ordem. A matriz $\mathbf{A} \cdot \mathbf{P}^{-1}$ é uma outra matriz de mistura que será encontrada pelo algoritmo de ICA. Em outras palavras, se estaria incluindo uma identidade na equação (3.5), resultando em:

$$\mathbf{x} = \mathbf{A} \cdot \mathbf{P}^{-1} \cdot \mathbf{P} \cdot \mathbf{s} \quad (3.23)$$

3.2 - ICA por Maximização da Não Gaussianidade

3.2.1 Não- Gaussianidade

Considere um vetor de dados observáveis \mathbf{x} distribuído de acordo com o modelo de ICA, ou seja, composto por uma mistura de componentes independentes:

$$\mathbf{x} = \mathbf{A} \cdot \mathbf{s} \quad (3.24)$$

Nesse caso, as componentes independentes podem ser completamente determinadas pela matriz de mistura \mathbf{A} :

$$\mathbf{s} = \mathbf{A}^{-1} \cdot \mathbf{x} \quad (3.25)$$

Isso significa que é possível estimar cada componente independente a partir de uma combinação linear dos x_i .

Suponha, então, uma combinação linear qualquer de x_i dada por $y = \mathbf{b}^T \cdot \mathbf{x}$.
 Como $\mathbf{x} = \mathbf{A} \cdot \mathbf{s}$, pode-se escrever $y = \mathbf{b}^T \cdot \mathbf{x} = \mathbf{b}^T \cdot \mathbf{A} \cdot \mathbf{s}$:

$$y = \underbrace{[b_1 \quad b_2 \quad \dots \quad b_n]}_{\mathbf{b}^T} \cdot \underbrace{\begin{bmatrix} a_{11} & a_{12} & & a_{1n} \\ a_{21} & \ddots & & \\ \vdots & & \ddots & \vdots \\ a_{n1} & \dots & & a_{nn} \end{bmatrix}}_{\mathbf{A}} \cdot \underbrace{\begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{bmatrix}}_{\mathbf{s}} \quad (3.26)$$

Se \mathbf{b}^T for uma linha da inversa da matriz \mathbf{A} , então $\mathbf{b}^T \cdot \mathbf{A}$ será um vetor-linha com um dos elementos igual a 1 e os demais iguais a zero, e y será uma das fontes originais.

Lembre-se, porém, que apenas o vetor de mistura \mathbf{x} é conhecido, e, portanto o vetor \mathbf{b} não pode ser determinado exatamente. Precisa-se, pois, encontrar um estimador que forneça uma boa aproximação de \mathbf{b} .

Pelo teorema do valor central (Apêndice A), a soma de variáveis aleatórias independentes tende a uma distribuição gaussiana, supondo que as n componentes tenham distribuições idênticas. Assim, a soma de duas variáveis aleatórias é mais gaussiana que as variáveis originais. Partindo desse pressuposto, y será mais gaussiano do que qualquer componente independente s_i e será o menos gaussiano possível quando, de fato, se igualar a uma das s_i .

Isso sugere que deve-se variar \mathbf{b} e observar a distribuição de y . Ao se encontrar um vetor \mathbf{b} que maximize a não-gaussianidade de y , ter-se-á encontrado uma componente independente.

Na realidade, o espaço de otimização do vetor \mathbf{b} tem $2n$ máximos locais, dois para cada componente independente, correspondentes a s_i e $-s_i$, visto que o sinal da componente independente não pode ser identificado univocamente.

Resta, então, apenas encontrar uma medida para gaussianidade da distribuição de y .

3.2.2 Kurtosis como medida de não-gaussianidade

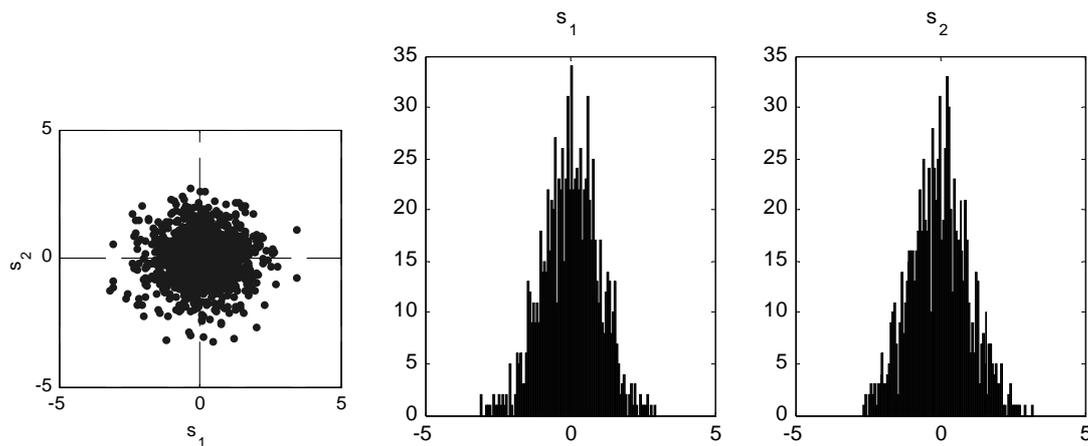
Kurtosis é o cumulante de quarta ordem de uma variável aleatória. A *kurtosis* de uma variável aleatória qualquer y de média zero é definida como:

$$kurt(y) = E\{y^4\} - 3 \cdot (E\{y^2\})^2 \quad (3.27)$$

Supondo que y tem variância unitária (busca-se sempre componentes independentes com variância unitária), a *kurtosis* pode ser simplificada para $kurt(y) = E\{y^4\} - 3$, que é uma versão normalizada do momento de 4ª ordem.

Se y é uma variável gaussiana, o 4º momento é igual a $3 \cdot (E\{y^2\})^2$. Portanto, a *kurtosis* é zero para variáveis gaussianas.

A Figura 3.8 revela uma distribuição conjunta de 2 variáveis gaussianas independentes. À medida que o número de pontos da distribuição cresce, a incerteza computacional diminui e a *kurtosis* tende a zero, como pode ser observado.



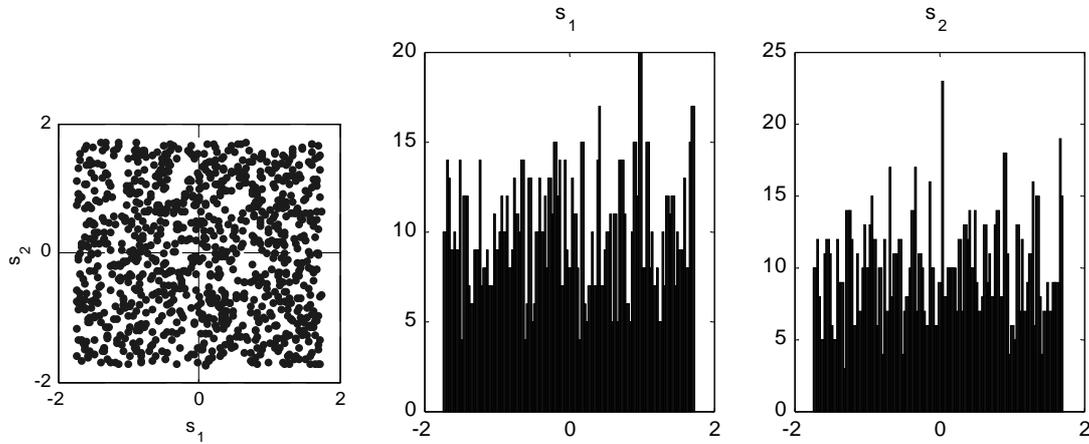
**Figura 3.8 – a) Distribuição conjunta de 2 variáveis gaussianas independentes
b),c) Histogramas de s_1 e s_2 , respectivamente.**

$$p(s_i) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{s_i^2}{2}\right) \quad i = 1,2$$

$$\begin{aligned} kurt(s_1) &= +0.013 \\ kurt(s_2) &= -0.012 \end{aligned}$$

Para a grande maioria das variáveis não-gaussianas, embora não todas, a *kurtosis* é diferente de zero.

Distribuições cuja *kurtosis* é negativa são chamadas subgaussianas, ou *platokúrticas*. Essas variáveis apresentam uma *pdf* mais distribuída pelo domínio da variável.

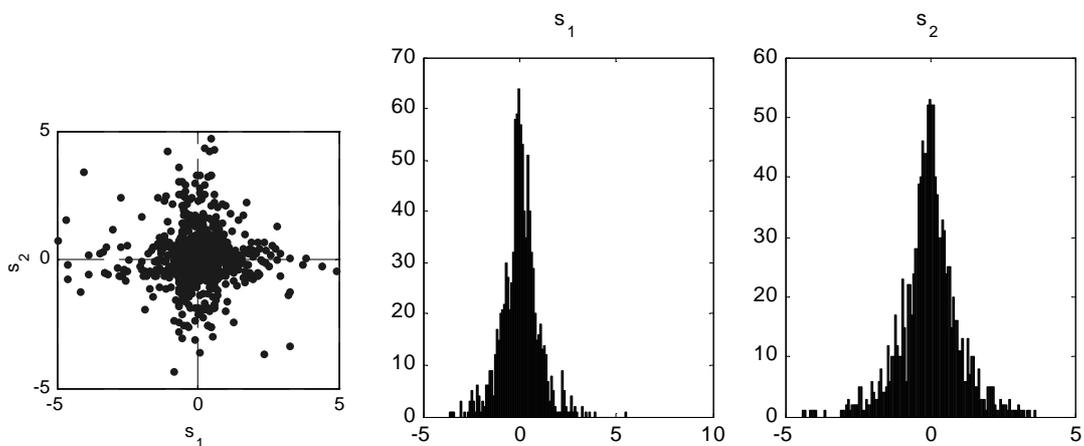


**Figura 3.9 – a) Distribuição conjunta de 2 variáveis platokúrticas independentes
b),c) Histogramas de s_1 e s_2 , respectivamente.**

$$p(s_i) = \begin{cases} \frac{1}{2\sqrt{3}}, & \text{se } s_i \leq \sqrt{3} \\ 0, & \text{caso contrário} \end{cases} \quad i = 1,2$$

$$\begin{aligned} \text{kurt}(s_1) &= -1.200 \\ \text{kurt}(s_2) &= -1.237 \end{aligned}$$

Distribuições cuja *kurtosis* é positiva são chamadas supergaussianas, ou *leptokúrticas*. Variáveis supergaussianas tem uma distribuição mais concentrada em torno da média e caudas mais longas que a distribuição gaussiana com mesma média e variância.



**Figura 3.10 – a) Distribuição conjunta de 2 variáveis leptokúrticas;
b),c) Histogramas de s_1 e s_2 , respectivamente.**

$$p(s_i) = \frac{1}{\sqrt{2}} \log(2 \cdot |s_i|) \quad i = 1, 2$$

$$\begin{aligned} \text{kurt}(s_1) &= + 5.350 \\ \text{kurt}(s_2) &= + 3.511 \end{aligned}$$

Assim, a *kurtosis*, ou ainda, seu valor absoluto, é amplamente utilizada como medida de não gaussianidade, devido a sua simplicidade teórica e computacional.

Modelo 2D

Assuma que as componentes s_1 e s_2 tenham valores de *kurtosis* diferentes de zero. Por definição, ambas têm variância unitária.

$$\text{Seja } y = \mathbf{b}^T \cdot \mathbf{x} = \mathbf{b}^T \cdot \mathbf{A} \cdot \mathbf{s} = \mathbf{q}^T \cdot \mathbf{s} = q_1 \cdot s_1 + q_2 \cdot s_2$$

Se as variâncias são unitárias, então:

$$E\{y^2\} = q_1^2 + q_2^2 = 1 \tag{3.28}$$

Geometricamente, isso significa que o vetor \mathbf{q} está restrito ao círculo unitário no plano 2D.

De acordo com a propriedade aditiva da *kurtosis*:

$$\text{kurt}(y) = \text{kurt}(q_1 s_1) + \text{kurt}(q_2 s_2) = q_1^4 \text{kurt}(s_1) + q_2^4 \text{kurt}(s_2) \tag{3.29}$$

A otimização do problema consiste em encontrar o valor máximo para a função $|\text{kurt}(y)| = |q_1^4 \cdot \text{kurt}(s_1) + q_2^4 \cdot \text{kurt}(s_2)|$ sobre o círculo unitário.

Supondo, inicialmente, que as *kurtosis* sejam iguais a 1. A função que se deseja maximizar é $F(\mathbf{q}) = q_1^4 + q_2^4$. As curvas de nível para essa função estão expressas na Figura 3.11.

A curva mais grossa representa a esfera unitária, e as curvas finas representam os contornos em que a função F é constante.

Verifica-se que os pontos de máximo da função correspondem aos pontos onde uma componente do vetor \mathbf{q} é zero e a outra é 1 ou -1 . Para essa configuração, y corresponde à componente independente!

Se as *kurtosis* forem -1 , o problema é análogo, pois, como se está utilizando módulo da função, a função a maximizar é a mesma.

Supondo que deseja-se encontrar as componentes independentes a partir da variável branqueada \mathbf{z} , conforme descrito pela equação (3.15), a variável y é tal que $y = \mathbf{w}^T \cdot \mathbf{z} = \mathbf{w}^T \cdot \mathbf{V} \cdot \mathbf{A} \cdot \mathbf{s} = \mathbf{q}^T \cdot \mathbf{s}$, ou seja, $\mathbf{q} = \mathbf{w}^T \cdot \mathbf{V} \cdot \mathbf{A}$.

Como demonstrado na equação (3.18), $\tilde{\mathbf{A}} \cdot \tilde{\mathbf{A}}^T = (\mathbf{V} \cdot \mathbf{A}) \cdot (\mathbf{V} \cdot \mathbf{A})^T = \mathbf{I}$, ou seja, a matriz de mistura após o branqueamento é ortogonal, de forma que:

$$\|\mathbf{q}\|^2 = (\mathbf{w}^T \mathbf{V} \mathbf{A}) \cdot (\mathbf{A}^T \mathbf{V}^T \mathbf{w}) = \|\mathbf{w}\|^2 \quad (3.30)$$

Isso significa que manter \mathbf{q} na esfera unitária é equivalente a manter \mathbf{w} na esfera unitária. Assim, é preciso maximizar o valor absoluto da *kurtosis* com a restrição simples de manter $\|\mathbf{w}\| = 1$. Ainda, após o branqueamento, as componentes independentes podem ser interpretadas como projeções de \mathbf{z} na linha (espaço unidimensional) criada pelo vetor \mathbf{w} .

Cada ponto da esfera unitária corresponde a uma projeção.

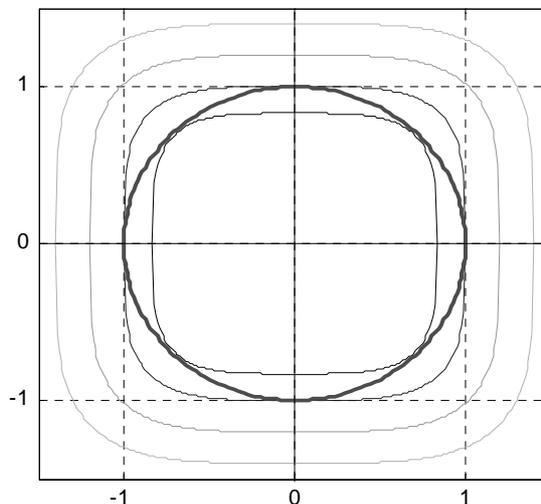


Figura 3.11 - Curvas de nível da função $F(\mathbf{q})$.

3.2.2.1 Gradiente

As formas mais populares para solucionar problemas de otimização baseiam-se no gradiente dessas funções de custo também chamadas funções-objetivo.

O gradiente de uma função multivariável $f(\mathbf{w})$ com relação ao vetor de variáveis $\mathbf{w} = (w_1, \dots, w_m)^T$ é definido como o vetor de derivadas parciais:

$$\frac{df}{d\mathbf{w}} = \begin{pmatrix} \frac{df}{dw_1} \\ \vdots \\ \frac{df}{dw_m} \end{pmatrix} \quad (3.31)$$

Suponha que deseja-se minimizar a função de custo $f(\mathbf{w})$ (ou maximizar $-f(\mathbf{w})$), com relação ao vetor de parâmetros \mathbf{w} . Pelo método do gradiente descendente, a função $f(\mathbf{w})$ é minimizada iterativamente a partir de um ponto inicial $\mathbf{w}(0)$. O gradiente de $f(\mathbf{w})$ é computado nesse ponto e então move-se para um outro ponto a uma dada distância na oposta direção à apontada pelo gradiente (*steepest descent*). O procedimento é repetido para o novo ponto, sugerindo o seguinte algoritmo de adaptação, onde o gradiente é calculado no ponto $\mathbf{w}(p-1)$:

$$\mathbf{w}(p) = \mathbf{w}(p-1) - \alpha(p) \frac{\partial f(\mathbf{w})}{\partial \mathbf{w}} \Big|_{\mathbf{w}=\mathbf{w}(p-1)} \quad (3.32)$$

O parâmetro $\alpha(p)$, também chamado de taxa de aprendizado, é o tamanho do passo na direção oposta à do gradiente. A iteração é repetida até a convergência, normalmente determinada quando a distância Euclidiana $\|\mathbf{w}(p) - \mathbf{w}(p-1)\|$ atinge um valor menor que um valor de tolerância.

Podes-se reescrever o algoritmo como:

$$\Delta \mathbf{w} = -\alpha \frac{\partial f(\mathbf{w})}{\partial \mathbf{w}} \quad (3.33)$$

ou ainda,

$$\Delta \mathbf{w} \propto \frac{\partial f(\mathbf{w})}{\partial \mathbf{w}} \quad (3.34)$$

Dessa forma, o gradiente do valor absoluto da *kurtosis* pode ser calculado como:

$$\frac{\partial |kurt(\mathbf{w}^T \cdot \mathbf{z})|}{\partial \mathbf{w}} = 4 \cdot \text{sign}(kurt(\mathbf{w}^T \cdot \mathbf{z})) \left[E\{\mathbf{z} \cdot (\mathbf{w}^T \cdot \mathbf{z})^3\} - 3 \cdot \mathbf{w} \cdot \|\mathbf{w}\|^2 \right] \quad (3.35)$$

Como essa função está sendo otimizada para a hiper-esfera unitária, deve-se garantir, a cada passo, que o vetor \mathbf{w} tenha norma 1.

Dada essa exigência, o último termo da equação (3.35) pode ser ignorado, uma vez que ele apenas muda a magnitude do vetor \mathbf{w} e o que nos interessa é sua direção.

Assim o algoritmo de iteração para o método do gradiente descendente pode ser, portanto, expresso como:

$$\Delta \mathbf{w} \propto \text{sign}(\text{kurt}(\mathbf{w}^T \cdot \mathbf{z})) \cdot [E\{\mathbf{z} \cdot (\mathbf{w}^T \cdot \mathbf{z})^3\}] \quad (3.36)$$

$$\mathbf{w} \leftarrow \mathbf{w} / \|\mathbf{w}\| \quad (3.37)$$

ou ainda, suprimindo-se a expectância da equação (3.35), pode-se atualizar o vetor \mathbf{w} a cada observação de \mathbf{z} apresentada:

$$\Delta \mathbf{w} \propto \text{sign}(\text{kurt}(\mathbf{w}^T \cdot \mathbf{z})) \cdot \mathbf{z} \cdot (\mathbf{w}^T \cdot \mathbf{z})^3 \quad (3.38)$$

$$\mathbf{w} \leftarrow \mathbf{w} / \|\mathbf{w}\| \quad (3.39)$$

Assim, cada observação $\mathbf{z}(p)$ pode ser usada por vez. No entanto, é preciso ressaltar que, ao computar $\text{sign}(\text{kurt}(\mathbf{w}^T \cdot \mathbf{z}))$, não é possível suprimir a expectância na definição da *kurtosis*. O que se pode fazer é estimá-la por uma média temporal. Denotando-se γ como uma estimativa da *kurtosis*, tem-se que:

$$\Delta \gamma \propto ((\mathbf{w}^T \cdot \mathbf{z})^4) - 3 - \gamma \quad (3.40)$$

Na maioria dos casos a natureza das distribuições das componentes independentes é conhecida previamente. Por exemplo, se é desejado separar misturas de sinais de áudio, é conhecido que as componentes apresentarão distribuições acentuadamente leptokúrticas. Pode-se, então, simplesmente colocar o sinal correto da *kurtosis* e evitar estimá-lo.

3.2.2.2 Ponto-Fixo (FastICA por kurtosis)

O método do gradiente descendente sempre se move na direção do mínimo, o que em muitos casos, pode conduzir a um mínimo local. Isso torna o método muito dependente das condições iniciais e da taxa de aprendizado. Um valor de α muito pequeno levará a uma convergência lenta, ao passo que um valor de α muito elevado tenderá a instabilizar o algoritmo

No caso, porém, a função $f(\mathbf{w})$ deve ser minimizada sob a restrição de manter a norma de \mathbf{w} unitária.

Em geral, o problema de otimização com restrições pode ser formulado como:

$$\min f(\mathbf{w}), \text{ dado que } h_k(\mathbf{w}) = 0, \quad k = 1, \dots, K \quad (3.41)$$

onde $h_k(\mathbf{w}) = 0, \quad k = 1, \dots, K$ são as equações que determinam as restrições sobre \mathbf{w} .

A solução clássica para esse problema utiliza o Lagrangiano, definido como:

$$\mathcal{L}(\mathbf{w}, \lambda_1, \dots, \lambda_K) = f(\mathbf{w}) + \sum_{k=1}^K \lambda_k \cdot h_k(\mathbf{w}) \quad (3.42)$$

O número K é mesmo número de equações de restrição. Os valores de λ_k são chamados *multiplicadores de Lagrange*. A solução do problema reside em encontrar o ponto mínimo do Lagrangiano, onde o gradiente é zero com relação a \mathbf{w} e todos os λ_k . Assim,

$$\frac{\partial f(\mathbf{w})}{\partial \mathbf{w}} + \sum_{k=1}^K \lambda_k \frac{\partial h_k(\mathbf{w})}{\partial \mathbf{w}} = 0 \quad (3.43)$$

Dessa forma, o problema de minimização foi reduzido a um conjunto de equações que são bem mais fáceis de resolver.

No caso, em que se deseja manter a norma de \mathbf{w} unitária, o Lagrangiano é dado por:

$$\mathcal{L}(\mathbf{w}, \lambda) = f(\mathbf{w}) + \lambda \cdot (\|\mathbf{w}\| - 1)$$

O gradiente do Lagrangiano, é, pois:

$$\frac{\partial f(\mathbf{w})}{\partial \mathbf{w}} + \frac{\partial [\lambda \cdot (\|\mathbf{w}\| - 1)]}{\partial \mathbf{w}} + \frac{\partial [\lambda \cdot (\|\mathbf{w}\| - 1)]}{\partial \lambda} = \frac{\partial f(\mathbf{w})}{\partial \mathbf{w}} + 2 \cdot \lambda \cdot \mathbf{w} = 0 \quad (3.44)$$

O gradiente de $\mathcal{L}(\mathbf{w}, \lambda)$ com relação a λ é a própria função $h(\mathbf{w}) = (\|\mathbf{w}\| - 1) = 0$. Logo, no ponto de máximo, o gradiente da função $f(\mathbf{w})$ aponta na direção de \mathbf{w} :

$$\frac{\partial f(\mathbf{w})}{\partial \mathbf{w}} \propto \mathbf{w} \quad (3.45)$$

Ou seja, no ponto estável do algoritmo, o gradiente deve ser igual a \mathbf{w} multiplicado por uma constante escalar. Apenas nesse caso, adicionar o gradiente a \mathbf{w} não irá mudar a direção de \mathbf{w} e o algoritmo terá convergido, uma vez que a norma de \mathbf{w} é sempre unitária. Assim:

$$\mathbf{w} \propto \left[E\{\mathbf{z} \cdot (\mathbf{w}^T \cdot \mathbf{z})^3\} - 3 \cdot \|\mathbf{w}\|^2 \cdot \mathbf{w} \right] \quad (3.46)$$

Logo, pode-se adaptar \mathbf{w} como:

$$\mathbf{w} \leftarrow \left[E\{\mathbf{z} \cdot (\mathbf{w}^T \cdot \mathbf{z})^3\} - 3 \cdot \mathbf{w} \right] \quad (3.47)$$

A convergência do algoritmo de ponto fixo requer que os valores antigos e novos de \mathbf{w} apontem na mesma direção, ou seja, o produto escalar é igual, ou quase igual, a 1. Não é necessário que o vetor convirja para um ponto único, pois \mathbf{w} e $-\mathbf{w}$ definem a mesma direção. Isso reside no fato de que as componentes independentes não podem ser determinadas a menos de um sinal.

[36] mostra que esse algoritmo tem convergência cúbica. Ainda, ao contrário do gradiente, não é necessário taxa de aprendizado ou outros parâmetros de ajuste ao algoritmo, o que o torna fácil e confiável.

3.2.3 Neg-entropia

A *kurtosis* é uma medida muito sensível a *outliers*, e, portanto, não é uma medida robusta de não-gaussianidade. Uma alternativa seria usar a *neg-entropia*, que mais é robusta, embora computacionalmente complicada.

A entropia de uma variável aleatória está relacionada à quantidade de informação que essa variável contém. A entropia será maior o quanto mais imprevisível for a variável. A entropia diferencial de um vetor \mathbf{y} com densidade $p_{\mathbf{y}}(\boldsymbol{\eta})$ é definida como:

$$H(\mathbf{y}) = - \int p_{\mathbf{y}}(\boldsymbol{\eta}) \log p_{\mathbf{y}}(\boldsymbol{\eta}) d\boldsymbol{\eta} \quad (3.48)$$

Uma variável gaussiana tem a maior entropia entre todas as variáveis aleatórias de mesma variância. Isso significa que uma distribuição gaussiana é a menos estruturada de todas as distribuições. A entropia é a maior em distribuições que são muito concentradas em alguns valores.

Para obter uma medida de não-gaussianidade que seja zero para uma variável gaussiana e sempre não-negativa, costuma-se usar uma versão normalizada da entropia diferencial, chamada **neg-entropia**, que é definida como:

$$J(y) = H(y_{gauss}) - H(y), \quad (3.49)$$

onde y_{gauss} é uma variável gaussiana com a mesma matriz de correlação (ou covariância) que y . A neg-entropia é invariante a transformações lineares inversíveis.

Estatisticamente, a neg-entropia é a medida ótima de não-gaussianidade [37]. No entanto, estimar a neg-entropia usando a definição requer uma estimativa da *pdf*, o que é computacionalmente muito difícil.

O método clássico de estimativa da neg-entropia é a expansão em densidade polinomial, usando cumulantes de alta ordem:

$$J(y) \approx \frac{1}{12} E\{y^3\}^2 + \frac{1}{48} kurt(y)^2, \quad (3.50)$$

onde y tem média zero e variância unitária. No entanto, essa aproximação usa a *kurtosis* e esse método sofre dos mesmos problemas que a estimação por *kurtosis*.

Outras abordagens usam momentos não polinomiais. Em geral, é possível substituir os momentos polinomiais y^3 e y^4 por outra função G . O método propõe a aproximação da neg-entropia baseado em expectâncias E :

$$J(y) \approx k_1 \cdot (E\{G_1\{y\}\})^2 + k_2 \cdot (E\{G_2\{y\}\} - E\{G_2\{v\}\})^2 \quad (3.51)$$

onde k_1 e k_2 são constantes positivas e v é uma variável gaussiana de média zero e variância unitária, assim como y . Mesmo nos casos em que essa medida não é muito acurada, ela serve com uma medida de não-gaussianidade. Ela será sempre não negativa e igual a zero se y tiver uma distribuição gaussiana. A aproximação baseada em momentos é uma particularização desse caso, onde $G_1\{y\} = y^3$ e $G_2\{y\} = y^4$.

Caso seja usada apenas uma função não quadrática G , pode-se escrever:

$$J(y) \propto [E\{G\{y\}\} - E\{G\{v\}\}]^2 \quad (3.52)$$

Escolhendo a função G de forma que ela não cresça muito rapidamente, obtêm-se estimadores mais robustos, computacionalmente simples e com propriedades estatísticas interessantes.

Alguns exemplos de funções com estas propriedades estão descritas abaixo, e os gráficos das mesmas estão representados na Figura 3.12.

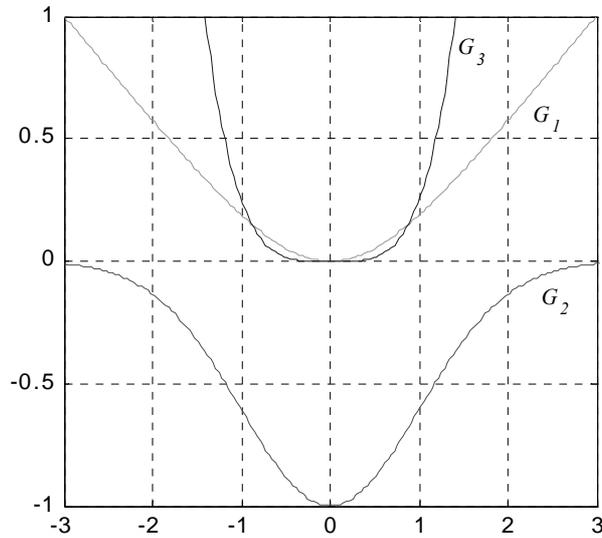


Figura 3.12 – Funções $G(y)$.

$$G_1(y) = \frac{1}{a_1} \log(\cosh(a_1 \cdot y)), \quad 1 \leq a_1 \leq 2 \quad (3.53)$$

$$G_2(y) = \exp\left(-\frac{y^2}{2}\right) \quad (3.54)$$

$$G_3(y) = \frac{y^4}{4} \quad (3.55)$$

3.2.3.1 Gradiente

O gradiente com relação a \mathbf{w} da aproximação da neg-entropia dada pela equação (3.69) é dado por:

$$\Delta \mathbf{w} \propto \gamma \cdot \left[E\{\mathbf{z} \cdot g(\mathbf{w}^T \cdot \mathbf{z})\} \right] \quad (3.56)$$

$$\mathbf{w} \leftarrow \mathbf{w} / \|\mathbf{w}\| \quad (3.57)$$

onde $\gamma = E\{G(\mathbf{w}^T \cdot \mathbf{z})\} - E\{G(v)\}$ e v é uma variável gaussiana padronizada.

A normalização é necessária para projetar \mathbf{w} na hiper-esfera unitária e manter a variância de $\mathbf{w}^T \cdot \mathbf{z}$ constante, uma vez que a variável \mathbf{z} é branca. A função G , usada na

aproximação da neg-entropia, pode ser uma das funções apresentadas em (3.53) ~ (3.55), cujas derivadas, expressas pela função g , são, respectivamente:

$$g_1(y) = \tanh(a_1 \cdot y) \quad (3.58)$$

$$g_2(y) = y \cdot \exp\left(-\frac{y^2}{2}\right) \quad (3.59)$$

$$g_3(y) = y^3 \quad (3.60)$$

O uso da função G_3 recai no algoritmo de gradiente por *kurtosis* apresentado na seção anterior.

A expectância pode ser omitida para proporcionar um algoritmo adaptativo.

A constante γ , que garante uma auto-adaptatividade do algoritmo pode ser estimada como:

$$\Delta\gamma \propto \left(G(\mathbf{w}^T \cdot \mathbf{z}) - E\{G(v)\}\right) - \gamma \quad (3.61)$$

O que seria o equivalente à estimativa do sinal da *kurtosis* em (3.40), e que pode ser substituída pelo sinal adequado (negativo para supergaussianas e positivo para subgaussianas), caso se tenha conhecimento prévio da distribuição das fontes.

3.2.3.2 Ponto-Fixo (FastICA por neg-entropia)

Deseja-se maximizar a aproximação da neg-entropia de $\mathbf{w}^T \cdot \mathbf{z}$ expressa em (3.69) respeitando a condição de $\|\mathbf{w}\| = 1$. O Lagrangiano da função é dado por:

$$\mathcal{L}(\mathbf{w}, \lambda) = \left[E\{G(\mathbf{w}^T \cdot \mathbf{z})\} - E\{G(v)\} \right]^2 + \lambda \cdot (\|\mathbf{w}\| - 1) \quad (3.62)$$

Note, porém que os máximos da aproximação da neg-entropia de $\mathbf{w}^T \cdot \mathbf{z}$ ocorrem para um certo valor ótimo de $E\{G(\mathbf{w}^T \cdot \mathbf{z})\}$. Assim é basta encontrar o ponto onde o gradiente do Lagrangiano da função $E\{G(\mathbf{w}^T \cdot \mathbf{z})\}$ é zero, ou seja, minimizar a função dada por:

$$f(\mathbf{w}) = E\{G(\mathbf{w}^T \cdot \mathbf{z})\} + \lambda \cdot (\|\mathbf{w}\| - 1) \quad (3.63)$$

Essa equação pode ser resolvida pelo método de Newton. O método usa informação da curvatura da função que se deseja otimizar, contida nas derivadas de segunda ordem, e portanto proporciona uma melhor direção para o próximo passo da

iteração. Expandindo a função $f(\mathbf{w})$ que se deseja minimizar em série de Taylor, tem-se que:

$$f(\mathbf{w}') = f(\mathbf{w}) + \left[\frac{\partial f(\mathbf{w})}{\partial \mathbf{w}} \right]^T \cdot (\mathbf{w}' - \mathbf{w}) + \frac{1}{2} \cdot (\mathbf{w}' - \mathbf{w})^T \cdot \frac{\partial^2 f(\mathbf{w})}{\partial^2 \mathbf{w}} \cdot (\mathbf{w}' - \mathbf{w}) + \dots \quad (3.64)$$

Assim ao se buscar o ponto mínimo da função $f(\mathbf{w})$, busca-se o novo ponto \mathbf{w}' que ofereça a maior queda no valor de $f(\mathbf{w})$. Portanto, deseja minimizar a função

$$f(\mathbf{w}') - f(\mathbf{w}) \approx \left[\frac{\partial f(\mathbf{w})}{\partial \mathbf{w}} \right]^T \cdot \Delta \mathbf{w} + \frac{1}{2} \cdot \Delta \mathbf{w}^T \cdot \frac{\partial^2 f(\mathbf{w})}{\partial^2 \mathbf{w}} \cdot \Delta \mathbf{w} \text{ com relação a } \Delta \mathbf{w}.$$

O gradiente dessa função com relação a $\Delta \mathbf{w}$ pode, pois, ser aproximado por:

$$\frac{\partial f(\Delta \mathbf{w})}{\partial \Delta \mathbf{w}} = \frac{\partial f(\mathbf{w})}{\partial \mathbf{w}} + \frac{\partial^2 f(\mathbf{w})}{\partial \mathbf{w}^2} \cdot \Delta \mathbf{w} \quad (3.65)$$

Para minimizar a função $f(\Delta \mathbf{w})$ basta igualar o gradiente a zero, resultando em:

$$\Delta \mathbf{w} = - \left[\frac{\partial^2 f(\mathbf{w})}{\partial \mathbf{w}^2} \right]^{-1} \cdot \frac{\partial f(\mathbf{w})}{\partial \mathbf{w}} \quad (3.66)$$

Obtém-se, assim, a iteração de segunda ordem:

$$\mathbf{w}' = \mathbf{w} - \left[\frac{\partial^2 f(\mathbf{w})}{\partial \mathbf{w}^2} \right]^{-1} \cdot \frac{\partial f(\mathbf{w})}{\partial \mathbf{w}} \quad (3.67)$$

Aplicando $f(\mathbf{w}) = E\{G(\mathbf{w}^T \cdot \mathbf{z})\} + \lambda \cdot (\|\mathbf{w}\| - 1)$ à equação acima, tem-se que:

$$\mathbf{w} \leftarrow \mathbf{w} - \left[E\{\mathbf{z} \cdot \mathbf{z}^T \cdot g'(\mathbf{w}^T \cdot \mathbf{z})\} + \lambda \cdot \mathbf{I} \right]^{-1} \cdot \left[E\{\mathbf{z} \cdot g(\mathbf{w}^T \cdot \mathbf{z})\} + \lambda \cdot \mathbf{w} \right] \quad (3.68)$$

Para simplificar a inversão da matriz $\left[E\{\mathbf{z} \cdot \mathbf{z}^T \cdot g'(\mathbf{w}^T \cdot \mathbf{z})\} + \lambda \cdot \mathbf{I} \right]$, é possível aproximar o termo $E\{\mathbf{z} \cdot \mathbf{z}^T \cdot g'(\mathbf{w}^T \cdot \mathbf{z})\} \approx E\{\mathbf{z} \cdot \mathbf{z}^T\} \cdot E\{g'(\mathbf{w}^T \cdot \mathbf{z})\} = E\{g'(\mathbf{w}^T \cdot \mathbf{z})\}$, dado que \mathbf{z} é branco. Assim, como a matriz resultante de $E\{g'(\mathbf{w}^T \cdot \mathbf{z})\} + \lambda$ é diagonal, o termo torna-se, portanto, facilmente inversível:

$$\mathbf{w} \leftarrow \mathbf{w} - \left[E\{g'(\mathbf{w}^T \cdot \mathbf{z})\} + \lambda \right]^{-1} \cdot \left[E\{\mathbf{z} \cdot g(\mathbf{w}^T \cdot \mathbf{z})\} + \lambda \cdot \mathbf{w} \right] \quad (3.69)$$

O algoritmo pode ser ainda mais simplificado multiplicando-se ambos os lados da equação acima por $\left[E\{g'(\mathbf{w}^T \cdot \mathbf{z})\} + \lambda \right]$, resultando em:

$$\mathbf{w} \leftarrow E\{\mathbf{z} \cdot g(\mathbf{w}^T \cdot \mathbf{z})\} - E\{g'(\mathbf{w}^T \cdot \mathbf{z})\} \cdot \mathbf{w} \quad (3.70)$$

$$\mathbf{w} \leftarrow \mathbf{w} / \|\mathbf{w}\| \quad (3.71)$$

A equação (3.70) é o algoritmo de adaptação de \mathbf{w} .

Em suma, primeiramente se escolhe uma não linearidade g , que é a derivada de uma das funções não quadráticas G usadas em (3.69). As derivadas de g são dadas por:

$$g'_1(y) = a_1(1 - \tanh^2(a_1 \cdot y)) \quad (3.72)$$

$$g'_2(y) = (1 - y^2) \cdot \exp\left(-\frac{y^2}{2}\right) \quad (3.73)$$

$$g'_3(y) = 3 \cdot y^2 \quad (3.74)$$

Como existe a restrição $E\{y^2\} = 1$, a derivada em (3.55) é reduzida para 3.

A convergência implica em que os valores antigo e novo de \mathbf{w} apontam na mesma direção, ou seja, o valor absoluto do seu produto escalar está próximo de 1.

A convergência é ao menos quadrática, bem mais rápida que a convergência linear do método de gradiente. Ainda, não há uma taxa de aprendizado ou outros parâmetros de ajuste no algoritmo, o que o torna mais confiável.

3.3 - ICA por Maximização da Função de Verossimilhança

Seja uma determinada variável aleatória x e conhecendo-se um conjunto de P observações de x , $\mathbf{x}_P = [x(1), x(2), \dots, x(P)]$, deseja-se descrever a variável x a partir de um conjunto de parâmetros θ , que forneçam informações mais palatáveis acerca da distribuição de x .

O vetor de parâmetros $\theta = [\theta_1, \theta_2, \dots, \theta_m]$, não é, porém, conhecido, e deve ser estimado. É preciso, pois, encontrar um vetor de parâmetros que melhor caracterize a distribuição de x .

Dado um conjunto de observações \mathbf{x}_P da variável aleatória x , uma das formas de estimar o vetor de parâmetros θ é encontrar um vetor de parâmetros tal que a probabilidade desse vetor seja máxima, ou seja, $\hat{\theta} = \arg \max(p_{\theta/x}(\theta / \mathbf{x}_P))$

Pelo teorema de Bayes tem-se que:

$$p_{\theta/x}(\theta / \mathbf{x}_p) = \frac{p_{x/\theta}(\mathbf{x}_p / \theta) p_{\theta}(\theta)}{p_x(\mathbf{x}_p)} \quad (3.75)$$

O denominador da função não depende de θ , e, portanto, é apenas um fator de escalamento na distribuição conjunta de \mathbf{x}_p e de θ . Supondo que todos os conjuntos de parâmetros são equiprováveis, maximizar a densidade *a posteriori* de θ equivale, pois, a maximizar a densidade *a priori*.

Nesse caso, então, maximizar a densidade a posteriori ou encontrar o estimador MAP (*maximum a posteriori estimator*) é equivalente a maximizar a função de verossimilhança ML (*maximum likelihood*):

$$\max(p_{\theta/x}(\theta / \mathbf{x}_p)) = \max(p_{x/\theta}(\mathbf{x}_p / \theta)) \quad (3.76)$$

Para exemplificar, suponha que se conheça um vetor de observações de uma variável escalar, e que essa variável obedeça a uma distribuição gaussiana:

$$p(\mathbf{x}_p / \mu, \sigma^2) = (2 \cdot \pi \cdot \sigma)^{-P/2} \cdot \exp \left[-\frac{1}{2\sigma^2} \sum_{p=1}^P [x(p) - \mu]^2 \right] \quad (3.77)$$

A função de verossimilhança é obtida calculando o logaritmo:

$$\ln p(\mathbf{x}_p / \mu, \sigma^2) = -\frac{P}{2} \ln(2 \cdot \pi \cdot \sigma^2) - \frac{1}{2\sigma^2} \cdot \sum_{p=1}^P [x(p) - \mu]^2 \quad (3.78)$$

Os estimadores de máxima verossimilhança podem ser obtidos derivando-se a função de verossimilhança com relação aos parâmetros que se deseja estimar:

$$\frac{\partial}{\partial \mu} \ln p(\mathbf{x}_p / \hat{\mu}_{ML}, \hat{\sigma}_{ML}^2) = \frac{1}{2\hat{\sigma}_{ML}^2} \cdot \sum_{p=1}^P [x(p) - \hat{\mu}_{ML}] = 0 \quad (3.79)$$

$$\therefore \hat{\mu}_{ML} = \frac{1}{P} \cdot \sum_{p=1}^P x(p) \quad (3.80)$$

e

$$\frac{\partial}{\partial \sigma^2} \ln p(\mathbf{x}_p / \hat{\mu}_{ML}, \hat{\sigma}_{ML}^2) = -\frac{P}{2\hat{\sigma}_{ML}^2} + \frac{1}{2\hat{\sigma}_{ML}^2} \cdot \sum_{p=1}^P [x(p) - \hat{\mu}_{ML}]^2 = 0 \quad (3.81)$$

$$\therefore \hat{\sigma}_{ML}^2 = \frac{1}{P} \cdot \sum_{p=1}^P [x(p) - \hat{\mu}_{ML}]^2 \quad (3.82)$$

Ou seja, para uma distribuição gaussiana, os parâmetros que maximizam a distribuição do vetor de observações \mathbf{x} são a média e o desvio padrão. Na realidade, a estimativa obtida para o desvio-padrão é polarizada. Isso ocorre porque ela usa a média estimada invés da verdadeira na equação (3.81). Esse procedimento reduz em uma amostra a quantidade informação nova que está de fato disponível para estimação. No entanto, essa polarização é, normalmente, pequena e a medida é assintoticamente despolarizada.

Agora, seja o modelo de mistura para o ICA:

$$\mathbf{x} = \mathbf{A} \cdot \mathbf{s}, \quad (3.83)$$

onde nem o vetor de fontes \mathbf{s} , nem a matriz de mistura \mathbf{A} são conhecidos.

Tendo observado o vetor de mistura \mathbf{x} , deseja-se encontrar uma matriz de parâmetros \mathbf{A} de forma que a probabilidade de que o vetor \mathbf{x} tenha sido gerado a partir da mistura de \mathbf{s} seja máxima.

$$\hat{\mathbf{A}} = \arg \max_{\mathbf{x}=\mathbf{A}\mathbf{s}} (p_{\mathbf{A}/\mathbf{x}}(\mathbf{A} / \mathbf{x}_p)) \quad (3.84)$$

A densidade de probabilidade do vetor de mistura \mathbf{x} é dada por [31]:

$$p_x(\mathbf{x} / \mathbf{B}) = |\det \mathbf{B}| \cdot p_s(\mathbf{s}), \text{ onde } \mathbf{B} = \mathbf{A}^{-1} \quad (3.85)$$

Como as componentes são independentes, pode-se escrever a probabilidade a priori de \mathbf{s} como o produtório das densidades de distribuição *a priori* das componentes independentes, p_i :

$$p_x(\mathbf{x} / \mathbf{B}) = |\det \mathbf{B}| \cdot \prod_i p_i(s_i) \quad (3.86)$$

Sabendo-se que $s_i = \mathbf{b}_i^T \mathbf{x}(p)$, onde $\mathbf{B} = [\mathbf{b}_1, \dots, \mathbf{b}_n]^T$, pode-se descrever a função de verossimilhança como o produto das densidades em P pontos:

$$L(\mathbf{B}) = \prod_{p=1}^P \prod_{i=1}^n p_i(\mathbf{b}_i^T \mathbf{x}(p)) \cdot |\det \mathbf{B}| \quad (3.87)$$

O produtório é convertido em somatório se extrairmos o logaritmo da função de verossimilhança:

$$\log L(\mathbf{B}) = \sum_{p=1}^P \sum_{i=1}^n \log p_i(\mathbf{b}_i^T \mathbf{x}(p)) + P \cdot \log |\det \mathbf{B}| \quad (3.88)$$

Ou, para manter a consistência:

$$\frac{1}{P} \log L(\mathbf{B}) = E \left\{ \sum_{i=1}^n \log p_i(\mathbf{b}_i^T \mathbf{x}(p)) \right\} + \log |\det \mathbf{B}| \quad (3.89)$$

Da forma como foi expressa, a verossimilhança é uma função dos parâmetros do modelo, ou seja, da matriz de mistura \mathbf{A} ou ainda, da sua inversa \mathbf{B} , uma vez que, para o modelo adotado, a matriz de mistura pode ser diretamente obtida a partir de sua inversa.

No entanto, a verossimilhança é também função das densidades das componentes independentes $p_i(\mathbf{s}(p)) = p_i(\mathbf{b}_i^T \mathbf{x}(p))$, o que resulta em um problema não paramétrico, ou seja, há uma infinidade de parâmetros que descrevem uma função de densidade de probabilidade, o que dificulta sua estimação. Uma forma de solucionar esse problema é aproximar as densidades das componentes independentes por uma família de densidades que sejam especificadas por um número limitado de parâmetros.

Para o problema de estimação de componentes independentes, pode-se provar [31] que é possível fazer a estimativa da distribuição das componentes independentes a partir de apenas duas aproximações. Para cada componente independente, é preciso apenas determinar qual das duas aproximações é a melhor estimativa para sua distribuição de densidade de probabilidade.

Isso seria equivalente a estimar se a componente independente teria uma distribuição sub ou supergaussiana, no caso de estimação pelo gradiente da *kurtosis*. Uma forma de fazer essa estimativa era usar uma função não linear cujo sinal proveria a informação desejada. No caso da estimativa por máxima verossimilhança, essa função não linear corresponde à aproximação da densidade da componente independente.

Seja \tilde{p}_i a densidade da *i*-ésima componente independente e que:

$$g_i(s_i) = \frac{\partial}{\partial s_i} \log \tilde{p}_i(s_i) = \frac{\tilde{p}'_i(s_i)}{\tilde{p}_i(s_i)} \quad (3.90)$$

Para que o estimador de máxima verossimilhança seja localmente consistente, dada a restrição de que as estimativas das componentes independentes são descorrelacionadas e têm variância unitária, pode-se provar [31] que é necessário que:

$$E\{s_i \cdot g_i(s_i) - g'_i(s_i)\} > 0, \quad \forall i. \quad (3.91)$$

O teorema significa que pequenos erros nas densidades de s_i não afetam a consistência local do estimador ML, desde que essas pequenas alterações não impliquem na mudança do sinal da equação (3.91).

Dessa forma, é possível construir famílias que consistam em apenas duas densidades, de forma que a condição expressa em (3.91) seja verdadeira para apenas uma dessas densidades de cada vez.. Um exemplo de funções seria:

$$\log \tilde{p}_i^+(s) = \alpha_1 - 2 \cdot \log \cosh(s) \quad (3.92)$$

$$\log \tilde{p}_i^-(s) = \alpha_2 - \left[\frac{s^2}{2} - \log \cosh(s) \right] \quad (3.93)$$

Onde α_1 e α_2 são constantes positivas escolhidas de forma a tornar as funções logaritmos de densidades de probabilidade.

Lembrando que $E\{s^2\} = 1$ e aplicando as funções acima no teorema expresso em (3.91), tem-se que:

$$C_1 : E\{s_i \cdot g_i(s_i) - g'(s_i)\} = 2 \cdot E\{-\tanh(s_i) \cdot s_i + (1 - \tanh(s_i)^2)\} \quad (3.94)$$

$$C_2 : E\{s_i \cdot g_i(s_i) - g'(s_i)\} = E\{\tanh(s_i) \cdot s_i - (1 - \tanh(s_i)^2)\} \quad (3.95)$$

Os sinais dessas duas expressões são sempre opostos, conforme verifica-se na Figura 3.13. Portanto, para praticamente qualquer distribuição de s_i , apenas uma dessas expressões satisfaz o critério expresso em (3.91), tornando a estimação consistente. Por sorte, são considerados muito raros os casos em que ambos zeram, o que corresponderia ao caso de distribuições não-gaussianas com *kurtosis* igual a zero na estimação por cumulantes.

Logo, é preciso apenas computar os momentos não polinomiais para as duas distribuições nas equações (3.92) e (3.93) e escolher a que satisfaz o critério de estabilidade expresso em (3.91). Isso pode ser feito durante a maximização da verossimilhança. Esse procedimento garante que o estimador será sempre consistente localmente e resolve o problema da estimação não paramétrica.

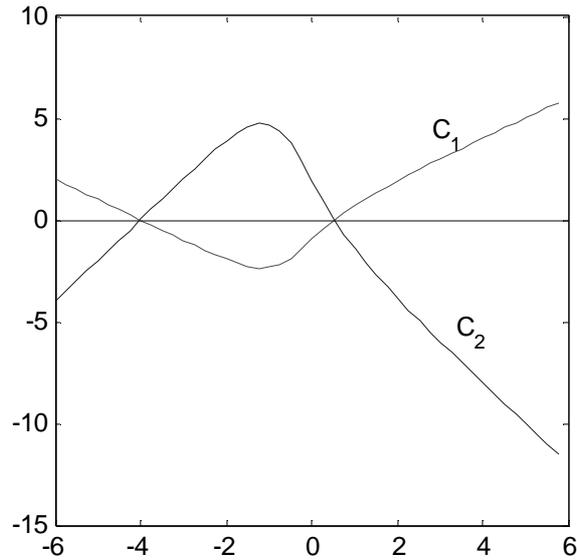


Figura 3.13 – Funções de condição de estabilidade.

Na realidade, o momento não-polinomial mede a forma da função de densidade da mesma forma que a *kurtosis*. Para $g(s) = -s^3$ obtém-se a *kurtosis*. Assim, a escolha da não linearidade pode ser comparada com a escolha de maximizar a *kurtosis*, baseada no valor do sinal da *kurtosis*, conforme realizado na seção 3.2.2.1.

Ainda, o momento não-polinomial usado pelo estimador ML é equivalente à escolha de uma função não linear usada para medir não gaussianidade, na seção 3.2.3. A escolha da não-linearidade para maximização da verossimilhança é, porém, mais restrita, uma vez que deve corresponder à derivada do logaritmo de uma função de densidade de probabilidade (*pdf*). Por exemplo, não é possível escolher $g(s) = s^3$ porque a *pdf* correspondente, de acordo com a equação (3.91), seria $\exp(s^4/4)$, visto que

$\frac{\partial \log[\exp(s^4/4)]}{\partial s} = \frac{\partial(s^4/4)}{\partial s} = s^3$. No entanto, $\int_{-\infty}^{\infty} \exp(s^4/4)$ é infinita, e, portanto, $\exp(s^4/4)$ não pode ser uma *pdf*.

3.3.1.1 Gradiente

Um dos métodos para obter a estimação da máxima verossimilhança, é o gradiente da função expressa na equação (3.89), resultando:

$$\frac{1}{P} \frac{\partial \log L}{\partial \mathbf{B}} = [\mathbf{B}^T]^{-1} + E\{g(\mathbf{B} \cdot \mathbf{x}) \cdot \mathbf{x}^T\} \quad (3.96)$$

onde $g(\mathbf{y}) = ((g_i(y_i), \dots, g_n(y_n)))$ são funções de custo da distribuição de s_i , definidas como na equação (3.90).

Assim, pode-se estimar a verossimilhança por:

$$\Delta \mathbf{B} \propto [\mathbf{B}^T]^{-1} + E\{\mathbf{g}(\mathbf{B} \cdot \mathbf{x}) \cdot \mathbf{x}^T\} \quad (3.97)$$

Para uma implementação adaptativa, basta suprimir o operador de expectância:

$$\Delta \mathbf{B} \propto [\mathbf{B}^T]^{-1} + \mathbf{g}(\mathbf{B} \cdot \mathbf{x}) \cdot \mathbf{x}^T \quad (3.98)$$

Esse algoritmo é conhecido como *Bell-Sejnowski* e foi derivado pela primeira vez em [38].

Em [39], Amari prova que o algoritmo apresenta melhor convergência se o gradiente usual é multiplicado por $\mathbf{W}^T \cdot \mathbf{W}$, levando ao algoritmo conhecido como **gradiente natural**:

$$\Delta \mathbf{W} \propto \frac{\partial f(\mathbf{W})}{\partial \mathbf{W}} \cdot \mathbf{W}^T \cdot \mathbf{W} \quad (3.99)$$

Dessa forma, multiplicando o gradiente encontrado na equação (3.97) por $\mathbf{B}^T \cdot \mathbf{B}$, chega-se ao seguinte algoritmo para o método de estimação das componentes independentes por maximização da verossimilhança:

$$\Delta \mathbf{B} \propto (\mathbf{I} + E\{\mathbf{g}(\mathbf{y}) \cdot \mathbf{y}^T\}) \cdot \mathbf{B} \quad (3.100)$$

Um exemplo de não-linearidade usada para componentes supergaussianas seria:

$$g^+(y) = -2 \cdot \tanh(y), \quad (3.101)$$

e para subgaussianas:

$$g^-(y) = \tanh(y) - y \quad (3.102)$$

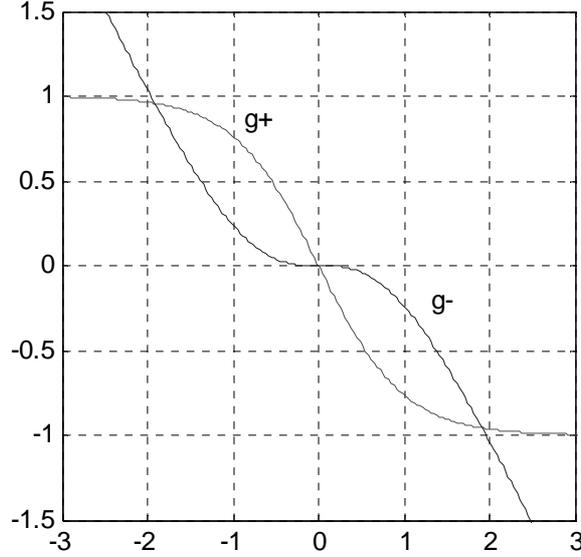


Figura 3.14 – Funções não lineares de aproximação.

A fim de satisfazer o teorema (3.91), a escolha entre as funções não lineares é feita a partir do momento não-polinomial:

$$\gamma = \text{sign} \left[E \left\{ -\tanh(s_i) \cdot s_i + (1 - \tanh(s_i))^2 \right\} \right] \quad (3.103)$$

Ou seja, se o momento não-polinomial γ for positivo, a não-linearidade usada é a da equação (3.101) e, caso seja negativo, é usada a expressão dada pela equação (3.102).

3.3.1.2 Ponto Fixo (FastICA por Maximização da Verossimilhança)

O algoritmo de ponto fixo desenvolvido na seção 3.2.3.2 resultou na iteração apresentada em (3.69), reproduzida aqui por conveniência:

$$\mathbf{w} \leftarrow \mathbf{w} - \left[E \{ g'(\mathbf{w}^T \cdot \mathbf{z}) \} + \lambda \right]^{-1} \cdot \left[E \{ \mathbf{z} \cdot g(\mathbf{w}^T \cdot \mathbf{z}) \} + \lambda \cdot \mathbf{w} \right] \quad (3.104)$$

Usando notação matricial:

$$\mathbf{W} \leftarrow \mathbf{W} + \text{diag}(\alpha_i) \cdot [\text{diag}(\beta_i) + E \{ g(\mathbf{y}) \cdot \mathbf{y}^T \}] \cdot \mathbf{W} \quad (3.105)$$

onde $\mathbf{y} = \mathbf{W} \cdot \mathbf{z}$ e $\alpha_i = -1 / (E \{ \beta_i + g'(y_i) \})$. Multiplicando ambos os lados por \mathbf{V} , onde $\mathbf{W} = \mathbf{V}^T \cdot \mathbf{B}$, é possível expressar esse resultado para dados não branqueados:

$$\mathbf{B} \leftarrow \mathbf{B} + \text{diag}(\alpha_i) \cdot [\text{diag}(\beta_i) + E \{ g(\mathbf{y}) \cdot \mathbf{y}^T \}] \cdot \mathbf{B} \quad (3.106)$$

onde $\mathbf{y} = \mathbf{B} \cdot \mathbf{x}$, $\beta_i = -E\{y_i \cdot g(y_i)\}$ e $\alpha_i = -1/(\beta_i + E\{g'(y_i)\})$

Para manter o algoritmo estável, após cada passo a matriz \mathbf{B} deve ser projetada no espaço das matrizes brancas. Isso pode ser feito pelo método da raiz quadrada de matrizes:

$$\mathbf{B} \leftarrow (\mathbf{B} \cdot \mathbf{C} \cdot \mathbf{B}^T)^{-1/2} \cdot \mathbf{B}, \quad (3.107)$$

onde $\mathbf{C} = E\{\mathbf{x} \cdot \mathbf{x}^T\}$.

Essa versão do FastICA pode ser considerada como uma versão computacionalmente otimizada do algoritmo do gradiente. A convergência é otimizada pela escolha das matrizes $diag(\alpha_i)$ e $diag(\beta_i)$. Essas duas matrizes dão o tamanho do passo ótimo para ser usado no algoritmo.

Outra vantagem do FastICA é que ele pode estimar ambas as componentes independentes sub e supergaussianas sem passos adicionais: pode-se fixar as não-linearidades para serem iguais à tangente hiperbólica para todas as componentes independentes. Isso porque a matriz $diag(\alpha_i)$ contém estimativas sobre a natureza das componentes independentes. A matriz $diag(\beta_i)$ pode ser considerada como um fator de escala das não-linearidades, uma vez que pode-se reescrever:

$$[diag(\beta_i) + E\{g(\mathbf{y}) \cdot \mathbf{y}^T\}] = diag(\beta_i) \cdot [\mathbf{I} + diag(\beta_i^{-1}) \cdot E\{g(\mathbf{y}) \cdot \mathbf{y}^T\}] \quad (3.108)$$

Assim, o FastICA pode ser entendido como uma parametrização mais rica de densidades do que a parametrização de apenas 2 densidades.

No FastICA as saídas y_i são descorrelacionadas e normalizadas para variância unitária após cada passo. Esse procedimento, essencial para a estabilidade do algoritmo, reduz o espaço de otimização consideravelmente.

3.4 - Estimação de várias componentes

Até agora, foram propostos métodos de estimação de apenas uma componente independente. Uma forma de encontrar as demais componentes independentes seria executar o algoritmo várias vezes a partir de condições iniciais diversas. No entanto, esse não é um método confiável e eficiente.

Os vetores \mathbf{w}_i correspondentes a diferentes componentes independentes são ortogonais no espaço branqueado, uma vez que a independência entre as componentes exige que eles sejam descorrelacionados. Assim, $E\{(w_i^T \cdot z) \cdot (w_j^T \cdot z)^T\} = w_i^T \cdot w_j$ e, portanto, descorrelação significa ortogonalidade. \mathbf{w}_i é, por definição, a i -ésima linha da inversa da matriz de mistura $\tilde{\mathbf{A}}$, que é igual à i -ésima coluna de $\tilde{\mathbf{A}}$, pois $\tilde{\mathbf{A}}^{-1} = \tilde{\mathbf{A}}^T$, onde $\tilde{\mathbf{A}} = \mathbf{V} \cdot \mathbf{A}$, conforme apresentado na equação (3.18).

Dessa forma, para estimar várias componentes independentes, basta impedir que os vetores \mathbf{w}_i convirjam para o mesmo máximo ortogonalizando os vetores $\mathbf{w}_1, \dots, \mathbf{w}_n$ em cada iteração.

Há dois métodos para realizar isso: a ortogonalização por deflação e a ortogonalização simétrica.

3.4.1 Ortogonalização por deflação

Um dos métodos para ortogonalizar os vetores \mathbf{w} é a deflação usando a ortogonalização de Gram-Schmidt. As componentes independentes serão estimadas uma a uma. Assim subtrai-se do vetor \mathbf{w}_p que acabou de ser estimado as projeções sobre os demais $p-1$ vetores que foram previamente estimados, ou seja $(\mathbf{w}_p^T \cdot \mathbf{w}_j) \cdot \mathbf{w}_j, j = 1, \dots, p-1$ e, então, se re-normaliza o vetor \mathbf{w}_p :

$$\mathbf{w}_p \leftarrow \mathbf{w}_p - \sum_{j=1}^{p-1} (\mathbf{w}_p^T \cdot \mathbf{w}_j) \cdot \mathbf{w}_j \quad (3.109)$$

$$\mathbf{w}_p \leftarrow \mathbf{w}_p / \|\mathbf{w}_p\| \quad (3.110)$$

Quando \mathbf{w}_p tiver convergido, passa-se a estimar a $(p+1)$ -ésima componente independente.

3.4.2 Ortogonalização simétrica

A ortogonalização de Gram-Schmidt provoca uma acumulação de erros para as componentes que são estimadas depois.

O método de ortogonalização simétrica propõe que todas as componentes sejam estimadas em paralelo, de forma que nenhuma componente seja privilegiada sobre as outras.

Assim, em um passo inicial as n componentes independentes são computadas em paralelo, usando um dos métodos de ICA, com os vetores \mathbf{w}_i inicializados randomicamente. A matriz $\mathbf{W} = (\mathbf{w}_1, \dots, \mathbf{w}_m)^T$ encontrada é então ortogonalizada. Um dos métodos clássicos para ortogonalização simétrica utiliza a inversa da raiz quadrada da matriz:

$$\mathbf{W} \leftarrow (\mathbf{W} \cdot \mathbf{W}^T)^{-1/2} \cdot \mathbf{W} \quad (3.111)$$

A inversa da raiz quadrada de \mathbf{W} pode ser obtida a partir da decomposição em autovalores de $\mathbf{W} \cdot \mathbf{W}^T = \mathbf{E} \cdot \text{diag}(d_1, \dots, d_m) \cdot \mathbf{E}^T$:

$$(\mathbf{W} \cdot \mathbf{W}^T)^{-1/2} = \mathbf{E} \cdot \text{diag}(d_1^{-1/2}, \dots, d_m^{-1/2}) \cdot \mathbf{E}^T \quad (3.112)$$

Outra forma alternativa é o algoritmo alternativo:

$$1. \mathbf{W} \leftarrow \mathbf{W} / \|\mathbf{W}\| \quad (3.113)$$

$$2. \mathbf{W} \leftarrow \frac{3}{2} \mathbf{W} - \frac{1}{2} \mathbf{W} \cdot \mathbf{W}^T \cdot \mathbf{W} \quad (3.114)$$

3. Se $\mathbf{W} \cdot \mathbf{W}^T$ não está suficientemente próxima da unidade, voltar ao passo 2.

4. Redes Neurais

Redes Neurais Artificiais são técnicas computacionais que apresentam um modelo matemático inspirado na estrutura neural de organismos inteligentes e que adquirem conhecimento através da experiência.

O sistema nervoso é formado por um conjunto extremamente complexo de células, os neurônios. Eles têm um papel essencial na determinação do funcionamento e comportamento do corpo humano e do raciocínio. Os neurônios são formados pelos dendritos, que são um conjunto de terminais de entrada, pelo corpo central, e pelos axônios que são longos terminais de saída.

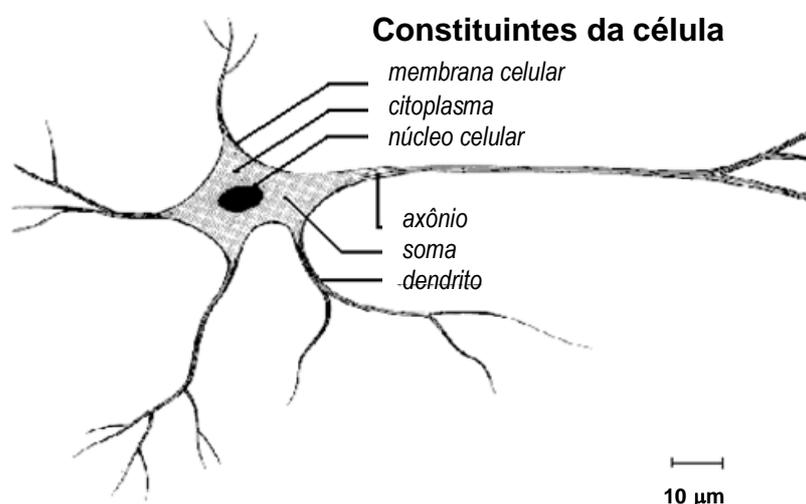


Figura 4.1 – Constituição da célula neuronal.

Os neurônios se comunicam através de sinapses, que correspondem à região onde dois neurônios entram em contato e através da qual, os impulsos nervosos são transmitidos entre eles. Os impulsos recebidos por um neurônio *A*, em um determinado momento, são processados, e atingindo um dado limiar de ação, o neurônio *A* dispara, produzindo uma substância neurotransmissora que flui do corpo celular para o axônio, que pode estar conectado a um dendrito de um outro neurônio *B*. O neurotransmissor pode diminuir ou aumentar a polaridade da membrana pós-sináptica, inibindo ou excitando a geração dos pulsos no neurônio *B*. Este processo depende de vários fatores, como a geometria da sinapse e o tipo de neurotransmissor.

Em média, cada neurônio forma entre mil e dez mil sinapses. O cérebro humano possui cerca de 10^{11} neurônios, e o número de sinapses é de mais de 10^{14} , possibilitando a formação de redes muito complexas.

O trabalho pioneiro sobre *Redes Neurais Artificiais* foi realizado no ano de 1943, por McCullock e Pitts. Neste, eles desenvolveram um estudo sobre o comportamento do neurônio biológico, com o objetivo de criar um modelo matemático para este.

No entanto, o primeiro modelo de rede neural – o **perceptron** – foi implementado somente em 1958, por Frank Rosenblatt.

O **perceptron** é uma rede neural simples: cada sinal de entrada é multiplicado por um número, ou **peso**, que indica a sua influência na saída da unidade. É então feita a soma ponderada dos sinais, produzindo um nível de atividade. Se este nível de atividade exceder um certo limite (*threshold*), a unidade produz uma determinada resposta de saída.

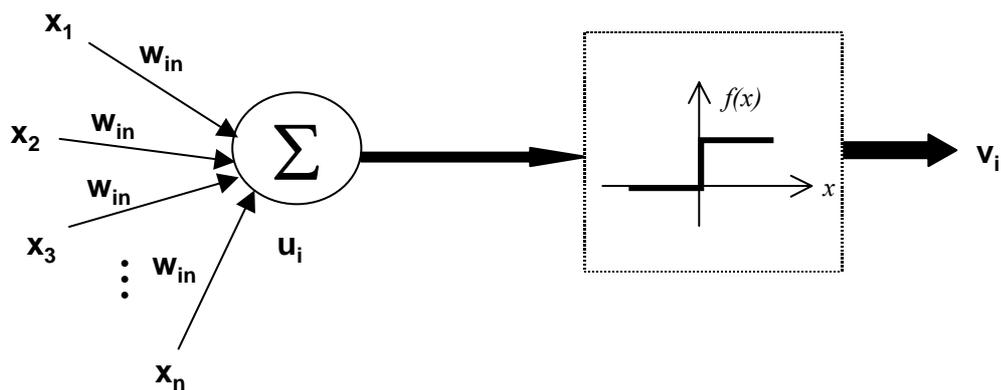


Figura 4.2 – Neurônio: elemento de processamento.

Matematicamente:

$$u_i = \sum_{j=1}^n w_{ij} \cdot x_j + b_i = \mathbf{w}_i^T \cdot \mathbf{x} + b_i \quad (4.1)$$

onde x_j são as entradas da rede e $v_i = f(u_i)$. $f(u_i)$ pode ser uma função de transição entre dois níveis de forma abrupta ou suave.

As arquiteturas neurais são tipicamente organizadas em camadas, com unidades que podem estar conectadas às unidades da camada posterior.

Uma rede neural é especificada, principalmente pela sua *topologia*, pelas *características dos nós* e pelas *regras de treinamento*.

A propriedade mais importante das redes neurais é a habilidade de aprender de seu ambiente e com isso melhorar seu desempenho. Isso é feito através de um processo iterativo de ajustes aplicado a seus pesos, o **treinamento**. O aprendizado ocorre quando a rede neural atinge uma solução generalizada para uma classe de problemas.

4.1 - Redes em *backpropagation*

4.1.1 Associadores e Classificadores

O objetivo da rede neural é encontrar um mapeamento unívoco de uma entrada \mathbf{x} para uma saída \mathbf{y} , ou seja, um determinado conjunto k de J entradas $\mathbf{x}^k = [x_1^k, x_2^k, \dots, x_J^k]$ deve produzir uma saída $\mathbf{y}^k = [y_1^k, y_2^k, \dots, y_L^k]$.

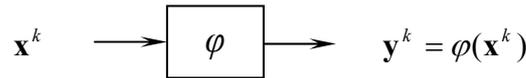


Figura 4.3 – Esquema de operação.

Sendo x_j as entradas da rede, para cada saída em uma rede com uma única camada intermediária, pode-se expressar matematicamente:

$$u_i = \sum_{j=1}^J w_{ij} \cdot x_j + b_i = \mathbf{w}_i^T \cdot \mathbf{x} + b_i \quad (4.2)$$

$$y_i = f(u_i) \quad (4.3)$$

Portanto, considerando-se que a saída esteja normalizada, ela pode assumir um valor qualquer compreendido no intervalo entre -1 e 1. Esse tipo de rede é chamado de *associador*:

$$\mathbf{x}^k \rightarrow \mathbf{y}^k, y_i^k \in (-1, +1) \quad (4.4)$$

No entanto, a rede pode ser usada como um *classificador*, que é o caso a que este trabalho se propõe. Dessa forma, para um vetor de entrada \mathbf{x} , a saída y_i assume um valor lógico que representa a classe C_i a que a entrada está associada.

$$\mathbf{x}^k \rightarrow \mathbf{y}^k, y_i^k \in \{-1, +1\} \quad (4.5)$$

Assim,

$$\tilde{y}_i = \text{sign}(u_i) = \begin{cases} +1 & u_i \geq 0 \\ -1 & u_i < 0 \end{cases} \quad (4.6)$$

Portanto, para um dado vetor de entradas \mathbf{x} ,

$$u_i > 0 \Rightarrow y_i > 0 \Rightarrow \mathbf{x} \in C_i \quad (4.7)$$

$$u_i < 0 \Rightarrow y_i < 0 \Rightarrow \mathbf{x} \notin C_i \quad (4.8)$$

A curva $u_i = 0$ define um separador entre as classes. Sejam duas classes dispersas no plano bi-dimensional conforme a Figura 4.4.

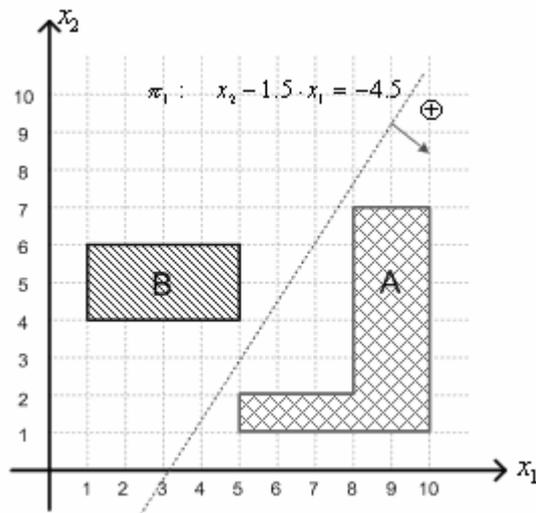


Figura 4.4 – Classes linearmente separáveis.

A reta separadora é definida por $u = 0.1 \cdot x_2 - 0.15 \cdot x_1 + 0.45$ e a saída é dada por $\tilde{y} = \text{sign}(u)$. Em outras palavras, uma entrada situada no semi-plano direito limitado pela reta $x_2 - 1.5 \cdot x_1 = -4.5$ pertence à classe **A**; caso contrário a entrada deverá pertencer à classe **B**.

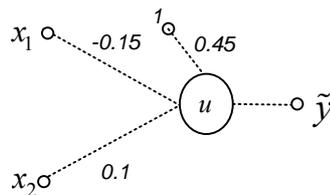


Figura 4.5 – Rede Neural para discriminação entre as classes A e B linearmente separáveis.

O vetor w_i define a direção ótima sobre a qual projetar x para discriminar a classe. O *bias* b_i é o nível ótimo de discriminação. A Figura 4.6 mostra os elementos das classes **A** e **B** projetados na direção de w . A interseção da reta $x_2 - 1.5 \cdot x_1 = -4.5$ com a direção de w é o ponto de separação entre as classes.

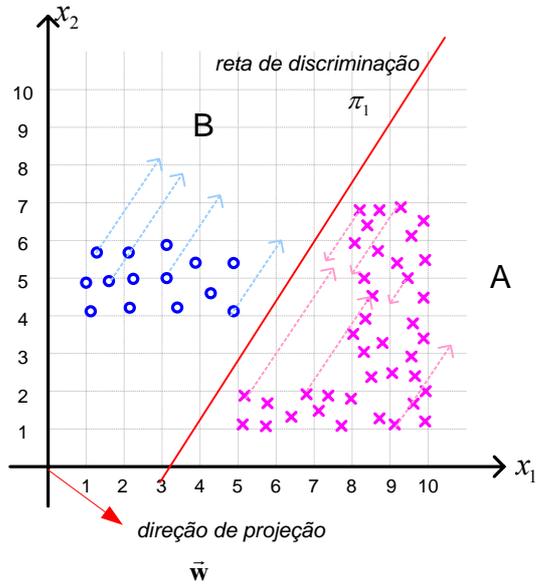


Figura 4.6 – Projeção dos elementos das classes na direção de w .

Se as classes não são linearmente separáveis, uma camada não é suficiente para separá-las. Sejam agora duas classes **A** e **B** conforme a Figura 4.7

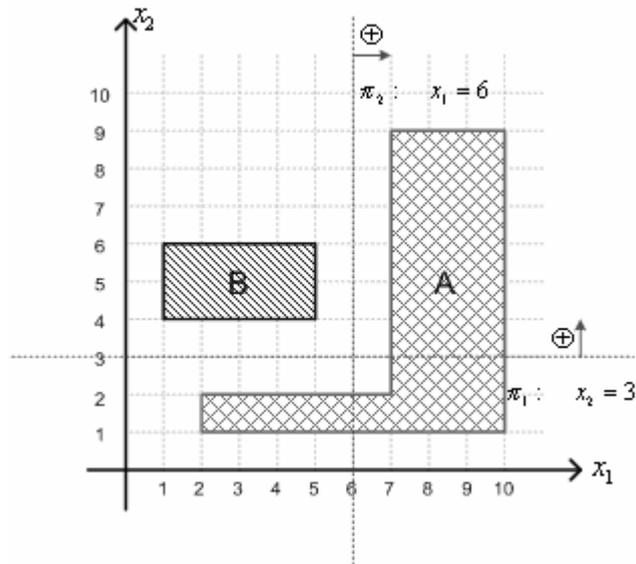


Figura 4.7 – Classes linearmente não-separáveis.

A camada de saída separa classes linearmente separáveis. A camada escondida mapeia classes não linearmente separáveis em classes linearmente separáveis. Caso haja um número suficiente de neurônios na camada escondida, ela projeta as classes no vértice de um hipercubo unitário.

Para as classes da Figura 4.7:

$$u_1 = 0.1 \cdot x_2 - 0.3 \quad v_1 = \text{sign}(u_1)$$

$$u_2 = 0.1 \cdot x_1 - 0.6 \quad v_2 = \text{sign}(u_2)$$

Uma entrada situada nos semi-planos positivos definidos pelas retas π_2 e π_1 mostradas na Figura 4.7 será mapeada no ponto $(v_1, v_2) = (+1, +1)$. Uma entrada situada no semi-plano positivo definido pela reta π_2 e no semi-plano negativo de π_1 será mapeada no ponto $(v_1, v_2) = (-1, +1)$. Uma entrada situada no semi-plano positivo definido pela reta π_1 e no semi-plano negativo de π_2 será mapeada no ponto $(v_1, v_2) = (+1, -1)$. Uma entrada situada nos semi-planos negativos definidos pelas retas π_2 e π_1 será mapeada no ponto $(v_1, v_2) = (-1, -1)$.

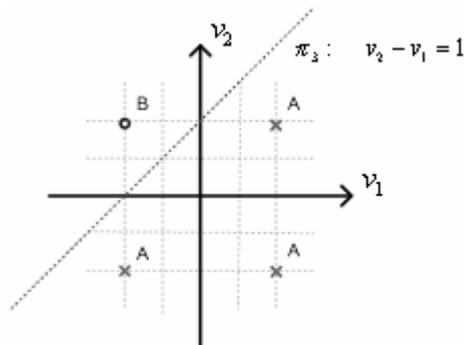


Figura 4.8 – Mapeamento das classes A e B no hipercubo unitário.

Assim, uma entrada irá pertencer a uma classe, de acordo com a tabela abaixo:

Tabela 4.1 – Classes A e B

v_1	v_2	classe	\tilde{y}
+	+	A	+
+	-	A	+
-	+	B	-

$$- \quad | \quad - \quad | \quad \mathbf{A} \quad | \quad +$$

Assim a reta de separação entre as classes pode ser definida como:

$$u_3 = 0.1 \cdot u_2 - 0.1 \cdot u_1 - 0.1 \quad \tilde{y} = \text{sign}(u_3)$$

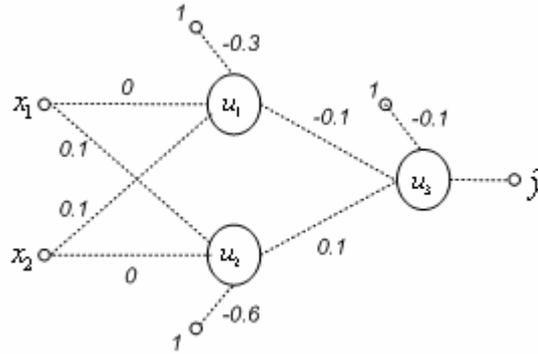


Figura 4.9 – Rede Neural de separação entre as classes A e B linearmente não separáveis.

4.1.2 Treinamento

Para um dado par de entradas-saídas $(\mathbf{x}^k, \mathbf{y}^k)$, o erro na saída da rede é dado por:

$$\varepsilon^{2k} = \|\mathbf{y}^k - \tilde{\mathbf{y}}^k\|^2 = \sum_{l=1}^L (y^k - \tilde{y}^k)^2 \quad (4.9)$$

O treinamento da rede consiste, pois, em minimizar o erro médio quadrático da saída, ou seja, $F_0 = E[|\mathbf{y} - \tilde{\mathbf{y}}|^2] = E(\varepsilon^{2k}) = \frac{1}{P} \sum_{k=1}^P \varepsilon^{2k}$ deve ser mínimo.

O método clássico de minimização de uma função é o gradiente descendente.

$$\nabla_{\mathbf{w}} F_0 = \frac{\partial F_0}{\partial w_{ij}} \quad \therefore \quad \Delta \mathbf{w} = -\alpha \cdot \nabla_{\mathbf{w}} F_0 \quad (4.10)$$

Como a rede é um classificador, tem-se que:

$$F = E[|\mathbf{y} - \tilde{\mathbf{y}}|^2] = E[|\mathbf{y} - \text{sign}(\mathbf{u})|^2] \quad (4.11)$$

No entanto a função $\text{sign}(u)$ não é uma função contínua e, portanto, tem derivada infinita quando u tende a 0.

Seja $\tilde{y} = f(u)$ uma função monótona crescente, limitada entre dois valores quaisquer a e b , pode-se provar que o minimante do valor esperado de $|y - \tilde{y}|^2$ é o mesmo para $|y - \text{sign}(u)|^2$ [40]. Matematicamente:

$$y \in \{a, b\} \text{ e } y = f(u), \text{ tal que } f(u) \in (a, b).$$

Supondo que $\frac{d}{du} f(u) \geq 0$ e, ainda, $\lim_{u \rightarrow -\infty} f(u) = a$ e $\lim_{u \rightarrow +\infty} f(u) = b$ então:

$$\arg(\min(F_0)) \Rightarrow \arg(\min(F))$$

$$\arg(\min(F_0)) \not\Leftarrow \arg(\min(F))$$

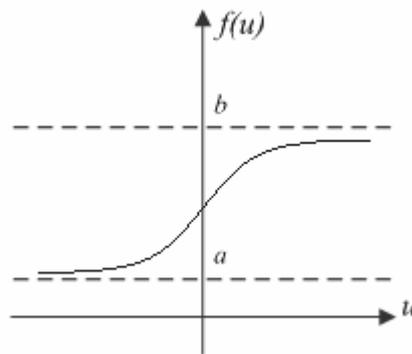


Figura 4.10 – Função de ativação do tipo tangente hiperbólica: função monótona crescente limitada.

Conclui-se, portanto, que minimizar F é equivalente a minimizar F_0 , desde que a saída seja uma função monótona crescente e limitada. Portanto, pode-se treinar a rede, usando como função de ativação da saída a tangente hiperbólica e operá-la com uma função sinal, sem qualquer prejuízo.

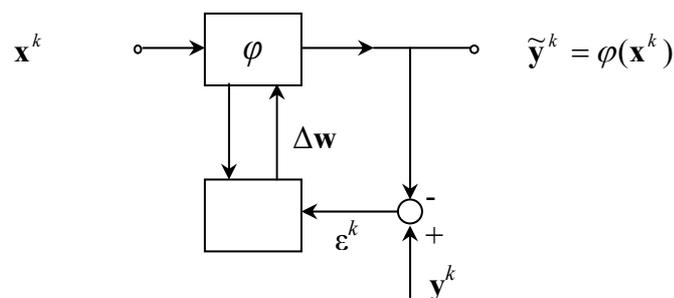


Figura 4.11 – Esquema de treinamento.

O gradiente de uma função aponta na direção oposta a do seu mínimo. Assim, basta incrementar \mathbf{w} um vetor proporcional ao oposto do gradiente da função que se deseja minimizar:

$$\mathbf{w} \leftarrow \mathbf{w} + \Delta \mathbf{w} \quad (4.12)$$

onde $\Delta \mathbf{w} = -\alpha \cdot \nabla_{\mathbf{w}} F_0$.

O gradiente da função erro F_0 é ser calculado derivando-se a função F_0 com relação a w_{ij}

$$\begin{aligned} \nabla_{\mathbf{w}} F_0 &= \nabla_{\mathbf{w}} E(\varepsilon^{2k}) = \frac{\partial F_0}{\partial w_{ij}} = \frac{\partial}{\partial w_{ij}} E(\varepsilon^{2k}) = \frac{\partial}{\partial w_{ij}} \frac{1}{P} \sum_{k=1}^P \varepsilon^{2k} = \\ &= \frac{1}{P} \sum_{k=1}^P \frac{\partial}{\partial w_{ij}} \varepsilon^{2k} = E(\nabla_{\mathbf{w}} \varepsilon^{2k}) \end{aligned} \quad (4.13)$$

Verifica-se que o gradiente do valor esperado do erro quadrático é igual ao valor esperado do gradiente do erro quadrático, enquanto \mathbf{w} for mantido constante. Assim, basta calcular a derivada do erro quadrático com relação a w_{ij} em cada passo:

$$\frac{\partial}{\partial w_{ij}} \varepsilon^{2k} = \frac{\partial}{\partial w_{ij}} \sum_l (\varepsilon_l^k)^2 = \sum_l 2 \cdot \varepsilon_l^k \cdot \frac{\partial \varepsilon_l^k}{\partial w_{ij}} = -2 \cdot \sum_l \varepsilon_l^k \cdot \frac{\partial \tilde{y}_l^k}{\partial w_{ij}} \quad (4.14)$$

Pode-se mostrar [40] que uma rede de 2 camadas é capaz de realizar qualquer mapeamento, e que, exceto em casos excepcionais, não há necessidade de inserir outra camada intermediária na rede. Uma rede genérica com J entradas, I neurônios na camada intermediária e L neurônios na camada de saída está desenhada na Figura 4.12:

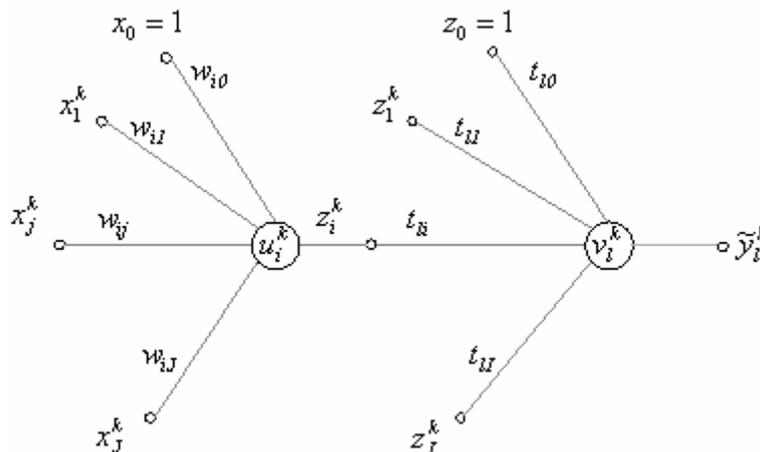


Figura 4.12 – Rede com 2 camadas.

Como cada saída só depende dos neurônios ligados a ela, pode-se expandir, pela regra da cadeia, para os pesos entre a camada intermediária e a camada de saída:

$$\frac{\partial F_0}{\partial t_{li}} = -2 \cdot \varepsilon_l^k \cdot \frac{\partial \tilde{y}_l^k}{\partial v_l^k} \cdot \frac{\partial v_l^k}{\partial w_{ij}} \quad (4.15)$$

$$v_l^k = \sum_{i=0}^I t_{li} \cdot z_i^k \Rightarrow \frac{\partial v_l^k}{\partial t_{li}} = z_i^k \quad (4.16)$$

Para redes usadas como classificadores, a camada de saída deverá do tipo tangente hiperbólica. :

$$y_l^k = \tanh(v_l^k) \Rightarrow \frac{\partial y_l^k}{\partial v_l^k} = 1 - (y_l^k)^2 \quad (4.17)$$

$$\text{Assim, } \frac{\partial F_0}{\partial t_{li}} = -2 \cdot E[z_i^k \cdot g_l^k], \text{ onde } g_l^k = \varepsilon_l^k \cdot (1 - (y_l^k)^2)$$

$$\boxed{\Delta t_{li} = 2 \cdot \alpha \cdot E[z_i^k \cdot g_l^k]} \quad (4.18)$$

Para os pesos entre a entrada e a camada intermediária:

$$\frac{\partial F_0}{\partial w_{ij}} = -2 \cdot \sum_l \varepsilon_l^k \cdot \frac{\partial \tilde{y}_l^k}{\partial u_i^k} \cdot \frac{\partial u_i^k}{\partial w_{ij}} \quad (4.19)$$

$$u_i^k = \sum_{j=0}^J w_{ij} \cdot x_j^k \Rightarrow \frac{\partial u_i^k}{\partial w_{ij}} = x_j^k \quad (4.20)$$

Assim:

$$\frac{\partial F_0}{\partial w_{ij}} = -2 \cdot \sum_l \varepsilon_l^k \cdot x_j^k \cdot \frac{\partial \tilde{y}_l^k}{\partial u_i^k} = -2 \cdot x_j^k \cdot \delta_i^k, \quad (4.21)$$

$$\text{e } \boxed{\Delta w_{ij} = 2 \cdot \alpha \cdot E[x_j^k \cdot \delta_i^k]}, \quad (4.22)$$

onde $\delta_i^k = \sum_{l=1}^L \varepsilon_l^k \cdot \frac{\partial \tilde{y}_l^k}{\partial u_i^k} = \sum_{l=1}^L g_l^k \cdot t_{li} \cdot (1 - (z_i^k)^2)$, uma vez que os neurônios da camada intermediária são do tipo tangente hiperbólica.

O treinamento como foi exposto chama-se *batelada*, ou em inglês, *batch*. Esse tipo de treinamento é extremamente lento, pois todos os exemplos do conjunto de

treinamento precisam ser apresentados à rede para que o erro médio seja calculado e a partir deles sejam feitas as correções dos pesos:

$$\Delta \mathbf{w} = -\alpha \cdot \nabla_{\mathbf{w}} F_0 = -\alpha \cdot E\left(\nabla_{\mathbf{w}} \varepsilon^{2k}\right) = -\alpha \cdot \frac{1}{P} \sum_{k=1}^P \frac{\partial}{\partial w_{ij}} \varepsilon^{2k} \quad (4.23)$$

Uma forma alternativa de treinamento é a chamada *regra delta*, pela qual a correção dos pesos acontece a cada apresentação à rede de um exemplo do conjunto de treinamento. Cada correção de pesos baseia-se somente no erro do exemplo apresentado naquela iteração. Assim, para uma rede com P pares entrada-saída, em cada ciclo ocorrem P correções. Esse tipo de treinamento, embora rápido, apresenta muita oscilação.

$$\Delta \mathbf{w} = -\alpha \cdot \nabla_{\mathbf{w}} F_0 = -\alpha \cdot \nabla_{\mathbf{w}} \varepsilon^{2k} = -\alpha \cdot \frac{\partial}{\partial w_{ij}} \varepsilon^{2k} \quad (4.24)$$

Outra forma de treinamento, denominada *época*, calcula o valor esperado de um subconjunto dos pares de entrada-saída. Assim, os P pares entrada-saída são agrupados em um subconjunto de P' pares. Após esses P' serem apresentados à rede, as sinapses são adaptadas como a média para esse subconjunto:

$$\Delta \mathbf{w} = -\alpha \cdot \frac{1}{P'} \sum_{k=1}^{P'} \frac{\partial}{\partial w_{ij}} \varepsilon^{2k} \quad (4.25)$$

Esse procedimento é útil quando $P > 200$.

Uma forma intermediária entre a regra delta e a época, chama-se *momento*. Por essa técnica, a sinapse é adaptada como uma soma ponderada dos deltas atual e anterior.

$$\Delta \mathbf{w}_{novo} = \beta \cdot \Delta \mathbf{w}_{anterior} - (1 - \beta) \cdot \alpha \cdot \frac{\partial}{\partial w_{ij}} \varepsilon^{2k} \quad (4.26)$$

Usualmente, β é escolhido em torno de 0.9.

4.2 - Componentes de Atuação Conjunta

É bastante comum que a informação que se deseja discriminar esteja dispersa em diversas variáveis ou entradas para a rede neural. Isso dificulta bastante o trabalho da rede neural. É, pois, usual comprimir a informação relevante em algumas poucas componentes antes de apresentá-las à rede. Existem vários critérios para seleção dessas componentes. Um dos critérios mais usuais são as chamadas componentes principais,

que são uma projeção do conjunto de variáveis originais em um subespaço ortogonal, de forma que as componentes estejam em ordem crescente de energia. Essas componentes podem ser obtidas pela fórmula fechada, conforme apresentado no Apêndice B, ou por uma rede neural composta por neurônios lineares que realizem a decomposição, ou *análise*, das entradas respeitando a ordenação por variância.

Essa abordagem, no entanto, não garante que as componentes de maior energia sejam as mais relevantes para discriminação das entradas.

Propõe-se então, uma forma alternativa de treinamento da Rede Neural, que pretende realizar um mapeamento linear das entradas, considerando sua capacidade discriminatória, acoplando-se à rede de decomposição linear uma rede com neurônios lineares que realizem a classificação da entrada.

4.2.1 Treinamento

O treinamento da rede foi resumido nos passos que se seguem.

Passo 1:

- uma rede neural com 3 camadas é estruturada conforme a Figura 4.13

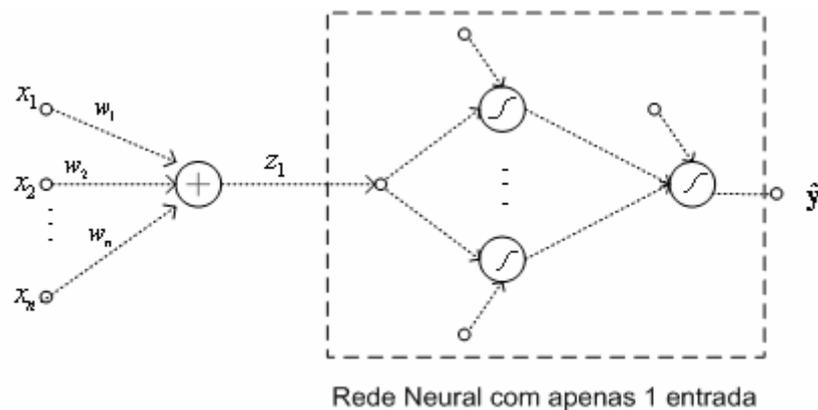


Figura 4.13 – Passo 1.

A primeira camada tem como objetivo comprimir a informação em uma única componente, considerando, no entanto, sua relevância para classificação da entrada.

A segunda camada mapeia a componente selecionada pela primeira no vértice do hiper-cubo unitário e a terceira encontra os separadores ótimos que classificarão a entrada, conforme apresentado na seção 4.1.1.

- todas as sinapses são treinadas conforme algum dos métodos descritos na seção 4.1.2 até convergência a um valor mínimo ou um número máximo de passos de treinamento.

Passo 2:

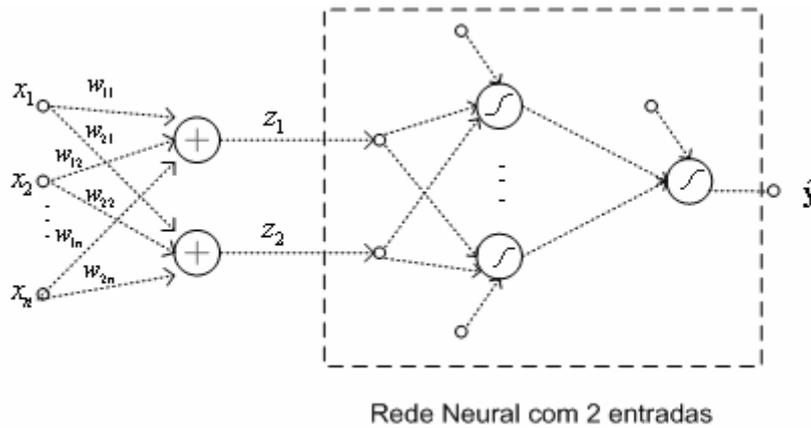


Figura 4.14 – Passo 2.

- subtrai-se da entrada sua projeção na primeira componente:

$$\mathbf{x} \leftarrow \mathbf{x} - \frac{1}{|\mathbf{w}_1|^2} \cdot \underbrace{\mathbf{w}_1^T \cdot \mathbf{x}}_{z_1} \cdot \mathbf{w}_1 \quad (4.27)$$

- \mathbf{w}_1 é congelado.
- acrescenta-se um novo neurônio linear à primeira camada, que representa a segunda componente de discriminação, gerando assim uma nova entrada para a rede neural não-linear.
- toda a rede neural não linear, cujas entradas são \mathbf{z}_1 e \mathbf{z}_2 , é submetida a um novo treinamento, inclusive as sinapses conectadas a \mathbf{z}_1 .

Passo 3:

- subtrai-se, então, da entrada sua projeção na segunda componente:

$$\mathbf{x} \leftarrow \mathbf{x} - \frac{1}{|\mathbf{w}_2|^2} \cdot \underbrace{\mathbf{w}_2^T \cdot \mathbf{x}}_{z_2} \cdot \mathbf{w}_2 \quad (4.28)$$

- w_1 e w_2 são congelados.
- acrescenta-se um novo neurônio linear na primeira camada, que representa a terceira componente de discriminação, gerando assim uma nova entrada para a rede neural não-linear.
- toda a rede neural não linear, cujas entradas são z_1 e z_2 e z_3 , é submetida a um novo treinamento, inclusive as sinapses conectadas a z_1 e z_2 .

E assim sucessivamente até um número máximo de neurônios na primeira camada igual ao número de entradas.

5. Sistema de diagnóstico

5.1 - Materiais

As imagens utilizadas no experimento foram cedidas pelo Instituto Nacional do Câncer (INCa - RJ), e somam um total de 85 mamografias, em incisões crânio-caudal e médio-lateral.

As mamografias foram digitalizadas por um scanner Labscan V4.01 da *Image Scanner*, com resolução de 1200 dpi e quantizadas a 8 bits por *pixel*.

A Tabela 5.1 traz o diagnóstico das mamografias utilizadas. O diagnóstico histológico, quando existente, provém de um exame realizado no INCa e anexado ao prontuário da paciente. O laudo radiológico foi dado pela Dra.Carolina Maria de Azevedo, radiologista do INCa.

Com o auxílio da radiologista, foram selecionadas 410 regiões de interesse de tamanho 1200x1200 *pixels*, o que corresponde a regiões de cerca de 1 polegada² na mamografia original.

O conjunto selecionado abrange:

-205 regiões saudáveis contendo amostras de parênquima, gordura, músculo peitoral;

-205 regiões contendo o cerne das lesões identificadas pela radiologista.

Tabela 5.1 – Diagnóstico Radiológico e Histológico das mamografias utilizadas

Mamografia	Diagnóstico		ROIs
	Radiológico	Histológico	
050601_CC_E	Suspeita	Nº incompatível	105,106,107,108,109,110,111,112,113,130,131
050601_ML_D	Calcificações parasitárias e mastite	Nº incompatível	114,115,116,132,133
1200_ML_E	Hematoma calcificado	Nº incompatível	58
121205_CC_E	Nódulo com mcs no interior, na RRA - ME, medindo 13 mm no maior eixo, apresentando contorno regular e limites pouco definidos	Doença fibrocística com hiperplasia ductal	181,182,183,184,185,186
123593_CC_E	Área de densidade assimétrica, associada à distorção arquitetural	Sem exame no INCa	100,101
123593_ML_E	Suspeita	Sem exame no INCa	103,104

1257924_CC_E	Suspeita	Sem exame no INCa	5,6,7,8,9,10,11,12,13,14,15,16,17,18,19
1269938_CC_E	Pele, tecido celular subcutâneo e complexo areolo-papilar de aspecto normal. Parênquima mamário quase inteiramente liposubstituído. Nódulo de bordo irregular e limites definidos em QSE medindo em seu maior eixo 1 cm. Calcificação grosseira na união dos quadrantes.	Sem exame no INCa	20
1292946_ML_E	Pele, tecido celular subcutâneo e complexo mamilo-areolar sem alterações. Parênquima mamário quase inteiramente liposubstituído. Ausência de nódulos atípicos e calcificações tumorais. Calcificações grosseiras polimórficas de aspecto benigno esparsas.	Sem exame no INCa	51,52,53
164429_ML_D	Carcinoma ductal infiltrativo	Segmento da mama direita com hiperplasia epitelial ductal com atipias, lesão esclerosante radial e alterações fibrocísticas. Presença de freqüentes microcalcificações.	35
164617_CC_E	Suspeita	Sem exame no INCa	96,97
180787_CC_D	Esteatonecrose (lesão grande) pequenas calcificações benignas	Sem exame no INCa	30,31
180787_ML_D	Esteatonecrose (lesão grande) pequenas calcificações benignas	Sem exame no INCa	34
184775_CC_E	Microcalcificações tipo IV e V no QII da mama esquerda	CDI com predomínio <i>in situ</i> , padrão cribiforme e comedo	147,148
184775_ML_E	Microcalcificações tipo IV e V no QII da mama esquerda	CDI com predomínio <i>in situ</i> , padrão cribiforme e comedo	149,150
195017_CC_D	Microcalcificações tipo II no QSE da mama direita	Mastopatia fibrocística com microcalcificações nos ductos normais	204
19709_CC_D	Cisto calcificado (lesão maior) esteatonecrose (lesão menor)	Nº incompatível	27,28,29
198244_CC_E	Microcalcificações tipo IV no UQQI - ME	Fibrose com Microcalcificações	177,178
207238_ML_E	Fibroadenoma com calcificações benignas	Mastopatia fibrocística da mama com áreas de adenose e microcalcificações	121,122,123,124,125
214201_CC_D	Densidade assimétrica e calcificações no QSE da mama direita	Alterações fibrocísticas com adenose esclerosante. Hiperplasia. Microcalcificações	164,165,166
214201_ML_D	Densidade assimétrica e calcificações no QSE da mama direita	Alterações fibrocísticas com adenose esclerosante. Hiperplasia. Microcalcificações	167,168,169,170,171,172,173,174,175,176
218550_CC_E	Nódulos com suspeitas de malignidade	Sem exame no INCa	102,129
220836_ML_D	Adenose e calcificação grosseira	Sem exame no INCa	57,70
221522_CC_E	Fibroadenoma com calcificações benignas	Sem exame no INCa	1,2,3,4
231051_ML_D	Calcificações benignas e mama extranumerária	Sem exame no INCa	54,55,56
233909_CC_D	Microcalcificações tipo III no QSE da mama esquerda	CID	145,146

233909_ML_D	Microcalcificações arenosas (suspeitas)	Alterações fibrocísticas da mama com hiperplasia ductal epitelial leve, trechos de hialinose do estroma e microcalcificações. Ausência de malignidade	81,82
241151_CC_D	Tumor maligno com microcalcificação	Sem exame no INCa	41,42,43
246047_CC_E	Suspeita	Sem exame no INCa	45
246443_CC_E	Microcalcificações tipo II mama esquerda	Mastopatia fibrocística. Adenose esclerosante. Hiperplasia ductal moderada	187,188,189,190,191,192,193,194
246851_CC_E	Calcificação grosseira	Tecido mamário exibindo esteatonecrose e focos de calcificação. Ausência de malignidade	32,33
248253_CC_E	Linfonodo intramamário e calcificações vasculares	Maligna (Paciente de 81 anos, cirurgia conservadora com esvaziamento axilar e radioterapia esquerda em 1995, exame de seguimento)	126,127
248253_ML_E	Linfonodo intramamário e calcificações vasculares	Maligna (Paciente de 81 anos, cirurgia conservadora com esvaziamento axilar e radioterapia esquerda em 1995, exame de seguimento)	45,47,48,49,50
249652_CC_E	Micro calcificações tipo II e III de LEGAL esparsos na mama esquerda	Mamas simétricas. Estroma fibro-adiposo normal. Microcalcificações tipo II e III de LEGAL esparsos na mama esquerda. As microcalcificações da mama esquerda aumentaram em numero em comparação ao exame anterior de 02.08.1998	60,61,62,63,64,65,66,67,68,69,138,139,140,141,142,143,144
249797_CC_D	Microcalcificações monomórficas e isodensas, no QSE - MD	Mastopatia fibrocística	155,156
249797_ML_D	Microcalcificações monomórficas e isodensas, no QSE - MD	Mastopatia fibrocística	152,153,154
251652_CC_D	Microcalcificações tipo II no QIE da mama direita	Mastopatia fibrocística com focos de microcalcificações nos ductos	201,202,203
251652_ML_D	Microcalcificações tipo II no QIE da mama direita	Mastopatia fibrocística com focos de microcalcificações nos ductos	158
252032_CC_D	Nódulo suspeito, calcificações (superior externo) calcificações vasculares e benignas	Presença de êmbolo de carcinoma ductal. Hiperplasia moderada. Fibroadenoma. Metaplasia apócrina. Microcalcificações	26
252032_CC_E	Calcificações com baixos índices de suspeitas para malignidade (externo) e benignas (interno)	Presença de êmbolo de carcinoma ductal. Hiperplasia moderada. Fibroadenoma. Metaplasia apócrina. Microcalcificações	90,91
252032_ML_D	Nódulo suspeito, calcificações (superior externo), calcificações vasculares e benignas	Presença de êmbolo de carcinoma ductal. Hiperplasia moderada. Fibroadenoma. Metaplasia apócrina. Microcalcificações	25
252032_ML_E	Nódulo sugestivo de linfonodo e microcalcificações (quadrante externo)	Presença de êmbolo de carcinoma ductal. Hiperplasia moderada. Fibroadenoma. Metaplasia apócrina. Microcalcificações	92,93

257829_CC_D	Suspeita	Alterações fibrocísticas da mama com hiperplasia epitelial ductal leve e área de esclerohialinose. Ausência de malignidade	78,79
257829_ML_D	Microcalcificações polimrficas (quadrante superior), Calcificação grosseira de aspecto benigno)	Alterações fibrocísticas da mama com hiperplasia epitelial ductal leve e área de esclerohialinose. Ausência de malignidade	75,76,77
265430_ML_D	Microcalcificações, no QSE - MD, puntiformes, isodensas	Alterações fibrocísticas com áreas de hiperplasia epitelial ductal e lobular, com atipias livres.Processo inflamatório crônico granulomatoso.	159,160161,162,163
265430_ML_E	Microcalcificações, no QSE - ME, puntiformes, isodensas	Alterações fibrocísticas com áreas de hiperplasia epitelial ductal e lobular, com atipias livres	157
266121_CC_D	Nódulo sugestivo de linfonodo e Microcalcificações suspeitas (quadrante externo)	Sem exame no INCa	83,84
266121_CC_E	Microcalcificações suspeitas (quadrante externo) e benignas (quadrante interno)	Sem exame no INCa	85,86,87
266121_ML_D	Linfonodo intramamário e calcificações vasculares e Microcalcificações arenosas (suspeitas)	Sem exame no INCa	71
266121_ML_E	Nódulo suspeito e calcificações grosseiras	Sem exame no INCa	59
267664_CC_D	Suspeita	Carcinoma intra ductal de padrão cribriforme e micropapilífero de alto grau, presente em região retroareolar e quadrante inferior interno. Doença de Paget do mamilo. Limites cirúrgicos e dezesseis linfonodos livres de neoplasia. Parênquima não neoplásico	98,99
267699_ML_D	Microcalcificações pleomórficas no QSI e QSE da MD tipo IV	Mama com atrofia das glândulas mamarias, com focos de mastopatia fibrocística. Ausência de malignidade (QSI). Mama com atrofia glandular e extensa fibrose do estroma. Ausência de malignidade (QSE)	179,180
271320_CC_D2	Suspeita	Mama sem malignidade. Limites livres de neoplasia. Presença de microcalcificações ao lado de ductos e lóbulos bem constituídos	94,95,117,118
271320_ML_D	Suspeita	Mama sem malignidade. Limites livres de neoplasia. Presença de microcalcificações ao lado de ductos e lóbulos bem constituídos	119,120
271320_ML_E	Microcalcificação suspeitas (superior) e calcificações vasculares	Carcinoma ductal infiltrante de mama esquerda moderadamente diferenciado (grau II de Elston)	128
271976_CC_E	Microcalcificações QSE - ME tipo IV e V		151
272327_CC_E	Tumor maligno com microcalcificação e calcificações benignas	Carcinoma ductal infiltrante, grau II de Elston	38,39,40
272327_ML_D	Nódulo suspeito e calcificações vasculares (inferior)	Linfonodo livre de neoplasia. Negativo para malignidade	24

272327_ML_E	Tumor maligno com microcalcificação e calcificações benignas	Carcinoma ductal infiltrante, grau II de Elston	36,37
275552_CC_E	Esteatonecrose e microcalcificações suspeitas no quadrante interno	Carcinoma ductal infiltrante grau I de Elston (grau nuclear 2 e baixo índice mitótico) com áreas intraductais do tipo sólido de segmento de mama esquerda medindo 1, 2x1, 0x0, 8cm com dois limites cirúrgicos comprometidos por carcinoma infiltrante e "in situ"	88
275552_ML_E	Esteatonecrose e microcalcificações suspeitas (superior)	Carcinoma ductal infiltrante grau I de Elston (grau nuclear 2 e baixo índice mitótico) com áreas intraductais do tipo sólido de segmento de mama esquerda medindo 1, 2x1, 0x0, 8cm com dois limites cirúrgicos comprometidos por carcinoma infiltrante e "in situ"	89
281447_CC_E	Pele, tecido celular subcutâneo e complexo areolo-mamilar sem alterações. Mama predominantemente densa. Ausência de nódulos atípicos. Calcificações arredondadas esparsas na mama, sem alterações significativas em relação ao exame anterior (27/ set 1999).	Sem exame no INCa	22,23,134,135,136,137
303083_CC_E	Suspeita	Maligna com mastectomia	198,199,200
303083_ML_E	Suspeita	Maligna com mastectomia	205
310438_CC_E	Mamas predominantemente densas, sem evidência de Microcalcificações tumorais. Revestimento cutâneo e complexo areolo-papilar esquerdo sem alterações. Calcificações de aspecto benigno na mama esquerda.	Sem exame no INCa	74
310438_ML_E	Mamas predominantemente densas, sem evidencia de microcalcificações tumorais. Revestimento cutâneo e complexo areolo-papilar esquerdo sem alterações. Calcificações de aspecto benigno na mama esquerda.	Sem exame no INCa	80
31500_ML_E	Suspeita	Tumor medular bi-lateral síncrono	73
316851_CC_E	Suspeita	Sem exame no INCa	44
611006_ML_D	Suspeita	Carcinoma ductal infiltrante, multifocal, com áreas <i>in situ</i> em meio a hiperplasia intraductal atípica. Limites cirúrgicos checados em congelação	195,196,197
76130_ML_E	Nódulo em mama esquerda	Carcinoma intraductal tipo comedo carcinoma (HUSF)	21
JMaria_ML_D	Suspeita	Tumor maligno superior retroglandular	72

5.2 - Intuição sobre o método

No capítulo 2 foram descritas diversas mastopatias e sua apresentação na imagem mamográfica. Uma análise comparativa revela que características bastante sutis podem denunciar a presença de uma lesão, e dificultam a generalização de um algoritmo de detecção de anormalidade no tecido.

Pretende-se construir um algoritmo de CAD capaz de percorrer a imagem e identificar quaisquer regiões que contenham lesões. Não há, porém, um padrão para o tipo de lesão que se quer identificar. As regiões lesionadas podem conter desde calcificações grosseiras e benignas a áreas de densidade assimétrica mascarando a presença de microcalcificações. Em outros casos, calcificações puntiformes confundem-se com o parênquima.

A Figura 5.1 revela exemplos de regiões de interesse selecionadas.

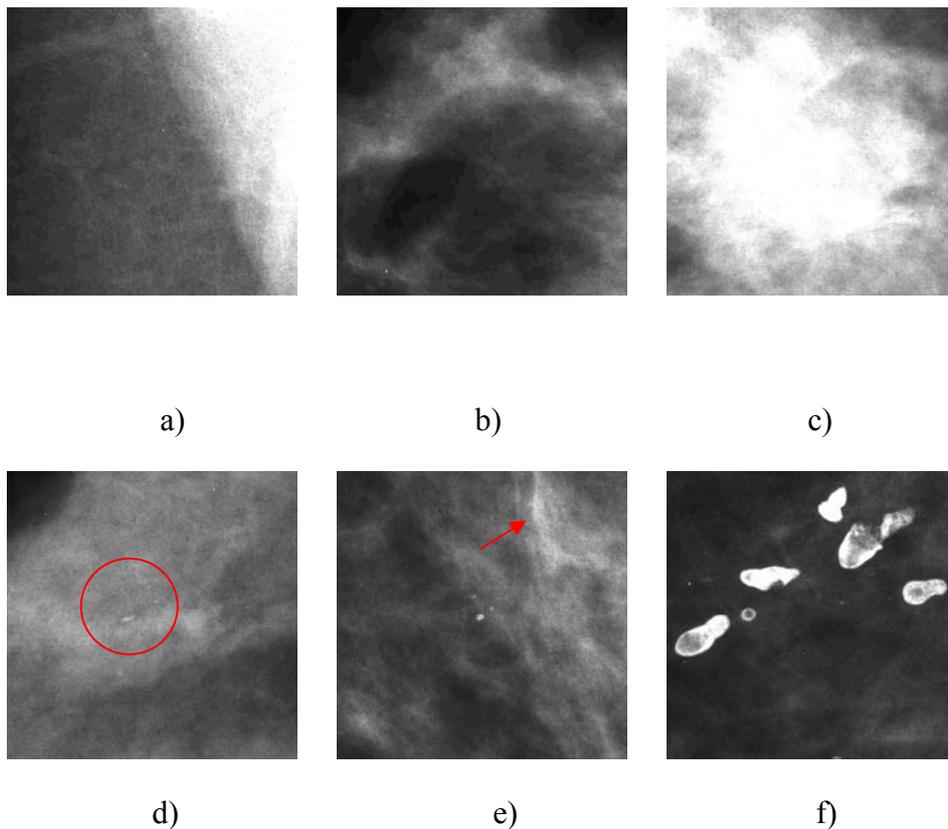


Figura 5.1 - ROIs a) normal contendo parte do músculo peitoral(r186); b) loja de gordura (r130); c) parênquima livre(r15); d) área contendo microcalcificações benignas (r82); e) carcinoma ductal infiltrante; r(128) f) calcificações grosseiras (r11).

Não foi considerada anormalidade, a presença de vasos calcificados na imagem, pois esta é uma característica sem relevância clínica na detecção de tumores da mama e bastante comum, sobretudo em mulheres de idade avançada.

Conforme apresentado no capítulo 1, a literatura contém diversos métodos para diferenciar regiões saudáveis e lesionadas em uma mamografia, baseados na extração de parâmetros da imagem que sejam capazes de caracterizar a região de interesse selecionada [19]. Em sua grande maioria, esses parâmetros são descritores quantitativos da textura da região, que são submetidos a algum método de discriminação, como redes neurais, para fornecer um diagnóstico acerca do potencial de malignidade da região analisada.

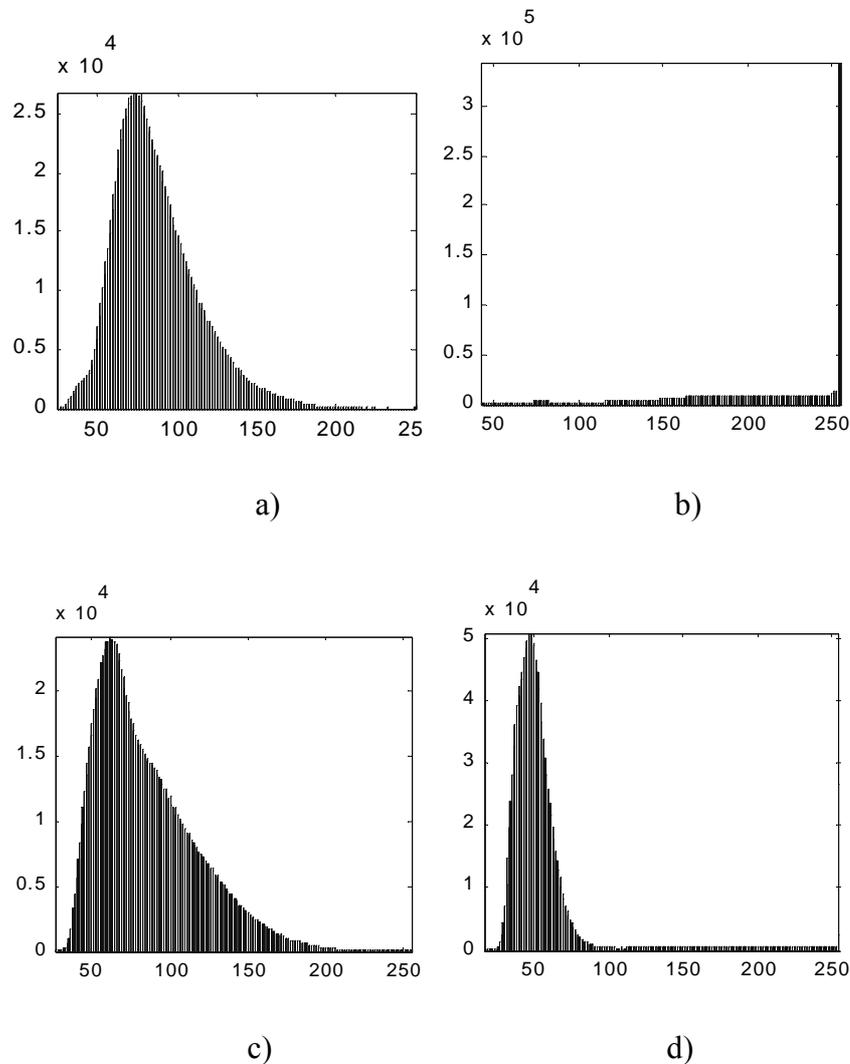


Figura 5.2 - Histogramas correspondentes a:
a), b) ROIs normais apresentadas na Figura 5.1 (b) e (c), respectivamente;
c), d) ROIs lesionadas apresentadas na Figura 5.1 (e) e (f), respectivamente.

Poderia se imaginar que uma análise dos histogramas seria capaz de fornecer informações relevantes acerca das características da imagem, conforme demonstram alguns estudos na literatura [25, 41, 42].

Os histogramas apresentados na Figura 5.2 mostram que a distribuição de probabilidade das ROI também não revela um potencial de separação aparente entre regiões saudáveis e lesionadas. Os resultados para a matriz de co-ocorrência, que se baseia em histogramas bidimensionais para descrição de textura da imagem, são, para o conjunto de imagens em questão, bem pobres em comparação com os demais métodos, conforme será apresentado no Capítulo 6.

Na verdade, dada a variabilidade dos tipos de lesão que podem ocorrer, aliada ao biomorfismo da mulher que altera as características da imagem mamográfica, é muito difícil definir um padrão de textura que caracterize uma anormalidade. Em algumas mulheres, a aparência do parênquima simula a presença de microcalcificações e induz a um diagnóstico falso. Em outros casos, mamas densas mascaram a existência de lesões, que podem ser descobertas apenas quando o tumor tiver evoluído a um estágio palpável.

O próprio tipo do filme usado na radiografia, e do mamógrafo utilizado influenciam na imagem.

Todos esses fatores não conseguem ser eliminados pelos métodos tradicionais de extração de parâmetros, que analisam a ROI como um todo.

Dessa forma, considerando que um mamograma típico consiste um conjunto de informações heterogêneas de diversas procedências, como diferentes tipos de tecido (parênquima, gordura, vasos, dutos, pele), a película do filme e características da mulher e do mamógrafo, imaginou-se que seria possível descrever uma imagem mamográfica como uma superposição de um determinado conjunto de imagens-base. Com isso, se conseguiria separar a imagem em “camadas” e avaliar a influência de cada “camada” na imagem que se observa.

Inicialmente, pode-se supor que essa superposição ocorra de forma linear, ou seja, como uma soma ponderada dessas imagens-base, cujos pesos dependeriam das condições expostas, ou seja, das diversas informações que contribuem para a formação da imagem mamográfica. Assim, o vetor de pesos de cada imagem-base na mistura poderia ser um descritor de anormalidade do tecido na região de interesse selecionada.

O método proposto consiste, então, em aproximar uma região de interesse extraída de um mamograma digitalizado por uma combinação linear de imagens-base que são obtidas a partir de um grupo de imagens selecionadas para o treinamento.

A técnica de ICA descrita no capítulo 3 foi empregada para encontrar as imagens-base em que uma imagem mamográfica genérica pode ser decomposta, a partir da simples suposição de independência estatística entre elas.

Os vetores-peso são, então, usados como parâmetros de entrada para uma rede neural com treinamento supervisionado em *backpropagation*, que irá classificar a região como normal ou lesionada.

5.3 - Metodologia

A descrição da metodologia empregada para classificação das ROIs em áreas normais e anormais será dividida em três partes: *pré-processamento* das imagens, *extração de parâmetros* e *classificação*.

5.3.1 Pré-processamento

Foram realizados alguns testes preliminares sobre as ROIs com 1200 x 1200 *pixels* e comparados com os resultados obtidos sobre as regiões de interesse subamostradas em 10 vezes por um filtro de média, resultando em áreas de 120 x 120 *pixels*. Os resultados foram equivalentes, inclusive para classificação das ROIs contendo microcalcificações como áreas lesionadas. Dessa forma, o procedimento de subamostragem foi adotado como padrão em todos os testes que se sucederam, visto que o processamento torna-se muito mais ágil. As ROI foram normalizadas para apresentarem faixa dinâmica entre 0 e 255 tons de cinza.

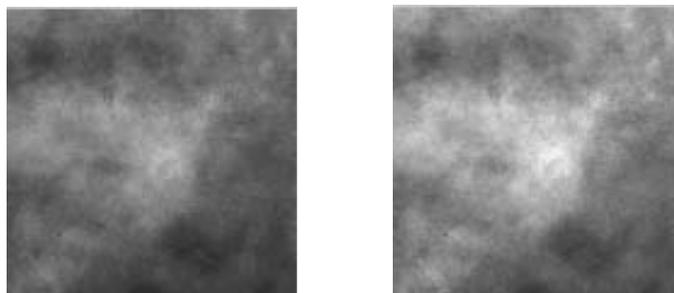
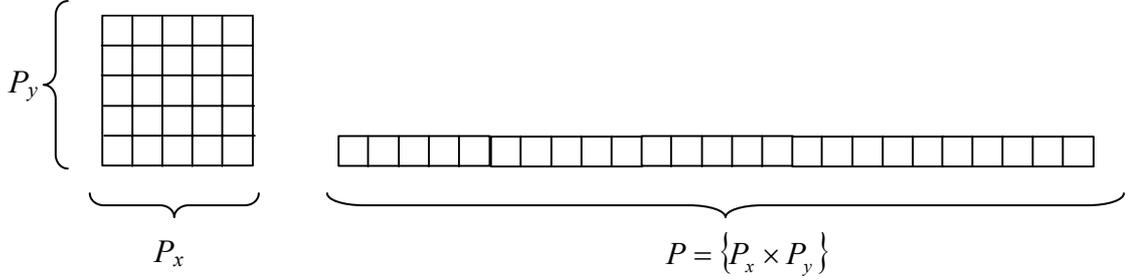


Figura 5.3 - a) ROI original;

b) ROI normalizada.

Para alguns testes, as imagens normalizadas foram submetidas a um algoritmo de realce por filtragem morfológica, conforme descrito na seção 5.3.1.1.

Realçadas ou não, as regiões de interesse originalmente com $P_x \times P_y$ pixels são convertidas em um vetor unidimensional \mathcal{G}_i^T de tamanho $1 \times P$, onde $P = \{P_x \times P_y\}$.



Conforme apresentado na seção 3.1.2, será retirada a média de cada imagem:

$$\mathcal{G}_i^T \leftarrow \mathcal{G}_i^T - E[\mathcal{G}_i^T] \quad (5.1)$$

Assim, empilhando-se o conjunto de N imagens dispostas em linha, obtém-se uma matriz de observação \mathbf{X} com $N \times P$ elementos:

$$\mathbf{X}_{N \times P} = \begin{bmatrix} \mathcal{G}_1^T \\ \mathcal{G}_2^T \\ \vdots \\ \mathcal{G}_N^T \end{bmatrix} \quad (5.2)$$

5.3.1.1 Filtragem morfológica

O objetivo da filtragem morfológica é extrair informações relativas à geometria e à topologia de um conjunto desconhecido de uma imagem.

O Apêndice B apresenta a Morfologia Matemática e os operadores morfológicos.

O algoritmo conhecido como *top-hat*, descrito na seção B.11, é uma alternativa para detectar certas características de interesse em uma imagem cujo *background* apresente intensidade bastante variável, como é o caso do estroma mamário.

Primeiramente, é aplicada uma abertura na imagem, compreendendo uma erosão seguida de uma dilatação. Durante a erosão, o valor do *pixel* no centro do elemento

estruturante é substituído pelo valor mínimo nos *pixels* ao redor (erosão em tons de cinza). Essa operação reduz as regiões com sinais de alta intensidade na imagem. Após a erosão, a operação complementar de dilatação é responsável por substituir o valor central do *pixel* no centro do elemento estruturante pelo valor máximo dos *pixels* vizinhos (dilatação em tons de cinza).

Essa operação restaura as regiões de alta intensidade que não desapareceram completamente como resultado da erosão. Mais precisamente, foi utilizado um algoritmo de reconstrução conforme descrito na seção B.9. Em outras palavras, a dilatação é feita na imagem resultante da interseção da imagem original com a imagem proveniente da erosão, sendo chamada de *marcadores*. A imagem final é obtida pela interseção do resultado desta dilatação com a imagem original ou condicionante. A iteração se repete n vezes, onde n é a maior dimensão da matriz da imagem.

Com isso, consegue-se proporcionar uma restauração mais completa da região de *background*, pois os objetos selecionados pela imagem condicionante voltarão à sua forma original. Dessa forma, não há prejuízo causado às características morfológicas das microcalcificações que serão de extrema relevância na decisão de benignidade.

O elemento estruturante utilizado foi um disco de raio pouco maior que o raio médio das microcalcificações descrito na literatura. [29, 30] Assim, foi selecionado um disco de raio igual a 5, o que corresponderia a um disco de pouco mais que 1 mm de raio, visto que as imagens estão sub-amostradas por um fator de 10.

A segunda parte do algoritmo de *top-hat* consiste em uma subtração da imagem resultante da operação de abertura pela imagem original. Como as microcalcificações foram removidas, e o estroma permaneceu relativamente inalterado na imagem modificada, as características predominantes na imagem resultante da subtração serão as áreas lesionadas.

As figuras abaixo apresentam algumas regiões de interesse filtradas pelo operador morfológico *top-hat*, com fechamento por reconstrução.

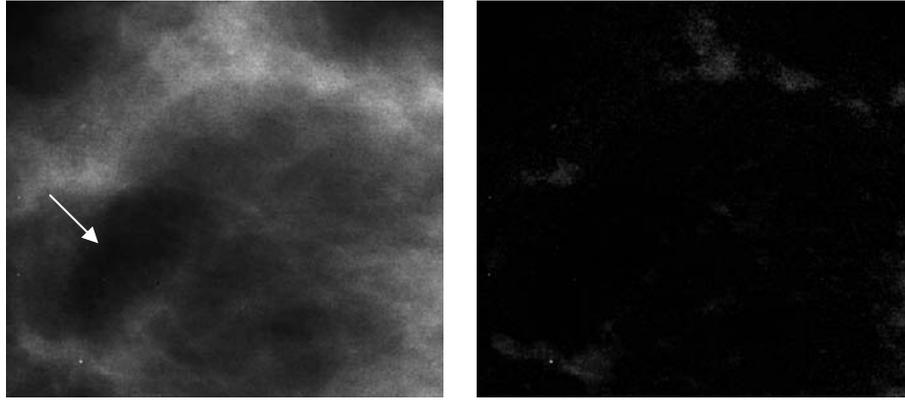


Figura 5.4

a) ROIs contendo loja de gordura (r130) b) ROI filtrada pelo *top-hat*

Observe que a loja de gordura (Figura 5.4) torna-se uma região única, pois não há muito contraste com as regiões adjacentes. O mesmo acontece com a região de parênquima livre (Figura 5.5), apesar de tratar-se de uma região originalmente de alta intensidade.

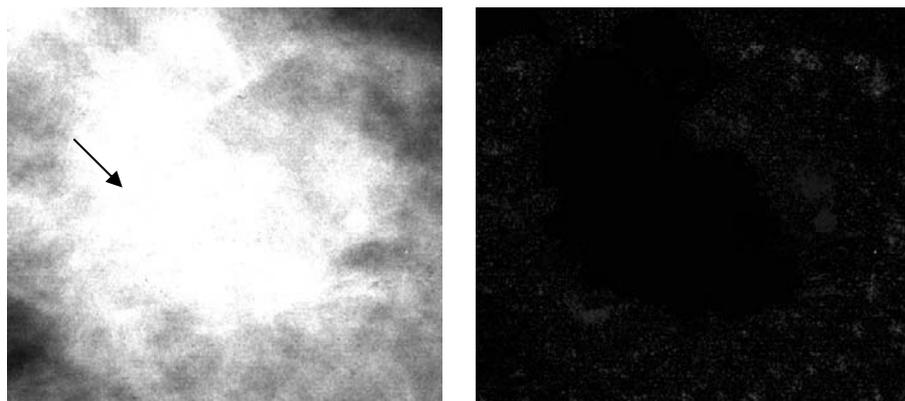


Figura 5.5

a) ROIs contendo parênquima livre (r15) b) ROI filtrada pelo *top-hat*

Pretende-se, com a filtragem, facilitar o trabalho do algoritmo de extração de parâmetros, retirando da imagem a textura do estroma que não correspondem a lesões.

As Figura 5.6 e Figura 5.7 apresentam regiões lesionadas submetidas ao algoritmo de *top-hat*. Verifica-se, na imagem filtrada, que apenas as lesões ficam mais evidentes após a subtração do estroma mamário.

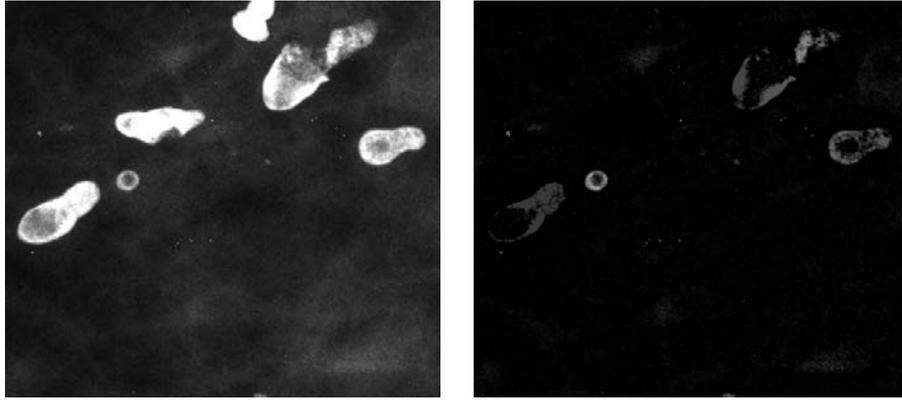


Figura 5.6

a) ROIs contendo calcificações grosseiras (r11) b) ROI filtrada pelo *top-hat*

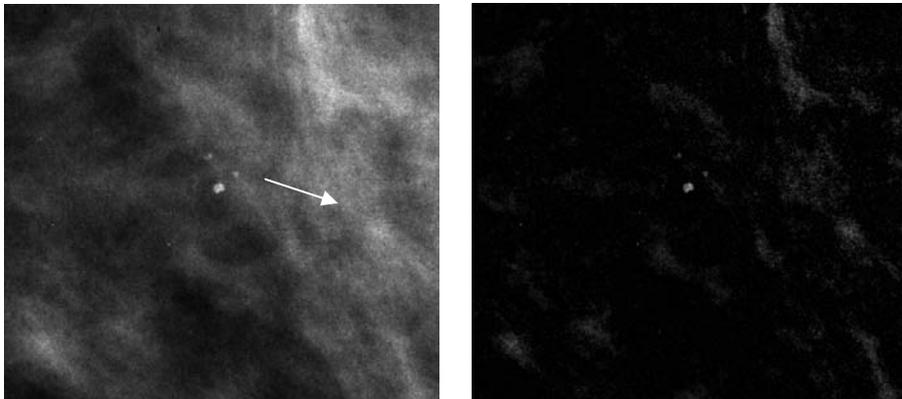


Figura 5.7

a) ROI com CDI (r128) b) ROI filtrada pelo *top-hat*

As operações morfológicas foram realizadas usando a *toolbox* de Morfologia Matemática do *Matlab*®.

5.3.2 Extração de parâmetros

5.3.2.1 ICA

O modelo de mistura adotado é generativo e linear, onde os dados observados x_i são gerados a partir de um processo de combinação linear de variáveis aleatórias latentes s_j , estatisticamente independentes entre si.

Matematicamente,

$$x_i = a_{i1} \cdot s_1 + a_{i2} \cdot s_2 + \dots + a_{in} s_n, \forall i = 1, \dots, n \quad (5.3)$$

No modelo, apenas as variáveis aleatórias x_i são conhecidas, e a partir delas serão estimados os coeficientes da mistura a_{ij} e as componentes independentes s_j .

Ainda, foi considerado um sistema de base completa, ou seja, supõe-se que as N variáveis observadas foram geradas a partir de N fontes.

Em notação matricial:

$$\mathbf{X} = \mathbf{A} \cdot \mathbf{S} \quad (5.4)$$

onde \mathbf{X} é a matriz com as imagens conhecidas, \mathbf{A} é a matriz de mistura e \mathbf{S} é a matriz de fontes ou imagens-base. Cada linha de \mathbf{X} corresponde a uma ROI e cada linha de \mathbf{S} corresponde a uma imagem-base, onde tanto a ROI como as imagens-base foram convertidas em vetores unidimensionais com P elementos.

Dessa forma, cada linha da matriz \mathbf{A} está associada a uma ROI, e cada coluna representa um peso atribuído a uma imagem-base, ou seja, um parâmetro de entrada para a rede neural.

A estimação das componentes independentes, e conseqüentemente da matriz de mistura \mathbf{A} , foi baseada nos três algoritmos apresentados no capítulo 3: *kurtosis*, *neg-entropia* e *máxima verossimilhança*, todos calculados para ponto fixo.

5.3.2.1.1 Redução da dimensionalidade das entradas

Da forma como foi proposto, para um conjunto de N ROIs, ter-se-ia uma rede neural com N observações e N entradas, representadas, respectivamente pelas N linhas e N colunas de \mathbf{A} . Uma rede treinada dessa forma não apresenta boa convergência, uma vez que a informação discriminatória está muito distribuída nos N parâmetros de entrada.

PCA

Pretende-se, pois, encontrar um conjunto menor de variáveis com menos informação redundante, que seja uma boa representação das variáveis originais. A técnica, conhecida como PCA, do inglês *Principal Component Analysis*, ou Análise de Componentes Principais está detalhada no Apêndice A e é reproduzida aqui, em linhas gerais, por conveniência.

Define-se o processo estocástico \mathbf{x} , cujas N variáveis aleatórias são as linhas de $\mathbf{X}_{N \times P}$, ou seja, cada uma das N imagens do conjunto. Cada variável de \mathbf{x} tem P realizações, que representam os *pixels* de uma imagem genérica \mathcal{G}_i , disposta em uma linha, conforme definido em (5.2). A matriz \mathbf{C}_x representa a autocovariância de \mathbf{x} .

A redundância entre as variáveis é medida pela correlação entre suas observações. Como a média foi retirada em (5.1) a correlação e covariância são equivalentes. As componentes principais são as variáveis descorrelacionadas que compõe um vetor \mathbf{z} , obtidas a partir de uma transformação linear do vetor \mathbf{x} , tal que a energia está maximamente concentrada.

$$\mathbf{z} = \mathbf{V}^T \cdot \mathbf{x} \quad (5.5)$$

A matriz de autocovariância de \mathbf{z} deve, pois, ser diagonal, de forma que:

$$\mathbf{C}_z = E[\mathbf{z} \cdot \mathbf{z}^T] = \mathbf{\Lambda} \quad (5.6)$$

Sabe-se, porém, que:

$$\mathbf{C}_z = E[\mathbf{z} \cdot \mathbf{z}^T] = E[\mathbf{V}^T \cdot \mathbf{z} \cdot \mathbf{z}^T \cdot \mathbf{V}] = \mathbf{V}^T \cdot E[\mathbf{x} \cdot \mathbf{x}^T] \cdot \mathbf{V} = \mathbf{V}^T \cdot \mathbf{C}_x \cdot \mathbf{V} \quad (5.7)$$

De (5.6) e (5.7), tem-se que:

$$\mathbf{V}^T \cdot \mathbf{C}_x \cdot \mathbf{V} = \mathbf{\Lambda} \quad (5.8)$$

Portanto, \mathbf{V}^T é a matriz ortogonal com $N \times N$ elementos que diagonaliza a matriz \mathbf{C}_x . Como resultado clássico da álgebra, \mathbf{V}^T é a matriz cujas linhas são os autovetores da matriz de \mathbf{C}_x , correspondentes aos autovalores em ordem crescente de variância. $\mathbf{\Lambda}$ é uma matriz diagonal $N \times N$ cujos elementos são os autovalores de \mathbf{C}_x , ou correspondentemente, as variâncias de \mathbf{z} , em ordem crescente de energia.

Em outras palavras, \mathbf{Z} é uma transformação linear da matriz \mathbf{X} , cujas linhas são ortogonais entre si e que estão em ordem crescente de variância:

$$\mathbf{Z}_{N \times P} = \mathbf{V}_{N \times N}^T \cdot \mathbf{X}_{N \times P} = \mathbf{V}_{N \times N}^T \cdot \mathbf{A}_{N \times N} \cdot \mathbf{S}_{N \times P} = \mathbf{W}_{N \times N} \cdot \mathbf{S}_{N \times P} \quad (5.9)$$

A técnica de extração de componentes principais é uma transformação linear das imagens de mistura, que branqueia as variáveis observadas e as coloca em ordem crescente de energia. A seção 3.1.3.1.1 demonstra que o branqueamento das variáveis

observadas é um pré-processamento recomendado para uma boa convergência do algoritmo de ICA.

A fim de reduzir o número de entradas para a rede, o método de ICA foi aplicado sobre um subconjunto de K componentes principais da matriz de observação \mathbf{X} .

Se forem utilizadas apenas as K primeiras componentes de maior energia, obtém-se, então, o modelo de mistura para as K componentes principais, aplicados sobre o conjunto de treinamento \mathbf{X}_{trein} :

$$\mathbf{Z}_{trein} = \mathbf{V}^T \cdot \mathbf{X}_{trein} = \mathbf{W}_{trein} \cdot \mathbf{S}_{trein} \quad (5.10)$$

$K \times P$ $K \times N$ $N \times P$ $K \times K$ $K \times P$

onde \mathbf{W}_{trein} é a matriz de mistura com $K \times K$ elementos, obtida pelo algoritmo de estimação de componentes independentes aplicado sobre a matriz \mathbf{Z}_{trein} com as K componentes principais da matriz de imagens \mathbf{X}_{trein} , $K \leq N$.

A matriz de mistura pode ser recuperada multiplicando-se \mathbf{V} pela matriz branqueada \mathbf{Z}_{trein} :

$$\mathbf{X}_{trein} \approx \mathbf{V} \cdot \mathbf{Z}_{trein} = \mathbf{V} \cdot \mathbf{W}_{trein} \cdot \mathbf{S}_{trein} \quad (5.11)$$

$N \times P$ $N \times K$ $K \times P$ $N \times K$ $K \times K$ $K \times P$

Como $\mathbf{X}_{trein} = \mathbf{A}_{trein} \cdot \mathbf{S}_{trein}$, então $\mathbf{V} \cdot \mathbf{W}_{trein} \approx \mathbf{A}_{trein}$. Ou seja, a matriz com pesos para as N ROI pode ser aproximada pelo produto transposta da matriz de projeção pela matriz de mistura encontrada pelo algoritmo de separação de fontes.

$$\mathbf{A}_{trein} \approx \mathbf{V} \cdot \mathbf{W}_{trein} \quad (5.12)$$

$N \times K$ $N \times K$ $K \times K$

As imagens reservadas para o teste podem ser descritas como uma combinação das fontes encontradas no treinamento, normalizadas para variância unitária:

$$\mathbf{X}_{teste} \approx \mathbf{A}_{teste} \cdot \sqrt{\mathbf{\Lambda}_{trein}^{-1}} \cdot \mathbf{S}_{trein} \quad (5.13)$$

$N' \times P$ $N' \times K$ $K \times K$ $K \times P$

Como, por definição as linhas de \mathbf{W}_{trein} têm variância unitária, as variâncias das linhas de \mathbf{S}_{trein} são iguais às variâncias das linhas de \mathbf{Z}_{trein} , e, portanto, $\mathbf{\Lambda}_{trein}$ são as $K \times K$ primeiras linhas e colunas de $\mathbf{\Lambda}$, conforme definido em (5.6):

$$\mathbf{S}_{trein} = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & & \\ \vdots & & \ddots & \vdots \\ 0 & & \dots & \sigma_N^2 \end{bmatrix} \cdot \mathbf{S}_{trein}^{\sigma^2=1}$$

$$\underset{\mathbf{\Lambda}}{=}$$

Isso equivaleria a projetar as imagens de teste no espaço das componentes principais obtidas no treinamento. Como \mathbf{Z}_{trein} é ortogonal:

$$\mathbf{A}_{teste} \approx \mathbf{X}_{teste} \cdot \mathbf{S}_{train}^{-1} \cdot \sqrt{\mathbf{\Lambda}_{train}} = \mathbf{X}_{teste} \cdot \mathbf{Z}_{train}^T \cdot \mathbf{W}_{train} \cdot \sqrt{\mathbf{\Lambda}_{train}} \quad (5.14)$$

As K colunas das matrizes \mathbf{A}_{trein} e \mathbf{A}_{teste} são as entradas da rede neural. São destinadas N regiões de interesse ou N pares entrada-saída para treinamento e N' regiões de interesse ou N' pares entrada-saída para o teste.

A Figura 5.8 apresenta uma descrição esquemática do sistema de diagnóstico por ICA com redução de dimensionalidade por PCA.

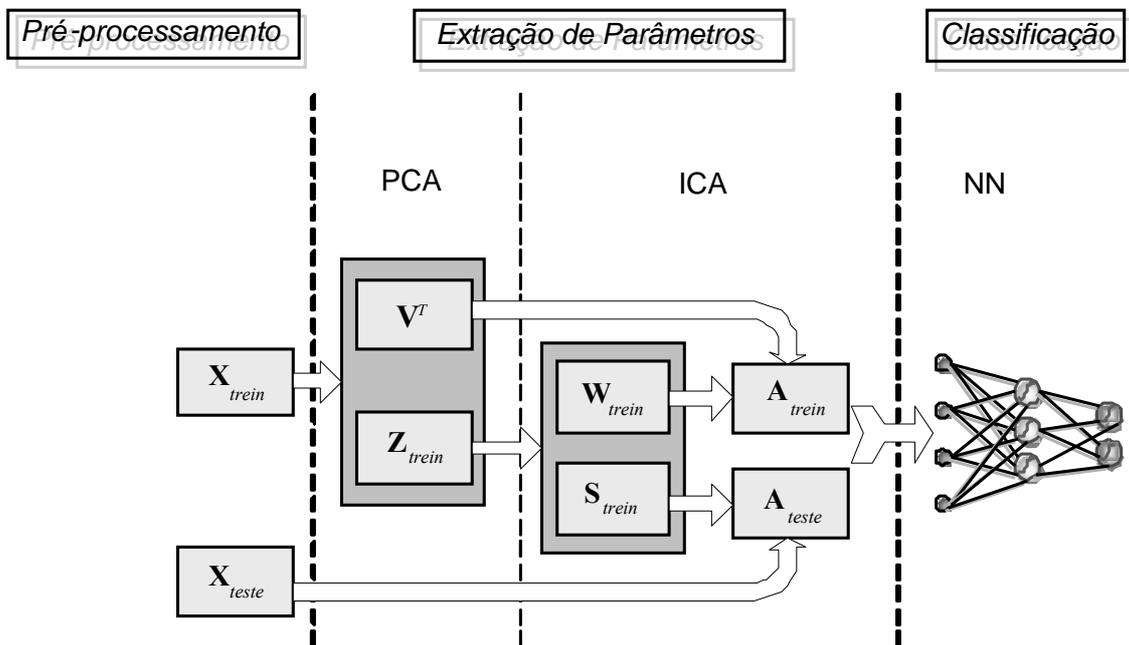


Figura 5.8 – Diagramação da metodologia empregada para o sistema de diagnóstico por ICA.

5.3.2.2 Outros métodos de extração de parâmetros

A fim de comparar o desempenho do esquema proposto, foram extraídos parâmetros do mesmo conjunto de imagens a partir de dois descritores de textura apresentados na literatura: a *matriz de co-ocorrência* e a *curva de complexidade*.

A extração de parâmetros que caracterizem a textura de uma imagem bidimensional é tradicionalmente feita por métodos estatísticos e estruturais.

No método estrutural, a textura é considerada como um conjunto de alguns padrões primitivos que se repetem de acordo com alguma regra. Em geral, a imagem transformada em frequência (DCT, DFT, etc.) é usada para determinar as primitivas (padrão básico de textura) e a regra de deslocamento, ou seja, como as primitivas se dispõem na imagem original.

Quando não é possível identificar padrões na imagem, consegue-se uma melhor descrição de textura através de métodos estatísticos, que caracterizam as propriedades estocásticas da distribuição de níveis de cinza. Dentro dessa abordagem, a matriz de co-ocorrência ou GLCM (*Gray-level Co-occurrence Matrix*) é o método mais usado.

Vários pesquisadores mostraram que a matriz de co-ocorrência é capaz de proporcionar resultados melhores que filtros de Gabor, geometria fractal, campos de Markov (MRF *Markov random fields*) e *wavelets* [[41-43]] para descrição de textura. No entanto, em [26] Baheerathan obteve uma melhor capacidade discriminatória entre padrões de textura usando a curva de complexidade do que a melhor combinação da matriz co-ocorrência.

Devido ao bom desempenho como classificadores de textura, a matriz de co-ocorrência e a curva de complexidade foram aplicadas às mesmas regiões de interesse. Para cada uma das técnicas, foram extraídos parâmetros que foram usados como entrada para a rede neural. A descrição de ambas as técnicas e dos parâmetros escolhidos para quantificação da textura encontram-se nos apêndices.

Os resultados obtidos foram comparados ao método proposto, a fim de avaliar o desempenho do ICA como discriminador de textura, com relação aos métodos clássicos descritos na literatura.

5.3.3 Classificação

5.3.3.1 Backpropagation

Para classificação das lesões foi utilizada uma rede neural treinada com *backpropagation*, utilizando como entrada os parâmetros extraídos do ICA para gerar uma saída lógica $y \in \{-1, +1\}$, indicando normalidade ou anormalidade.

Foi desenvolvido um programa para implementar o algoritmo *backpropagation* no *Matlab*® na sua forma matricial para que o processamento fosse mais veloz, uma vez que o *Matlab*® é otimizado para a realização de operações matriciais.

Foram selecionadas 205 regiões para o treinamento, sendo 103 saudáveis e 102 lesionadas, reservando-se a mesma distribuição para o teste.

Os parâmetros de entrada para rede neural foram normalizados para:

$$x_i \rightarrow \frac{(x_i - \bar{x}_i)}{2\sigma_{x_i}} \therefore x_i \in [-1, +1] \quad (5.15)$$

onde

$$\bar{x}_i = \frac{1}{n} \cdot \sum_{i=1}^n x_i \quad (5.16)$$

e

$$\sigma_{x_i} = \frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x}_i) \quad (5.17)$$

sendo n o número de amostras para cada entrada da rede neural.

O treinamento foi repetido por 10 vezes a partir de condições iniciais diversas. Em cada repetição, o conjunto de teste e treinamento era sorteado aleatoriamente.

Experimentou-se implementar taxa de aprendizado adaptativa, o que levou a rede a resultados um pouco mais rápidos, porém sem ganhos de acurácia. O uso de momento também trouxe apenas uma diminuição na oscilação da curva de treinamento.

Os pesos iniciais foram escolhidos aleatoriamente. A faixa dinâmica foi variada diversas vezes apresentando o melhor resultado entre -1.2 e 1.2.

Cada treinamento prosseguia até que atingisse um erro máximo especificado ou um número máximo de épocas. A cada ciclo de épocas, ou passo, todo o conjunto de

ROIs (treinamento e teste) era apresentado à rede e registrava-se o erro médio na operação. No entanto, também era registrado o número de diagnósticos errados a cada passo para que fosse possível balizar o que o erro médio representava para o total das imagens, o que se chamou de erro lógico da operação. A Figura 5.9 apresenta os gráficos com as curvas de erro médio e lógico, para o treinamento e teste, plotados ao final do treinamento.

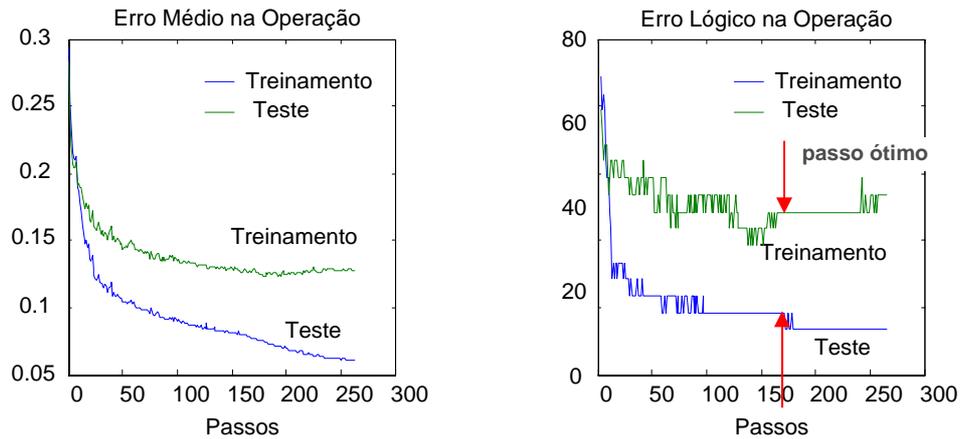


Figura 5.9 - Curvas de Erro de Aprendizado para o Treinamento e Teste.

Na maioria das vezes, ao final do treinamento, a rede havia se tornado especializada para o conjunto de treinamento, resultando em um número de acertos menor para o conjunto total de imagens. Foi então considerado como passo ótimo de treinamento, o ponto de divergência das curvas de erro lógico de treinamento e teste, conforme a Figura 5.9.

5.3.3.2 Componentes de atuação conjunta (CAC)

A idéia de usar uma rede neural de extração de componentes de atuação conjunta para reduzir a dimensionalidade é bastante interessante: ao invés de calcular as componentes principais para, a partir delas, extrair as componentes independentes, comprime-se a matriz de mistura \mathbf{X} nas componentes de atuação conjunta com uso de uma rede neural de 3 camadas (vide seção 4.2 -) e calculam-se as componentes independentes a partir da saída da 1ª camada.

No entanto, da forma como foi exposta, a técnica esbarra em uma questão. A saída da rede para separação de componentes de atuação conjunta, em seu conceito, é o diagnóstico da região de interesse apresentada à entrada, com P elementos. Assim, a saída da primeira camada reduz a entrada a $P' < P$ elementos que serão classificados

pela 2ª e 3ª camadas. Esse procedimento, porém, não resolve o problema desejado, visto que se pretende reduzir o número N de imagens, e não a quantidade P de pontos por imagem, para que a matriz de pesos das componentes independentes A tenha um número de colunas (que serão as entradas para a rede de classificação) menor que o número de linhas (pares entrada-saída). A solução seria transpor a matriz de entrada X , resultando em uma rede com N entradas, e P ocorrências ou pares entrada-saída. Nesse caso, porém, o conjunto de entradas não corresponde a um diagnóstico, pois cada entrada provém de uma ROI, que pode ser normal ou lesionada, e perde-se o significado da técnica.

Para solucionar o problema, as componentes independentes, e, conseqüentemente, as matrizes A_{trein} e A_{teste} foram calculadas conforme a seção 5.3.2.1.1. A rede neural de classificação, porém, foi treinada conforme apresentado na seção 4.2.1. Ou seja, as N entradas da rede, representadas pelas N colunas das matrizes A_{trein} e A_{teste} , são comprimidas em K componentes de atuação conjunta (saída da 1ª camada), que são apresentadas à parte da rede que efetivamente executa a classificação (2ª e 3ª camadas).

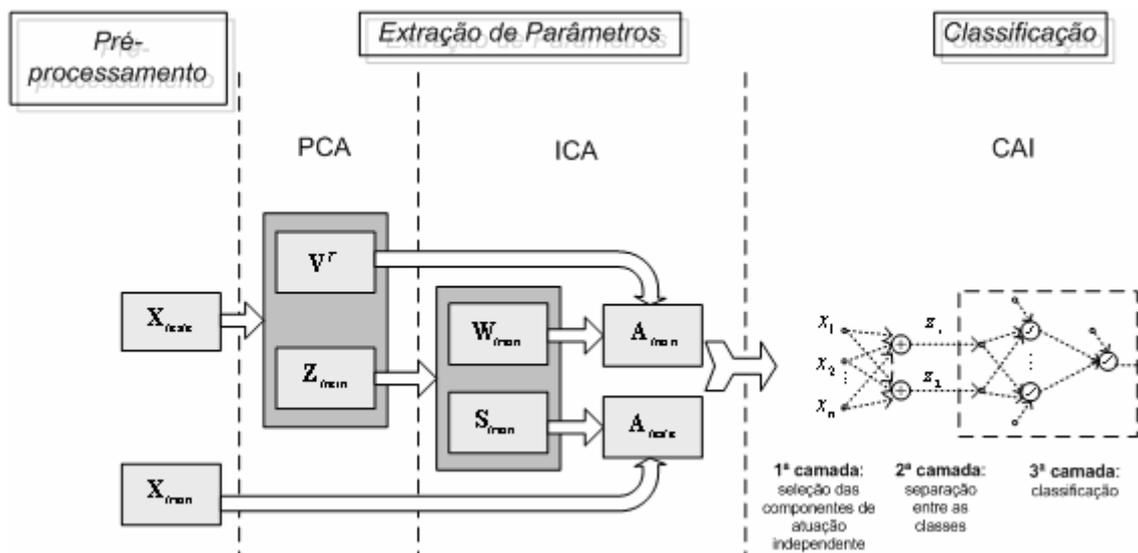


Figura 5.10 – Diagramação da metodologia empregada para o sistema de diagnóstico por CAC.

Como o branqueamento é um passo útil, conforme apresentado na seção 3.1.3.1.1, o PCA foi mantido, uma vez que, mantendo-se todas as componentes principais, obtém-se a projeção de todas as imagens no espaço branqueado. No entanto, para efeitos de teste do sistema, o desempenho da rede de seleção de componentes de

atuação conjunta também foi avaliado reduzindo-se o número de componentes principais apresentadas ao algoritmo que extrai as componentes independentes.

6. Resultados

Foram realizados diversos experimentos. A extração de parâmetros foi feita por análise de componentes independentes (ICA), matriz de co-ocorrência (SGLD) e curva de complexidade (CClx). Para a extração por ICA foram testados os métodos de *kurtosis*, *neg-entropia* e máxima verossimilhança. Como a maior taxa de acertos foi obtida para o método da *neg-entropia*, para esse método o teste foi repetido utilizando-se uma rede classificatória em 3 camadas, com seleção de parâmetros por componentes de atuação conjunta. O método da *neg-entropia* com rede classificatória em 3 camadas também foi experimentado sobre imagens pré-processadas pelo algoritmo de *top-hat*, assim como sobre as imagens transladadas.

O organograma abaixo apresenta um resumo dos testes realizados.

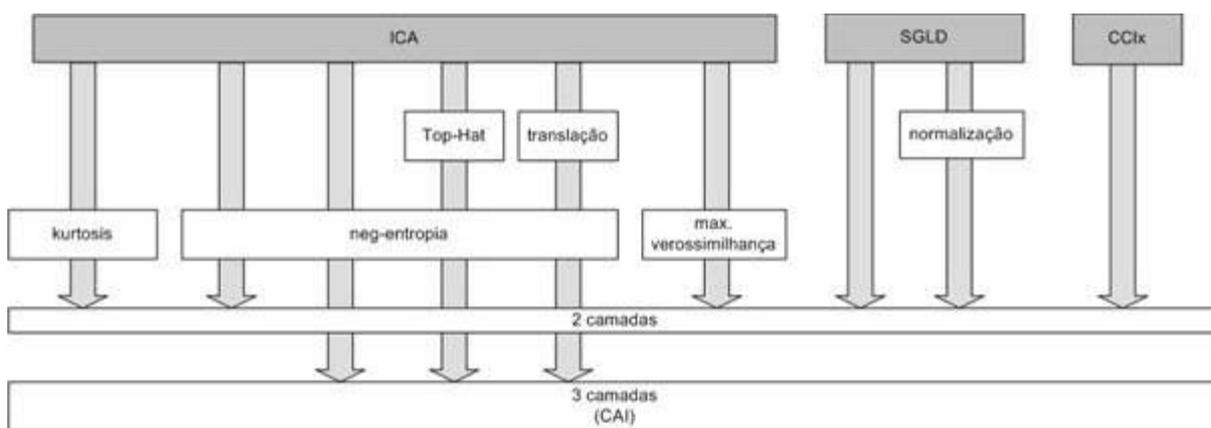


Figura 6.1 – Organograma opções de processamento aplicadas.

6.1 - Extração de parâmetros por ICA

6.1.1 Classificação por *backpropagation*

O primeiro experimento realizado utilizou a extração de parâmetros por ICA conforme os três métodos descritos no Capítulo 3, com classificação por uma rede neural em *backpropagation*.



Figura 6.2 – Desenho Esquemático: ICA + *backpropagation*

Para a extração de parâmetros por ICA, um conjunto de treinamento com 205 regiões de interesse, escolhidas aleatoriamente dentre as ROIs selecionadas pela radiologista, foi separado para o cálculo das imagens-base, sendo 103 regiões normais e 102 lesionadas.

As ROIs foram sub-amostradas pelo filtro de média, resultando em áreas de 120×120 *pixels*, e, então, normalizadas para apresentarem faixa dinâmica entre 0 e 255 tons de cinza. As ROIs foram convertidas em um vetor unidimensional de tamanho 1×14.400 e embaralhadas em uma matriz \mathbf{X}_{trein} de 205×14.400 .

O mesmo procedimento foi repetido para gerar a matriz de mistura \mathbf{X}_{teste} .

A matriz de mistura de treinamento é branqueada pelo algoritmo de PCA, conforme a seção 5.3.2.1.1. O número de componentes principais K foi variado em 20, 40, 50, 60, 65, 70, 75, 80, 85, 100, 150 e 200. A matriz branqueada \mathbf{Z}_{trein} é apresentada a uma das 3 metodologias de estimação das componentes independentes empregadas: *kurtosis*, *neg-entropia* e *máxima verossimilhança*, gerando uma matriz de coeficientes \mathbf{W}_{trein} .

A matriz de mistura \mathbf{A}_{trein} foi calculada conforme apresentado na equação (5.12). O cálculo dos parâmetros sobre as imagens restantes, reservadas para o teste, foi feito conforme descrito em (5.14), gerando a matriz \mathbf{A}_{teste} .

Os parâmetros calculados foram submetidos a uma rede neural com treinamento supervisionado. O número de neurônios na camada intermediária foi incrementado de 5 a 90. Para cada número de neurônios na camada intermediária, dado um número de K componentes independentes, o treinamento era repetido 10 vezes, a partir de condições iniciais diferentes, e para conjuntos de treinamento e teste sorteados aleatoriamente. O desempenho da rede foi avaliado como a média dos resultados apresentados nas 10 repetições.

O procedimento repetido para diferentes números de componentes principais, e para as 3 metodologias de estimação das componentes independentes estão apresentados nas Tabela F.1 a Tabela F.12 (*kurtosis*), Tabela F.13 a Tabela F.24 (*neg-entropia*) e Tabela F.25 a Tabela F.37 (*máxima verossimilhança*).

A Figura 6.3 apresenta um conjunto de 56 imagens- base obtidas a partir das 56 primeiras componentes principais do conjunto de 205 imagens de treinamento.

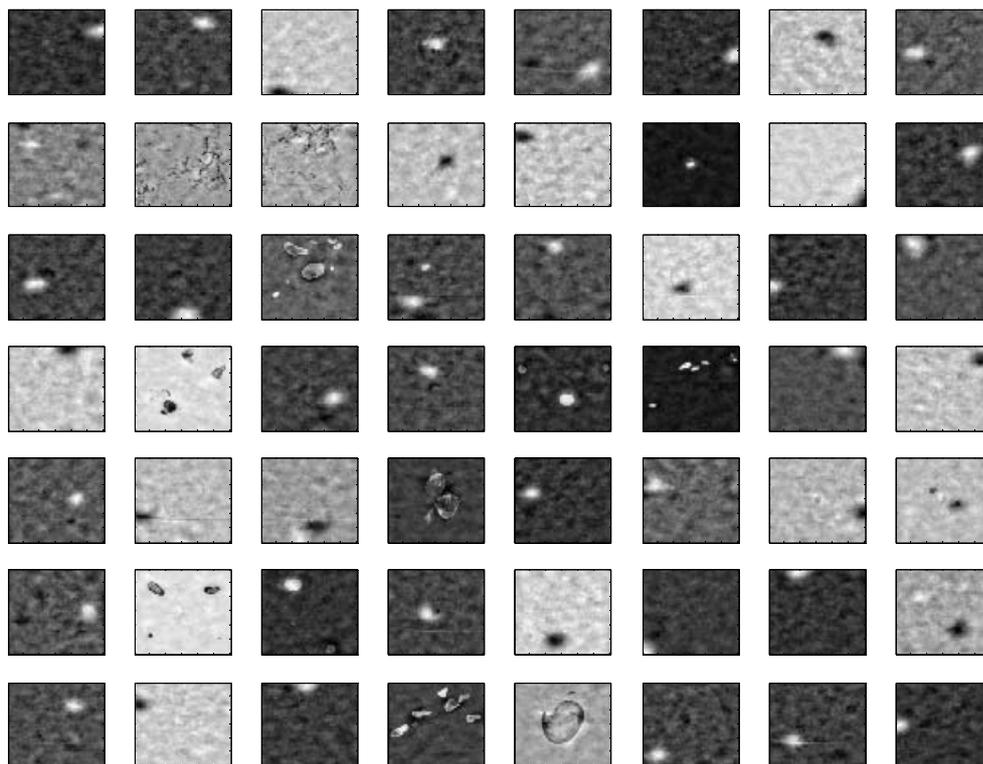


Figura 6.3 – Imagens-base (método: *neg-entropia*).

O resultado com cálculo das componentes independentes por *neg-entropia* foi, em geral, melhor que os demais, não só porque apresentou uma maior taxa de acerto na maioria dos testes, mas também, porque esse aumento na taxa de acertos foi causado por uma redução do número de falsos negativos, ou seja, um número menor de regiões lesionadas foi interpretado como saudável. A rede com 100 neurônios na camada escondida e 60 entradas (ou 60 componentes principais) apresentou o melhor

desempenho com 83,22% de acertos e sensibilidade de $S = \left(\frac{VP}{VP + FN} \cdot 100 \right) = 87,52\%$ e especificidade de $E = \left(\frac{VN}{VN + FP} \cdot 100 \right) = 78,91\%$

Em geral, os melhores desempenhos da rede neural encontram-se entre 20 e 60 neurônios. Ao aumentar o número de neurônios da rede, tende-se a especializar a rede ao conjunto de treinamento prematuramente, reduzindo a performance para o conjunto de teste, e prejudicando o resultado geral. Apenas no caso em que um número muito pequeno de parâmetros de entrada é apresentado à rede (5 a 20 componentes principais), há uma tendência a aumentar o número de acertos com um número muito maior de neurônios (mais que 160), uma vez que é preciso um maior esforço da rede para extrair informações de um número tão reduzido de parâmetros. No entanto, em nenhum desses casos, a taxa de acertos alcança o resultado obtido para mais que 20 componentes principais, não compensando o esforço computacional de aumentar o número de neurônios na camada intermediária. Para o algoritmo de estimação de componentes independentes por máxima verossimilhança, isso fica ainda mais evidente, pois com 150 a 200 componentes principais há uma queda significativa na taxa de acertos, mesmo com um maior número de neurônios na camada intermediária.

6.1.1.1 Translação da imagem

A fim de avaliar a robustez do sistema a translações na imagem, foi realizado novo teste com as ROIs de interesse transladadas. Assim, à rede com melhor desempenho obtida no experimento da seção 6.1.1 foram apresentadas novas regiões de interesse resultantes da translação das ROIs originais.

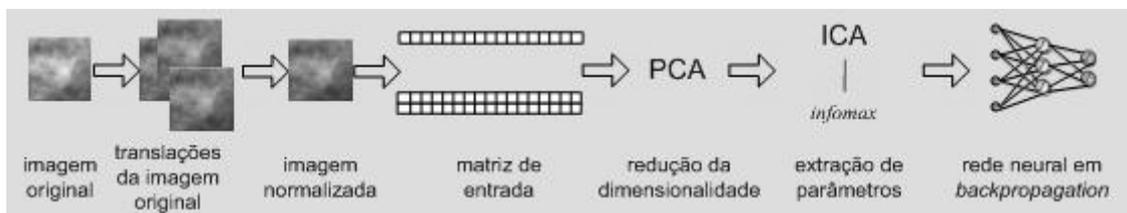


Figura 6.4 – Desenho Esquemático: translação + ICA + *backpropagation*

O teste foi realizado com extração de parâmetros por ICA, através do método de maximização da não-gaussianidade por neg-entropia. Foi utilizada para classificação a

rede neural com 100 neurônios na camada escondida e 70 entradas, tal como descrito na seção 6.1.1

As ROIs foram transladadas em passos de 100 *pixels* nos eixos horizontal e vertical, assegurando-se que a nova ROI permanecia dentro do mamograma original.

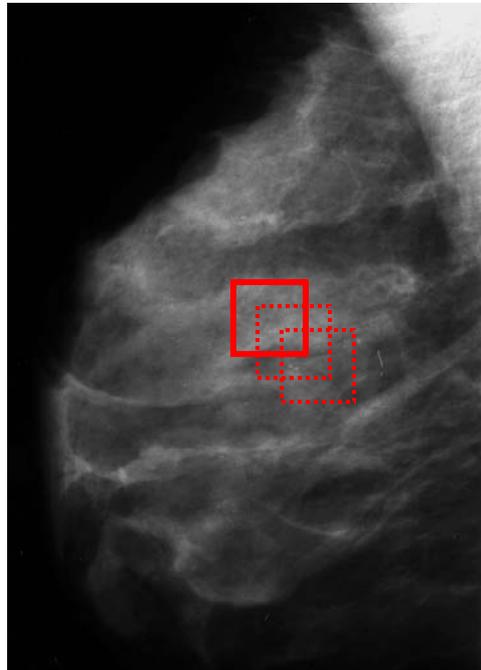


Figura 6.5 – ROI transladada.

Para validar o experimento, seria necessária a supervisão do radiologista para assegurar o diagnóstico de cada nova ROI de interesse obtida a partir da translação da ROI original. Isso significaria avaliar mais de 10.000 ROIs. A fim de agilizar e automatizar os testes, cada ROI transladada era submetida a uma nova bateria de translações horizontais e verticais, e o diagnóstico da ROI transladada (região normal ou anormal) era aceito como o que ocorresse em maior número para a bateria de translações.

Os resultados obtidos não revelam uma robustez do sistema a translações, decaindo de uma média de 83,22% para 55,37% de acertos.

6.1.2 Classificação por Componentes de Atuação Conjunta

Um novo experimento foi realizado usando ICA por neg-entropia como extrator de parâmetros. Nesse caso, porém, à rede classificadora foi anexada uma rede linear que separa componentes relevantes à discriminação, tal como descrito na seção 6.1.2.

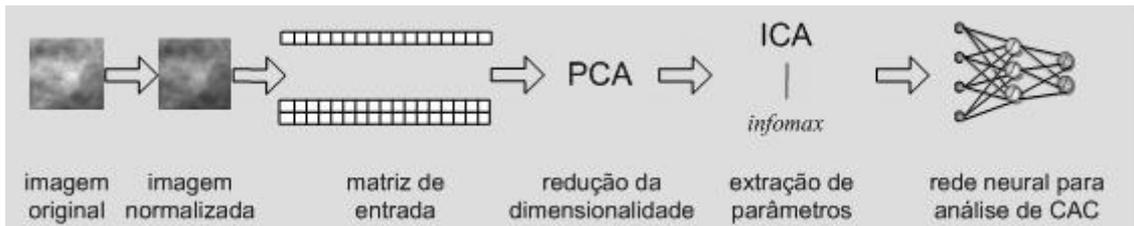


Figura 6.6 – Desenho Esquemático: ICA + CAC

A compressão da informação que a rede neural usará para classificar uma entrada por análise de componentes principais apenas prioriza as componentes de maior energia, o que não garante que elas sejam as mais relevantes para discriminação das entradas.

O experimento pretende realizar um mapeamento linear das entradas, considerando sua capacidade discriminatória, acoplando-se à rede de decomposição linear uma rede com neurônios lineares que realizem a classificação da entrada, conforme exposto na seção 5.3.3.2.

Como o melhor resultado para classificação foi obtido para rede com 50 neurônios, foram testados de 70 a 90 neurônios na 2ª camada, uma vez que se imagina que a rede terá um esforço maior para classificação a partir de menos componentes.

Foram realizados experimentos com 3, 10, 20, 40 e 50 componentes de atuação conjunta (ou seja, o número de neurônios na 1ª camada). Para cada condição, o treinamento era repetido 10 vezes, a partir de condições iniciais diferentes, e para conjuntos de treinamento e teste sorteados aleatoriamente.

Foi avaliado o desempenho da rede tendo como entradas a matriz de pesos obtida pelas componentes independentes calculadas a partir de todas as imagens do conjunto de treinamento (205), branqueadas pelo algoritmo de extração de componentes principais ou sobre apenas algumas componentes principais (50, 70, 90 150 e 180).

As taxas de acerto apresentam um aumento generalizado com relação às obtidas pela rede em *backpropagation*, conforme verificado nas Tabela F.38 a Tabela F.72. Esse resultado valida o conceito de comprimir a informação que será apresentada à rede, considerando sua relevância para a discriminação desejada.

A maior taxa de acertos, 85,68%, ocorreu para 20 componentes de atuação conjunta, usadas como entrada pra uma rede neural classificadora com 90 neurônios na

camada intermediária. Esse número representa sensibilidade de

$$S = \left(\frac{VP}{VP + FN} \cdot 100 \right) = 86,92\% \text{ e especificidade de } E = \left(\frac{VN}{VN + FP} \cdot 100 \right) = 84,44\% .$$

6.1.2.1 ICA sobre imagens pré-processadas

Imaginou-se que algum tipo de realce nas estruturas do estroma mamário poderia facilitar o trabalho do algoritmo de separação de fontes, e conseqüentemente a classificação da rede neural. Assim, as ROIs apresentadas ao algoritmo de extração de parâmetros por ICA foram submetidas a uma filtragem morfológica por Top-Hat, conforme descrito na seção B.11 do Apêndice. A rede classificatória usou componentes de atuação conjunta conforme a seção anterior.

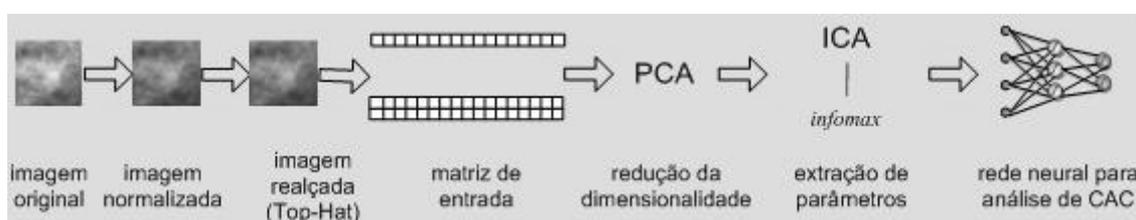


Figura 6.7 – Desenho Esquemático: Top-Hat + ICA + CAC

O elemento estruturante utilizado foi um disco de raio igual a 5, o que corresponderia a um disco de pouco mais que 1 mm de raio, visto que as imagens estão sub-amostradas por um fator de 10. Esse valor foi escolhido para ser um pouco maior que o raio médio das microcalcificações descrito na literatura como 1 mm [44, 45].

As imagens resultantes do pré-processamento foram submetidas ao mesmo procedimento exposto na seção (6.1.2).

Foi avaliado o desempenho da rede tendo como entradas a matriz de pesos obtida pelas componentes independentes calculadas a partir de algumas componentes principais (50, 70, 90, 150). Foram testadas redes com 70, 90 e 110 neurônios na 2ª camada. O número de componentes de atuação conjunta (ou seja, o número de neurônios na 1ª camada) foi variado entre 10, 20 e 30. Para cada condição, o treinamento era repetido 10 vezes, a partir de condições iniciais diferentes, e para conjuntos de treinamento e teste sorteados aleatoriamente.

O resultado apresentou conforme verificado na Tabela F.73 a Tabela F.80.

O melhor resultado, 81,66% de acertos, ocorreu para rede com 90 neurônios na camada intermediária com 90 componentes principais. Esse número representa sensibilidade de $S = \left(\frac{VP}{VP + FN} \cdot 100 \right) = 85,86\%$ e especificidade de $E = \left(\frac{VN}{VN + FP} \cdot 100 \right) = 77,46\%$.

Assim, não houve ganho ao se realizar um pré-processamento das imagens, apesar de ser este um campo ainda bastante frutífero.

6.2 - Extração de parâmetros por Matriz de co-ocorrência

A fim de comparar o desempenho do método proposto com outras técnicas empregadas na literatura, as ROIs foram classificadas usando os parâmetros extraídos pela matriz de co-ocorrência, conforme descrito no Apêndice D.

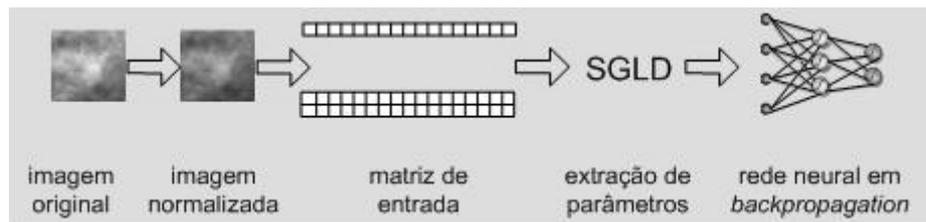


Figura 6.8 – Desenho Esquemático: GLCM + *backpropagation*

O custo computacional para extração de parâmetros pela GLCM depende no número de níveis de cinza da imagem. Foram realizados testes preliminares que não revelaram qualquer ganho na porcentagem de acertos ao calcular-se a matriz de co-ocorrência a partir da imagem com 256 níveis de cinza, com relação ao cálculo com 64 níveis de cinza. A segunda opção foi, portanto, adotada como padrão para os testes que se sucederam.

Foram realizadas duas baterias de testes: a primeira a partir das imagens quantizadas para 64 tons de cinza e a segunda a partir das mesmas imagens submetidas à equalização do histograma de primeira ordem a fim de distribuir os tons de cinza por toda faixa dinâmica.

Assim como para a extração de parâmetros por ICA, 205 regiões de interesse foram reservadas para o treinamento, escolhidas aleatoriamente dentre as ROIs

selecionadas pela radiologista, sendo 103 regiões normais e 102 lesionadas. As demais 205 regiões de interesse foram deixadas para o teste de desempenho da rede neural.

Para cada ROI, a matriz de co-ocorrência foi calculada para 1, 3 e 5 *pixels* de distância, com angulações de 0°, 45° e 90°, e com a média dos três.

Em cada caso, foi obtido um vetor com 5 parâmetros, segundo descrito no Apêndice D.

Esses vetores foram apresentados a uma rede neural em *backpropagation* conforme a seção 4.1.2. O número de neurônios na camada intermediária foi incrementado de 5 a 160.

A média dos resultados obtidos para extração de parâmetros pela matriz de co-ocorrência está representada nas Tabela F.81 a Tabela F.92. Os resultados obtidos sobre imagens com histogramas equalizados encontram-se nas Tabela F.93 a Tabela F.104.

O melhor desempenho foi obtido a partir da matriz de co-ocorrência calculada sobre as imagens originais, sem equalização do histograma, para uma distância de 3 *pixels* a 90°, com 100 neurônios na camada escondida da rede neural. A taxa de acertos, para esse caso ficou em 63.78%, com sensibilidade de $S = \left(\frac{VP}{VP + FN} \cdot 100 \right) = 51,76\%$ e especificidade de $E = \left(\frac{VN}{VN + FP} \cdot 100 \right) = 75,80\%$.

Nota-se que, em geral, o resultado para histogramas equalizados é pior que o resultado sobre as imagens originais, revelando uma contribuição da concentração dos tons de cinza para a discriminação. Ainda, há, de maneira geral, uma melhora na taxa de acertos para matrizes de co-ocorrência com distância de 3 *pixels*, onde o algoritmo encontrou uma melhor descrição do padrão de textura das microcalcificações.

No entanto, apesar de ser um método clássico na literatura, o resultado da matriz de co-ocorrência é bastante pobre para identificação de lesões no conjunto de mamogramas utilizado.

6.3 - Extração de parâmetros por Curva de Complexidade

Da mesma forma que na seção 6.2 -, as ROIs foram classificadas usando os parâmetros extraídos pela curva de complexidade, conforme descrito no Apêndice E.

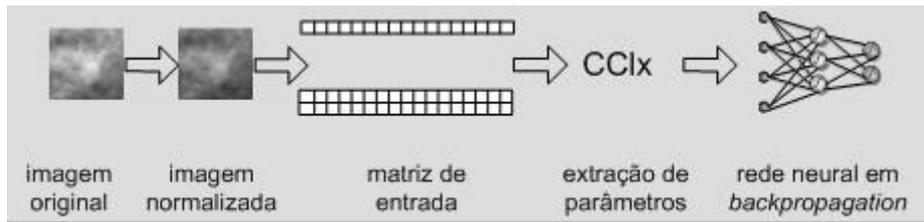


Figura 6.9 – Desenho Esquemático: CClx + *backpropagation*

Também para esse caso, foi selecionado um conjunto de treinamento com 205 regiões de interesse, escolhidas aleatoriamente dentre as ROIs selecionadas pela radiologista, sendo 103 regiões normais e 102 lesionadas. As demais 205 regiões de interesse foram reservadas para o teste. Para cada região de interesse foi levantada uma curva de complexidade conforme descrito em (E.1) e extraídos 5 parâmetros conforme definidos no Apêndice E, organizados em vetores correspondentes a cada ROI.

Esses vetores foram apresentados a uma rede neural em *backpropagation* conforme a seção 4.1.2. O número de neurônios na camada intermediária foi incrementado de 5 a 160. A média dos resultados obtidos para curva de complexidade está representada na Tabela F.105.

A rede com 140 neurônios na camada escondida apresentou o melhor desempenho, com 75,41% de acertos, sensibilidade de $S = \left(\frac{VP}{VP + FN} \cdot 100 \right) = 84,24\%$ e especificidade de $E = \left(\frac{VN}{VN + FP} \cdot 100 \right) = 66,58\%$. A taxa de acertos é, sem dúvida, maior que o melhor resultado para Matriz de Co-Ocorrência, mas ainda se situa muito aquém dos resultados obtidos para qualquer dos métodos de extração de parâmetros por ICA.

7. Conclusão

A idéia de um algoritmo capaz de encontrar as fontes que originaram uma dada mistura, sem qualquer conhecimento prévio sobre as fontes, desperta o interesse e curiosidade da comunidade científica há décadas. No entanto, ICA é uma técnica relativamente nova. Foi introduzida no início dos anos 80 como um modelo de rede neural e conquistou a atenção da comunidade científica por seus expressivos resultados na separação de sinais da fala. Em pouco tempo, a técnica ganhou novas aplicações como *data mining*, séries financeiras, sinais de áudio e telecomunicações.

Na área biomédica, ICA tem apresentado resultados surpreendentes para separação de sinais de EEG.

O emprego de ICA no domínio do processamento de imagens é ainda mais inovador. Em 2001, Hyvärinen demonstra a utilização do algoritmo de ICA para remoção de ruído em imagens naturais [46]. No mesmo ano, Christoyianni apresentou a idéia de separar um mamograma em imagens-base [19] e usar os parâmetros do modelo de mistura de ICA para classificar regiões da mama em áreas normais e anormais.

Inicialmente, este trabalho foi idealizado como um classificador de lesões da mama, dando prosseguimento ao trabalho desenvolvido em [10]. Nesse trabalho, era usada Morfologia Matemática para extrair parâmetros a serem apresentados a uma rede neural que classificaria a lesão de acordo com a Classificação de Le Gal [6], usando o banco de imagens público MIAS-UK (The Mammographic Images Analysis Society database).

Com o banco de imagens do INCa, poder-se-ia desenvolver um sistema aplicado à casuística nacional. As imagens foram digitalizadas em altíssima resolução (1200 dpi), permitindo uma sub-amostragem em prol de um algoritmo mais veloz. No entanto, apenas um pequeno número de radiografias contava com o diagnóstico histológico, essencial para validar o experimento, uma vez que em diversos casos o diagnóstico radiológico apenas aponta a presença de áreas suspeitas no mamograma.

A idéia original foi substituída por um sistema que identificasse áreas lesionadas nas mamografias. Foi desenvolvido, assim, um sistema com objetivo de avaliar o desempenho de algoritmos de separação cega de fontes na identificação de regiões

lesionadas contidas em um conjunto de regiões de interesse selecionadas por um radiologista experiente.

Comparando com os outros métodos empregados, o algoritmo de extração de parâmetros por ICA foi bem sucedido, com cerca de 86% de acertos. Os métodos de extração de parâmetros largamente empregados na literatura não atingem taxas de acerto tão elevadas para o conjunto de imagens utilizado.

A Tabela 7.1 apresenta um resumo com os melhores resultados obtidos em todos os experimentos realizados onde a sensibilidade está definida como

$$S = \left(\frac{VP}{VP + FN} \cdot 100 \right) \text{ e a especificidade como } E = \left(\frac{VN}{VN + FP} \cdot 100 \right).$$

Tabela 7.1 – Resumo dos melhores resultados obtidos.

Método	% acertos	S %	E %	# neurônios 1ª camada	# neurônios 2ª camada	# comp. principais
Kurtosis + BP	82,07	82,44	81,70	-	35	75
Neg-entropia +BP	83,22	87,52	78,91	-	100	60
Bell Sejnowsky +BP	82,24	84,88	79,60	-	35	40
Translação + Neg-entropia + BP	55,37	-	-	-	100	60
Neg-entropia +CAC	85,68	86,92	84,44	20	90	180
Top-Hat + Neg-entropia +CAC	81,66	85,86	77,46	20	90	90
GLCM +BP (3 pixels / 90°)	63,78	51,76	75,80	-	100	-
CClx +BP	75,41	84,24	66,58	-	140	-

Em [19], Christoyianni alcançou 88,23% de acertos para separar regiões normais e anormais. Verifica-se, porém, que as regiões de interesse usadas em [19] apresentam uma separabilidade aparente nos histogramas, o que também pode ser inferido a partir dos resultados obtidos usando a matriz de co-ocorrência como extrator de parâmetros. Com a GLCM, Christoyianni obteve 84,87% de acertos, contra 63,78 % obtido neste trabalho. A maioria das lesões selecionadas no banco do INCa é quase imperceptível a olho nu, como se pode identificar nas Figura 5.1 (d) e (e). Da mesma forma, a

distribuição de probabilidade das ROIs também não revela um potencial de separação aparente entre regiões normais e lesionadas, conforme verifica-se na Figura 5.2.

Foi introduzido, ainda, o conceito de componentes de atuação conjunta, a fim de realizar um mapeamento linear das entradas, considerando sua capacidade discriminatória. Com o emprego da técnica, conseguiu-se eliminar possíveis ruídos na entrada que estariam atrapalhando o desempenho da rede neural em seu trabalho de classificar os parâmetros apresentados a ela. A técnica é também bastante recente [47, 48] e tem apresentado resultados promissores no campo de redes neurais.

O trabalho ainda comporta uma proposta de implementação clínica, como parte de um sistema de CAD. Seria interessante validar o experimento em um hospital, e comparar com a taxa de acertos de diversos radiologistas, com diferentes níveis de experiência. Uma vez digitalizada a mamografia, com um clique do radiologista, o algoritmo selecionaria uma região de interesse e informaria se a região selecionada contém uma lesão. A rede neural pode ser modificada também para indicar qual a chance percentual de ter ou não uma lesão na região de interesse selecionada. Ainda, o algoritmo poderia ser adaptado para indicar automaticamente as regiões suspeitas de lesão em um mamograma digitalizado.

Seria igualmente interessante avaliar o desempenho com modelos de mistura não lineares, visto que os modelos biológicos tendem a se comportar de maneira não linear.

Dispondo de um banco com diversas imagens contendo calcificações classificadas segundo BIRADS ou Le Gal, poderia ser avaliado o potencial classificador do sistema, conforme idealizado inicialmente.

Apêndices

A Teorema do Limite Central

Seja $x_k = \sum_{i=1}^k z_i$ a soma de uma seqüência de variáveis aleatórias identicamente distribuídas e independentes entre si. Como a média e a variância de x_k pode crescer sem limites à medida que $k \rightarrow \infty$, considere as variáveis normalizadas:

$$y_k = \frac{x_k - m_{x_k}}{\sigma_{x_k}}, \text{ onde } m_{x_k} \text{ e } \sigma_{x_k} \text{ são, respectivamente, a média e a variância de } x_k.$$

Pode-se provar que a distribuição de y_k converge para uma distribuição gaussiana com média zero e variância unitário quando $k \rightarrow \infty$.

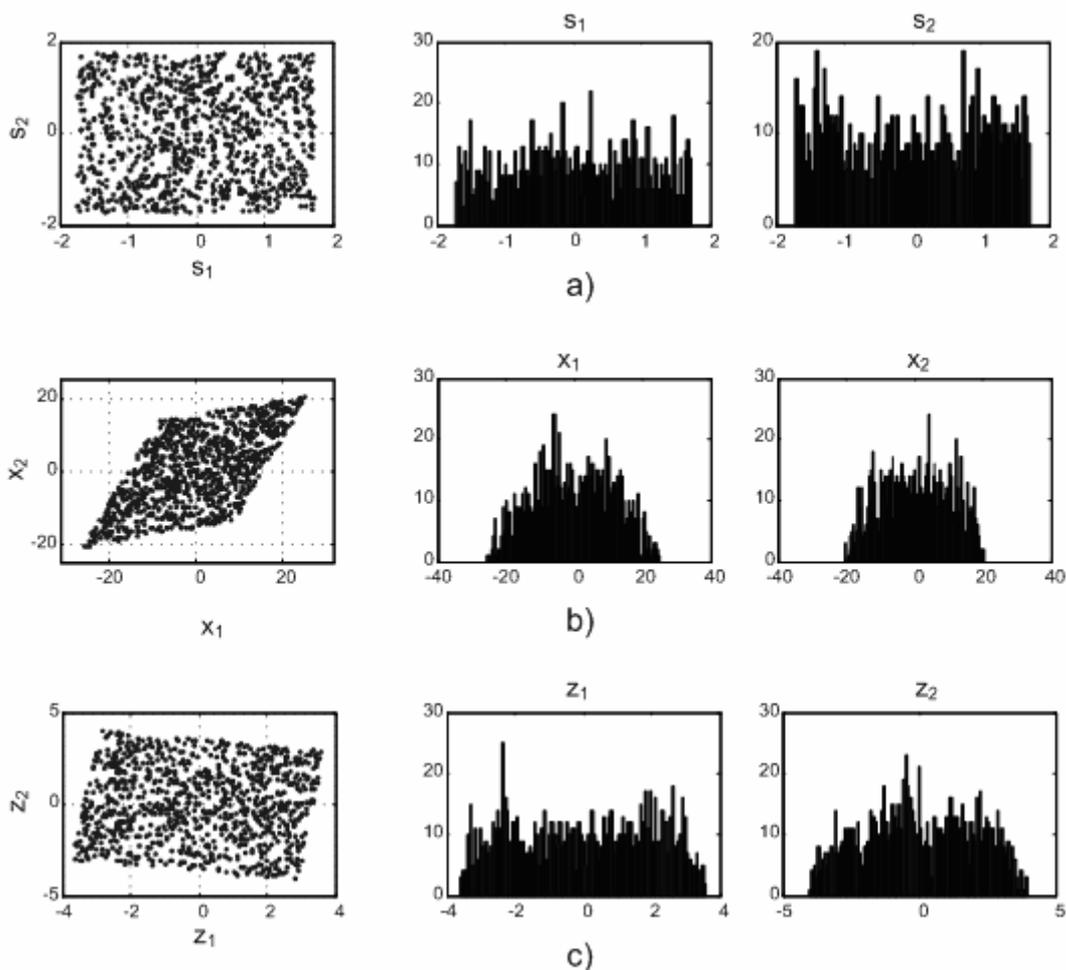


Figura A.1

Para exemplificar o teorema, considere duas componentes independentes que apresentem distribuição uniforme. A distribuição conjunta das variáveis originais, bem como os histogramas para as densidades uniformes, são apresentados na Figura A.1 a). A Figura A.1 b) mostra a distribuição conjunta das variáveis misturadas. Observe como os histogramas da mistura são mais “gaussianos” que os histogramas das variáveis originais. A mistura é branqueada e o resultado é mostrado na Figura A.1 c). A distribuição conjunta é apenas uma rotação da densidade conjunta original.

B Morfologia Matemática

A morfologia matemática é um ramo do processamento de imagens cujo desenvolvimento foi iniciado no *Centre de Morphologie Mathematique* (CMM), na França. Consiste no estudo da estrutura geométrica das entidades presentes em uma imagem [49, 50].

A grande chave da morfologia matemática é o **elemento estruturante**. O elemento estruturante é um conjunto definido e conhecido (forma, tamanho), que é comparado, a partir de uma transformação, ao conjunto desconhecido da imagem. O formato e o tamanho do elemento estruturante possibilitam testar e quantificar de que maneira o elemento estruturante "está ou não está contido" na imagem. Marcando os resultados das posições onde o elemento estruturante inclui-se na imagem, tem-se uma primeira resposta sobre a estrutura geométrica das entidades dessa imagem.

Existem dois tipos de morfologia matemática: a *morfologia binária*, que se aplica sobre imagens binárias, e a *morfologia em níveis de cinza*, que se aplica sobre imagens em níveis de cinza.

Abaixo, serão apresentados alguns operadores morfológicos básicos. Em [10], podem ser encontradas maiores informações sobre Morfologia Matemática e as operações de filtragem morfológica.

B.1 Erosão

A operação de **erosão** pode ser definida por:

$$A \ominus K = \{x \in \mathfrak{R}^N / x+k \in A, \forall k \in K\},$$

onde A e K são conjuntos do espaço \mathfrak{R}^N e K é chamado elemento estruturante.

Assim, durante a erosão, o valor do *pixel* na origem do elemento estruturante é substituído pelo valor mínimo dos *pixels* ao redor.

As Figuras B.1 e B.2 exemplificam a erosão de uma imagem binária. K é o elemento estruturante no processo. Ao ser posicionado em um determinado ponto da imagem original A , o elemento estruturante somente aparecerá na imagem resultante se estiver incluído em A .

Em ambas as figuras, tem-se em **c)** e **d)** o andamento do processo da erosão. Em **c)** tem-se que todos os *pixels* do elemento estruturante coincidem com um *pixel* da imagem binária A , de forma que o *pixel* correspondente à origem do elemento estruturante (assinalado com as setas $\downarrow \rightarrow$) estará presente na imagem resultante. O mesmo não ocorre em **d)**. Há pelo menos 1 *pixel* do elemento estruturante sem correspondência na imagem original. Assim, ele não aparecerá na imagem resultante.

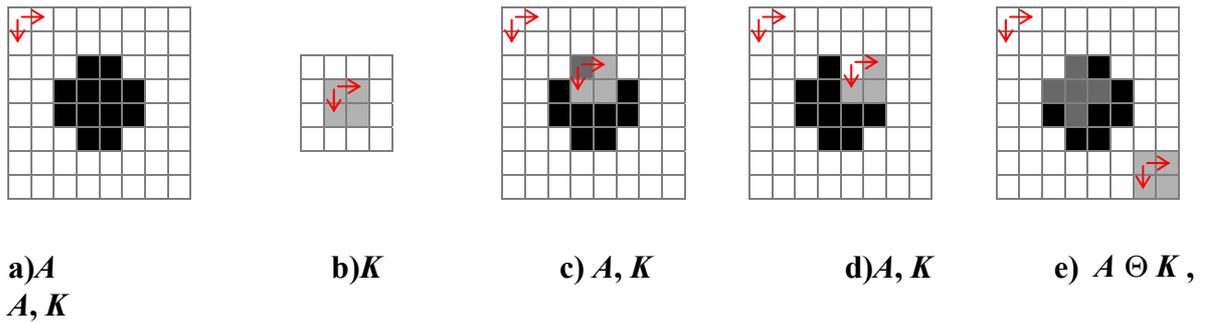


Figura B.1

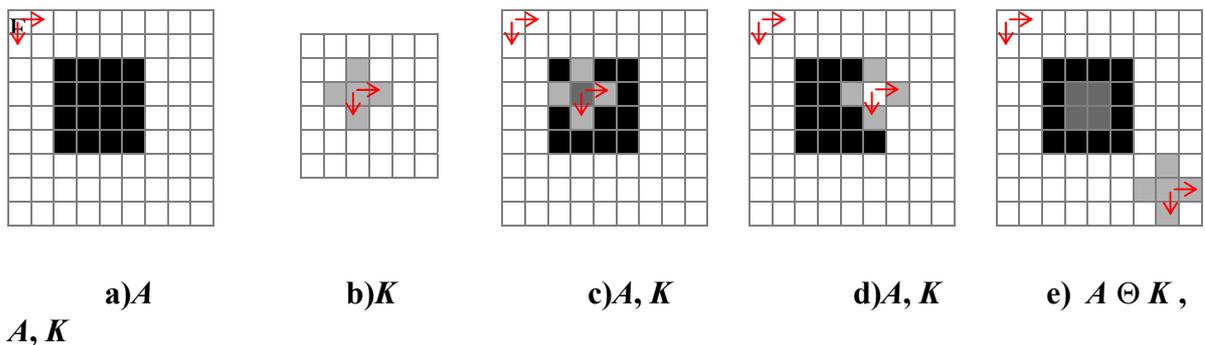


Figura B.2

B.2 Dilatação

A **dilatação** é a operação dual da erosão. Sua interpretação é, pois, complementar à interpretação da erosão: o valor na origem do elemento estruturante será substituído pelo valor máximo dos pixels vizinhos. Dessa forma:

$$A \oplus K = \{x \in \mathcal{R}^N / x = a + k \in A, a \in A \wedge k \in K\},$$

onde A e K são conjuntos do espaço \mathcal{R}^N e K é chamado elemento estruturante.

A imagem dilatada é sempre maior do que a imagem original. A dilatação permite conectar conjuntos separados e também preencher pequenos buracos.

As Figuras B.3 e B.4 mostram o objeto binário original e o resultado da sua dilatação pelo elemento estruturante K . O elemento estruturante aparecerá na imagem resultante se algum de seus *pixels* coincidir com a imagem original.

Em ambas as figuras, tem-se em **c)** e **d)** o andamento do processo da dilatação. Verifica-se que basta que pelo menos 1 *pixel* do elemento estruturante coincida com a imagem original para que o *pixel* correspondente à origem do elemento estruturante (assinalado com as setas $\searrow \rightarrow$) esteja presente na imagem resultante.

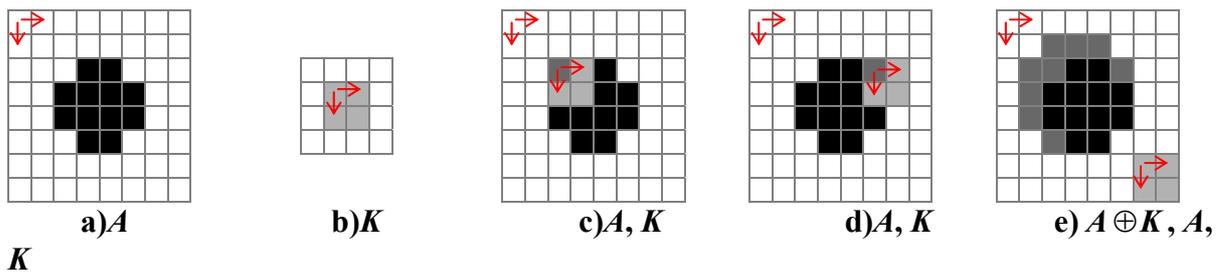


Figura B.3

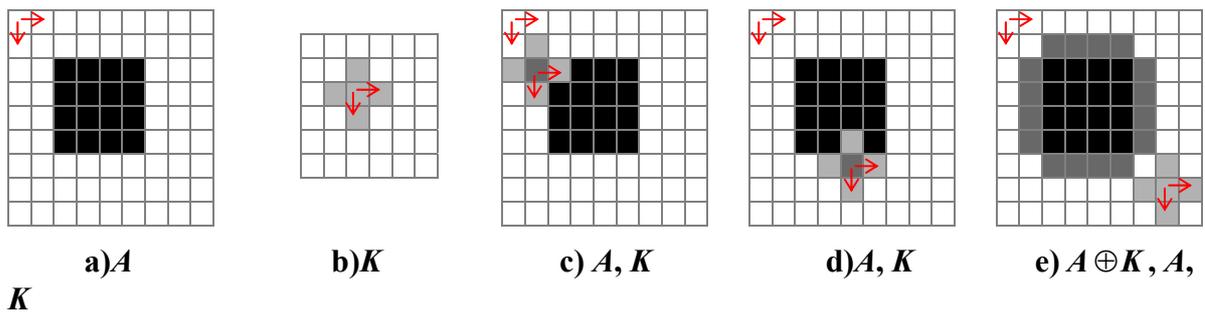


Figura B.4

B.3 Abertura

Duas outras importantes operações da morfologia matemática são a **abertura** e o **fechamento**.

Essas operações permitem que a imagem seja filtrada, corrigida ou retocada sem modificar as características de forma e tamanho dos elementos que restam após essas operações. Os elementos inferiores ao elemento estruturante, com relação ao tamanho, são eliminados. A operação de abertura pode ser expressa como uma composição de uma erosão seguida por uma dilatação, ambas realizadas com o mesmo elemento estruturante:

$$A \circ K = (A \ominus K) \oplus K$$

onde A e K são conjuntos do espaço \mathcal{R}^N e K é chamado elemento estruturante.

A Figura B.5 mostra um exemplo de abertura binária. Em **d)** verifica-se a dilatação da imagem obtida pela erosão de A pelo elemento estruturante K . Dessa forma, ao ser posicionado em um determinado ponto, o elemento estruturante aparecerá completamente na imagem se *todos* os seus *pixels* estiverem inclusos na imagem original.

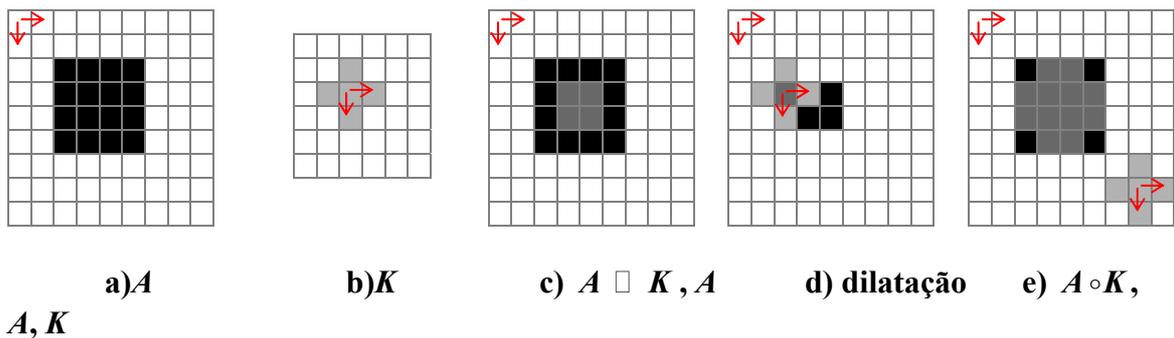


Figura B.5

As principais características da abertura morfológica binária são:

- a) a imagem resultante é de forma geral mais regular e menos rica em detalhes;
- f) a abertura nivela os contornos pelo interior;
- g) separa as partículas;
- h) elimina pequenas partículas inferiores em tamanho em relação ao elemento estruturante;
- i) as entidades restantes após abertura ficam quase idênticas às originais.

A Figura B.6 apresenta uma imagem em que se pode melhor analisar as propriedades da operação de abertura.

Os pontos isolados na imagem original menores que o elemento estruturante, desaparecem na imagem resultante. A operação também abre os contornos fechados cuja linha demarcatória é menor que o elemento estruturante. Os elementos ligados ao corpo principal da imagem por uma linha de dimensões menores que o elemento

estruturante são desconectadas. O restante da imagem permanece inalterado. O deslocamento de 1 *pixel* para a esquerda deve-se ao fato de o elemento estruturante não ser simétrico em torno da sua origem (assinalada pelo símbolo $\downarrow \rightarrow$).

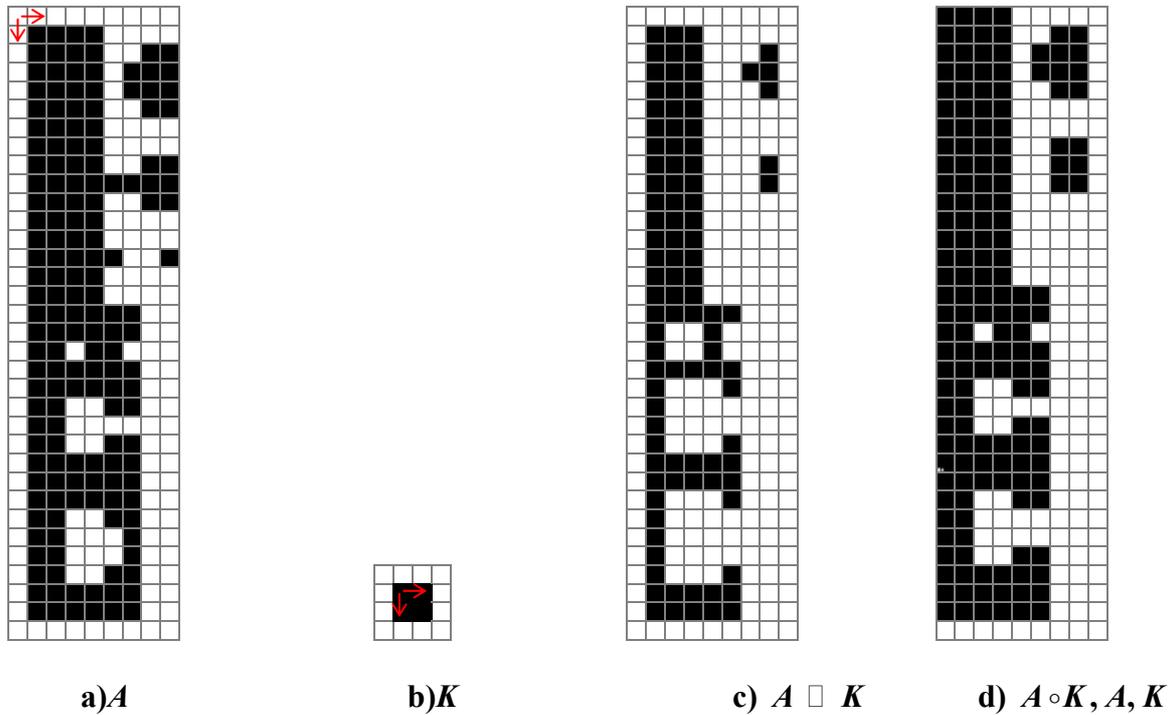


Figura B.6

B.4 Fechamento

A operação inversa da abertura é o fechamento. Esse operador morfológico pode ser expresso como uma composição de uma dilatação seguida de uma erosão.

$$A \bullet K = (A \oplus B) \ominus K$$

onde A e K são conjuntos do espaço \mathcal{R}^N e K é chamado elemento estruturante.

A Figura B.7 mostra um exemplo de um fechamento binário. Ao ser posicionado em um dado ponto da imagem original, o elemento estruturante somente aparecerá na imagem resultante se *todos* os seus *pixels* estiverem inclusos no complemento da imagem original.

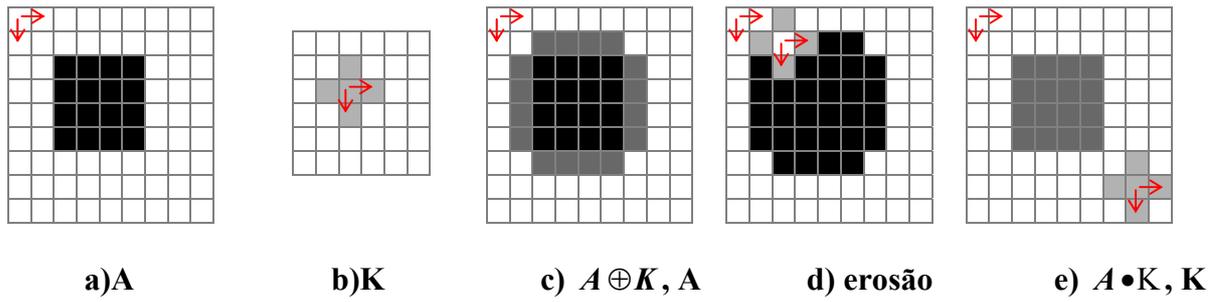


Figura B.7

Os efeitos de um fechamento morfológicos binários são:

- o fechamento suaviza as fronteiras pelo exterior;
- preenche buracos no interior das partículas inferiores em tamanho em relação ao elemento estruturante;
- emenda partículas próximas;
- as entidades restantes após o fechamento ficam quase idênticas;
- a imagem fechada é mais regular e menos rica em detalhes do que a imagem original.

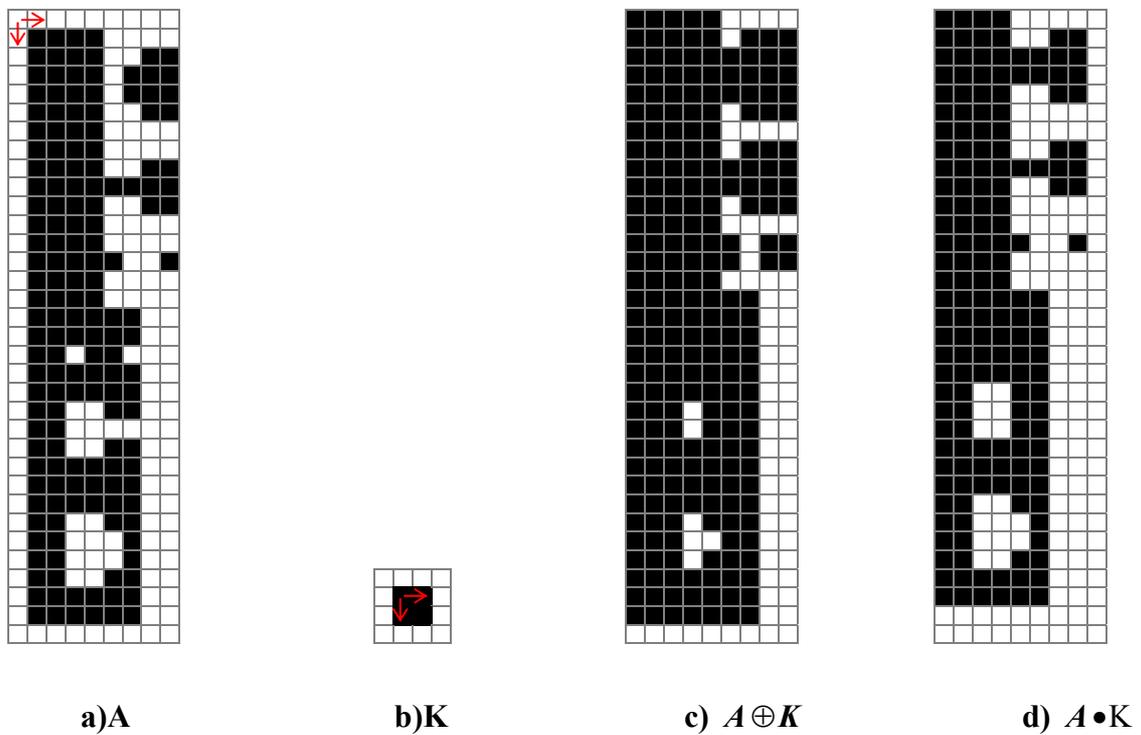


Figura B.8

A Figura B.8 apresenta uma imagem em que se pode melhor analisar as propriedades da operação de fechamento.

Os pontos isolados na imagem original menores que o elemento estruturante se mantêm na imagem resultante. A operação também fecha elementos cuja linha demarcatória seja menor que o elemento estruturante. Os elementos desconectados ao corpo principal da imagem por uma distância menor que o elemento estruturante são ligados a ele. O restante da imagem permanece inalterado. O deslocamento de 1 *pixel* para a esquerda deve-se ao fato de o elemento estruturante não ser simétrico em torno da origem (assinalada pelo símbolo $\downarrow\rightarrow$).

B.5 Dilatação condicional

A imagem resultante de uma dilatação condicional provém de uma dilatação da imagem original pelo elemento estruturante, dada uma determinada condição imposta por uma outra imagem. Ou seja, a dilatação é feita sobre a imagem resultante da interseção entre a imagem original e a imagem condicionante, chamada *marcadores*. A imagem final é obtida pela interseção do resultado desta dilatação com a imagem condicionante.

Para n iterações serão operadas sucessivas dilatações sobre a imagem resultante do passo anterior. O resultado da interseção de uma imagem em tons de cinza com os marcadores é o valor mínimo resultante da comparação *pixel a pixel*.

Neste exemplo, deseja-se extrair os dígitos 1 e 4 da imagem original. Como ambos apresentam um traço vertical (|) em comum, é possível extrair marcadores na imagem (a partir de erosões, por exemplo).



a) Imagem original

b) Marcadores

c) Imagem resultante da dilatação condicional

Figura B.9

A Figura B.9 b) apresenta os marcadores. A interseção dessa imagem com a imagem original (condicionante) é dilatada e a interseção do resultado dessa dilatação com a imagem original irá sofrer nova dilatação. O processo se repete em 10 iterações. O resultado final é mostrado na Figura B.9 c).

B.6 Erosão condicional

A imagem resultante de uma erosão condicional provém de uma erosão da imagem original pelo elemento estruturante, dada uma determinada condição imposta por uma outra imagem. Em outras palavras, a erosão é feita sobre a imagem resultante da união entre a imagem original e a imagem condicionante, sendo esta imagem chamada *marcadores*.

Para n iterações serão operadas sucessivas erosões sobre a união da imagem resultante no passo anterior com a imagem condicionante. O resultado da união de uma imagem em tons de cinza com os marcadores é o valor máximo resultante da comparação *pixel a pixel*.



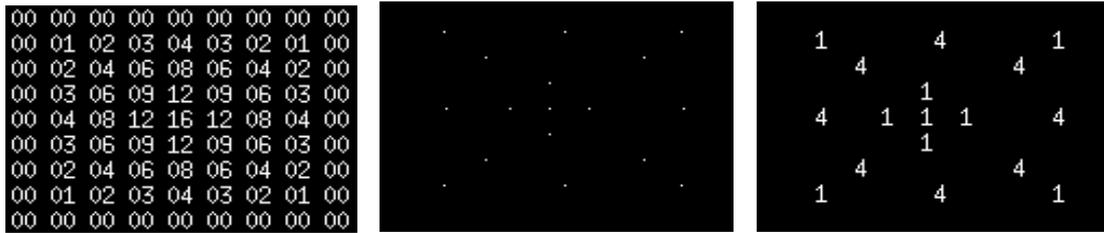
Figura B.10

O exemplo é dual ao exemplo da seção anterior.

B.7 Reconstrução Inferior

O operador de reconstrução inferior cria uma imagem a partir de um número de iterações recursivas (até a estabilização da saída) de dilatação da imagem original, pelo elemento estruturante, condicionada por uma outra imagem.

Em outras palavras, a *inf-reconstrução* da imagem g a partir da imagem marcadora f nada mais é que n dilatações de f condicionadas pela imagem g .



a) Imagem original

b) Marcadores

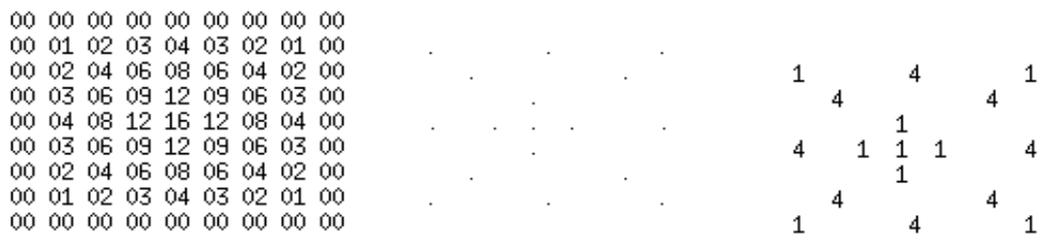
c) Imagem resultante da reconstrução inferior

Figura B.11

O exemplo acima revela que o operador *inf-reconstrução* executa a mesma função que um número suficiente de iterações da função de dilatação condicional, uma vez que a imagem resultante está em pleno acordo com o resultado da dilatação condicional apresentada na Figura B.9 c).

B.8 Reconstrução superior

O operador de reconstrução superior cria uma imagem a partir de um número de iterações recursivas (até a estabilização da saída) de erosão da imagem original, chamada *marcadores*, pelo elemento estruturante, condicionada por uma outra imagem.



a) Imagem original

b) Marcadores

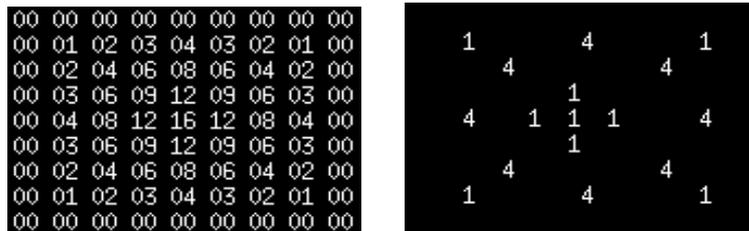
c) Imagem resultante da reconstrução superior

Figura B.12

Em outras palavras, a *sup-reconstrução* da imagem *g* a partir da imagem marcadora *f* nada mais é que *n* erosões de *f* condicionadas pela imagem *g*. O exemplo é dual ao exemplo da seção anterior.

B.9 Abertura por reconstrução

Esse operador morfológico realiza a reconstrução de uma imagem a partir da imagem original erodida que funciona como marcadores. O operador aplica uma série de dilatações na imagem original erodida (marcadores) usando a imagem original como imagem condicionante.



a) Imagem original

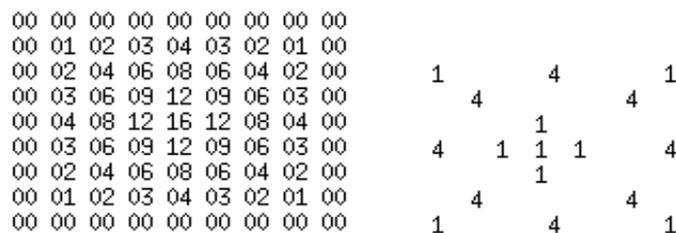
b) Imagem resultante da abertura por reconstrução

Figura B.13

A operação de abertura por reconstrução consegue assim recuperar elementos da imagem original que apresentem um padrão semelhante ao obtido pelo resultado da erosão da imagem original por um dado elemento estruturante. O exemplo apresentado equivale aos exemplos apresentados Figuras B.9 e B.11.

B.10 Fechamento por reconstrução

Esse operador morfológico realiza a reconstrução de uma imagem a partir da imagem original dilatada que funciona como marcadores.



a) Imagem original

c) Imagem resultante do fechamento por reconstrução

Figura B.14

O operador aplica uma série de erosões na imagem marcadora usando a imagem original dilatada como imagem condicional. O exemplo é dual ao exemplo da seção anterior.

B.11 Top-Hat

O algoritmo de *top-hat* é um método de filtragem morfológica usado para excluir padrões indesejáveis na imagem que possuam características comuns. A imagem resultante de uma determinada operação morfológica é subtraída da imagem original. Para excluir objetos com intensidade maior que o fundo da imagem, deve-se realizar abertura na imagem original, antes da subtração. Analogamente, para retirar objetos de menor intensidade, é preciso previamente realizar um fechamento.

Em outras palavras o *top-hat* constitui-se de uma abertura ou um fechamento (com ou sem reconstrução) seguidos da subtração da imagem resultante dessa operação morfológica pela imagem original. Com isso, na imagem final aparecem apenas os elementos que não foram selecionados pela operação anterior à subtração.

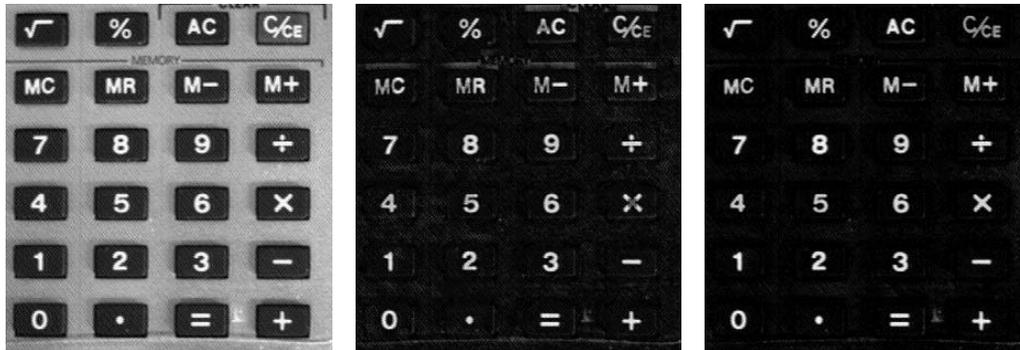
O algoritmo pode ser melhor compreendido se a variação espacial da intensidade do sinal da imagem for interpretado como uma superfície topográfica. Essa superfície seria percorrida por um “chapéu” e apenas os objetos na imagem com alto contraste conseguiriam ultrapassar o topo deste e aparecer na imagem resultante. Dessa forma, controlando-se as dimensões do chapéu, ou seja, a abertura e a altura, objetos de diferentes diâmetros e brilho seriam detectados.

Os exemplos abaixo ilustram o uso do *top-hat*.

No primeiro exemplo, deseja-se ressaltar os números da calculadora, ignorando o restante da imagem (botões, inscrições, relevos, etc.). Como os números são mais brilhantes que o restante da imagem basta extraí-los da imagem original com uma operação de abertura usando um elemento estruturante que eroda os números de forma que a dilatação não seja capaz de recuperá-los.

A imagem resultante da abertura será semelhante à original, sem os números. Para obter uma imagem que os contenha, basta subtrair essa imagem resultante da abertura pela imagem original. Em **b)** tem-se o resultado do *top-hat* sobre a imagem aberta; **c)** ilustra o *top-hat* por reconstrução, ou seja, a subtração da imagem original pela imagem aberta por reconstrução.

Em outras palavras, nesse caso, foi feita uma série de dilatações na imagem marcadora (imagem original erodida) usando a imagem original como imagem condicionante. A imagem em **c)** é, portanto, menos “borrada” que a imagem **b)** pois foi condicionada à imagem original.



a) f -Imagem original **b) g** – Abertura com Top-Hat **c) h** – Abertura por reconstrução com Top-Hat

Figura B.15

O exemplo abaixo é dual ao exemplo anterior. Deseja-se ressaltar as bactérias da amostra, ignorando o restante da imagem. Como elas são menos brilhantes que o restante da imagem, basta extraí-las da imagem original com uma operação de fechamento usando um elemento estruturante que dilate o *background* de forma que a erosão subsequente não seja capaz de recuperá-las.



a) f -Imagem original **b) g** – Abertura com Top-Hat **c) h** – Abertura por reconstrução com Top-Hat

Figura B.16

A imagem resultante do fechamento será apenas o *background* sem as bactérias. Para obter uma imagem que as contenha, basta subtrair essa imagem resultante do

fechamento pela imagem original. Em **b)** tem-se o resultado do *top-hat* sobre a imagem fechada; **c)** ilustra o *top-hat* por reconstrução, ou seja, a subtração da imagem original pela imagem fechada por reconstrução. Em outras palavras, nesse caso, foi feita uma série de erosões na imagem marcadora (imagem original dilatada) usando a imagem original como imagem condicionante. A imagem em **c)** é, portanto, menos “borrada” que a imagem **b)** pois foi condicionada à imagem original.

C Componentes Principais

A redundância entre variáveis é medida pela correlação entre suas observações. Dado um conjunto de medidas multi-variáveis, pretende-se encontrar um grupo de variáveis com menos redundância, que poderia representar o conjunto por um número menor de variáveis.

Seja um vetor aleatório \mathbf{x} com N variáveis dispostas nas linhas e P ocorrências.

Deseja-se encontrar uma transformação linear \mathbf{z} de \mathbf{x} , com K elementos, tais que $K < N$, de forma a remover a redundância devido à correlação entre as variáveis. Para tanto, é preciso encontrar um sistema de coordenadas ortogonais de forma que os elementos de \mathbf{x} projetados nas novas coordenadas estejam descorrelacionados entre si. Ao mesmo tempo, as variâncias das projeções de \mathbf{x} nas novas coordenadas são maximizadas de maneira que o primeiro eixo corresponde à máxima variância, o segundo eixo corresponde à variância máxima na direção ortogonal ao primeiro eixo, e assim por diante.

Matematicamente, considere a combinação linear:

$$\mathbf{z} = \mathbf{V}^T \cdot \mathbf{x} \quad (\text{C.1})$$

A matriz de autocovariância de \mathbf{z} deve, pois, ser diagonal, de forma que:

$$\mathbf{C}_z = E[\mathbf{z} \cdot \mathbf{z}^T] = \mathbf{\Lambda} \quad (\text{C.2})$$

Sabe-se, porém, que:

$$\mathbf{C}_z = E[\mathbf{z} \cdot \mathbf{z}^T] = E[\mathbf{V}^T \cdot \mathbf{z} \cdot \mathbf{z}^T \cdot \mathbf{V}] = \mathbf{V}^T \cdot E[\mathbf{x} \cdot \mathbf{x}^T] \cdot \mathbf{V} = \mathbf{V}^T \cdot \mathbf{C}_x \cdot \mathbf{V} \quad (\text{C.3})$$

De (C.2) e (C.3), tem-se que:

$$\mathbf{V}^T \cdot \mathbf{C}_x \cdot \mathbf{V} = \mathbf{\Lambda} \quad (\text{C.4})$$

Portanto, \mathbf{V}^T é a matriz ortogonal com $N \times P$ elementos que diagonaliza a matriz \mathbf{C}_x . Como resultado clássico da álgebra, \mathbf{V}^T é a matriz cujas linhas são os autovetores da matriz de \mathbf{C}_x , correspondentes aos autovalores em ordem crescente de variância. $\mathbf{\Lambda}$ é uma matriz diagonal $N \times N$ cujos elementos são os autovalores de \mathbf{C}_x , ou correspondentemente, as variâncias de \mathbf{z} , em ordem crescente de energia.

A fim de exemplificar o conceito da análise em componentes principais considere duas variáveis aleatórias x_1 e x_2 que apresentam as seguintes ocorrências:

$$[x_1, x_2] = \begin{bmatrix} 1 & 1 \\ 2 & 2 \\ 3 & 3 \\ 4 & 4 \\ 5 & 5 \\ 6 & 6 \\ 7 & 7 \\ 8 & 8 \end{bmatrix}$$

Claramente, verifica-se que $x_1 = x_2$, e que, portanto, a variável x_2 não introduz nenhuma novidade acerca da variável x_1 . A dispersão das variáveis pode ser verificada pela Figura C.1.

Assim, a informação pode ser comprimida em apenas uma variável. A primeira componente principal é a direção ótima de projeção indicada pelo vetor u_1 na Figura C.1. Nessa direção, toda a energia do vetor \mathbf{x} está concentrada.

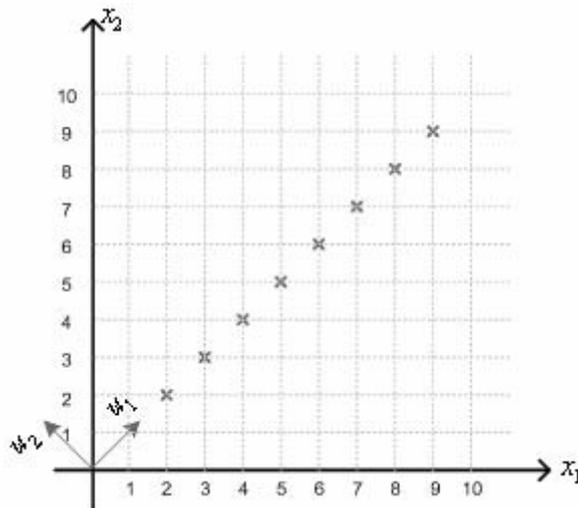


Figura C.1

Rotacionando os eixos formados por u_1 e u_2 tem-se a Figura C.2

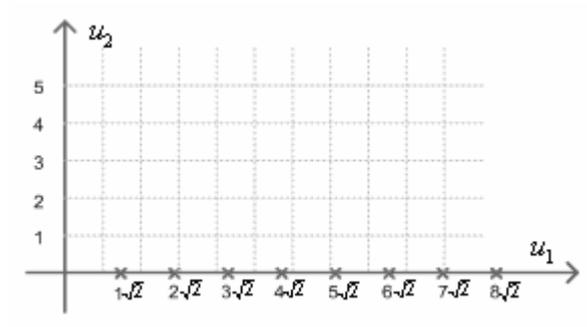


Figura C.2.

D Matriz de Co-ocorrência

A matriz de co-ocorrência é um histograma bidimensional que descreve a ocorrência de pares de *pixels* espaçados por uma determinada distância em uma dada direção. [25]

Seja uma imagem **I**, com N_x pixels de resolução na direção horizontal e N_y pixels na direção vertical, sendo quantizada com \mathcal{G} tons de cinza.

$P_{x,y}(i,j)$ é uma matriz com $\mathcal{G} \times \mathcal{G}$, cujos elementos representam a contagem da ocorrência em conjunto dos níveis de cinza (i,j) na imagem **I**, separados por um vetor (x,y) . Matematicamente,

$$P_{x,y}(i,j) = \# \left\{ [(u,v), (s,t)] \in [(N_x \times N_y) \times (N_x \times N_y)] \mid \left[((\mathbf{I}(u,v) = i) \wedge (\mathbf{I}(s,t) = j)) \right] \wedge [(s-u = x) \wedge (t-v = y)] \right\} \quad (\text{D.1})$$

É comum expressar a matriz de co-ocorrência em função de uma distância d em *pixels* em um determinado ângulo θ .

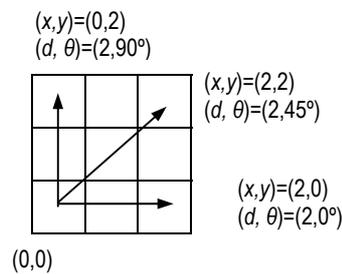


Figura D.1.

Para exemplificar o cálculo da matriz de co-ocorrência, considere a seguinte matriz com 3 níveis de cinza:

0	0	1	1	2
0	0	1	1	1
0	2	2	2	1
2	2	0	1	2

Figura D.2.

As matrizes de co-ocorrência para uma distância de 1 *pixel*, e ângulos de 0°, 45° e 90° são respectivamente dadas por:

$$GLCM_{1,90^\circ} = \begin{bmatrix} 3 & 0 & 2 \\ 0 & 3 & 3 \\ 1 & 2 & 1 \end{bmatrix} \quad GLCM_{1,0^\circ} = \begin{bmatrix} 2 & 3 & 1 \\ 0 & 3 & 2 \\ 1 & 1 & 3 \end{bmatrix} \quad GLCM_{1,45^\circ} = \begin{bmatrix} 1 & 1 & 3 \\ 0 & 3 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Em geral, utiliza-se a matriz normalizada para computar a frequência relativa:

$$p_{x,y}(i, j) = \frac{P_{x,y}(i, j)}{\sum_{i,j} P_{x,y}(i, j)} \quad (D.2)$$

Haralick [42] sugere um conjunto de 28 parâmetros que podem ser usados para descrição de textura a partir da matriz de co-ocorrência. Vários autores [Gupta, Undrill, 1996][Meersman, 1998] desenvolveram algoritmos de detecção de microcalcificações e massas suspeitas em mamografias a partir da análise de textura através da matriz de co-ocorrência. A seleção de quantos e quais parâmetros utilizar foi baseada nos estudos apresentados na literatura. Dessa forma, de acordo com a literatura, os 5 parâmetros com melhor desempenho para classificação de textura em imagens mamográficas foram:

Entropia: medida da homogeneidade da imagem; quanto maior a entropia, mais desordenada é a imagem, em uma determinada direção, a uma dada distância

$$H = \sum_{i,j} p(i, j) \cdot \log[p(i, j)] \quad (D.3)$$

Contraste: medida da quantidade de variação dos tons de cinza da imagem, em uma determinada direção, a uma dada distância:

$$C = \sum_{i,j} (i - j)^2 \cdot p(i, j) \quad (D.4)$$

Desvio padrão: medida do quanto o histograma de pares de pixels se espalha em torno do valor médio, em uma determinada direção, a uma dada distância:

$$\sigma = \sqrt{\frac{1}{G^2} \sum_{i,j} (p(i,j) - \bar{p})^2} \quad (D.5)$$

Segundo momento angular:

$$SMA = \sum_{i,j} p(i,j)^2 \quad (D.6)$$

Correlação: medida da ordenação local dos tons de cinza da imagem

$$R = \frac{\sum_{i,j} i \cdot j \cdot p(i,j) - \mu_x \mu_y}{\sqrt{\sigma_x^2 \cdot \sigma_y^2}} \quad (D.7)$$

onde $\sigma_x^2 = \sum_i i \sum_j p(i,j) - \mu_x^2$ e $\sigma_y^2 = \sum_j j \sum_i p(i,j) - \mu_y^2$,

para $\mu_x = \sum_i i \sum_j p(i,j)$ e $\mu_y = \sum_j j \sum_i p(i,j)$

Exemplo:

Seja uma imagem tal qual a Figura C.3:

0	0	1	1	2	2	1	0	2	0
0	0	1	1	2	2	1	0	2	0
0	0	1	1	2	2	1	0	2	0
0	0	1	1	2	2	1	0	2	0

Figura D.3.

$$GLCM_{1,0^\circ} = \begin{bmatrix} 4 & 4 & 4 \\ 4 & 4 & 4 \\ 4 & 4 & 4 \end{bmatrix}$$

$$H=2.1972 \quad C=0.3333 \quad \sigma=0 \quad R=0.1111 \quad SMA=0.1111$$

Todos os tons de cinza de *pixels* vizinhos na horizontal são equiprováveis aos pares, resultando em uma matriz de co-ocorrência homogênea. O valor da entropia é,

portanto, alto, o que significa que a ocorrência de um tom de cinza nesta direção não nos permite uma dedução sobre o tom do vizinho.

Nesse caso, o contraste é não nulo, o que, de fato se verifica, pois há uma variação de tons de cinza na horizontal, porém o desvio padrão zera, indicando que a ocorrência conjunta dos níveis de cinza apresenta uma total concentração.

$$GLCM_{1,90^\circ} = \begin{bmatrix} 12 & 0 & 0 \\ 0 & 9 & 0 \\ 0 & 0 & 9 \end{bmatrix}$$

$$H=1.0889 \quad C=0 \quad \sigma=0.1691 \quad R=0.3400 \quad SMA=0.3400$$

O valor da entropia é mais baixo que para o calculado a partir da matriz de co-ocorrência para um ângulo de 90°, ou seja, existe uma certa predição vertical, o que se verifica pelo fato de ser um padrão de listras verticais.

O contraste é zero para esse caso, pois a matriz de co-ocorrência é diagonal. Isso significa que não existem transições entre tons de cinza na vertical. O desvio padrão é não-nulo, uma vez que há os pares de tons de cinza ocorrem com probabilidades diversas.

Há, porém, um aumento da correlação e do segundo momento angular, indicando um padrão vertical.

E Curva de Complexidade

A curva de complexidade $C(\alpha)$ é definida como o número de transições na imagem binária [26]:

$$C(\alpha) = \frac{T_{1,0}^\alpha + T_{0,1}^\alpha}{N_x \times (N_y - 1) + N_y \times (N_x - 1)}, \alpha \in [0, 1, \dots, G - 1] \quad (\text{E.1})$$

onde $T_{i,j}^\alpha$ é o número total de transições na imagem binária $f^\alpha(x, y)$, para um passo fixo (i, j) e $N_x \times N_y$ é o tamanho da imagem.

$$f^\alpha(x, y) = \begin{cases} 1, & \forall f(x, y) \geq \alpha \\ 0, & \text{caso contrário} \end{cases} \quad (\text{E.2})$$

onde $\alpha \in [0, 1, \dots, G - 1]$,

G é o número de níveis de cinza da imagem binária

Assim como para a matriz de co-ocorrência, foram selecionados 5 parâmetros sobre a curva de complexidade, de acordo com a literatura. [51]

$$\underline{\text{Valor máximo:}} \max(C(\alpha)), 0 \leq \alpha \leq G - 1 \quad (\text{E.3})$$

$$\underline{\text{Valor médio das transições:}} \frac{\sum_{\alpha=0}^{G-1} C(\alpha)}{G - 1} \quad (\text{E.4})$$

$$\underline{\text{Média dos níveis de cinza:}} \frac{1}{\sum_{\alpha=0}^{G-1} C(\alpha)} \cdot \sum_{\alpha=0}^{G-1} \alpha \cdot C(\alpha) \quad (\text{E.5})$$

$$\underline{\text{Desvio-padrão:}} \sqrt{\frac{1}{\sum_{\alpha=0}^{G-1} C(\alpha)} \cdot \sum_{\alpha=0}^{G-1} (\alpha - sm)^2 \cdot C(\alpha)} \quad (\text{E.6})$$

$$\underline{\textit{Entropia}}: \sum_{\alpha=0}^{G-1} C(\alpha) \cdot \log_2[C(\alpha)], C(\alpha) > 0 \quad (\text{E.7})$$

F Tabelas de Resultados

KURTOSIS

Tabela F.1

20 componentes independentes (85.66% de energia)					
#	% de acertos	VP	VN	FP	FN
5	72.32	38.49	33.83	16.17	11.51
15	75.46	39.76	35.71	14.29	10.24
25	77.02	40.02	37.00	13.00	9.98
35	77.27	40.20	37.07	12.93	9.80
45	77.93	38.93	39.00	11.00	11.07
50	78.34	40.46	37.88	12.12	9.54
55	78.80	41.85	36.95	13.05	8.15
60	77.85	41.10	36.76	13.24	8.90
70	79.15	40.41	38.73	11.27	9.59
80	78.88	40.88	38.00	12.00	9.12
90	79.12	40.88	38.24	11.76	9.12
100	79.73	40.22	39.51	10.49	9.78
110	78.00	39.24	38.76	11.24	10.76
120	80.00	42.32	37.68	12.32	7.68
140	80.44	41.17	39.27	10.73	8.83
160	80.46	40.05	40.41	9.59	9.95

Tabela F.3

40 componentes independentes (91.96% de energia)					
#	% de acertos	VP	VN	FP	FN
5	76.32	40.80	35.51	14.49	9.20
15	80.15	41.51	38.63	11.37	8.49
25	80.34	41.46	38.88	11.12	8.54
35	80.29	41.12	39.17	10.83	8.88
45	81.02	41.07	39.95	10.05	8.93
50	80.34	40.71	39.63	10.37	9.29
55	81.07	41.20	39.88	10.12	8.80
60	80.85	41.59	39.27	10.73	8.41
70	80.49	41.24	39.24	10.76	8.76
80	81.17	42.20	38.98	11.02	7.80
90	80.80	41.76	39.05	10.95	8.24
100	79.93	40.76	39.17	10.83	9.24
110	81.51	42.17	39.34	10.66	7.83
120	80.41	41.61	38.80	11.20	8.39
140	80.63	41.80	38.83	11.17	8.20
160	80.22	41.05	39.17	10.83	8.95

Tabela F.2

50 componentes independentes (93.62% de energia)					
#	% de acertos	VP	VN	FP	FN
5	77.83	39.39	38.44	11.56	10.61
15	80.10	41.15	38.95	11.05	8.85
25	80.93	41.68	39.24	10.76	8.32
35	81.39	40.73	40.66	9.34	9.27
45	81.44	41.88	39.56	10.44	8.12
50	82.00	42.54	39.46	10.54	7.46
55	81.02	41.66	39.37	10.63	8.34
60	81.54	42.73	38.80	11.20	7.27
70	80.83	40.17	40.66	9.34	9.83
80	81.17	41.46	39.71	10.29	8.54
90	81.59	41.83	39.76	10.24	8.17
100	80.76	41.20	39.56	10.44	8.80
110	81.44	40.27	41.17	8.83	9.73
120	80.41	40.27	40.15	9.85	9.73
140	81.93	40.54	41.39	8.61	9.46
160	81.15	41.49	39.66	10.34	8.51

Tabela F.4

60 componentes independentes (94.79% de energia)					
#	% de acertos	VP	VN	FP	FN
5	79.41	41.39	38.02	11.98	8.61
15	81.66	42.12	39.54	10.46	7.88
25	80.83	41.37	39.46	10.54	8.63
35	81.29	41.29	40.00	10.00	8.71
45	81.68	42.12	39.56	10.44	7.88
50	81.49	42.10	39.39	10.61	7.90
55	80.95	41.07	39.88	10.12	8.93
60	81.56	40.66	40.90	9.10	9.34
70	81.39	40.85	40.54	9.46	9.15
80	81.56	42.24	39.32	10.68	7.76
90	80.61	40.93	39.68	10.32	9.07
100	81.02	41.46	39.56	10.44	8.54
110	80.34	40.80	39.54	10.46	9.20
120	80.78	40.27	40.51	9.49	9.73
140	79.85	40.63	39.22	10.78	9.37
160	79.81	40.55	39.26	10.74	9.45

Tabela F.5

65 componentes independentes (95.25% de energia)					
#	% de acertos	VP	VN	FP	FN
5	79.24	41.98	37.27	12.73	8.02
15	81.27	41.34	39.93	10.07	8.66
25	81.24	40.88	40.37	9.63	9.12
35	81.56	41.78	39.78	10.22	8.22
45	80.54	40.73	39.80	10.20	9.27
50	81.49	41.73	39.76	10.24	8.27
55	80.95	40.63	40.32	9.68	9.37
60	81.66	40.95	40.71	9.29	9.05
70	81.37	40.68	40.68	9.32	9.32
80	80.93	41.63	39.29	10.71	8.37
90	80.73	41.61	39.12	10.88	8.39
100	80.93	41.22	39.71	10.29	8.78
110	80.98	41.34	39.63	10.37	8.66
120	80.68	41.76	38.93	11.07	8.24
140	80.95	41.88	39.07	10.93	8.12
160	80.90	41.93	38.97	11.03	8.07

Tabela F.7

70 componentes independentes (95.69% de energia)					
#	% de acertos	VP	VN	FP	FN
5	80.73	41.71	39.02	10.98	8.29
15	80.76	41.88	38.88	11.12	8.12
25	80.88	41.63	39.24	10.76	8.37
35	81.24	42.22	39.02	10.98	7.78
45	81.29	42.02	39.27	10.73	7.98
50	81.41	41.46	39.95	10.05	8.54
55	81.85	42.07	39.78	10.22	7.93
60	81.63	40.85	40.78	9.22	9.15
70	81.41	41.46	39.95	10.05	8.54
80	80.46	40.10	40.37	9.63	9.90
90	80.32	39.83	40.49	9.51	10.17
100	81.00	41.17	39.83	10.17	8.83
110	81.93	42.63	39.29	10.71	7.37
120	80.37	41.27	39.10	10.90	8.73
140	81.32	41.59	39.73	10.27	8.41
160	80.78	41.02	39.76	10.24	8.98

Tabela F.6

75 componentes independentes (96.10% de energia)					
#	% de acertos	VP	VN	FP	FN
5	80.63	41.63	39.00	11.00	8.37
15	82.05	42.37	39.68	10.32	7.63
25	81.12	41.22	39.90	10.10	8.78
35	82.07	41.22	40.85	9.15	8.78
45	81.80	41.20	40.61	9.39	8.80
50	81.56	41.63	39.93	10.07	8.37
55	80.98	41.76	39.22	10.78	8.24
60	80.66	41.68	38.98	11.02	8.32
70	81.44	40.71	40.73	9.27	9.29
80	80.63	39.98	40.66	9.34	10.02
90	80.17	42.10	38.07	11.93	7.90
100	81.46	41.85	39.61	10.39	8.15
110	81.22	40.68	40.54	9.46	9.32
120	80.20	40.95	39.24	10.76	9.05
140	80.24	42.34	37.90	12.10	7.66
160	80.18	41.90	38.28	11.72	8.10

Tabela F.8

80 componentes independentes (96.45% de energia)					
#	% de acertos	VP	VN	FP	FN
5	80.56	41.05	39.51	10.49	8.95
15	81.54	41.83	39.71	10.29	8.17
25	81.98	42.10	39.88	10.12	7.90
35	80.88	41.24	39.63	10.37	8.76
45	81.93	40.85	41.07	8.93	9.15
50	80.95	40.83	40.12	9.88	9.17
55	81.68	40.83	40.85	9.15	9.17
60	80.98	41.27	39.71	10.29	8.73
70	80.80	41.32	39.49	10.51	8.68
80	81.29	40.68	40.61	9.39	9.32
90	81.46	42.05	39.41	10.59	7.95
100	80.68	40.80	39.88	10.12	9.20
110	81.02	41.29	39.73	10.27	8.71
120	80.32	40.80	39.51	10.49	9.20
140	80.27	39.80	40.46	9.54	10.20
160	79.21	38.88	40.33	9.67	11.12

Tabela F.9

85 componentes independentes (96.78% de energia)					
#	% de acertos	VP	VN	FP	FN
5	80.37	40.71	39.66	10.34	9.29
15	81.73	42.02	39.71	10.29	7.98
25	81.41	43.07	38.34	11.66	6.93
35	81.66	42.66	39.00	11.00	7.34
45	81.24	41.71	39.54	10.46	8.29
50	81.46	41.02	40.44	9.56	8.98
55	80.93	41.37	39.56	10.44	8.63
60	81.10	41.51	39.59	10.41	8.49
70	81.98	43.17	38.80	11.20	6.83
80	81.37	41.02	40.34	9.66	8.98
90	80.68	41.37	39.32	10.68	8.63
100	80.78	41.15	39.63	10.37	8.85
110	80.90	41.71	39.20	10.80	8.29
120	79.63	40.49	39.15	10.85	9.51
140	79.44	40.49	38.95	11.05	9.51
160	78.97	40.78	38.19	11.81	9.22

Tabela F.11

100 componentes independentes (97.59% de energia)					
#	% de acertos	VP	VN	FP	FN
5	81.00	41.32	39.68	10.32	8.68
15	81.41	41.76	39.66	10.34	8.24
25	81.46	42.22	39.24	10.76	7.78
35	80.85	41.32	39.54	10.46	8.68
45	81.80	42.68	39.12	10.88	7.32
50	80.88	41.46	39.41	10.59	8.54
55	81.51	42.61	38.90	11.10	7.39
60	80.98	41.34	39.63	10.37	8.66
70	81.22	42.78	38.44	11.56	7.22
80	80.88	41.27	39.61	10.39	8.73
90	80.78	42.49	38.29	11.71	7.51
100	81.20	41.20	40.00	10.00	8.80
110	80.22	40.24	39.98	10.02	9.76
120	80.63	41.15	39.49	10.51	8.85
140	80.44	41.07	39.37	10.63	8.93
160	80.49	41.09	39.40	10.60	8.91

Tabela F.10

150 componentes independentes (99.20% de energia)					
#	% de acertos	VP	VN	FP	FN
5	80.56	41.56	39.00	11.00	8.44
15	81.17	41.61	39.56	10.44	8.39
25	81.37	40.54	40.83	9.17	9.46
35	81.37	41.17	40.20	9.80	8.83
45	80.39	40.66	39.73	10.27	9.34
50	81.05	41.85	39.20	10.80	8.15
55	80.49	41.10	39.39	10.61	8.90
60	80.81	41.71	39.10	10.90	8.29
70	80.29	40.83	39.46	10.54	9.17
80	81.17	42.37	38.80	11.20	7.63
90	80.02	41.71	38.32	11.68	8.29
100	78.91	39.98	38.93	11.07	10.02
110	79.53	40.63	38.90	11.10	9.37
120	80.62	40.29	40.34	9.66	9.71
140	80.10	41.20	38.90	11.10	8.80
160	79.58	41.02	38.56	11.44	8.98

Tabela F.12

200 componentes independentes (99.97% de energia)					
#	% de acertos	VP	VN	FP	FN
5	81.49	41.85	39.63	10.37	8.15
15	81.91	41.98	39.93	10.07	8.02
25	81.24	42.27	38.98	11.02	7.73
35	81.46	42.12	39.34	10.66	7.88
45	81.37	41.07	40.29	9.71	8.93
50	80.83	40.66	40.17	9.83	9.34
55	81.15	41.46	39.68	10.32	8.54
60	80.52	41.37	39.15	10.85	8.63
70	81.66	42.59	39.07	10.93	7.41
80	80.64	41.32	39.32	10.68	8.68
90	79.58	41.12	38.46	11.54	8.88
100	79.68	41.07	38.61	11.39	8.93
110	79.71	40.44	39.27	10.73	9.56
120	79.39	41.73	37.66	12.34	8.27
140	79.51	41.93	37.59	12.41	8.07
160	79.22	41.95	37.27	12.73	8.05

NEG-ENTROPIA

Tabela F.13

20 componentes independentes (85.67% de energia)					
#	% de acertos	VP	VN	FP	FN
5	72.12	38.32	33.80	16.20	11.68
15	75.44	40.54	34.90	15.10	9.46
25	77.68	41.12	36.56	13.44	8.88
35	78.07	40.90	37.17	12.83	9.10
45	78.07	39.20	38.88	11.12	10.80
50	78.10	40.34	37.76	12.24	9.66
55	78.27	39.73	38.54	11.46	10.27
60	78.46	40.80	37.66	12.34	9.20
70	79.27	40.73	38.54	11.46	9.27
80	79.07	41.66	37.41	12.59	8.34
90	79.07	41.10	37.98	12.02	8.90
100	79.90	39.71	40.20	9.80	10.29
110	80.49	42.49	38.00	12.00	7.51
120	79.41	41.17	38.24	11.76	8.83
140	79.51	41.90	37.61	12.39	8.10
160	79.41	40.54	38.88	11.12	9.46

Tabela F.15

40 componentes independentes (91.98% de energia)					
#	% de acertos	VP	VN	FP	FN
5	75.59	40.56	35.02	14.98	9.44
15	80.15	40.71	39.44	10.56	9.29
25	80.61	41.32	39.29	10.71	8.68
35	80.24	41.27	38.98	11.02	8.73
45	81.24	41.51	39.73	10.27	8.49
50	81.15	41.24	39.90	10.10	8.76
55	81.95	42.34	39.61	10.39	7.66
60	80.49	41.73	38.76	11.24	8.27
70	81.37	40.41	40.95	9.05	9.59
80	80.61	41.24	39.37	10.63	8.76
90	81.10	42.27	38.83	11.17	7.73
100	78.63	38.49	40.15	9.85	11.51
110	80.44	41.76	38.68	11.32	8.24
120	80.73	42.63	38.10	11.90	7.37
140	81.32	42.05	39.27	10.73	7.95
160	81.37	41.85	39.51	10.49	8.15

Tabela F.14

50 componentes independentes (93.62% de energia)					
#	% de acertos	VP	VN	FP	FN
5	77.56	39.29	38.27	11.73	10.71
15	80.76	41.49	39.27	10.73	8.51
25	81.46	42.29	39.17	10.83	7.71
35	81.76	41.61	40.15	9.85	8.39
45	81.63	41.80	39.83	10.17	8.20
50	81.71	42.12	39.59	10.41	7.88
55	81.12	42.24	38.88	11.12	7.76
60	81.39	41.71	39.68	10.32	8.29
70	82.00	42.10	39.90	10.10	7.90
80	81.63	42.22	39.41	10.59	7.78
90	82.20	42.29	39.90	10.10	7.71
100	81.22	41.90	39.32	10.68	8.10
110	82.10	41.61	40.49	9.51	8.39
120	80.63	41.66	38.98	11.02	8.34
140	80.68	42.05	38.63	11.37	7.95
160	82.00	42.63	39.37	10.63	7.37

Tabela F.16

60 componentes independentes (94.78% de energia)					
#	% de acertos	VP	VN	FP	FN
5	79.05	41.66	37.39	12.61	8.34
15	81.76	41.98	39.78	10.22	8.02
25	80.44	40.73	39.71	10.29	9.27
35	81.59	41.37	40.22	9.78	8.63
45	81.37	41.98	39.39	10.61	8.02
50	81.39	41.49	39.90	10.10	8.51
55	81.44	41.73	39.71	10.29	8.27
60	81.59	41.17	40.41	9.59	8.83
70	81.29	41.66	39.63	10.37	8.34
80	81.85	41.17	40.68	9.32	8.83
90	80.00	40.37	39.63	10.37	9.63
100	83.22	43.76	39.46	10.54	6.24
110	81.46	43.07	38.39	11.61	6.93
120	80.83	40.24	40.59	9.41	9.76
140	80.29	41.17	39.12	10.88	8.83
160	80.88	43.51	37.37	12.63	6.49

Tabela F.17

65 componentes independentes (95.26% de energia)					
#	% de acertos	VP	VN	FP	FN
5	78.93	41.34	37.59	12.41	8.66
15	81.49	41.76	39.73	10.27	8.24
25	81.61	41.02	40.59	9.41	8.98
35	81.83	42.39	39.44	10.56	7.61
45	81.32	41.29	40.02	9.98	8.71
50	81.37	41.71	39.66	10.34	8.29
55	81.76	42.24	39.51	10.49	7.76
60	81.02	40.05	40.98	9.02	9.95
70	81.49	40.63	40.85	9.15	9.37
80	81.02	41.56	39.46	10.54	8.44
90	80.88	42.37	38.51	11.49	7.63
100	80.24	40.93	39.32	10.68	9.07
110	79.90	41.27	38.63	11.37	8.73
120	79.95	41.51	38.44	11.56	8.49
140	80.10	41.22	38.88	11.12	8.78
160	80.00	40.39	39.61	10.39	9.61

Tabela F.19

70 componentes independentes (95.72% de energia)					
#	% de acertos	VP	VN	FP	FN
5	79.88	41.56	38.32	11.68	8.44
15	81.24	41.98	39.27	10.73	8.02
25	81.44	42.59	38.85	11.15	7.41
35	81.39	42.27	39.12	10.88	7.73
45	81.49	42.00	39.49	10.51	8.00
50	81.71	42.44	39.27	10.73	7.56
55	81.39	41.88	39.51	10.49	8.12
60	81.49	40.59	40.90	9.10	9.41
70	80.17	41.73	38.44	11.56	8.27
80	82.00	42.34	39.66	10.34	7.66
90	80.07	40.27	39.80	10.20	9.73
100	80.83	41.95	38.88	11.12	8.05
110	82.34	43.56	38.78	11.22	6.44
120	80.78	39.37	41.41	8.59	10.63
140	80.29	40.05	40.24	9.76	9.95
160	81.85	41.71	40.15	9.85	8.29

Tabela F.18

75 componentes independentes (96.09% de energia)					
#	% de acertos	VP	VN	FP	FN
5	80.20	42.24	37.95	12.05	7.76
15	81.49	41.78	39.71	10.29	8.22
25	81.24	41.93	39.32	10.68	8.07
35	82.29	42.27	40.02	9.98	7.73
45	81.24	40.63	40.61	9.39	9.37
50	81.22	40.95	40.27	9.73	9.05
55	80.61	40.34	40.27	9.73	9.66
60	81.71	42.07	39.63	10.37	7.93
70	81.00	41.51	39.49	10.51	8.49
80	81.73	41.83	39.90	10.10	8.17
90	81.37	42.22	39.15	10.85	7.78
100	81.46	41.66	39.80	10.20	8.34
110	81.51	40.15	41.37	8.63	9.85
120	81.07	40.39	40.68	9.32	9.61
140	80.34	41.02	39.32	10.68	8.98
160	81.56	41.66	39.90	10.10	8.34

Tabela F.20

80 componentes independentes (96.45% de energia)					
#	% de acertos	VP	VN	FP	FN
5	80.34	40.85	39.49	10.51	9.15
15	80.88	41.68	39.20	10.80	8.32
25	81.88	42.20	39.68	10.32	7.80
35	81.10	41.71	39.39	10.61	8.29
45	81.34	41.05	40.29	9.71	8.95
50	81.00	40.76	40.24	9.76	9.24
55	80.56	42.46	38.10	11.90	7.54
60	81.37	41.90	39.46	10.54	8.10
70	80.78	41.10	39.68	10.32	8.90
80	81.46	41.90	39.56	10.44	8.10
90	81.93	41.20	40.73	9.27	8.80
100	80.63	41.37	39.27	10.73	8.63
110	80.10	40.34	39.76	10.24	9.66
120	80.63	42.15	38.49	11.51	7.85
140	80.98	41.12	39.85	10.15	8.88
160	80.29	40.05	40.24	9.76	9.95

Tabela F.21

85 componentes independentes (96.78% de energia)					
#	% de acertos	VP	VN	FP	FN
5	80.46	40.85	39.61	10.39	9.15
15	82.00	42.24	39.76	10.24	7.76
25	80.90	42.00	38.90	11.10	8.00
35	81.71	42.56	39.15	10.85	7.44
45	81.12	42.05	39.07	10.93	7.95
50	81.90	41.10	40.80	9.20	8.90
55	81.34	43.05	38.29	11.71	6.95
60	80.37	40.78	39.59	10.41	9.22
70	80.66	41.63	39.02	10.98	8.37
80	81.98	41.27	40.71	9.29	8.73
90	81.15	41.34	39.80	10.20	8.66
100	80.78	41.37	39.41	10.59	8.63
110	81.71	42.93	38.78	11.22	7.07
120	81.17	41.85	39.32	10.68	8.15
140	80.20	40.78	39.41	10.59	9.22
160	80.20	40.98	39.22	10.78	9.02

Tabela F.23

100 componentes independentes (97.59% de energia)					
#	% de acertos	VP	VN	FP	FN
5	80.76	41.56	39.20	10.80	8.44
15	81.83	41.95	39.88	10.12	8.05
25	81.37	41.71	39.66	10.34	8.29
35	81.20	41.41	39.78	10.22	8.59
45	82.37	42.80	39.56	10.44	7.20
50	81.39	42.27	39.12	10.88	7.73
55	81.12	41.15	39.98	10.02	8.85
60	81.56	41.66	39.90	10.10	8.34
70	81.29	40.83	40.46	9.54	9.17
80	81.83	42.12	39.71	10.29	7.88
90	80.41	41.34	39.07	10.93	8.66
100	80.88	41.95	38.93	11.07	8.05
110	80.49	41.02	39.46	10.54	8.98
120	82.24	40.49	41.76	8.24	9.51
140	80.05	40.49	39.56	10.44	9.51
160	80.68	41.76	38.93	11.07	8.24

Tabela F.22

150 componentes independentes (99.20% de energia)					
#	% de acertos	VP	VN	FP	FN
5	80.07	41.05	39.02	10.98	8.95
15	81.37	41.93	39.44	10.56	8.07
25	80.98	40.54	40.44	9.56	9.46
35	81.07	41.90	39.17	10.83	8.10
45	79.90	41.17	38.73	11.27	8.83
50	81.29	42.73	38.56	11.44	7.27
55	80.83	43.15	37.68	12.32	6.85
60	81.63	42.88	38.76	11.24	7.12
70	81.07	42.05	39.02	10.98	7.95
80	81.59	41.49	40.10	9.90	8.51
90	80.27	43.15	37.12	12.88	6.85
100	79.27	41.66	37.61	12.39	8.34
110	78.78	39.71	39.07	10.93	10.29
120	80.20	39.90	40.29	9.71	10.10
140	80.34	41.12	39.22	10.78	8.88
160	79.27	41.12	38.15	11.85	8.88

Tabela F.24

200 componentes independentes (99.97% de energia)					
#	% de acertos	VP	VN	FP	FN
5	81.27	41.41	39.85	10.15	8.59
15	82.56	42.59	39.98	10.02	7.41
25	81.41	41.98	39.44	10.56	8.02
35	81.46	42.10	39.37	10.63	7.90
45	81.10	41.88	39.22	10.78	8.12
50	80.68	40.80	39.88	10.12	9.20
55	81.02	42.88	38.15	11.85	7.12
60	79.95	42.24	37.71	12.29	7.76
70	79.95	41.49	38.46	11.54	8.51
80	80.90	42.00	38.90	11.10	8.00
90	80.41	42.27	38.15	11.85	7.73
100	80.68	42.39	38.29	11.71	7.61
110	80.68	40.54	40.15	9.85	9.46
120	80.00	40.83	39.17	10.83	9.17
140	79.12	41.66	37.46	12.54	8.34
160	79.02	41.22	37.80	12.20	8.78

MÁXIMA VEROSSIMILHANÇA (BELL-SEJNOWSKI)

Tabela F.25

20 componentes independentes (85.47% de energia)					
#	% de acertos	VP	VN	FP	FN
5	73.66	37.66	36.00	14.00	12.34
15	77.85	40.49	37.37	12.63	9.51
25	78.78	39.41	39.37	10.63	10.59
35	78.29	39.41	38.88	11.12	10.59
45	79.51	41.66	37.85	12.15	8.34
50	79.32	40.78	38.54	11.46	9.22
55	79.46	40.78	38.68	11.32	9.22
60	80.20	41.56	38.63	11.37	8.44
70	79.02	41.12	37.90	12.10	8.88
80	79.17	39.76	39.41	10.59	10.24
90	79.61	41.85	37.76	12.24	8.15
100	80.29	41.41	38.88	11.12	8.59
110	80.24	40.49	39.76	10.24	9.51
120	80.39	42.20	38.20	11.80	7.80
140	80.54	40.29	40.24	9.76	9.71
160	81.90	41.80	40.10	9.90	8.20

Tabela F.27

40 componentes independentes (92.02% de energia)					
#	% de acertos	VP	VN	FP	FN
5	77.27	39.61	37.66	12.34	10.39
15	80.00	41.22	38.78	11.22	8.78
25	80.93	40.63	40.29	9.71	9.37
35	82.24	42.44	39.80	10.20	7.56
45	81.02	41.02	40.00	10.00	8.98
50	80.15	41.17	38.98	11.02	8.83
55	80.98	40.73	40.24	9.76	9.27
60	80.44	40.98	39.46	10.54	9.02
70	79.80	40.05	39.76	10.24	9.95
80	81.66	41.80	39.85	10.15	8.20
90	78.98	39.85	39.12	10.88	10.15
100	81.22	40.54	40.68	9.32	9.46
110	79.56	40.63	38.93	11.07	9.37
120	79.22	41.37	37.85	12.15	8.63
140	80.59	39.85	40.73	9.27	10.15
160	80.05	40.20	39.85	10.15	9.80

Tabela F.26

50 componentes independentes (93.78% de energia)					
#	% de acertos	VP	VN	FP	FN
5	78.00	38.63	39.37	10.63	11.37
15	81.76	42.39	39.37	10.63	7.61
25	81.37	41.61	39.76	10.24	8.39
35	81.85	42.10	39.76	10.24	7.90
45	81.27	41.66	39.61	10.39	8.34
50	80.68	41.17	39.51	10.49	8.83
55	80.05	39.76	40.29	9.71	10.24
60	80.39	40.05	40.34	9.66	9.95
70	80.68	41.32	39.37	10.63	8.68
80	80.39	39.85	40.54	9.46	10.15
90	81.22	41.95	39.27	10.73	8.05
100	80.10	41.22	38.88	11.12	8.78
110	79.95	41.02	38.93	11.07	8.98
120	81.41	41.71	39.71	10.29	8.29
140	78.88	41.27	37.61	12.39	8.73
160	78.98	40.88	38.10	11.90	9.12

Tabela F.28

60 componentes independentes (94.84% de energia)					
#	% de acertos	VP	VN	FP	FN
5	80.93	40.73	40.20	9.80	9.27
15	81.37	40.83	40.54	9.46	9.17
25	80.83	41.41	39.41	10.59	8.59
35	79.80	38.98	40.83	9.17	11.02
45	80.10	40.59	39.51	10.49	9.41
50	79.12	40.20	38.93	11.07	9.80
55	79.22	39.85	39.37	10.63	10.15
60	79.41	40.59	38.83	11.17	9.41
70	80.54	40.44	40.10	9.90	9.56
80	80.39	41.61	38.78	11.22	8.39
90	80.24	41.41	38.83	11.17	8.59
100	80.05	41.37	38.68	11.32	8.63
110	78.39	39.07	39.32	10.68	10.93
120	79.71	41.37	38.34	11.66	8.63
140	78.73	40.24	38.49	11.51	9.76
160	79.51	42.24	37.27	12.73	7.76

Tabela F.29

65 componentes independentes (95.25% de energia)					
#	% de acertos	VP	VN	FP	FN
5	79.37	40.39	38.98	11.02	9.61
15	81.22	42.00	39.22	10.78	8.
25	81.32	41.80	39.51	10.49	8.20
35	79.46	39.76	39.71	10.29	10.24
45	81.07	40.93	40.15	9.85	9.07
50	80.78	40.98	39.80	10.20	9.02
55	80.24	40.59	39.66	10.34	9.41
60	81.07	42.00	39.07	10.93	8.
70	79.90	40.34	39.56	10.44	9.66
80	79.56	41.71	37.85	12.15	8.29
90	78.88	41.12	37.76	12.24	8.88
100	80.34	41.17	39.17	10.83	8.83
110	78.93	40.00	38.93	11.07	10.00
120	77.66	38.54	39.12	10.88	11.46
140	78.88	39.46	39.41	10.59	10.54
160	78.34	39.61	38.73	11.27	10.39

Tabela F.31

70 componentes independentes (95.69% de energia)					
#	% de acertos	VP	VN	FP	FN
5	80.88	41.07	39.80	10.20	8.93
15	79.22	39.85	39.37	10.63	10.15
25	81.02	40.88	40.15	9.85	9.12
35	79.41	40.68	38.73	11.27	9.32
45	79.07	41.17	37.90	12.10	8.83
50	81.17	40.88	40.29	9.71	9.12
55	78.98	41.66	37.32	12.68	8.34
60	81.07	43.17	37.90	12.10	6.83
70	80.54	42.10	38.44	11.56	7.90
80	79.90	40.93	38.98	11.02	9.07
90	78.10	40.93	37.17	12.83	9.07
100	79.07	40.73	38.34	11.66	9.27
110	77.27	38.29	38.98	11.02	11.71
120	77.17	39.46	37.71	12.29	10.54
140	78.00	41.12	36.88	13.12	8.88
160	78.49	40.24	38.24	11.76	9.76

Tabela F.30

75 componentes independentes (96.38% de energia)					
#	% de acertos	VP	VN	FP	FN
5	82.10	41.76	40.34	9.66	8.24
15	81.37	40.88	40.49	9.51	9.12
25	79.71	39.95	39.76	10.24	10.05
35	79.71	40.59	39.12	10.88	9.41
45	80.39	41.12	39.27	10.73	8.88
50	80.10	40.63	39.46	10.54	9.37
55	78.83	39.76	39.07	10.93	10.24
60	81.07	41.90	39.17	10.83	8.10
70	78.05	41.61	36.44	13.56	8.39
80	78.15	41.02	37.12	12.88	8.98
90	78.00	40.78	37.22	12.78	9.22
100	78.15	41.07	37.07	12.93	8.93
110	78.68	39.51	39.17	10.83	10.49
120	78.63	39.56	39.07	10.93	10.44
140	77.61	40.54	37.07	12.93	9.46
160	77.95	39.02	38.93	11.07	10.98

Tabela F.32

80 componentes independentes (96.70% de energia)					
#	% de acertos	VP	VN	FP	FN
5	80.49	42.15	38.34	11.66	7.85
15	80.78	39.71	41.07	8.93	10.29
25	80.05	40.54	39.51	10.49	9.46
35	80.24	40.73	39.51	10.49	9.27
45	79.71	39.95	39.76	10.24	10.05
50	79.27	40.24	39.02	10.98	9.76
55	81.95	40.98	40.98	9.02	9.02
60	81.22	41.56	39.66	10.34	8.44
70	78.78	40.59	38.20	11.80	9.41
80	79.37	39.51	39.85	10.15	10.49
90	78.39	39.71	38.68	11.32	10.29
100	79.41	42.00	37.41	12.59	8.
110	77.27	40.34	36.93	13.07	9.66
120	78.24	40.24	38.00	12.00	9.76
140	78.73	40.98	37.76	12.24	9.02
160	78.39	40.05	38.34	11.66	9.95

Tabela F.33

85 componentes independentes (96.80% de energia)					
#	% de acertos	VP	VN	FP	FN
5	79.51	40.15	39.37	10.63	9.85
15	80.15	39.07	41.07	8.93	10.93
25	81.02	40.98	40.05	9.95	9.02
35	81.12	41.56	39.56	10.44	8.44
45	80.98	41.61	39.37	10.63	8.39
50	78.63	39.07	39.56	10.44	10.93
55	79.90	41.27	38.63	11.37	8.73
60	80.15	40.98	39.17	10.83	9.02
70	79.32	41.41	37.90	12.10	8.59
80	78.73	40.63	38.10	11.90	9.37
90	78.15	39.80	38.34	11.66	10.20
100	78.34	39.95	38.39	11.61	10.05
110	78.78	40.24	38.54	11.46	9.76
120	78.00	40.29	37.71	12.29	9.71
140	78.10	39.95	38.15	11.85	10.05
160	75.17	38.44	36.73	13.27	11.56

Tabela F.35

100 componentes independentes (97.61% de energia)					
#	% de acertos	VP	VN	FP	FN
5	79.61	40.98	38.63	11.37	9.02
15	81.41	40.54	40.88	9.12	9.46
25	78.98	40.54	38.44	11.56	9.46
35	79.51	40.73	38.78	11.22	9.27
45	80.05	39.71	40.34	9.66	10.29
50	77.61	39.07	38.54	11.46	10.93
55	78.78	39.51	39.27	10.73	10.49
60	78.68	40.78	37.90	12.10	9.22
70	78.44	39.66	38.78	11.22	10.34
80	78.34	39.02	39.32	10.68	10.98
90	76.24	38.49	37.76	12.24	11.51
100	76.10	40.63	35.46	14.54	9.37
110	75.90	39.76	36.15	13.85	10.24
120	78.10	41.56	36.54	13.46	8.44
140	75.80	38.78	37.02	12.98	11.22
160	75.76	38.24	37.51	12.49	11.76

Tabela F.34

150 componentes independentes (99.20% de energia)					
#	% de acertos	VP	VN	FP	FN
5	77.56	39.46	38.10	11.90	10.54
15	78.29	40.54	37.76	12.24	9.46
25	78.98	38.88	40.10	9.90	11.12
35	77.85	41.17	36.68	13.32	8.83
45	76.59	39.17	37.41	12.59	10.83
50	75.61	38.54	37.07	12.93	11.46
55	78.39	40.83	37.56	12.44	9.17
60	77.07	39.41	37.66	12.34	10.59
70	77.90	39.22	38.68	11.32	10.78
80	76.78	39.76	37.02	12.98	10.24
90	74.98	38.68	36.29	13.71	11.32
100	74.24	38.00	36.24	13.76	12.
110	74.98	39.66	35.32	14.68	10.34
120	71.66	36.68	34.98	15.02	13.32
140	73.41	39.22	34.20	15.80	10.78
160	72.78	37.85	34.93	15.07	12.15

Tabela F.36

200 componentes independentes (99.97% de energia)					
#	% de acertos	VP	VN	FP	FN
5	77.12	39.37	37.76	12.24	10.63
15	77.37	38.44	38.93	11.07	11.56
25	74.10	37.51	36.59	13.41	12.49
35	76.59	39.51	37.07	12.93	10.49
45	75.22	38.29	36.93	13.07	11.71
50	73.85	37.80	36.05	13.95	12.20
55	75.37	39.02	36.34	13.66	10.98
60	73.02	38.29	34.73	15.27	11.71
70	72.98	37.76	35.22	14.78	12.24
80	73.17	36.78	36.39	13.61	13.22
90	73.71	38.54	35.17	14.83	11.46
100	71.46	35.27	36.20	13.80	14.73
110	70.44	35.12	35.32	14.68	14.88
120	70.93	36.44	34.49	15.51	13.56
140	70.88	35.32	35.56	14.44	14.68
160	70.54	37.02	33.51	16.49	12.98

NEG-ENTROPIA + CAC

3 componentes de atuação conjunta

Tabela F.37

50 componentes principais (93.34% de energia)					
#	% de acertos	VP	VN	FP	FN
70	80.88	40.24	40.24	9.76	9.76
90	80.24	40.98	39.27	10.73	9.02
110	80.73	40.98	39.76	10.24	9.02

Tabela F.40

70 componentes principais (96.20% de energia)					
#	% de acertos	VP	VN	FP	FN
70	80.73	43.66	37.07	12.93	6.34
90	80.24	43.90	36.34	13.66	6.10
110	81.46	43.90	37.56	12.44	6.10

Tabela F.38

90 componentes principais (97.01% de energia)					
#	% de acertos	VP	VN	FP	FN
70	82.80	42.56	40.24	9.76	7.44
90	81.46	40.30	41.16	8.84	9.70
110	81.30	44.55	36.75	13.25	5.45

Tabela F.41

150 componentes principais (99.25% de energia)					
#	% de acertos	VP	VN	FP	FN
70	81.30	41.30	40.00	10.00	8.70
90	81.16	40.18	40.98	9.02	9.82
110	80.98	41.95	39.02	10.98	8.05

Tabela F.39

180 componentes principais (99.74% de energia)					
#	% de acertos	VP	VN	FP	FN
70	82.38	41.46	40.91	9.09	8.54
90	80.49	40.85	39.63	10.37	9.15
110	81.83	41.89	39.94	10.06	8.11

Tabela F.42

205 componentes principais (100% de energia)					
#	% de acertos	VP	VN	FP	FN
70	82.11	40.98	41.14	8.86	9.02
90	79.40	40.93	38.47	11.53	9.06
110	79.91	40.41	39.50	10.49	9.59

10 componentes de atuação conjunta

Tabela F.43

50 componentes principais (93.87% de energia)					
#	% de acertos	VP	VN	FP	FN
70	79.45	40.62	38.83	11.17	9.38
90	80.73	41.46	39.27	10.73	8.54
110	79.74	40.18	39.56	10.44	9.82

Tabela F.46

70 componentes principais (95.70% de energia)					
#	% de acertos	VP	VN	FP	FN
70	80.61	40.73	39.88	10.12	9.27
90	80.73	40.24	40.49	9.51	9.76
110	80.98	41.10	39.88	10.12	8.90

Tabela F.44

90 componentes principais (97.18% de energia)					
#	% de acertos	VP	VN	FP	FN
70	82.07	42.20	39.88	10.12	7.80
90	81.83	41.22	40.61	9.39	8.78
110	80.37	40.24	40.12	9.88	9.76

Tabela F.47

150 componentes principais (99.22% de energia)					
#	% de acertos	VP	VN	FP	FN
70	82.32	43.05	39.27	10.73	6.95
90	82.00	42.20	39.80	10.20	7.80
110	80.89	41.79	39.11	10.89	8.21

Tabela F.45

180 componentes principais (99.75% de energia)					
#	% de acertos	VP	VN	FP	FN
70	81.83	40.79	41.04	8.96	9.21
90	83.05	41.71	41.34	8.66	8.29
110	81.46	42.07	39.39	10.61	7.93

Tabela F.48

205 componentes principais (100% de energia)					
#	% de acertos	VP	VN	FP	FN
70	79.02	40.83	38.19	11.81	9.17
90	81.59	41.52	40.06	9.94	8.48
110	81.34	42.32	39.02	10.98	7.68

20 componentes de atuação conjunta

Tabela F.49

50 componentes principais (93.79% de energia)					
#	% de acertos	VP	VN	FP	FN
70	80.06	41.36	38.70	11.30	8.64
90	81.44	42.69	38.75	11.25	7.31
110	80.34	41.66	38.68	11.32	8.34

Tabela F.52

70 componentes principais (95.86% de energia)					
#	% de acertos	VP	VN	FP	FN
70	79.01	39.42	39.59	10.41	10.58
90	81.70	42.55	39.16	10.84	7.45
110	80.53	41.73	38.81	11.19	8.27

Tabela F.50

90 componentes principais (97.20% de energia)					
#	% de acertos	VP	VN	FP	FN
70	81.43	41.59	39.83	10.17	8.41
90	82.48	42.11	40.37	9.63	7.89
110	81.62	42.79	38.83	11.17	7.21

Tabela F.53

150 componentes principais (99.22% de energia)					
#	% de acertos	VP	VN	FP	FN
70	82.14	42.57	39.57	10.43	7.43
90	83.74	42.14	41.60	8.40	7.86
110	80.57	41.37	39.20	10.80	8.63

Tabela F.51

180 componentes principais (99.74% de energia)					
#	% de acertos	VP	VN	FP	FN
70	83.20	42.30	40.90	9.10	7.70
90	85.68	43.46	42.22	7.78	6.54
110	83.01	42.89	40.12	9.88	7.11

Tabela F.54

205 componentes principais (100% de energia)					
#	% de acertos	VP	VN	FP	FN
70	81.49	41.19	40.30	9.70	8.81
90	81.96	42.52	39.43	10.57	7.48
110	81.57	41.72	39.85	10.15	8.28

30 componentes de atuação conjunta

Tabela F.55

50 componentes principais (93.45% de energia)					
#	% de acertos	VP	VN	FP	FN
70	80.66	41.34	39.32	10.68	8.66
90	79.90	38.57	41.34	8.66	11.43
110	79.20	39.63	39.57	10.43	10.37

Tabela F.58

70 componentes principais (96.27% de energia)					
#	% de acertos	VP	VN	FP	FN
70	80.70	41.40	39.30	10.70	8.60
90	80.31	41.84	38.47	11.53	8.16
110	80.41	41.66	38.76	11.24	8.34

Tabela F.56

90 componentes principais (97.05% de energia)					
#	% de acertos	VP	VN	FP	FN
70	80.37	41.46	38.90	11.10	8.54
90	83.25	42.88	40.38	9.62	7.12
110	82.66	42.37	40.30	9.70	7.63

Tabela F.59

150 componentes principais (99.26% de energia)					
#	% de acertos	VP	VN	FP	FN
70	80.73	40.98	39.76	10.24	9.02
90	82.81	42.70	40.10	9.90	7.30
110	81.57	42.70	38.87	11.13	7.30

Tabela F.57

180 componentes principais (99.75% de energia)					
#	% de acertos	VP	VN	FP	FN
70	80.98	40.49	40.49	9.51	9.51
90	83.47	43.06	40.40	9.60	6.94
110	82.70	42.98	39.72	10.28	7.02

Tabela F.60

205 componentes principais (100% de energia)					
#	% de acertos	VP	VN	FP	FN
70	81.85	42.17	39.71	10.29	7.83
90	83.66	43.28	40.38	9.62	6.72
110	81.51	41.78	39.73	10.27	8.22

40 componentes de atuação conjunta

Tabela F.61

50 componentes principais (93.38% de energia)					
#	% de acertos	VP	VN	FP	FN
70	81.34	41.71	39.63	10.37	8.29
90	81.45	41.93	39.51	10.49	8.07
110	80.85	41.59	39.26	10.74	8.41

Tabela F.64

70 componentes principais (96.22% de energia)					
#	% de acertos	VP	VN	FP	FN
70	80.98	43.17	37.80	12.20	6.83
90	81.85	41.53	40.32	9.68	8.47
110	81.82	42.64	39.18	10.82	7.36

Tabela F.62

90 componentes principais (97.05% de energia)					
#	% de acertos	VP	VN	FP	FN
70	80.81	41.01	39.80	10.20	8.99
90	81.02	41.75	39.27	10.73	8.25
110	81.34	42.29	39.05	10.95	7.71

Tabela F.65

150 componentes principais (99.22% de energia)					
#	% de acertos	VP	VN	FP	FN
70	80.14	41.20	38.94	11.06	8.80
90	82.30	42.19	40.11	9.89	7.81
110	81.85	41.53	40.32	9.68	8.47

Tabela F.63

180 componentes principais (99.73% de energia)					
#	% de acertos	VP	VN	FP	FN
70	81.82	42.66	39.16	10.84	7.34
90	83.27	43.25	40.02	9.98	6.75
110	82.68	42.38	40.30	9.70	7.62

Tabela F.66

205 componentes principais (100% de energia)					
#	% de acertos	VP	VN	FP	FN
70	79.95	40.72	39.23	10.77	9.28
90	80.85	41.59	39.26	10.74	8.41
110	80.38	41.86	38.52	11.48	8.14

50 componentes de atuação conjunta

Tabela F.67

50 componentes principais (93.32% de energia)					
#	% de acertos	VP	VN	FP	FN
70	81.22	41.83	39.39	10.61	8.17
90	80.81	41.01	39.80	10.20	8.99
110	80.06	41.35	38.71	11.29	8.65

Tabela F.70

70 componentes principais (96.20% de energia)					
#	% de acertos	VP	VN	FP	FN
70	79.82	40.38	39.44	10.56	9.62
90	80.20	41.60	38.59	11.41	8.40
110	79.94	40.92	39.02	10.98	9.08

Tabela F.68

90 componentes principais (96.99% de energia)					
#	% de acertos	VP	VN	FP	FN
70	80.27	41.20	39.07	10.93	8.80
90	80.84	41.63	39.21	10.74	8.37
110	79.87	41.86	38.02	11.98	8.14

Tabela F.71

150 componentes principais (99.24% de energia)					
#	% de acertos	VP	VN	FP	FN
70	80.15	41.70	38.45	11.55	8.30
90	80.73	41.41	39.31	10.69	8.59
110	80.19	41.68	38.51	11.49	8.32

Tabela F.69

180 componentes principais (99.74% de energia)					
#	% de acertos	VP	VN	FP	FN
70	81.50	42.90	38.60	11.40	7.10
90	82.68	42.38	40.30	9.70	7.62
110	81.04	41.61	39.42	10.58	8.39

Tabela F.72

205 componentes principais (100% de energia)					
#	% de acertos	VP	VN	FP	FN
70	80.38	41.86	38.52	11.48	8.14
90	81.61	41.02	40.59	9.41	8.98
110	79.92	41.18	38.74	11.26	8.82

NEG-ENTROPIA + CAC + TOP-HAT

20 componentes de atuação conjunta

Tabela F.73

50 componentes principais (93.44% de energia)					
#	% de acertos	VP	VN	FP	FN
70	78.59	39.90	38.68	11.32	10.10
90	77.37	39.27	38.10	11.90	10.73
110	78.05	39.85	38.20	11.80	10.15

Tabela F.75

70 componentes principais (96.01% de energia)					
#	% de acertos	VP	VN	FP	FN
70	79.07	41.56	37.51	12.49	8.44
90	78.05	38.73	39.32	10.68	11.27
110	78.98	41.17	37.80	12.20	8.83

Tabela F.74

90 componentes principais (97.00% de energia)					
#	% de acertos	VP	VN	FP	FN
70	80.73	40.88	39.85	10.15	9.12
90	81.66	42.93	38.73	11.27	7.07
110	79.95	40.88	39.07	10.93	9.12

Tabela F.76

150 componentes principais (99.25% de energia)					
#	% de acertos	VP	VN	FP	FN
70	81.07	41.90	39.17	10.83	8.10
90	81.52	42.66	38.86	11.14	7.34
110	79.66	40.88	38.78	11.22	9.12

30 componentes de atuação conjunta

Tabela F.77

50 componentes principais (93.87% de energia)					
#	% de acertos	VP	VN	FP	FN
70	78.05	39.85	38.20	11.80	10.15
90	76.98	40.73	36.24	13.76	9.27
110	76.05	39.34	36.71	13.29	10.66

Tabela F.79

70 componentes principais (95.80% de energia)					
#	% de acertos	VP	VN	FP	FN
70	78.98	41.17	37.80	12.20	8.83
90	79.22	40.29	38.93	11.0	9.71
110	79.18	39.41	39.77	10.23	10.59

Tabela F.78

90 componentes principais (97.06% de energia)					
#	% de acertos	VP	VN	FP	FN
70	78.63	40.00	38.63	11.37	10.00
90	79.32	40.39	39.41	10.59	9.61
110	79.41	39.89	39.52	10.48	10.11

Tabela F.80

150 componentes principais (99.20% de energia)					
#	% de acertos	VP	VN	FP	FN
70	79.95	40.88	39.07	10.93	9.12
90	79.80	40.39	39.41	10.49	9.61
110	79.62	39.79	39.82	10.18	10.21

MATRIZ DE CO-OCORRÊNCIA DESNORMALIZADA

Tabela F.81

SGLD Desnormalizado 01					
#	% de acertos	VP	VN	FP	FN
5	56.95	22.49	34.46	15.54	27.51
15	57.20	24.85	32.34	17.66	25.15
25	57.93	24.83	33.10	16.90	25.17
35	59.12	23.02	36.10	13.90	26.98
45	58.32	24.51	33.80	16.20	25.49
50	58.34	23.63	34.71	15.29	26.37
55	58.00	24.90	33.10	16.90	25.10
60	57.54	23.95	33.59	16.41	26.05
70	58.20	23.41	34.78	15.22	26.59
80	58.44	25.17	33.27	16.73	24.83
90	59.15	25.44	33.71	16.29	24.56
100	58.39	24.32	34.07	15.93	25.68
110	58.24	23.83	34.41	15.59	26.17
120	58.78	25.98	32.80	17.20	24.02
140	59.22	24.61	34.61	15.39	25.39
160	57.76	25.17	32.59	17.41	24.83

Tabela F.83

SGLD Desnormalizado 10					
#	% de acertos	VP	VN	FP	FN
5	56.71	24.07	32.63	17.37	25.93
15	56.95	25.49	31.46	18.54	24.51
25	57.73	26.66	31.07	18.93	23.34
35	58.61	23.29	35.32	14.68	26.71
45	57.73	25.41	32.32	17.68	24.59
50	58.22	27.59	30.63	19.37	22.41
55	57.61	23.98	33.63	16.37	26.02
60	57.07	25.17	31.90	18.10	24.83
70	57.61	24.49	33.12	16.88	25.51
80	57.76	26.78	30.98	19.02	23.22
90	58.32	25.00	33.32	16.68	25.00
100	58.37	25.73	32.63	17.37	24.27
110	58.44	24.83	33.61	16.39	25.17
120	57.78	26.51	31.27	18.73	23.49
140	58.41	22.54	35.88	14.12	27.46
160	57.80	23.20	34.61	15.39	26.80

Tabela F.82

SGLD Desnormalizado 11					
#	% de acertos	VP	VN	FP	FN
5	56.98	24.51	32.46	17.54	25.49
15	57.44	23.76	33.68	16.32	26.24
25	57.88	25.51	32.37	17.63	24.49
35	58.85	21.41	37.44	12.56	28.59
45	58.39	25.59	32.80	17.20	24.41
50	58.27	24.90	33.37	16.63	25.10
55	58.15	25.49	32.66	17.34	24.51
60	57.39	23.32	34.07	15.93	26.68
70	58.12	23.22	34.90	15.10	26.78
80	58.05	23.02	35.02	14.98	26.98
90	59.05	24.54	34.51	15.49	25.46
100	58.41	24.41	34.00	16.00	25.59
110	59.29	25.41	33.88	16.12	24.59
120	59.20	22.59	36.61	13.39	27.41
140	58.71	23.17	35.54	14.46	26.83
160	57.39	22.18	35.21	14.79	27.82

Tabela F.84

SGLD Desnormalizado 1 média					
#	% de acertos	VP	VN	FP	FN
5	56.95	23.37	33.59	16.41	26.63
15	57.24	25.44	31.80	18.20	24.56
25	57.68	25.80	31.88	18.12	24.20
35	58.85	22.63	36.22	13.78	27.37
45	58.07	24.88	33.20	16.80	25.12
50	58.15	26.37	31.78	18.22	23.63
55	57.95	24.83	33.12	16.88	25.17
60	57.39	24.20	33.20	16.80	25.80
70	57.95	22.46	35.49	14.51	27.54
80	57.98	24.71	33.27	16.73	25.29
90	58.88	25.07	33.80	16.20	24.93
100	58.32	25.17	33.15	16.85	24.83
110	57.90	23.15	34.76	15.24	26.85
120	58.63	25.63	33.00	17.00	24.37
140	59.12	22.49	36.63	13.37	27.51
160	57.59	25.12	32.46	17.54	24.88

Tabela F.85

SGLD Desnormalizado 03					
#	% de acertos	VP	VN	FP	FN
5	62.56	23.44	39.12	10.88	26.56
15	62.59	24.61	37.98	12.02	25.39
25	63.02	24.93	38.10	11.90	25.07
35	62.90	22.68	40.22	9.78	27.32
45	62.68	25.85	36.83	13.17	24.15
50	63.02	27.05	35.98	14.02	22.95
55	62.80	24.78	38.02	11.98	25.22
60	63.54	24.73	38.80	11.20	25.27
70	63.12	24.59	38.54	11.46	25.41
80	62.95	25.73	37.22	12.78	24.27
90	63.63	25.27	38.37	11.63	24.73
100	62.95	26.90	36.05	13.95	23.10
110	63.22	24.10	39.12	10.88	25.90
120	63.41	26.07	37.34	12.66	23.93
140	63.49	23.98	39.51	10.49	26.02
160	62.59	21.29	41.29	8.71	28.71

Tabela F.87

SGLD Desnormalizado 30					
#	% de acertos	VP	VN	FP	FN
5	62.37	23.34	39.02	10.98	26.66
15	62.71	24.20	38.51	11.49	25.80
25	63.17	25.27	37.90	12.10	24.73
35	62.76	22.49	40.27	9.73	27.51
45	62.71	25.80	36.90	13.10	24.20
50	63.49	27.88	35.61	14.39	22.12
55	63.00	25.90	37.10	12.90	24.10
60	63.15	25.20	37.95	12.05	24.80
70	62.85	26.51	36.34	13.66	23.49
80	63.44	28.20	35.24	14.76	21.80
90	63.44	27.46	35.98	14.02	22.54
100	63.78	25.88	37.90	12.10	24.12
110	63.22	27.12	36.10	13.90	22.88
120	63.37	25.66	37.71	12.29	24.34
140	63.02	28.76	34.27	15.73	21.24
160	63.17	25.44	37.73	12.27	24.56

Tabela F.86

SGLD Desnormalizado 33					
#	% de acertos	VP	VN	FP	FN
5	62.02	23.34	38.68	11.32	26.66
15	62.20	24.37	37.83	12.17	25.63
25	62.56	24.73	37.83	12.17	25.27
35	62.54	24.61	37.93	12.07	25.39
45	62.22	25.51	36.71	13.29	24.49
50	62.76	27.59	35.17	14.83	22.41
55	62.78	27.88	34.90	15.10	22.12
60	62.49	25.51	36.98	13.02	24.49
70	62.49	24.00	38.49	11.51	26.00
80	62.66	23.80	38.85	11.15	26.20
90	62.93	25.17	37.76	12.24	24.83
100	62.51	24.00	38.51	11.49	26.00
110	62.15	25.71	36.44	13.56	24.29
120	61.32	26.59	34.73	15.27	23.41
140	62.41	25.98	36.44	13.56	24.02
160	62.41	24.27	38.15	11.85	25.73

Tabela F.88

SGLD Desnormalizado 3 média					
#	% de acertos	VP	VN	FP	FN
5	62.12	23.34	38.78	11.22	26.66
15	62.32	24.02	38.29	11.71	25.98
25	63.05	25.15	37.90	12.10	24.85
35	62.66	23.15	39.51	10.49	26.85
45	62.39	24.73	37.66	12.34	25.27
50	62.98	27.22	35.76	14.24	22.78
55	62.66	24.20	38.46	11.54	25.80
60	62.90	24.93	37.98	12.02	25.07
70	62.85	24.32	38.54	11.46	25.68
80	62.68	26.56	36.12	13.88	23.44
90	63.20	26.05	37.15	12.85	23.95
100	62.66	27.66	35.00	15.00	22.34
110	63.12	27.88	35.24	14.76	22.12
120	63.10	26.44	36.66	13.34	23.56
140	63.17	27.29	35.88	14.12	22.71
160	62.71	27.20	35.51	14.49	22.80

Tabela F.89

SGLD Desnormalizado 05					
#	% de acertos	VP	VN	FP	FN
5	58.07	23.71	34.37	15.63	26.29
15	58.02	24.76	33.27	16.73	25.24
25	59.51	25.24	34.27	15.73	24.76
35	59.46	22.00	37.46	12.54	28.00
45	59.05	24.95	34.10	15.90	25.05
50	59.07	23.00	36.07	13.93	27.00
55	59.07	24.20	34.88	15.12	25.80
60	58.34	23.27	35.07	14.93	26.73
70	58.73	22.78	35.95	14.05	27.22
80	58.85	23.39	35.46	14.54	26.61
90	60.15	25.10	35.05	14.95	24.90
100	59.73	24.66	35.07	14.93	25.34
110	59.95	25.90	34.05	15.95	24.10
120	59.46	23.07	36.39	13.61	26.93
140	59.32	24.15	35.17	14.83	25.85
160	59.34	25.20	34.15	15.85	24.80

Tabela F.91

SGLD Desnormalizado 50					
#	% de acertos	VP	VN	FP	FN
5	59.05	23.68	35.37	14.63	26.32
15	58.66	24.32	34.34	15.66	25.68
25	59.85	26.68	33.17	16.83	23.32
35	59.93	24.02	35.90	14.10	25.98
45	59.63	26.00	33.63	16.37	24.00
50	60.00	26.37	33.63	16.37	23.63
55	59.66	24.44	35.22	14.78	25.56
60	58.90	25.34	33.56	16.44	24.66
70	59.49	24.93	34.56	15.44	25.07
80	58.83	24.24	34.59	15.41	25.76
90	59.88	26.63	33.25	16.75	23.37
100	58.90	22.80	36.10	13.90	27.20
110	59.68	27.73	31.95	18.05	22.27
120	59.44	27.44	32.00	18.00	22.56
140	60.15	25.49	34.66	15.34	24.51
160	60.10	26.22	33.88	16.12	23.78

Tabela F.90

SGLD Desnormalizado 55					
#	% de acertos	VP	VN	FP	FN
5	55.78	22.73	33.05	16.95	27.27
15	56.63	23.46	33.17	16.83	26.54
25	57.44	24.73	32.71	17.29	25.27
35	57.78	22.59	35.20	14.80	27.41
45	58.63	22.63	36.00	14.00	27.37
50	58.44	25.02	33.41	16.59	24.98
55	58.61	25.59	33.02	16.98	24.41
60	58.80	22.88	35.93	14.07	27.12
70	59.59	23.41	36.17	13.83	26.59
80	59.73	24.27	35.46	14.54	25.73
90	58.85	23.88	34.98	15.02	26.12
100	59.51	26.71	32.80	17.20	23.29
110	60.00	22.88	37.12	12.88	27.12
120	59.71	24.83	34.88	15.12	25.17
140	59.51	26.02	33.49	16.51	23.98
160	59.05	24.93	34.12	15.88	25.07

Tabela F.92

SGLD Desnormalizado 5 média					
#	% de acertos	VP	VN	FP	FN
5	58.44	22.68	35.76	14.24	27.32
15	58.59	24.59	34.00	16.00	25.41
25	59.49	25.17	34.32	15.68	24.83
35	59.88	24.17	35.71	14.29	25.83
45	59.27	24.93	34.34	15.66	25.07
50	59.46	26.17	33.29	16.71	23.83
55	59.59	25.27	34.32	15.68	24.73
60	58.63	24.78	33.85	16.15	25.22
70	59.76	23.51	36.24	13.76	26.49
80	58.98	23.51	35.46	14.54	26.49
90	60.44	23.29	37.15	12.85	26.71
100	59.68	23.44	36.24	13.76	26.56
110	59.63	24.61	35.02	14.98	25.39
120	59.56	23.93	35.63	14.37	26.07
140	60.12	24.20	35.93	14.07	25.80
160	59.80	26.44	33.37	16.63	23.56

MATRIZ DE CO-OCORRÊNCIA NORMALIZADO

Tabela F.93

SGLD Normalizado 01					
#	% de acertos	VP	VN	FP	FN
5	58.27	23.22	35.05	14.95	26.78
15	58.34	24.02	34.32	15.68	25.98
25	59.46	25.29	34.17	15.83	24.71
35	59.83	24.46	35.37	14.63	25.54
45	59.44	24.73	34.71	15.29	25.27
50	59.32	23.51	35.80	14.20	26.49
55	59.41	24.44	34.98	15.02	25.56
60	58.49	23.68	34.80	15.20	26.32
70	59.41	23.78	35.63	14.37	26.22
80	58.90	25.83	33.07	16.93	24.17
90	60.02	23.54	36.49	13.51	26.46
100	59.07	24.29	34.78	15.22	25.71
110	59.05	22.54	36.51	13.49	27.46
120	58.98	21.20	37.78	12.22	28.80
140	60.39	24.59	35.80	14.20	25.41
160	58.83	22.34	36.49	13.51	27.66

Tabela F.95

SGLD Normalizado 10					
#	% de acertos	VP	VN	FP	FN
5	55.59	24.10	31.49	18.51	25.90
15	55.83	24.05	31.78	18.22	25.95
25	57.05	25.54	31.51	18.49	24.46
35	57.39	23.46	33.93	16.07	26.54
45	57.34	23.05	34.29	15.71	26.95
50	57.39	22.61	34.78	15.22	27.39
55	57.80	25.63	32.17	17.83	24.37
60	57.68	26.27	31.41	18.59	23.73
70	58.46	22.78	35.68	14.32	27.22
80	58.59	26.63	31.95	18.05	23.37
90	57.68	23.37	34.32	15.68	26.63
100	58.22	24.22	34.00	16.00	25.78
110	58.37	21.85	36.51	13.49	28.15
120	58.10	24.66	33.44	16.56	25.34
140	58.34	25.10	33.24	16.76	24.90
160	58.12	24.37	33.76	16.24	25.63

Tabela F.94

SGLD Normalizado 11					
#	% de acertos	VP	VN	FP	FN
5	56.24	21.24	35.00	15.00	28.76
15	56.41	23.78	32.63	17.37	26.22
25	57.10	25.34	31.76	18.24	24.66
35	58.02	22.41	35.61	14.39	27.59
45	58.07	22.17	35.90	14.10	27.83
50	58.00	20.49	37.51	12.49	29.51
55	58.46	25.02	33.44	16.56	24.98
60	58.46	23.61	34.85	15.15	26.39
70	59.07	22.61	36.46	13.54	27.39
80	59.71	23.80	35.90	14.10	26.20
90	58.76	22.41	36.34	13.66	27.59
100	59.29	20.29	39.00	11.00	29.71
110	59.29	21.61	37.68	12.32	28.39
120	59.66	22.07	37.59	12.41	27.93
140	59.12	19.83	39.29	10.71	30.17
160	59.71	25.00	34.71	15.29	25.00

Tabela F.96

SGLD Normalizado 1 média					
#	% de acertos	VP	VN	FP	FN
5	55.95	22.00	33.95	16.05	28.00
15	56.39	23.66	32.73	17.27	26.34
25	57.24	24.07	33.17	16.83	25.93
35	57.71	21.63	36.07	13.93	28.37
45	57.98	22.73	35.24	14.76	27.27
50	57.98	23.59	34.39	15.61	26.41
55	58.27	24.20	34.07	15.93	25.80
60	58.27	22.93	35.34	14.66	27.07
70	59.05	22.63	36.41	13.59	27.37
80	59.37	22.95	36.41	13.59	27.05
90	58.44	22.44	36.00	14.00	27.56
100	58.88	23.93	34.95	15.05	26.07
110	58.73	23.37	35.37	14.63	26.63
120	59.10	21.78	37.32	12.68	28.22
140	59.90	23.90	36.00	14.00	26.10
160	58.93	23.61	35.32	14.68	26.39

Tabela F.97

SGLD Normalizado 03					
#	% de acertos	VP	VN	FP	FN
5	57.59	21.90	35.68	14.32	28.10
15	57.78	24.05	33.73	16.27	25.95
25	57.85	24.93	32.93	17.07	25.07
35	58.39	21.71	36.68	13.32	28.29
45	58.98	23.39	35.59	14.41	26.61
50	59.20	24.54	34.66	15.34	25.46
55	59.20	25.17	34.02	15.98	24.83
60	59.49	23.61	35.88	14.12	26.39
70	60.20	22.12	38.07	11.93	27.88
80	60.49	24.93	35.56	14.44	25.07
90	59.73	22.54	37.20	12.80	27.46
100	59.95	22.37	37.59	12.41	27.63
110	60.05	23.98	36.07	13.93	26.02
120	60.37	21.71	38.66	11.34	28.29
140	60.61	24.22	36.39	13.61	25.78
160	59.68	21.02	38.66	11.34	28.98

Tabela F.99

SGLD Normalizado 30					
#	% de acertos	VP	VN	FP	FN
5	57.07	22.54	34.54	15.46	27.46
15	57.37	24.73	32.63	17.37	25.27
25	57.73	25.05	32.68	17.32	24.95
35	58.07	22.41	35.66	14.34	27.59
45	58.27	23.15	35.12	14.88	26.85
50	58.22	24.12	34.10	15.90	25.88
55	58.76	23.32	35.44	14.56	26.68
60	58.46	26.02	32.44	17.56	23.98
70	59.15	23.32	35.83	14.17	26.68
80	59.41	25.85	33.56	16.44	24.15
90	58.90	23.37	35.54	14.46	26.63
100	58.85	24.02	34.83	15.17	25.98
110	58.88	23.85	35.02	14.98	26.15
120	59.00	23.80	35.20	14.80	26.20
140	59.41	25.02	34.39	15.61	24.98
160	59.15	23.80	35.34	14.66	26.20

Tabela F.98

SGLD Normalizado 33					
#	% de acertos	VP	VN	FP	FN
5	57.41	21.32	36.10	13.90	28.68
15	57.73	23.10	34.63	15.37	26.90
25	58.12	25.22	32.90	17.10	24.78
35	58.32	21.10	37.22	12.78	28.90
45	58.85	22.41	36.44	13.56	27.59
50	58.63	21.76	36.88	13.12	28.24
55	58.98	25.07	33.90	16.10	24.93
60	59.05	22.63	36.41	13.59	27.37
70	59.85	23.32	36.54	13.46	26.68
80	59.90	24.20	35.71	14.29	25.80
90	59.54	23.07	36.46	13.54	26.93
100	60.12	22.37	37.76	12.24	27.63
110	59.73	22.80	36.93	13.07	27.20
120	60.15	23.24	36.90	13.10	26.76
140	59.49	21.88	37.61	12.39	28.12
160	59.44	22.05	37.39	12.61	27.95

Tabela F.100

SGLD Normalizado 3 média					
#	% de acertos	VP	VN	FP	FN
5	57.39	21.27	36.12	13.88	28.73
15	57.66	23.83	33.83	16.17	26.17
25	57.80	25.10	32.71	17.29	24.90
35	58.27	22.98	35.29	14.71	27.02
45	58.80	24.44	34.37	15.63	25.56
50	58.76	23.27	35.49	14.51	26.73
55	58.93	23.76	35.17	14.83	26.24
60	59.20	22.61	36.59	13.41	27.39
70	59.93	23.00	36.93	13.07	27.00
80	60.00	22.07	37.93	12.07	27.93
90	59.49	21.93	37.56	12.44	28.07
100	58.27	22.98	35.29	14.71	27.02
110	58.80	24.44	34.37	15.63	25.56
120	58.76	23.27	35.49	14.51	26.73
140	57.66	23.83	33.83	16.17	26.17
160	57.80	25.10	32.71	17.29	24.90

Tabela F.101

SGLD Normalizado 05					
#	% de acertos	VP	VN	FP	FN
5	57.66	22.44	35.22	14.78	27.56
15	57.93	24.15	33.78	16.22	25.85
25	58.17	24.90	33.27	16.73	25.10
35	58.59	21.12	37.46	12.54	28.88
45	58.88	22.68	36.20	13.80	27.32
50	59.17	22.46	36.71	13.29	27.54
55	59.29	24.17	35.12	14.88	25.83
60	59.41	22.37	37.05	12.95	27.63
70	59.95	21.34	38.61	11.39	28.66
80	60.59	24.24	36.34	13.66	25.76
90	59.95	20.93	39.02	10.98	29.07
100	60.20	22.02	38.17	11.83	27.98
110	59.80	19.10	40.71	9.29	30.90
120	60.15	21.46	38.68	11.32	28.54
140	60.54	25.80	34.73	15.27	24.20
160	60.05	20.37	39.68	10.32	29.63

Tabela F.103

SGLD Normalizado 50					
#	% de acertos	VP	VN	FP	FN
5	58.12	22.51	35.61	14.39	27.49
15	58.73	24.78	33.95	16.05	25.22
25	58.76	24.63	34.12	15.88	25.37
35	58.83	23.49	35.34	14.66	26.51
45	59.12	24.80	34.32	15.68	25.20
50	59.07	23.68	35.39	14.61	26.32
55	59.59	24.39	35.20	14.80	25.61
60	59.10	23.85	35.24	14.76	26.15
70	59.61	25.56	34.05	15.95	24.44
80	59.51	26.51	33.00	17.00	23.49
90	59.41	23.02	36.39	13.61	26.98
100	59.63	25.24	34.39	15.61	24.76
110	59.56	27.02	32.54	17.46	22.98
120	59.44	22.71	36.73	13.27	27.29
140	59.80	24.27	35.54	14.46	25.73
160	59.44	24.83	34.61	15.39	25.17

Tabela F.102

SGLD Normalizado 55					
#	% de acertos	VP	VN	FP	FN
5	58.07	22.17	35.90	14.10	27.83
15	58.39	24.02	34.37	15.63	25.98
25	58.80	24.71	34.10	15.90	25.29
35	58.68	22.20	36.49	13.51	27.80
45	59.29	23.34	35.95	14.05	26.66
50	59.15	22.24	36.90	13.10	27.76
55	59.15	25.37	33.78	16.22	24.63
60	59.27	24.15	35.12	14.88	25.85
70	60.29	24.15	36.15	13.85	25.85
80	60.17	25.41	34.76	15.24	24.59
90	59.66	25.63	34.02	15.98	24.37
100	59.54	24.10	35.44	14.56	25.90
110	60.37	26.88	33.49	16.51	23.12
120	59.88	23.15	36.73	13.27	26.85
140	59.68	21.88	37.80	12.20	28.12
160	60.20	24.95	35.24	14.76	25.05

Tabela F.104

SGLD Normalizado 5 média					
#	% de acertos	VP	VN	FP	FN
5	52.10	25.61	26.49	23.51	24.39
15	52.90	26.59	26.32	23.68	23.41
25	53.93	27.22	26.71	23.29	22.78
35	53.41	26.32	27.10	22.90	23.68
45	54.41	27.32	27.10	22.90	22.68
50	54.29	25.93	28.37	21.63	24.07
55	55.00	28.80	26.20	23.80	21.20
60	54.83	30.61	24.22	25.78	19.39
70	55.44	27.85	27.59	22.41	22.15
80	55.34	30.07	25.27	24.73	19.93
90	55.41	27.73	27.68	22.32	22.27
100	54.73	31.22	23.51	26.49	18.78
110	55.20	29.10	26.10	23.90	20.90
120	55.46	31.49	23.98	26.02	18.51
140	54.85	23.83	31.02	18.98	26.17
160	55.34	28.00	27.34	22.66	22.00

CURVA DE COMPLEXIDADE

Tabela F.105

CClx					
#	% de acertos	VP	VN	FP	FN
5	72.10	42.07	30.02	19.98	7.93
15	73.10	41.66	31.44	18.56	8.34
25	73.90	42.27	31.63	18.37	7.73
35	74.05	42.24	31.80	18.20	7.76
45	74.27	42.12	32.15	17.85	7.88
50	73.68	41.17	32.51	17.49	8.83
55	74.71	42.85	31.85	18.15	7.15
60	74.71	42.49	32.22	17.78	7.51
70	74.44	41.90	32.54	17.46	8.10
80	74.78	42.20	32.59	17.41	7.80
90	74.56	42.56	32.00	18.00	7.44
100	74.93	41.37	33.56	16.44	8.63
110	74.54	42.10	32.44	17.56	7.90
120	75.24	41.98	33.27	16.73	8.02
140	75.41	42.12	33.29	16.71	7.88
160	75.32	42.73	32.59	17.41	7.27

Bibliografia

- [1] PARKIN, D. M., PISANI, P., FERLAY, J., "Global Cancer Statistics", *CA - A Cancer Journal for Clinicians*, v. 49, n. 1, pp. 33-64, 1999.
- [2] "Estimativa 2005: Incidência de Câncer no Brasil." Rio de Janeiro: Ministério da Saúde. Instituto Nacional de Câncer – INCA, 2004, pp. 94
- [3] "Instituto Nacional do Câncer - INCa," 2005 <http://www.inca.gov.br> [10/02/2005].
- [4] "National Alliance of Breast Cancer Organizations," 2003. <http://www.nabco.com> [03/10/2003].
- [5] HALL, F. M., STORELLA, J.M., SILVERSTONE, D.Z., WYSHAK, G., "Nonpalpable breast lesions for biopsy based on suspicion of carcinoma mammography", *Radiology*, v. 167, pp. 353-358, 1988.
- [6] GOUMOT, P. A., *Le Sein - Son Image*. 1 ed. Paris, 1993.
- [7] BOCCHI, L., COPPINI, G. ET AL., "Tissue characterization from X-ray images", *Med. Eng. Phis*, v. 19, n. 4, pp. 336-342, 1997.
- [8] LAFONTAN, B., DAURES, J.P. ET AL., "Isolated clustered microcalcifications: diagnostic values of mammography - Series of 400 cases with surgical verification", *Breast Imaging*, v. 190, n. 2, pp. 479 - 483, 1994.
- [9] ZIMMER, Y. E. A., "The distribution of the local entropy in ultrasound images", *Ultrasound in Med. & Biol*, v. 22, n. 4, pp. 431-439, 1996.
- [10] LEITE, L., *Software Identificador e Classificador de Microcalcificações em Mamografias*, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2001.
- [11] LEITE, L., MACIEL, C. D., PEREIRA, W. C. A., "Phantoms Computacionais para Imagens Ultra-Sônicas", In: *Proceedings of IV Fórum Nacional de Ciência e Tecnologia em Saúde*, pp. 382-382, Curitiba - PR - Brasil, 1998.
- [12] LEITE, L., MACIEL, C. D., PEREIRA, W. C. A., "Phantoms Computacionais para Imagens Ultra-Sônicas", In: *Proceedings of XVII Congresso Brasileiro de Engenharia Biomédica*, pp. 136, Florianópolis - SC - Brasil, 2000.
- [13] LEITE, L., "Análise de Componentes Independentes aplicada à identificação de regiões lesionadas em mamogramas", In: *Proceedings of II Congresso Latino-Americano de Engenharia Biomédica*, pp. 105, João Pessoa - PA - Brasil, 2004.
- [14] LEITE, L., GARCIA, L. G., CALÔBA, L. P., "Classificação de Lesões da Mama usando Parametrização e Redes Neurais", In: *Proceedings of Congresso Brasileiro de Automática*, Gramado - PR - Brasil, 2004.
- [15] ZHANG, L., SANKAR, R., QIAN, W., "Advances in micro-calcification clusters detection in mammography", *Comp. in Biology & Medicine*, v. 32, pp. 515-528, 2002.
- [16] BARLOW, W. E., ET AL., "Performance of Diagnostic Mammography for Women With Signs or Symptoms of Breast Cancer", *Journal of the National Cancer Institute*, v. 94, pp. 1151-1159, 2002.

- [17] RAJI, A., THAIBAOU, A., ET AL.. "A Gray-Level Transformation-Based Method for Image Enhancement", *Pattern Recognition Letters*, v. 19, pp. 1207-1212, 1998.
- [18] NUNES, F. L. S., SCHIABEL, H., BENATTI, "Detecção de Agrupamentos de Microcalcificações em Mamas Densas: Avaliação de duas Técnicas para Realce de Contraste", In: *Proceedings of Anais do XVIII Congresso de Engenharia Biomédica*, pp. 1154-1159, São José dos Campos, SP, 2002.
- [19] CHRISTOYIANNI, I., KOUTRAS, A., KOKKINAKIS, G., "Computer aided diagnosis of breast cancer in digitized mammograms", *Comp. Med. Imag. & Graph*, v. 26, pp. 309-319, 2002.
- [20] FOGEL, D. B., WASSON III, E.C., BOUGHTON, E.M. , PORTO, V.W., "Evolving artificial neural networks for screening features from mammograms", *Artificial Intelligence in Medicine - Elsevier*, v. 14, pp. 317-326, 1998.
- [21] FOGEL, D. B., WASSON III, E.C. , BOUGHTON, E.M., PORTO, V.W., "A step toward computer-assisted mammography using evolutionary programming and neural networks", *Cancer Letters - Elsevier*, v. 119, pp. 93-97, 1997.
- [22] AMADO, C., PIRES, A.M., "Uso de Técnicas de Reamostragem para Seleção de Variáveis em Análise Discriminante", In: *Proceedings of V Congresso Anual da Sociedade Portuguesa de Estatística*, Cúria, Portugal, 1997.
- [23] BANDYOPADHYAY, S., MURTHY, C. A. , SANKAR K. P., "Pattern classification using genetic algorithms: Determination of H", *Pattern Recognition Letters, Elsevier*, v. 19, pp. 1171-1181, 1998.
- [24] GARRELL I GUIU, J. M., GOLOBARDES I RIBÉ, E. , BERNADÓ I MANSILLA, E., LLORÀ I FÀBREGA, X., "Automatic diagnosis with genetic algorithms and case-based reasoning", *Artificial Intelligence in Engineering - Elsevier*, v. 13, pp. 367-372, 1999.
- [25] AL-JANOBI, A., "Performance evaluation of cross-diagonal textures matrix method of texture analysis", *Pattern Recognition*, v. 34, pp. 171-180, 1998.
- [26] BAHEERATHAN, S., ALBREGTSEN, F., DANIELSEN, H.E., "New texture features based on the complexity curve", *Pattern Recognition*, v. 32, pp. 605-618, 1999.
- [27] AZEVEDO, C. M., *Radiologia da Mama*. Rio de Janeiro, INCA, 1994.
- [28] LANYI, M., "Diagnosis and differential diagnosis of breast calcifications", In: *Proceedings of Springer-Verlag*, New York, 1986.
- [29] BASSETT, L. W., "Mammographic analysis of calcifications", *Radiol. Clin. North Am*, v. 30, pp. 93-105, 1992.
- [30] SICKLES, E. A., "Breast calcifications: mammographic evaluation", *Radiology*, v. 60, pp. 289-293, 1986.
- [31] HYVÄRINEN, A., KARHUNEN, J., OJA, E., *Independent Component Analysis*. New York, 2001.
- [32] PAJUNEN, A. H. A. P., "Nonlinear independent component analysis: Existence and uniqueness results", *Neural Networks*, v. 12, n. 3, pp. 429-439, 1999.

- [33] VALPOLA, H., GIANNAKOPOULOS, X. , HONKELA, A. , KARHUNEN, J., "Independent component analysis using ensemble learning: Experiments and discussion", In: *Proceedings of Proc. Int. Workshop on Independent Component Analysis and Blind Signal Separation (ICA2000)*, pp. 351-356, Helsinki, Finland, 2000.
- [34] PAPOULIS, A., *Probability, random variables and stochastic processes*. 3rd ed, 1991.
- [35] LEE, T. W., LEWICKI, M.S., GIROLAMI, M., SEJNOWSKI, T.J., "Blind source separation of more sources than mixtures using overcomplete representations", *IEEE Signal Processing Letters*, v. 4, n. 5, 1999.
- [36] HYVÄRINEN, A., OJA, E., "A fast fixed-point algorithm for independent component analysis", *Neural Computations*, v. 9, n. 7, pp. 1483-1492, 1997.
- [37] HYVÄRINEN, A., "One-unit contrast functions for independent component analysis: A statistical analysis", In: *Proceedings of Neural Networks for Signal Processing VII (Proc. IEEE Workshop on Neural Networks for Signal Processing)*, pp. 388-397, Amelia Island, Florida, 1998.
- [38] BELL, A. J., SEJNOWSKI, T.J., "An information-maximization approach to blind separation and blind deconvolution", *Neural Computations*, v. 7, pp. 1129-1159, 1995.
- [39] AMARI, S. I., "Natural gradient works efficiently in learning", *Neural Computation*, v. 10, n. 2, pp. 251-276, 1998.
- [40] HAYKIN, S., *Redes Neurais - Princípio e Práticas*. Porto Alegre, 2001.
- [41] OHANIAN, P. P., DUBES, R.C., "Performance evaluation for four classes of textural features", *Pattern Recognition*, v. 25, n. 8, pp. 819-833, 1992.
- [42] HARALICK, R. M., SHANMUGAM, K., DINSTEN, I., "Textural features for image classification", *IEEE Trans. Systems Man Cybernet*, v. 3, pp. 610-621, 1973.
- [43] WESZKA, J. S., DYER, C.R., ROSEVELD, A., "A comparative study of texture measures for terrain classification", *IEEE Trans. Systems Man Cybernet*, v. 5, pp. 269-285, 1976.
- [44] GAVRIELIDES, M. A., "Parameter optimization of a computer-aided diagnosis scheme for the segmentation of microcalcification clusters in mammograms", *Med Phys*, v. 29, 475- 483 2002.
- [45] NAGEL, R. H., NISHIKAWA, R.M., PAPAIOANNOU J., DOI K., "Analysis of methods for reducing false positives in the automated detection of clustered microcalcifications in mammograms", *Medical Physics*, v. 25, n. 8, pp. 1502-1506, 1998.
- [46] HYVÄRINEN, A., HOYER, P.O., OJA, E., 2001, *Image denoising by sparse code shrinkage*. In: *Intelligent Signal Processing* IEEE Press.
- [47] SILVA, R. R., SIQUEIRA, M.H.S., CALÔBA, L.P., REBELLO, J.M.A., "Contribution to the development of a radiographic inspection automated system." In: *Proceedings of 8th European Conference on Non Destructive Testing*, pp. 17–21, 2002.

- [48] SILVA, R. R., SIQUEIRA, M.H.S., CALÔBA, L.P., REBELLO, J.M.A., "Evaluation of the relevant characteristic parameters of welding defects and probability of correct classification using linear classifiers." *Insight 2002*, v. 44(10), pp. 616–622, 2003.
- [49] BOOMGAARD, R., *Mathematical Morphology: Extensions Towards Computer Vision*, Universiteit van Amsterdam, 1992.
- [50] FACON, J., *Morfologia Matemática: teoria e exemplos*. Curitiba, Editora Universitária Champagnat da Pontifícia Universidade Católica do Paraná, 1996.
- [51] ALVARENGA, A. V., INFANTOSI, A.F.C., AZEVEDO, C.M., PEREIRA, W.C.A., "Classifying Fatty and Glandular Tissues from Breast Ultrasound Images with the Complexity Curve and Genetic Algorithm", In: *Proceedings of IFMBE Proceedings of World Congress on Medical Physics and Biomedical Engineering*, pp. 01-04, Sydney, Australia, 2003.