

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

FACULDADE DE ENGENHARIA ELÉTRICA

PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA ELÉTRICA



Separação em Duas ou Mais Classes Utilizando o Classificador Polinomial

ALESSANDRO SANTANA MARTINS

Orientadora: Prof.^a Dr.^a Edna Lúcia Flôres

Co-orientador: Prof. Dr. Gilberto Arantes Carrijo

Uberlândia

2013

Alessandro Santana Martins

Separação em Duas ou Mais Classes Utilizando o Classificador Polinomial

Tese apresentada ao Programa de Pós-Graduação Stricto Sensu da Faculdade de Engenharia da Universidade Federal de Uberlândia, como recurso parcial para obtenção do título de **Doutor em Ciências**.

Área de Concentração: Processamento da Informação

Linha de Pesquisa: Processamento Digital de Sinais

Orientadora: Prof.^a Dr.^a Edna Lúcia Flôres

Co-orientador: Prof. Dr. Gilberto Arantes Carrijo

Uberlândia

2013

Alessandro Santana Martins

Separação em Duas ou Mais Classes Utilizando o Classificador Polinomial

Tese apresentada ao Programa de Pós-Graduação Stricto Sensu da Faculdade de Engenharia da Universidade Federal de Uberlândia, como recurso parcial para obtenção do título de **Doutor em Ciências**.

Área de Concentração: Processamento da Informação

Linha de Pesquisa: Processamento Digital de Sinais

Profa. Dr.^a Edna Lúcia Flores
Orientadora

Prof. Dr. Alexandre Cardoso
Coordenador do curso de Pós-Graduação

Banca Examinadora

Edna Lúcia Flôres, Dr.^a (UFU) - Orientadora

Gilberto Arantes Carrijo, Dr. (UFU) - Co-orientador

Cláudio Afonso Fleury, Dr. (PUCGO)

Wagner Fernando Delfino Angelotti, Dr. (UFTM)

Antônio Cláudio Paschoarelli Veiga, Dr. (UFU)

Uberlândia

01/03/2013

*Aos meus pais, Geraldo e Candida
e aos meus irmãos, Edemilson e Mario Marcos,
ofereço.*

*À minha esposa, Lucimar
e às minhas filhas Ingrid e Emily,
dedico.*

Agradecimentos

Primeiramente, devo agradecer a Deus por ter me dado a vida e oportunidade de poder conviver ao lado de pessoas maravilhosas.

Aos meus pais Geraldo e Candida, que sempre me apoiaram e me incentivaram em tudo que decidi realizar e que enchem minha vida de muito amor, carinho e bons exemplos.

À minha esposa Lucimar, a quem pude confiar todas as minhas preocupações, carinho, atenção, paciência, compreensão e que foi uma das principais fontes de incentivo.

Às minhas filhas Ingrid e Emily, que são a cara do pai e me trazem muitas felicidades.

Agradecimentos especiais à minha madrinha Marilda, padrinho Mário Lúcio, Tia Rosa e primo Maykeel por terem me acolhido no início de minha carreira acadêmica.

A todos os meus familiares por compreenderem a minha ausência constante. Em especial, aos meus irmãos Edemilson e Mário Marcos, Estela, Fátima, Luiz, Vânia, Tânia, Jânio, Luciene, Rubem, Wander Romes, Victor, Augusto, Karina, Larissa, Fillipy que me ajudou na formatação das figuras, Matheus, Bruno, Yasmim, Sônia, Edna, Itamar e Dona Mariinha.

Um agradecimento especial à Profa. Dra. Edna Lúcia Flôres que tornou possível a realização deste trabalho, pela dedicação e paciência que sempre teve comigo, pela amizade e companheirismo.

Ao Prof. Dr. Gilberto Arantes Carrijo, pelo apoio, orientação e confiança na minha pessoa.

Aos Profs. Dr. Leandro Alves Neves e Dr. Marcelo Zanchetta do Nascimento, pela

amizade, profissionalismo, dedicação e atenção na realização do artigo.

Aos meus compadres Leonardo e Franciele, Humberto e Andréia, afilhados Júlia e Otávio, pela amizade sincera. Aos meus amigos Daniela e Roberto, Celso e Ana Paula, Roberto e Telma, Josney, e outros que não foram citados mas são muito importantes na minha vida pelo apoio em todas as horas.

Aos amigos da UEMG que souberam entender a minha saída, mas que sempre estarão comigo onde eu estiver. Do IFTM que em pouco tempo de trabalho fui muito bem acolhido, em especial meu amigo de república Luis que me ajudou muito na consolidação deste trabalho. A todos amigos da UFMG, em especial Dona Maria José, uma pessoa iluminada por Deus e que muito admiro, ao professor Dr. Allynson, um amigo de verdade. Ao grande amigo Osmar e sua esposa Dora pelo apoio dedicado a mim num dos momentos mais difíceis de minha vida.

À Universidade Federal de Uberlândia (UFU) por permitir tornar possível a realização do meu trabalho e aos professores e funcionários da Engenharia Elétrica, pelo apoio constante e informação adquiridos durante os meus estudos.

“ Feliz de quem entende que é preciso
mudar muito para ser sempre o mesmo. ”

Mário Quintana

Resumo

O aperfeiçoamento dos sistemas computacionais tem beneficiado o desenvolvimento de muitas áreas de pesquisa na medicina. A interpretação e a análise de imagens médicas representam uma parte importante em visão computacional e no reconhecimento de padrões. Está se tornando essencial para muitos pesquisadores e centros clínicos o desenvolvimento de um sistema de diagnóstico auxiliado por computador para doenças como o câncer de mama, com a finalidade de auxiliar médicos em hospitais. O classificador polinomial desenvolvido neste trabalho é um método de classificação supervisionado e pode classificar duas ou mais classes. Esse classificador expande o vetor de características projetado no espaço \mathbb{R}^d para um espaço de dimensão superior onde é possível a classificação. O classificador polinomial se mostra um importante método de classificação principalmente em tratamento com classes não linearmente separáveis. Nesta tese, esse classificador foi utilizado para reconhecer padrões da base de dados IRIS de Fisher de 1936, que é composta das flores setosa, versicolor e virginica. Vários testes foram realizados com essa base e ao usar três ou quatro características, o classificador polinomial conseguiu classificar todas essas flores. Também foi realizada uma aplicação com o classificador polinomial na identificação de pixels parciais em regiões de interesse de imagens mamográficas. Comparados com os classificadores SVM e árvore de decisão, os resultados obtidos foram melhores utilizando o classificador polinomial.

Palavras-chave: Classificador polinomial, reconhecimento de padrões, processamento digital de imagens, vetor de características, imagens médicas.

Abstract

The improvement of computer systems has benefited the development of many research areas in medicine. Analysis and interpretation of medical images represent an important part in computer vision and pattern recognition. It's essential for many researchers and medical centers to develop a diagnosis system aided by computer for diseases such as breast cancer, in order to assist doctors in hospitals. The polynomial classifier developed in this work is a supervised classification method and can sort two or more classes. This classifier expands the feature vector projected into the space \mathbb{R}^d into a higher dimensional space where the classification is possible. The polynomial classifier proves to be an important method of classification mainly on treatment with non-linearly separable classes. In this thesis, the classifier was used to recognize patterns from the database IRIS Fisher 1936, which is composed by the flowers setosa, versicolor and virginica. Several tests were performed on this basis and by using three or four features the polynomial classifier managed to sort all these flowers. An application was also performed with the polynomial classifier for the identification of partial pixels in regions of interest in mammographic images. Compared with SVM classifiers and decision tree, the results were better using the polynomial classifier.

Keywords: polynomial classifier, pattern recognition, digital image processing, feature vector, medical images.

Sumário

1	Introdução	1
1.1	Introdução	1
1.2	Objetivos Deste Trabalho	6
1.3	Estrutura Deste Trabalho	6
1.4	Considerações Finais Deste Capítulo	7
2	Fundamentos de Processamento Digital de Imagens, Extração de Características, Dimensão Fractal e Lacunaridade	8
2.1	Introdução	8
2.2	Fundamentos de Processamento Digital de Imagens	9
2.3	Extração de Características	11
2.4	Dimensão Fractal Multiescala	12
2.4.1	Massa-raio	14
2.4.2	Análise por intersecção acumulativa	15
2.4.3	<i>Boxcounting</i>	16
2.5	Lacunaridade Multiescala	17
2.6	Considerações Finais Deste Capítulo	18
3	Classificadores	19
3.1	Introdução	19
3.2	Padrões e Classes de Padrões	20
3.3	Funções Discriminantes	21

3.3.1	Funções Discriminantes Lineares	23
3.3.2	Classificação em duas classes	23
3.3.3	Classificação em várias classes	24
3.3.4	Funções discriminantes lineares generalizadas	26
3.4	K Vizinhos Mais Próximos (K-NN)	27
3.5	Discriminante Linear de Fisher	29
3.6	Máquina de Vetores Suporte	33
3.7	Árvores de Decisão	36
3.7.1	Métodos de poda	42
3.7.2	Algoritmos de indução de árvores de decisão	43
3.8	Considerações Finais Deste Capítulo	45
4	Método Proposto	47
4.1	Introdução	47
4.2	Classificação Polinomial de Duas Classes	48
4.3	Classificação Polinomial em Várias Classes	55
4.4	Considerações Finais Deste Capítulo	60
5	Resultados Obtidos	62
5.1	Introdução	62
5.2	Classificação Polinomial da Base de Dados Iris	62
5.2.1	Classificação polinomial da base de dados Iris utilizando duas caracte- rísticas	63
5.2.2	Classificação polinomial da base de dados Iris utilizando três caracte- rísticas	88
5.2.3	Classificação polinomial da base de dados Iris utilizando quatro caracte- rísticas	101
5.2.4	Classificação polinomial da base de dados Iris utilizando o método de validação cruzada	103

5.3	Classificação Polinomial para Identificação de Pixels Parciais em Regiões de Interesse de Imagens Mamográficas	105
5.4	Conclusões Deste Capítulo	120
6	Conclusões, Contribuições Deste Trabalho e Trabalhos Futuros	122
6.1	Introdução	122
6.2	Conclusões	122
6.3	Contribuições Deste Trabalho	124
6.4	Trabalhos Futuros	125
6.5	Publicação Deste Trabalho	125
6.6	Considerações Finais Deste Capítulo	125
	Referências Bibliográficas	127

Lista de Figuras

2.1	Passos fundamentais em processamento digital de imagens [23].	9
2.2	Ilustração esquemática da extração de características em imagens de pé e de mão [24]. . .	12
3.1	Estrutura funcional de um classificador de padrões estatístico.	22
3.2	Representação das fronteiras de decisão para um classificador de Bayes bidimensional para duas classes onde as funções de densidades de probabilidade são Gaussianas [42].	23
3.3	Limites de decisão para um problema com quatro classes, dicotomizador ω_i /não ω_i	25
3.4	Limites de decisão para um problema com quatro classes, dicotomizador ω_i/ω_j	25
3.5	Classificação de duas classes utilizando o método K-NN para K=1, K=3 e K=5 [43].	28
3.6	Projeção de um mesmo conjunto de amostras em duas retas diferentes na direção de \mathbf{w} [42].	30
3.7	Identificação da margem ρ e dos vetores suporte sobre a linha pontilhada.	34
3.8	Distância entre os hiperplanos e os vetores suporte.	36
3.9	(a) Conjunto linearmente separável; (b) conjunto não-linearmente separável.	36
3.10	Exemplo de árvore de decisão univariada [42].	37
3.11	Exemplo de árvore de decisão oblíqua [42].	38
4.1	Distribuição de duas classes no espaço de características.	49
4.2	Classificação das duas classes da Figura 4.1 utilizando o discriminante linear de Fisher.	49
4.3	Diagrama de blocos do classificador polinomial.	51
4.4	Classificação polinomial de duas classes utilizando uma base polinomial de ordem 5.	55
4.5	Classificação polinomial de três classes utilizando uma base polinomial de ordem 1.	57
4.6	Classificação polinomial de três classes utilizando uma base polinomial de ordem 2.	58
4.7	Classificação polinomial de quatro classes utilizando uma base polinomial de ordem 1.	59

4.8	Classificação polinomial de quatro classes utilizando uma base polinomial de ordem 2. . .	60
5.1	Imagem das três flores Iris da base de dados obtida Fisher.	63
5.2	Distribuição das características das flores setosa, versicolor e virginica no espaço de características comprimento da sépala (x_1) versus largura da sépala (x_2).	64
5.3	Curva de nível do classificador polinomial de grau 40 para as flores setosa, versicolor e virginica.	66
5.4	Comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando as características comprimento da sépala (x_1) e largura da sépala (x_2) . .	67
5.5	Distribuição das características das flores setosa, versicolor e virginica no espaço de características comprimento da sépala (x_1) versus comprimento da pétala (x_3).	68
5.6	Curva de nível do classificador polinomial de grau 29 para as flores setosa, versicolor e virginica.	70
5.7	Comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando as características comprimento da sépala (x_1) e comprimento da pétala (x_3)	71
5.8	Distribuição das características das flores setosa, versicolor e virginica no espaço de características comprimento da sépala (x_1) versus largura da pétala (x_4).	72
5.9	Curva de nível do classificador polinomial de grau 8 para as flores setosa, versicolor e virginica.	74
5.10	Comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando as características comprimento da sépala (x_1) e largura da pétala (x_4)	75
5.11	Distribuição das características das flores setosa, versicolor e virginica no espaço de características largura da sépala (x_2) versus comprimento da pétala (x_3).	76
5.12	Curva de nível do classificador polinomial de grau 17 para as flores setosa, versicolor e virginica.	78
5.13	Comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando as características largura da sépala (x_2) e comprimento da pétala (x_3). . .	79

5.14	Distribuição das características das flores setosa, versicolor e virginica no espaço de características largura da sépala (x_2) versus largura da pétala (x_4).	80
5.15	Curva de nível do classificador polinomial de grau 3 para as flores setosa, versicolor e virginica.	82
5.16	Comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando as características largura da sépala (x_2) e largura da pétala (x_4)	83
5.17	Distribuição das características das flores setosa, versicolor e virginica no espaço de características comprimento da pétala (x_3) versus largura da pétala (x_4).	84
5.18	Curva de nível do classificador polinomial de grau 8 para as flores setosa, versicolor e virginica.	86
5.19	Comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando as características comprimento da pétala (x_3) e largura da pétala (x_4). . .	87
5.20	Comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando as características comprimento da sépala (x_1), largura da sépala (x_2) e comprimento da pétala (x_3).	91
5.21	Comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando as características comprimento da sépala (x_1), largura da sépala (x_2) e largura da pétala (x_4).	94
5.22	Comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando as características comprimento da sépala (x_1), comprimento da pétala (x_3) e largura da pétala (x_4).	97
5.23	Comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando as características largura da sépala (x_2), comprimento da pétala (x_3) e largura da pétala (x_4).	100
5.24	Comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando as características comprimento da sépala (x_1), largura da sépala (x_2), comprimento da pétala (x_3) e largura da pétala (x_4).	103
5.25	Método de validação cruzada.	104

5.26	Diagrama de blocos das etapas utilizadas no processamento das imagens mamográficas: segmentação, quantificação, análise e classificação.	106
5.27	Imagens mamográficas das visões CC e MLO do caso C-0100.	107
5.28	Visualização das ROIs utilizadas para a quantificação dos comportamentos das texturas. (a) e (d) ROIs do grupo NM, nas visões CC e MLO, respectivamente; (b) e (e) estruturas segmentadas do grupo NMS; (c) e (f) estruturas segmentadas do grupo PP.	108
5.29	Assinaturas das texturas das ROIs representadas em gráficos log-log das dimensões fractais.	109
5.30	Assinaturas das texturas das ROIs representadas em gráficos log-log das lacunaridades. .	109
5.31	Comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando a característica dimensão fractal para o grupo NMS versus ES.	112
5.32	Comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando a característica dimensão fractal para o grupo NMS versus PP.	112
5.33	Comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando a característica dimensão fractal para o grupo PP versus ES.	113
5.34	Comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando a característica lacunaridade para o grupo NMS versus ES.	115
5.35	Comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando a característica lacunaridade para o grupo NMS versus PP.	116
5.36	Comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando a característica lacunaridade para o grupo PP versus ES.	116
5.37	Comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando as características dimensão fractal e lacunaridade para o grupo NMS versus ES.	118
5.38	Comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando as características dimensão fractal e lacunaridade para o grupo NMS versus PP.	119
5.39	Comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando as características dimensão fractal e lacunaridade para o grupo PP versus ES.	119

Lista de Tabelas

1.1	Recursos computacionais utilizados neste trabalho.	3
5.1	Resultados obtidos da classificação polinomial utilizando as características comprimento da sépala (x_1) e largura da sépala (x_2).	65
5.2	Resultados obtidos pelos classificadores SVM e árvore de decisão (J48) utilizando as características comprimento da sépala (x_1) e largura da sépala (x_2).	67
5.3	Resultados obtidos da classificação polinomial utilizando as características comprimento da sépala (x_1) e comprimento da pétala (x_3).	69
5.4	Resultados obtidos pelos classificadores SVM e árvore de decisão (J48) utilizando as características comprimento da sépala (x_1) e comprimento da pétala (x_3).	70
5.5	Resultados obtidos da classificação polinomial utilizando as características comprimento da sépala (x_1) e largura da pétala (x_4).	73
5.6	Resultados obtidos pelos classificadores SVM e árvore de decisão (J48) utilizando as características comprimento da sépala (x_1) e largura da pétala (x_4).	75
5.7	Resultados obtidos da classificação polinomial com as características largura da sépala (x_2) e comprimento da pétala (x_3).	77
5.8	Resultados obtidos pelos classificadores SVM e árvore de decisão (J48) utilizando as características largura da sépala (x_2) e comprimento da pétala (x_3).	79
5.9	Resultados obtidos da classificação polinomial utilizando as características largura da sépala (x_2) e largura da pétala (x_4).	81
5.10	Resultados obtidos pelos classificadores SVM e árvore de decisão (J48) utilizando as características largura da sépala (x_2) e largura da pétala (x_4).	82

5.11	Resultados obtidos da classificação polinomial com as características comprimento da pétala (x_3) e largura da pétala (x_4).	85
5.12	Resultados obtidos pelos classificadores SVM e árvore de decisão (J48) utilizando as características comprimento da pétala (x_3) e largura da pétala (x_4).	87
5.13	Resumo dos resultados obtidos da classificação polinomial utilizando duas características: comprimento da sépala (x_1), largura da sépala (x_2), comprimento da pétala (x_3) e largura da pétala (x_4).	88
5.14	Resultados obtidos da classificação polinomial utilizando as características comprimento da sépala (x_1), largura da sépala (x_2) e comprimento da pétala (x_3).	89
5.15	Resultados obtidos pelos classificadores SVM e árvore de decisão (J48) utilizando as características comprimento da sépala (x_1), largura da sépala (x_2) e comprimento da pétala (x_3).	90
5.16	Resultados obtidos da classificação polinomial utilizando as características comprimento da sépala (x_1), largura da sépala (x_2) e largura da pétala (x_4).	92
5.17	Resultados obtidos pelos classificadores SVM e árvore de decisão (J48) utilizando as características comprimento da sépala (x_1), largura da sépala (x_2) e largura da pétala (x_4).	93
5.18	Resultados obtidos da classificação polinomial utilizando as características comprimento da sépala (x_1), comprimento da pétala (x_3) e largura da pétala (x_4).	95
5.19	Resultados obtidos pelos classificadores SVM e árvore de decisão (J48) utilizando as características comprimento da sépala (x_1), comprimento da pétala (x_3) e largura da pétala (x_4).	96
5.20	Resultados obtidos da classificação polinomial utilizando as características largura da sépala (x_2), comprimento da pétala (x_3) e largura da pétala (x_4).	98
5.21	Resultados obtidos pelos classificadores SVM e árvore de decisão (J48) utilizando as características largura da sépala (x_2), comprimento da pétala (x_3) e largura da pétala (x_4).	99
5.22	Resultados obtidos da classificação polinomial utilizando as características comprimento da sépala (x_1), largura da sépala (x_2), comprimento da pétala (x_3) e largura da pétala (x_4).	101

5.23	Resultados obtidos pelos classificadores SVM e árvore de decisão (J48) utilizando as características comprimento da sépala (x_1), largura da sépala (x_2), comprimento da pétala (x_3) e largura da pétala (x_4).	102
5.24	Resultados obtidos nos testes realizados com as características comprimento da sépala (x_1), largura da sépala (x_2), comprimento da pétala (x_3) e largura da pétala (x_4) utilizando o método de validação cruzada.	105
5.25	Resultados obtidos pelos classificadores polinomial (CP), SVM e árvore de decisão (J48) utilizando a característica dimensão fractal.	110
5.26	Resultados obtidos pelos classificadores polinomial (CP), SVM e árvore de decisão (J48) utilizando a característica lacunaridade.	114
5.27	Resultados obtidos pelos classificadores polinomial (CP), SVM e árvore de decisão (J48) utilizando as características dimensão fractal e lacunaridade.	117

Lista de Abreviaturas

SVM	<i>Support Vector Machine</i>
WEKA	<i>Waikato Environment for Knowledge Analysis</i>
CADe	<i>Computer-Aided Detection</i>
CADx	<i>Computer-Aided Diagnosis</i>
ROIs	<i>Regions of Interest</i>
OMS	<i>Organização Mundial de Saúde</i>
K-NN	<i>K Nearest Neighbours</i>
PDI	<i>Processamento Digital de Imagens</i>
D	<i>Dimensão Fractal</i>
OBCT	<i>Ordinary Binary Classification Tree</i>
ODT	<i>Oblique Decision Tree</i>
TDIDT	<i>Top-Down Induction of Decision Tree</i>
CART	<i>Classification and Regression Trees</i>
L	<i>Lacunaridade</i>
DDSM	<i>Digital Database for Screening Mammography</i>
MGH	<i>Massachusetts General Hospital</i>
WFUSM	<i>Wake Forest University School of Medicine</i>
SHH	<i>Sacred Heart Hospital</i>
WU	<i>Washington University</i>

CC	<i>Crânio-Caudal</i>
MLO	<i>Médio-Lateral Oblíqua</i>
NM	<i>Nódulos Mamários</i>
ES	<i>Estruturas Sadias</i>
NMS	<i>Nódulos Malignos Segmentados</i>
PP	<i>Pixels Parciais</i>
CP	<i>Classificador Polinomial</i>

Capítulo 1

Introdução

1.1 Introdução

O processo de reconhecimento de padrões faz parte do estudo da percepção, pois é por meio da percepção que o ser humano estabelece seu relacionamento com o mundo e desenvolve sua inteligência. O ser humano possui a capacidade de reconhecer e classificar padrões que o leva a ter uma percepção única do mundo. Quanto mais o ser humano experimenta, mais ele é capaz de compreender e classificar o que reconheceu em estruturas de acordo com a similaridade de suas características e padrões [1].

Como ciência, o reconhecimento de padrões vem sendo estudado desde o início de 1950, devido ao avanço da tecnologia que levou ao advento dos computadores. Ele pode ser definido como “o processo de identificar estruturas em dados por meio de comparações com as estruturas conhecidas” [2].

Diversos métodos de aprendizado são utilizados no estudo de reconhecimento de padrões. Estes métodos podem ser divididos em supervisionados, não-supervisionados e semi-supervisionados, de acordo com seus algoritmos de aprendizado [3]. Os métodos não-supervisionados não possuem conhecimento algum sobre as classes das amostras de treinamento, enquanto essa informação é conhecida em métodos supervisionados. Nos métodos semi-supervisionados são conhecidos apenas parte dos rótulos das amostras de treinamento

[4].

Como exemplo, considere um conjunto de amostras que pode ser particionado em c classes. A finalidade da classificação de padrões é encontrar a classe de cada amostra usando um vetor de características.

Os métodos de classificação apresentados nesta tese se baseiam em métricas de distância que associam os dados por meio de sua disposição espacial. Esses métodos partem do princípio que os dados que estão próximos pertencem a uma mesma classe, enquanto os dados distantes pertencem a classes diferentes. Neste caso, os dados são projetados no espaço \mathbb{R}^d a partir de suas d características, que correspondem às suas coordenadas nesse espaço.

O classificador polinomial desenvolvido neste trabalho é um método de classificação supervisionado e pode classificar em duas ou mais classes. Esse classificador expande o vetor de características projetado no espaço \mathbb{R}^d para um espaço de dimensão superior onde é possível a classificação. O classificador polinomial mostra ser um importante método de classificação principalmente em tratamento com classes não linearmente separáveis. Sua estrutura é simples e de fácil implementação, pois os polinômios são compostos por operações matemáticas simples tais como soma e produto.

Neste trabalho são realizados vários testes utilizando o classificador polinomial na:

- classificação da flores setosa, versicolor e virginica usando a base de dados IRIS, introduzida em 1936 por R. A. Fisher [5]; e
- identificação de pixels parciais em regiões de interesse de imagens mamográficas.

A base de dados IRIS, introduzida em 1936 por R. A. Fisher [5] talvez seja o banco de dados mais utilizado na literatura em reconhecimento de padrões. Neste trabalho utilizou-se essa base para realizar testes usando o classificador polinomial. Os resultados obtidos utilizando esse classificador são comparados com as técnicas de classificação máquina de vetor suporte (SVM) e árvore de decisão usando o algoritmo J48.

A árvore de decisão é uma técnica muito poderosa e muito utilizada em problemas de classificação [6]. Ela é diferente do SVM, que é um modelo de classificação que apresenta

grande eficiência em trabalhar com dados de alta dimensionalidade e é uma técnica altamente robusta ao ruído [7].

Os resultados das técnicas SVM e árvore de decisão foram encontradas usando a ferramenta Waikato Environment for Knowledge Analysis (WEKA) [8] e o computador pessoal conforme a configuração mostrada na Tabela 1.1.

Tabela 1.1: Recursos computacionais utilizados neste trabalho.

Processador	Intel i7
Memória RAM	8 Gb
Disco rígido	750 Gb
Ferramenta de programação	MATLAB R2011a

A textura é uma representação que se refere à repetição de elementos básicos chamados texels. O texel contém vários pixels, cuja colocação pode ser aleatória ou periódica. As texturas naturais são geralmente aleatórias, ao passo que as texturas artificiais são frequentemente determinísticas ou periódicas. A textura pode ser áspera, fina, suave, granulada, ondulada, irregular, regular ou linear [9].

A extração de características utilizando as informações de textura vem sendo usada na análise e na interpretação de imagens médicas. Essas informações podem ser obtidas por meio de uma variedade de métodos de processamento de imagens, incluindo medidas estatísticas obtidas por meio de uma matriz de co-ocorrência ou por meio de técnicas de multiresolução como a transformada wavelet. Em mamografia, as informações de textura permitem distinguir os padrões de densidade na análise das classes mais comuns de lesões mamográficas [10].

O aperfeiçoamento dos sistemas computacionais tem beneficiado o desenvolvimento de muitas áreas de pesquisas na medicina. Então, uma das alternativas adotadas para a obtenção de uma segunda leitura ou uma estratégia de “pré-leitura” mais utilizada atualmente é a aplicação de sistemas computacionais de apoio à detecção, o *Computer-Aided Detection* (CAdE) e ao diagnóstico, o *Computer-Aided Diagnosis* (CAdx) [10]. Esses sistemas vêm sendo desenvolvidos com a finalidade de aumentar a quantidade de detalhes reconhecidos em

uma imagem mamográfica. Esses detalhes são construídos com métodos de processamento de imagens, redes neurais, lógica fuzzy e outras abordagens, com o objetivo de realçar, segmentar, extrair características e classificar possíveis anormalidades [11, 12, 13].

A interpretação e a análise de imagens médicas representam uma parte importante em visão computacional e no reconhecimento de padrões. Está se tornando essencial para muitos pesquisadores e centros clínicos o desenvolvimento de um sistema de diagnóstico auxiliado por computador para doenças como câncer de mama, com a finalidade de auxiliar médicos em hospitais.

Apesar dos sistemas CAD terem proporcionado um grande número de pesquisas e aumentado a sensibilidade, a grande maioria desses sistemas analisa as visões crânio-caudal (CC) e médio-lateral oblíqua (MLO) de forma independente. Em algumas situações, o sistema detecta as anormalidades em apenas uma das visões. Os especialistas consideram que existe uma inconsistência se uma determinada lesão é semelhante em ambas as visões e o sistema não possui a capacidade de encontrá-la. Com isso, existe uma redução na confiabilidade e nos benefícios oferecidos pelo sistema. Estudos vem mostrando que essas limitações têm provocado mudanças e os especialistas estão ignorando os resultados proporcionados por esses sistemas [14]. Portanto, torna-se necessário um estudo teórico e prático para o desenvolvimento de novas técnicas que possam extrair as informações relevantes das regiões de interesses do inglês *Regions of Interest* (ROIs), assim como um estudo dos métodos de seleção e classificação para melhorar as análises dos tecidos normais e com anormalidades provocadas por nódulo mamário [10].

Segundo a Organização Mundial de Saúde (OMS), o câncer é uma das principais causas de morte no mundo, com uma projeção de mais de 11 milhões de mortes até 2030 [15]. O câncer de mama é um dos tipos de câncer mais comuns. Ele é o segundo tipo de câncer com maior incidência no mundo em mulheres com idade superior a 35 anos [16]. A causa do câncer de mama permanece incerta, porém alterações genéticas, influências hormonais e variáveis ambientais são fatores que colaboram para o desenvolvimento dessa doença.

A detecção precoce do câncer de mama é fundamental para um prognóstico melhor. A mamografia é o principal método para diagnosticar esse tipo de câncer em estágio inicial.

Ela é capaz de detectar alterações ainda não palpáveis, mas existem controvérsias sobre sua efetividade em mulheres abaixo de 50 anos de idade, pois a mama é mais densa nas mulheres mais jovens [17]. Ao envelhecer o tecido mamário fibroso é substituído por tecido adiposo e o rastreio se torna mais sensível [18]. Outro complicador é a superposição de estruturas e efeitos físicos presentes no processo de aquisição das imagens influenciando a exatidão e a confiabilidade do diagnóstico. Com isso, entre 10 a 30% das anormalidades do tipo maligno visíveis em mamogramas não são detectadas pelos especialistas durante os exames de rotina [19].

Para minimizar essa situação, existem propostas de modelos ou técnicas computacionais para apoiar o diagnóstico de câncer de mama com quantificações realizadas em ROIs [20]. No entanto, a separação e a quantificação de estruturas parciais (tecidos entre os estados sadio e doente) é um processo importante para a análise de ROIs e pode garantir informações mais precisas para a distinção e a compreensão dos diferentes estágios das estruturas investigadas. Esse procedimento é ignorado conforme o estudo apresentado em [21].

Neste trabalho, os comportamentos das estruturas mamárias sadias e doentes são quantificados com dimensão fractal e lacunaridade multiescalas, calculados por meio de assinaturas das texturas [8-12]. A separação de pixels parciais é obtida com a aplicação de um método de segmentação multinível baseado em entropia máxima, permitindo quantificações e comparações entre os grupos estudados. O classificador polinomial, método proposto nesta tese, faz a separabilidade dos grupos mostrando que os pixels parciais não devem ser incluídos em quantificações de estruturas doentes. Também é realizada uma comparação dos resultados obtidos por esse classificador com os classificadores SVM e árvore de decisão.

Este capítulo mostra uma breve introdução do sistema de visão computacional relacionando a classificação de padrões, principalmente para imagens médicas. Depois são apresentados os objetivos e a estrutura deste trabalho.

1.2 Objetivos Deste Trabalho

O objetivo deste trabalho é desenvolver um classificador polinomial de grau n , $n \geq 1$ que expande o vetor de características projetado no espaço \mathbb{R}^d para um espaço de dimensão superior onde é possível a classificação de duas ou mais classes. Esse classificador mapeia um conjunto de amostras de treinamento e classifica as amostras previamente identificadas.

Os objetivos específicos deste trabalho são:

- mostrar como o classificador polinomial se comporta ao variar o número de características e o grau da base polinomial;
- analisar o classificador polinomial como um método de otimização utilizado na seleção das melhores características;
- identificar como o classificador polinomial se comporta com as classes linearmente separáveis e não linearmente separáveis;
- verificar o desempenho do classificador polinomial utilizando a base de dados IRIS, introduzida em 1936 por R. A. Fisher [5] e comparar os resultados obtidos usando essa base com os classificadores SVM e árvore de decisão; e
- usar o classificador polinomial para classificar as mamografias digitalizadas por meio da aplicação de técnicas de processamento de imagens em regiões suspeitas que caracterizam a textura subjacente das regiões da mama.

1.3 Estrutura Deste Trabalho

Este trabalho consiste de seis capítulos.

O primeiro capítulo mostra uma introdução do sistema de visão computacional relacionado a classificação de padrões, principalmente para imagens médicas. Depois são apresentados os objetivos e a estrutura deste trabalho.

O Capítulo 2 apresenta um breve estudo sobre o sistema de processamento digital de imagens, mostrando um resumo de todas as etapas desse sistema. É realizada uma apresentação mais detalhada de alguns dos principais descritores usados em processamento digital de imagens. Estes descritores são importantes neste trabalho para a utilização das características na separação das classes determinadas pelo classificador polinomial.

O Capítulo 3 apresenta os padrões e as classes de padrões. Depois são descritos os classificadores: funções discriminantes, K vizinhos mais próximos (K-NN), discriminante linear de Fisher, máquina de vetores suporte e árvores de decisão.

O Capítulo 4 descreve o classificador polinomial, método proposto neste trabalho, formalmente em sua estrutura e organização para classificar duas ou mais classes .

O Capítulo 5 mostra os resultados obtidos nos testes realizados utilizando o classificador polinomial na classificação da base de dados IRIS [5] e na identificação de pixels parciais em ROIs de imagens mamográficas. Com o objetivo de comparar os resultados obtidos com esse classificador foram realizados também testes com os classificadores máquina de vetores suporte e árvore de decisão utilizando o algoritmo J48.

O Capítulo 6 apresenta as conclusões e as contribuições deste trabalho, os trabalhos futuros que poderão ser desenvolvidos a partir desta tese e a publicação deste trabalho.

1.4 Considerações Finais Deste Capítulo

Este capítulo mostrou uma breve introdução do sistema de visão computacional relacionado a classificação de padrões, principalmente para imagens médicas. Depois foram apresentados os objetivos e a estrutura deste trabalho.

O próximo capítulo apresenta um breve estudo sobre o sistema de processamento digital de imagens, mostrando um resumo de todas as etapas desse sistema. É realizada uma apresentação mais detalhada de alguns dos principais descritores usados em processamento digital de imagens. Estes descritores são importantes neste trabalho para a utilização das características na separação das classes determinadas pelo classificador polinomial.

Capítulo 2

Fundamentos de Processamento Digital de Imagens, Extração de Características, Dimensão Fractal e Lacunaridade

2.1 Introdução

Desde 1964 até o presente momento, a área de processamento de imagens vem apresentando crescimento expressivo e com aplicações em quase todos os ramos da atividade humana. Na Medicina, a utilização de imagens no diagnóstico médico tornou-se rotineira e os avanços em processamento de imagens vêm permitindo tanto o desenvolvimento de novos equipamentos quanto a maior facilidade de interpretação de imagens produzidas por equipamentos mais antigos, como por exemplo, o de raio X. Na Biologia, a capacidade de processar automaticamente imagens obtidas de microscópios facilita a execução de tarefas laboratoriais com alto grau de precisão. O processamento e a interpretação automática de imagens captadas por satélites auxiliam os trabalhos nas áreas de Geografia, Sensoriamento Remoto, Geoprocessamento e Meteorologia, dentre outras. Técnicas de restauração de imagens auxiliam os arqueologistas a recuperar fotos borradas de artefatos raros já destruídos. O uso de robôs com visão artificial em tarefas tais como controle de qualidade em linhas de

produção aumenta a cada ano, em um cenário de crescente automação industrial. Inúmeras outras áreas tão distintas como Astronomia, Segurança, Publicidade e Direito vêm sendo beneficiadas com os avanços nas áreas de processamento de imagens e visão por computador [22].

Este capítulo apresenta um breve estudo sobre o sistema de processamento digital de imagens, mostrando um resumo de todas as etapas desse sistema. É realizada uma apresentação mais detalhada de alguns dos principais descritores em processamento digital de imagens. Estes descritores são importantes neste trabalho para a utilização das características na separação de classes determinadas pelo classificador polinomial.

2.2 Fundamentos de Processamento Digital de Imagens

A Figura 2.1 mostra os passos fundamentais em processamento digital de imagens [23].

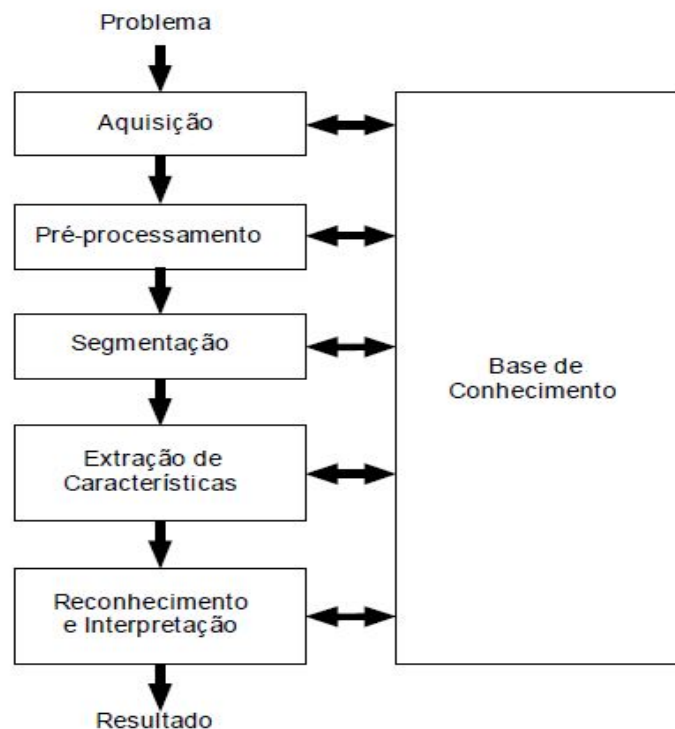


Figura 2.1: Passos fundamentais em processamento digital de imagens [23].

O primeiro passo em processamento digital de imagens (PDI) é a aquisição da imagem. Para isso são necessários um sensor que converte a informação óptica em sinal elétrico e um digitalizador que transforma a imagem analógica em imagem digital. Nessa etapa pode-se destacar vários aspectos utilizados para melhorar a aquisição da imagem, tais como: a escolha do tipo de sensor, o conjunto de lentes a utilizar, as condições de iluminação da cena, os requisitos de velocidade de aquisição, a resolução e o número de níveis de cinza da imagem digitalizada, dentre outros.

Após a obtenção de uma imagem digital, o próximo passo em PDI é pré-processá-la com o objetivo de melhorá-la para aumentar as possibilidades de bons resultados nos próximos passos. A fase de pré-processamento inclui técnicas para o realce de contraste, remoção de ruído e isolamento de regiões. As operações efetuadas nessa fase são consideradas de baixo nível porque trabalham diretamente com os valores de intensidade dos pixels. A imagem resultante do pré-processamento é uma imagem digitalizada de melhor qualidade do que a original.

O próximo passo em processamento digital de imagens é a segmentação. Esta é um processo de dividir uma imagem digital em múltiplas regiões ou objetos, com o objetivo de simplificar a representação de uma imagem para facilitar a sua análise. Um procedimento de segmentação robusto favorece a solução de forma satisfatória de um problema de imageamento. Por outro lado, algoritmos de segmentação fracos ou irregulares quase sempre resultam em falhas no processamento.

A saída do estágio de segmentação é constituída tipicamente por dados em forma de pixels, correspondendo tanto a fronteira de uma região como a todos os pontos dentro da mesma. Em ambos os casos é necessário converter os dados para uma forma adequada ao processamento computacional. A primeira decisão que precisa ser realizada é se os dados devem ser representados como fronteiras ou como regiões completas. A representação por fronteira é adequada quando o interesse se concentra nas características da forma externa, tais como cantos ou pontos de inflexão. A representação por região é adequada quando o interesse está em propriedades internas, tais como textura ou esqueleto.

O processo de descrição em PDI, também chamado de seleção de características, pro-

cura extrair características que resultem em alguma informação quantitativa de interesse ou que sejam básicas para a discriminação entre as classes de objetos. É importante observar que nesse passo a entrada ainda é uma imagem, mas a saída é um conjunto de dados correspondentes àquela imagem.

O último passo em processamento digital de imagens consiste no reconhecimento e na interpretação. O reconhecimento é o processo que atribui um rótulo ao objeto, baseado na informação fornecida pelo seu descritor. A interpretação consiste na atribuição de significado a um conjunto de objetos reconhecidos.

Todos os passos descritos acima pressupõem a existência de um conhecimento sobre o problema a ser resolvido, armazenado em uma base de conhecimento, cujos tamanho e complexidade podem variar. Além de direcionar a operação de cada etapa de processamento, essa base também controla a interação entre os passos [23].

2.3 Extração de Características

A extração de características é um processo de determinação de um vetor \mathbf{x} representativo dos componentes principais de uma imagem, denominado vetor de características com dimensão d , ou seja, $\mathbf{x} \in \mathbb{R}^d$. As características são valores extraídos de uma imagem usando técnicas específicas.

O vetor de características deve ser o mais descritivo possível da imagem para distinguir classes de imagens entre si. A Figura 2.2 ilustra um esquema de extração de características em imagens radiográficas de pé e de mão.

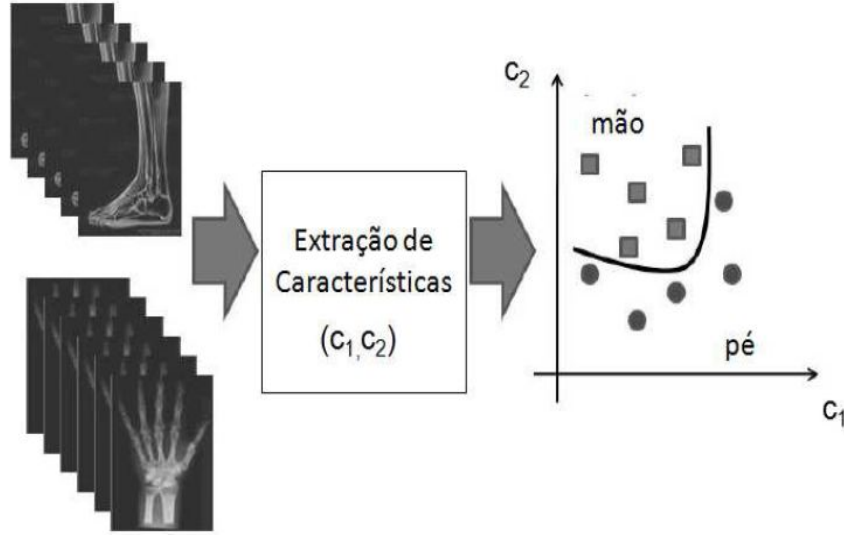


Figura 2.2: Ilustração esquemática da extração de características em imagens de pé e de mão [24].

Em geral, as características comuns extraídas de uma imagem são: cor, textura e forma [25]. Essas características fornecem bons resultados em aplicações específicas, como classificação de imagens de mama em patológico ou não patológico [26].

Em imagens médicas é comum existirem problemas com: contraste, escala, rotação e translação. Portanto, é necessário pensar em técnicas que possam extrair características que sejam invariantes a contraste, escala, rotação e translação.

2.4 Dimensão Fractal Multiescala

A dimensão fractal pode ser definida como uma medida da complexidade dos objetos. Aplicada às texturas, ela permite quantificar a complexidade da organização de seus pixels, onde esse nível de complexidade está diretamente relacionado com o aspecto visual e com a homogeneidade da textura. Assim, essa dimensão permite quantificar uma textura em termos da homogeneidade, possibilitando sua comparação com outras texturas [27].

Atualmente a dimensão fractal vem sendo utilizada em diversas áreas de conhecimento, tais como: estudo de sistemas caóticos; caracterização dos objetos; análise e reconhecimento dos padrões em imagens; análise de texturas e medição do comprimento de curvas

[28, 29, 30, 31].

Fractais são formas geométricas incapazes de serem classificadas na geometria Euclidiana devido principalmente a três características fundamentais que os definem e distinguem de outras formas: auto-semelhança em diferentes níveis de escala, dimensão fractal e sua complexidade infinita [32].

Das características que definem um fractal, a mais importante é a dimensão fractal. Ao contrário do que é observado na Geometria Euclidiana, onde o valor da dimensão representa a dimensionalidade do espaço em que determinado objeto está inserido, a dimensão fractal (ou dimensão dos fractais) representa o nível de irregularidade de um fractal [33].

Essa diferença nas definições faz com que a dimensão fractal venha a assumir valores fracionários, algo diferente dos valores inteiros encontrados para as dimensões Euclidianas. O fato da dimensão fractal ser um valor fracionário é porque ela representa o nível de ocupação do espaço pela forma e não o espaço em si onde a forma está inserida. Sendo assim, quanto maior é a irregularidade de uma forma, maior é a sua dimensão fractal. Essa característica da dimensão a torna uma ferramenta muito útil para a comparação de duas formas fractais [33].

A literatura fornece diversas abordagens para se estimar a dimensão fractal de um objeto ou uma imagem. No entanto, a grande maioria delas se baseia na dimensão de Hausdorff.

Para entender a dimensão de Hausdorff, considere uma linha de comprimento L e outra de comprimento u , onde $L > u$. Sobrepondo a linha u sobre a linha L até cobri-la completamente, encontra-se um valor $N = \frac{L}{u}$, que nada mais é do que uma medida da linha em unidades de u . Do mesmo modo que foi realizado para a linha, pode-se medir um quadrado de lado L cobrindo-o com pequenos quadrados de lado u , obtendo-se a mesma relação $N = \left(\frac{L}{u}\right)^2$ [34, 35].

A continuidade desse processo de cobertura de objetos por linhas, quadrados, cubos e etc, resulta em uma relação como mostrado na Equação (2.1)

$$N = \left(\frac{L}{u}\right)^D. \quad (2.1)$$

onde:

u - unidade de medida.

Aplicando-se o logaritmo em ambos os lados na Equação (2.1), obtém-se a Equação (2.2)

$$D = \frac{\ln N}{\ln (L/u)} \quad (2.2)$$

onde:

D - dimensão fractal de Hausdorff do objeto analisado.

Para um objeto uniforme e compacto, D é um inteiro igual à dimensão topológica. Mas, para um fractal, D pode ser um número fracionário [34, 35].

Vários são os métodos de estimativa da dimensão fractal existentes na literatura atual. No entanto, nem todos os métodos podem ser aplicados a qualquer tipo de estrutura. Isso ocorre porque cada método se baseia em um tipo de medição diferente do objeto analisado, gerando resultados diferentes para um mesmo objeto.

A seguir são descritos alguns dos principais métodos de estimativa da dimensão fractal que pode-se encontrar na literatura.

2.4.1 Massa-raio

O método de cálculo da massa-raio consiste em sobrepor círculos sobre uma imagem e contar a quantidade de pontos dessa imagem (pixels) presentes no interior do círculo. Esse método se baseia na Equação (2.3)

$$M_r(A) = \mu r^D \quad (2.3)$$

onde:

$M_r(A)$ - número de pontos de uma imagem A (pixels) presentes em um círculo de raio r ; e

μ - constante.

Da Equação (2.3) obtém-se a dimensão fractal como mostrado na Equação (2.4)

$$D = \lim_{r \rightarrow 0} \frac{\ln(M_r(A))}{\ln(r)}. \quad (2.4)$$

Um ou mais círculos podem ser utilizados no cálculo de D . Se usar apenas um círculo, é melhor escolher o centro do círculo como sendo o centro de massa da imagem analisada. Utilizando-se vários círculos, um mecanismo de sorteio pode ser usado para se escolher os diferentes centros dos círculos. Também, pode-se considerar a massa média calculada desses vários círculos.

Outra análise da massa-raio refere-se ao raio máximo (r_{max}) que é utilizado na expansão dos círculos. Os raios r usados nesse método variam então entre $1 < r < r_{max}$, onde a cada nova iteração o raio é incrementado de 1 pixel e o cálculo da massa é realizado novamente [36]. A quantificação para diferentes valores de r caracteriza a abordagem multiescala.

Traçando-se o gráfico de log-log entre a massa acumulada para um raio r ($M_r(A)$) e o raio do círculo utilizado (r), obtém-se a aproximação de uma reta com coeficiente angular α , onde $D = \alpha$ é a dimensão fractal de A [34, 37].

2.4.2 Análise por intersecção acumulativa

A análise por intersecção acumulativa é um método similar ao método massa-raio, no entanto, ao invés de calcular a massa de uma região esse método calcula o número acumulado de intersecções de segmentos de linha da imagem com círculos de raios crescentes ou subdivisões existentes nessa região [36]. O método de análise por intersecção acumulativa baseia-se na Equação (2.5)

$$I_c(A) = \mu r^D \quad (2.5)$$

onde:

$I_c(A)$ - número de intersecções existentes na imagem A quando sobreposta por um círculo de raio r .

Da Equação (2.5), obtém-se a dimensão fractal (D) como mostrado na Equação (2.6)

$$D = \lim_{r \rightarrow 0} \frac{\ln(I_c(A))}{\ln(r)}. \quad (2.6)$$

Como no método massa-raio, na análise da intersecção acumulativa também pode-se utilizar um ou mais círculos, desde que sejam observadas as mesmas restrições impostas no método massa-raio.

Traçando-se o gráfico de log-log entre o número de intersecções para um raio r ($I_c(A)$) e o raio do círculo utilizado (r) obtém-se a aproximação de uma reta com coeficiente angular α , onde $D = \alpha$ é a dimensão fractal de A [34, 37].

2.4.3 *Boxcounting*

O *boxcounting* é um dos métodos mais conhecidos e utilizados para estimar a dimensão fractal de uma imagem. Esse método consiste em sobrepor à imagem uma malha de quadrados e contar o número de quadrados necessários para cobrir toda a imagem [38]. O *boxcounting* se baseia na Equação (2.7)

$$N_r(A) = \mu r^{-D} \quad (2.7)$$

onde:

$N_r(A)$ - número de quadrados de lado r que contém toda a imagem A .

A partir da Equação (2.7), obtém-se a dimensão fractal (D) pela Equação (2.8)

$$D = -\lim_{r \rightarrow 0} \frac{\ln(N_r(A))}{\ln(r)}. \quad (2.8)$$

Para a execução do método *boxcounting* é necessário definir um conjunto B de tamanhos de lados r a ser utilizado nas diversas iterações desse método. Mesmo que B possa ser definido inicialmente por um usuário é padrão calculá-lo com base nas dimensões da imagem, conforme a Equação (2.9)

$$\forall r_i \in B \begin{cases} r_0 &= \max(\text{altura}, \text{largura}) \\ r_{i+1} &= \frac{r_i}{2} \end{cases} \quad (2.9)$$

Traçando-se um gráfico log-log do número de caixas ocupadas ($N_r(A)$) pelo tamanho do lado do quadrado (r), obtém-se a aproximação de uma reta com coeficiente angular α . Deste modo, é possível definir $D = -\alpha$ como dimensão fractal de A [34, 35].

2.5 Lacunaridade Multiescala

A lacunaridade é uma medida de como os pixels da imagem estão distribuídos e organizados em uma determinada região da imagem, ou seja, ela quantifica como o espaço está preenchido. A lacunaridade é diferente da dimensão fractal, que mede o quão preenchido está o espaço.

A lacunaridade é obtida medindo-se a distribuição espacial dos “gaps” ou buracos existentes na imagem. Por meio dela é possível quantificar o quão homogênea é a textura dessa imagem, de maneira que se possa compará-la com outras texturas [39, 40].

O algoritmo do *gliding-box* é similar ao algoritmo do *boxcounting* utilizado para estimar a dimensão fractal. Nele, uma caixa de lado r é colocada sobre o canto superior esquerdo da imagem e o número de pixels da imagem é contado. Esse processo é repetido para todas as linhas e colunas da imagem, resultando em uma distribuição de frequência da massa da imagem. O número de caixas de lado r contendo uma massa M da imagem é designado por $n(M, r)$ e o total de caixas contadas por $N(r)$. Então, essa distribuição de frequência é convertida para uma distribuição de probabilidade $F(M, r)$, como mostrado na Equação (2.10)

$$F(M, r) = \frac{n(M, r)}{N(r)}. \quad (2.10)$$

O primeiro $A^{(1)}$ e o segundo $B^{(2)}$ momentos da distribuição de probabilidade obtidos pela Equação (2.10) são determinados pela Equação (2.11)

$$\begin{aligned} A^{(1)} &= \sum M F(M, r) \\ B^{(2)} &= \sum M^2 F(M, r). \end{aligned} \quad (2.11)$$

A lacunaridade para uma caixa de tamanho r é obtida utilizando-se a Equação (2.12)

$$L(r) = \frac{B^{(2)}}{(A^{(1)})^2}. \quad (2.12)$$

Considerando-se a mesma abordagem multiescala para a dimensão fractal, a contagem $N(r)$ para L também pode ser realizada para diferentes valores de r , onde o r inicial (r_0) é

considerado o valor da maior dimensão da imagem, ou seja, $r_0 = \max(\text{altura}, \text{largura})$; e, nas demais iterações, o tamanho de r é reduzido sucessivamente pela metade, ou seja, $r_{i+1} = \frac{r_i}{2}$, [41].

O gráfico log-log da lacunaridade ($L(r)$) pelo tamanho do lado do quadrado (r), obtém-se a aproximação de uma reta com coeficiente angular ϕ . Desse modo, é possível definir $L = \phi$ como a lacunaridade multiescala da imagem.

Outras características relativas a lacunaridade podem ser obtidas alterando-se o tamanho da caixa utilizada no *gliding-box* [40].

2.6 Considerações Finais Deste Capítulo

Este capítulo apresentou um breve estudo sobre o sistema de processamento digital de imagens, mostrando um resumo de todas as etapas desse sistema. Foi realizada uma apresentação mais detalhada de alguns dos principais descritores usados em processamento digital de imagens. Estes descritores são importantes neste trabalho para a utilização das características na separação de classes determinadas pelo classificador polinomial.

O próximo capítulo apresenta os padrões e as classes de padrões. Depois são descritos os classificadores: funções discriminantes, K vizinhos mais próximos (K-NN), discriminante linear de Fisher, máquina de vetores suporte do inglês *Support Vector Machine* (SVM) e árvores de decisão.

Capítulo 3

Classificadores

3.1 Introdução

A análise discriminante e a classificação são técnicas multivariadas que procuram separar conjuntos distintos (classes) e identificar novos objetos de classes previamente definidas.

Considere um conjunto de amostras (pixels, voxels, contornos, regiões, imagens), que pode ser particionado em c classes (rótulos); o objetivo da classificação de padrões é encontrar a classe de cada amostra utilizando um vetor de atributos (características). Os métodos normalmente aprendem como as amostras se distribuem em diferentes classes no espaço de características, usando um subconjunto de treinamento e encontrando as superfícies ou as regras de decisão para o particionamento desse conjunto [42].

A classificação automática tem sido um grande desafio em uma variedade de problemas científicos e comerciais. De acordo com os algoritmos de aprendizagem, os métodos de aprendizagem podem ser divididos em supervisionados, não-supervisionados e semi-supervisionados [3].

O reconhecimento por decisões teóricas baseia-se na representação dos padrões na forma de um vetor, seguida da busca por abordagens para o agrupamento e para a atribuição dos vetores de padrões e classes de padrões. As principais abordagens para o reconhecimento por decisões teóricas são os classificadores por distância mínima, classificadores bayesianos

e redes neurais [23].

Este capítulo apresenta os padrões e as classes de padrões. Depois são descritos os classificadores: funções discriminantes, K vizinhos mais próximos (K-NN), discriminante linear de Fisher, máquina de vetores suporte e árvores de decisão.

3.2 Padrões e Classes de Padrões

Cada padrão ou objeto é representado por um vetor de características ou vetor de atributos

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix}, \quad (3.1)$$

onde cada x_i , $i = 1, 2, \dots, d$, representa uma característica do padrão. O espaço formado pelos vetores de características é chamado espaço de características e d determina a dimensão desse espaço.

Uma classe ou categoria ω_i é um conjunto de padrões que possui alguma relação ou peculiaridade em comum. O conjunto formado por todas as classes denota-se $\Omega = \{\omega_1, \omega_2, \dots, \omega_c\}$, onde c é o número de classes.

O classificador toma decisões baseando-se no aprendizado realizado a partir de um conjunto de treinamento que contém exemplos de padrões de todas as classes existentes no sistema.

A tarefa do classificador é utilizar o vetor de características e atribuir o objeto a uma categoria. O grau de dificuldade da classificação depende da variabilidade nos valores das características para os objetos na mesma categoria, relativo a diferença entre os valores das características para os objetos em categorias diferentes. A variabilidade dos valores de características para os objetos em uma mesma categoria pode ser devido a complexidade ou ao ruído [42].

Os métodos de classificação são divididos em supervisionados, não-supervisionados

e semi-supervisionados. No aprendizado supervisionado assume-se possuir um conjunto de amostras de treinamento previamente identificado, onde a tarefa é construir uma função que irá prever corretamente as classes das amostras futuras. No aprendizado não-supervisionado a ideia é segmentar as amostras não previamente identificadas em grupos, de tal forma a agrupar os elementos similares e separar as amostras que não possuam atributos em comum. Na aprendizagem semi-supervisionada, assume-se ter tanto amostras previamente identificadas ou não no conjunto de treinamento. Desta forma, a tarefa é prever a classe dos elementos futuros utilizando-se esses dois tipos de amostras [4].

O classificador polinomial é o método proposto neste trabalho. Esse classificador é um método de classificação supervisionada.

3.3 Funções Discriminantes

Existem várias maneiras de representar os classificadores de padrões. Uma das maneiras mais simples de representá-los é por meio de funções discriminantes $g_i(\mathbf{x})$, $i = 1, \dots, c$, onde c representa o número de classes. O classificador informa que um vetor de características \mathbf{x} pertence a classe ω_i se

$$g_i(\mathbf{x}) > g_j(\mathbf{x}) \quad \forall \quad j \neq i. \quad (3.2)$$

Dessa forma o classificador é considerado como uma máquina ou rede que calcula c funções discriminantes e seleciona a classe correspondente a maior de todas as funções discriminantes [42].

A Figura 3.1 ilustra uma rede de um classificador de padrões estatístico que inclui d entradas e c funções discriminantes $g_i(\mathbf{x})$.

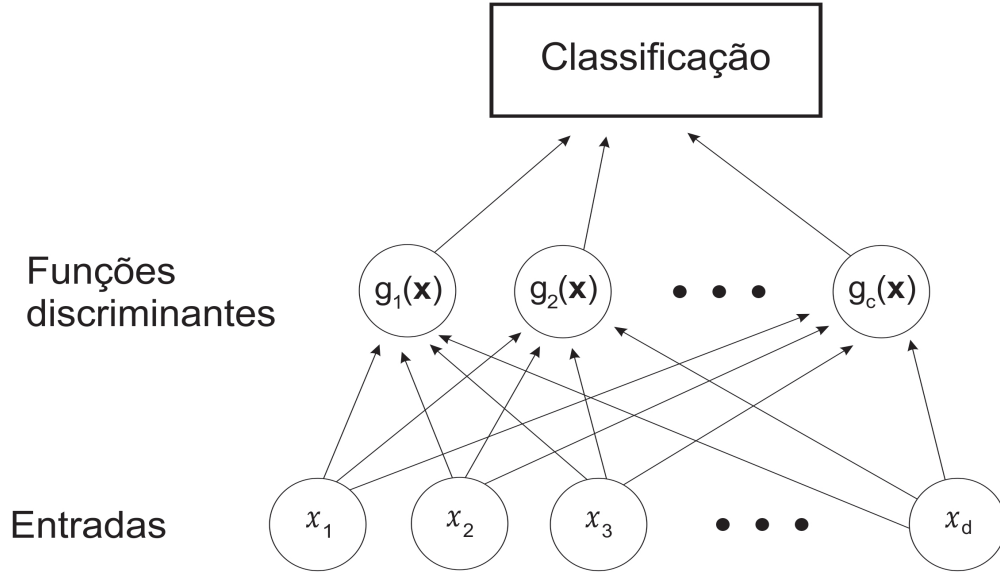


Figura 3.1: Estrutura funcional de um classificador de padrões estatístico.

Um classificador de Bayes pode ser representado pela estrutura da Figura 3.1. Como a escolha de uma função discriminante não é única, esse classificador pode ser representado pela função discriminante mostrada na Equação (3.3)

$$g_i(\mathbf{x}) = p(\mathbf{x}|\omega_i)P(\omega_i), \quad (3.3)$$

onde:

$p(\mathbf{x}|\omega_i)$ - função densidade de probabilidade de \mathbf{x} dado a classe ω_i ; e

$P(\omega_i)$ - probabilidade a priori da classe ω_i .

A regra de decisão para todas as formas de funções discriminantes é dividir o espaço de características em c regiões de decisões, R_1, \dots, R_c . Se $g_i(\mathbf{x}) > g_j(\mathbf{x})$ para todo $j \neq i$, então \mathbf{x} está em R_i e a regra de decisão atribui \mathbf{x} para a classe ω_i . As regiões de decisão são separadas por superfícies de decisão no espaço de características onde ocorrem os laços entre as maiores funções discriminantes.

A Figura 3.2 mostra um classificador de Bayes bidimensional para duas classes onde as funções de densidades de probabilidade são Gaussianas. A fronteira de decisão é constituída por duas hipérboles, onde a região de decisão R_2 não é simplesmente ligada.

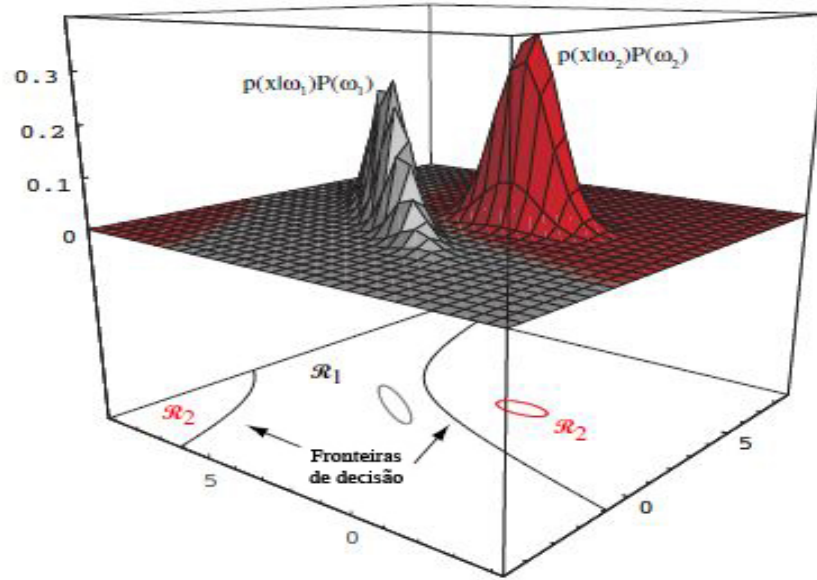


Figura 3.2: Representação das fronteiras de decisão para um classificador de Bayes bidimensional para duas classes onde as funções de densidades de probabilidade são Gaussianas [42].

3.3.1 Funções Discriminantes Lineares

Uma função discriminante linear é escrita como uma combinação linear dos componentes de \mathbf{x} como mostrado na Equação (3.4)

$$g(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + w_0, \quad (3.4)$$

onde:

\mathbf{w} - vetor peso; e

w_0 - bias ou limiar (*threshold*).

3.3.2 Classificação em duas classes

A classificação em duas classes recebe um tratamento especial em relação à classificação em várias classes. Segundo [42], um classificador que identifica um padrão como pertencente a uma das duas classes é denominado de *dicotomizador*. O classificador para mais de duas classes é apresentado logo a seguir neste capítulo.

No lugar de utilizar duas funções discriminantes $g_1(\mathbf{x})$ e $g_2(\mathbf{x})$ para designar \mathbf{x} à classe ω_1 se $g_1(\mathbf{x}) > g_2(\mathbf{x})$, é mais comum definir uma única função discriminante como mostrado na Equação (3.5)

$$g(\mathbf{x}) = g_1(\mathbf{x}) - g_2(\mathbf{x}) \quad (3.5)$$

e usar a seguinte regra de decisão: decide por ω_1 se $g(\mathbf{x}) > 0$; caso contrário decide por ω_2 .

O dicotomizador pode ser visto como uma máquina que calcula uma única função discriminante $g(\mathbf{x})$ e classifica \mathbf{x} de acordo com o sinal algébrico do resultado [42].

A equação $g(\mathbf{x}) = 0$ define a superfície de decisão que separa os padrões pertencentes à classe ω_1 dos padrões pertencentes à classe ω_2 . Quando $g(\mathbf{x})$ é linear, a superfície de decisão é um hiperplano. A orientação e a localização da superfície são determinadas pelo vetor normal \mathbf{w} e pelo bias w_0 , respectivamente. A função discriminante $g(\mathbf{x})$ é proporcional à distância de \mathbf{x} ao hiperplano com $g(\mathbf{x}) > 0$, quando \mathbf{x} está do lado positivo, e $g(\mathbf{x}) < 0$, quando \mathbf{x} está do lado negativo [42].

3.3.3 Classificação em várias classes

Existe mais de uma maneira de utilizar as funções discriminantes lineares na classificação de várias classes como, por exemplo, reduzir o problema de várias classes para c problemas de duas classes. Então, o i -ésimo problema é resolvido encontrando a função discriminante que separa os padrões pertencentes à classe ω_i dos padrões não pertencentes à classe ω_i . Esta maneira é mostrada na Figura 3.3.

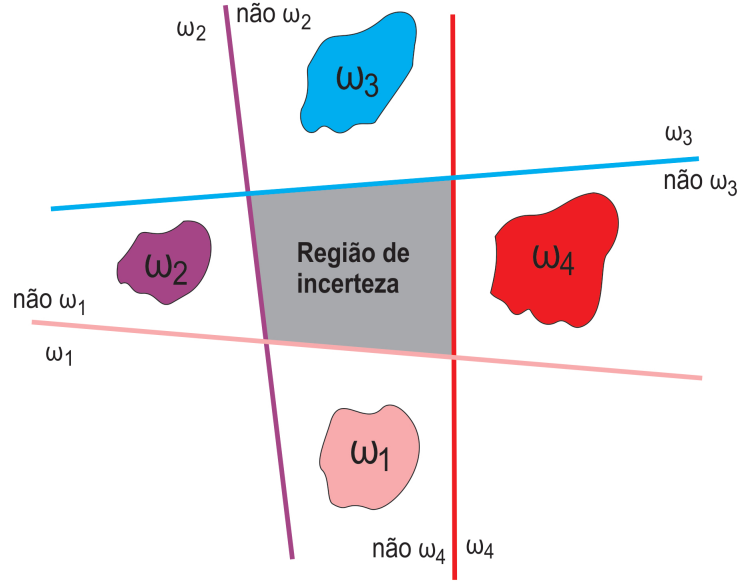


Figura 3.3: Limites de decisão para um problema com quatro classes, dicotomizador $\omega_i/\text{não } \omega_i$.

Uma outra maneira de utilizar as funções discriminantes lineares para várias classes é usar $c(c-1)/2$ funções discriminantes lineares, uma para cada par de classes, como mostrado na Figura 3.4 [42].

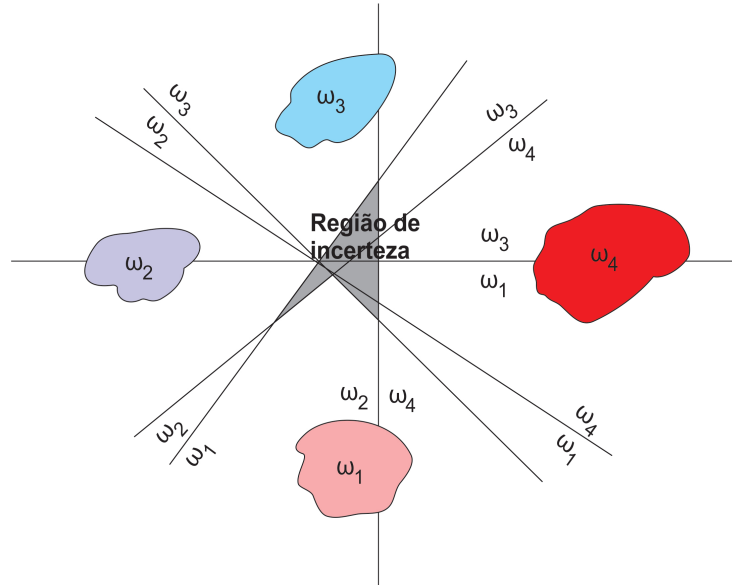


Figura 3.4: Limites de decisão para um problema com quatro classes, dicotomizador ω_i/ω_j .

As Figuras 3.3 e 3.4 ilustram que essas abordagens podem resultar em regiões na qual não é possível a classificação, denominadas como regiões de incerteza. O problema com

essas regiões pode ser resolvido definindo c funções discriminantes lineares como mostrado na Equação (3.6) [42]

$$g_i(\mathbf{x}) = \mathbf{w}_i^T \mathbf{x} + w_{i0} \quad 1 = 1, 2, \dots, c, \quad (3.6)$$

e assumir na regra de decisão que o padrão \mathbf{x} pertence a classe ω_i se $g_i(\mathbf{x}) > g_j(\mathbf{x})$ para todo $j \neq i$. Quando $g_i(\mathbf{x}) = g_j(\mathbf{x})$, o classificador linear é indefinido.

O classificador resultante é chamado de *máquina linear*. Esta máquina divide o espaço de características em c regiões de decisão, onde $g_i(\mathbf{x})$ é a função discriminante que assume o maior valor se \mathbf{x} está na região R_i . Se R_i e R_j são adjacentes, a fronteira entre elas é uma parte do hiperplano definido pela Equação (3.7)

$$g_i(\mathbf{x}) = g_j(\mathbf{x}). \quad (3.7)$$

3.3.4 Funções discriminantes lineares generalizadas

A função discriminante linear $g(\mathbf{x})$ da Equação (3.4) pode ser escrita como mostrado na Equação (3.8)

$$g(\mathbf{x}) = w_0 + \sum_{i=1}^d w_i x_i, \quad (3.8)$$

onde:

w_i - componentes do vetor peso \mathbf{w} .

A função discriminante quadrática pode ser escrita adicionando produtos de pares dos componentes de \mathbf{x} como mostrado na Equação (3.9) [42]

$$g(\mathbf{x}) = w_0 + \sum_{i=1}^d w_i x_i + \sum_{i=1}^d \sum_{j=1}^d w_{ij} x_i x_j. \quad (3.9)$$

Como $x_i x_j = x_j x_i$ pode-se assumir sem perda de generalidade que $w_{ij} = w_{ji}$.

Pode-se continuar adicionando produtos dos termos de \mathbf{x} tais como $w_{ijk} x_i x_j x_k$ e obter a classe das *funções discriminantes polinomiais*. Estas funções generalizam as funções discriminantes lineares como mostrado nas Equações (3.10) ou (3.11)

$$g(\mathbf{x}) = \sum_{i=1}^{\hat{d}} a_i y_i(\mathbf{x}) \quad (3.10)$$

ou

$$g(\mathbf{x}) = \mathbf{a}^T \mathbf{y} \quad (3.11)$$

onde:

\mathbf{a} - vetor peso \hat{d} -dimensional ($\hat{d} = d + 1$);

$y_i(\mathbf{x})$ - funções arbitrárias de \mathbf{x} .

O objetivo deste trabalho são as \hat{d} funções $y_i(\mathbf{x})$ mostrada na Equação (3.10). Elas são apresentadas com maiores detalhes no Capítulo 4 desta tese.

3.4 K Vizinhos Mais Próximos (K-NN)

O classificador K-NN é um método de classificação que não possui processamento na fase de treinamento. Nesse classificador é necessário um grande número de padrões de treinamento cuja classe é conhecida a priori. Ele é um classificador sub-ótimo que cria fronteiras de decisão complexas.

Considere um padrão de teste desconhecido \mathbf{x} . Sua classificação é realizada da seguinte maneira:

- inicialmente, calcula-se a distância entre \mathbf{x} e todos os padrões de treinamento;
- verifica-se quais são as classes que os K padrões mais próximos pertencem; e
- a classificação é realizada associando-se o padrão de teste à classe que é mais frequente entre os K padrões mais próximos de \mathbf{x} .

A Figura 3.5 mostra um exemplo do classificador K-NN para K=1, K=3 e K=5.

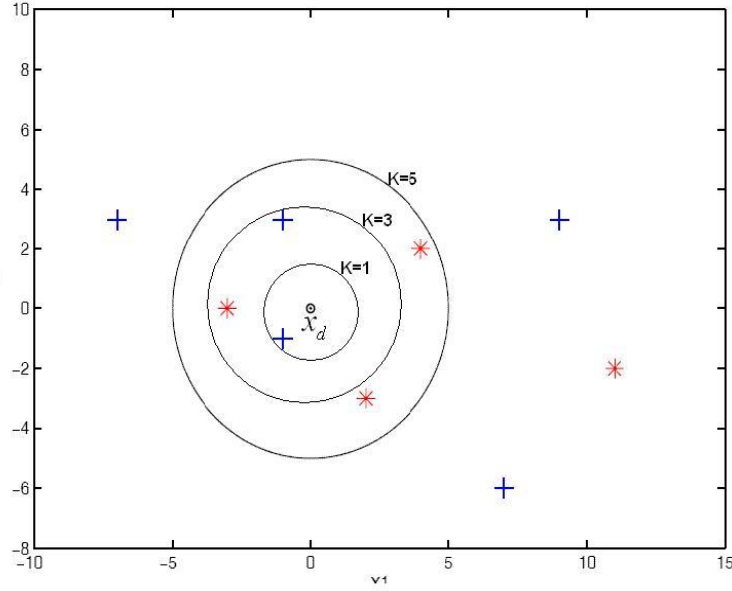


Figura 3.5: Classificação de duas classes utilizando o método K-NN para K=1, K=3 e K=5 [43].

Existem várias formas de medir a distância entre os conjuntos de classes diferentes no espaço de características. Dentre elas, pode-se citar:

Distância Euclidiana - é a mais simples das distâncias. Essa distância é encontrada como mostrado na Equação (3.12)

$$D(\mathbf{a}, \mathbf{b}) = \left(\sum_{i=1}^d (a_i - b_i)^2 \right)^{1/2}. \quad (3.12)$$

Distância de Manhattan - esta distância é obtida utilizando-se a Equação (3.13)

$$D(\mathbf{a}, \mathbf{b}) = \sum_{i=1}^d |a_i - b_i|. \quad (3.13)$$

Distância de Minkowski - a Equação (3.14) calcula essa distância

$$D(\mathbf{a}, \mathbf{b}) = \left(\sum_{i=1}^d (a_i - b_i)^p \right)^{1/p}. \quad (3.14)$$

A distância de Minkowski é a generalização das duas distâncias Euclidiana e de Manhattan. Quando $p = 1$, a distância de Minkowski representa a distância de Manhattan e quando $p = 2$, ela representa a distância Euclidiana.

Uma das dificuldades na utilização do classificador K-NN está na escolha do valor de K. Nesta escolha é sempre melhor optar por um valor ímpar para evitar que duas classes tenham o mesmo resultado. Se K é muito pequeno, a classificação fica sensível a pontos de ruído, se K é muito grande, a vizinhança pode incluir elementos de outras classes [44].

A precisão da classificação utilizando o algoritmo K-NN depende fortemente do modelo de dados. Na maioria das vezes os atributos precisam ser normalizados para evitar que as medidas de distância sejam dominadas por uma única característica.

Como todo classificador, o método de classificação K-NN possui vantagens e desvantagens. Como vantagens a técnica é simples e fácil de ser implementada. Ela é bastante flexível e em alguns casos apresenta ótimos resultados. Como desvantagens pode-se elucidar que o processo computacional pode ser muito complexo ao classificar um exemplo desconhecido. Ele requer um cálculo de distância para cada exemplo de treinamento, podendo consumir muito tempo quando o conjunto de treinamento é muito grande e a precisão do classificador pode ser severamente comprometida pela presença de ruído ou características irrelevantes [43].

Considerando $K=1$ no classificador de K vizinhos mais próximos, obtém-se o classificador vizinho mais próximo (1-NN). Este classificador é muito comum em aplicações de reconhecimento de face.

3.5 Discriminante Linear de Fisher

Considere o problema em projetar os dados de d -dimensões em uma reta. Se as amostras formarem aglomerações (*clusters*) compactas no espaço d -dimensional, a projeção sobre uma reta arbitrária costuma resultar em uma mistura confusa das amostras de todas as classes e, portanto, resultam em um desempenho pobre no reconhecimento. O objetivo da análise discriminante linear clássica é encontrar uma orientação para a reta que melhor separa a projeção das amostras [42].

Suponha um conjunto de n amostras d -dimensional $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$, seja n_1 o número de amostras pertencentes a classe ω_1 em um subconjunto D_1 e n_2 o número de amostras per-

tencentes a classes ω_2 em um subconjunto D_2 , com $n_1 + n_2 = n$. Se formar uma combinação linear dos componentes de \mathbf{x} , obtém-se o produto escalar como mostrado na Equação (3.15)

$$y = \mathbf{w}^T \cdot \mathbf{x} \quad (3.15)$$

e um conjunto correspondente de n amostras y_1, y_2, \dots, y_n dividido nos subconjuntos \mathbf{Y}_1 e \mathbf{Y}_2 . Geometricamente, se $\|\mathbf{w}\| = 1$, cada y_i é a projeção do correspondente \mathbf{x}_i em uma reta na direção de \mathbf{w} . A direção de \mathbf{w} é importante, pois o objetivo é projetar em uma reta as amostras pertencentes a classe ω_1 separadas das amostras pertencentes a classe ω_2 [42].

A Figura 3.6 ilustra o efeito de escolher diferentes valores para \mathbf{w} em um exemplo bidimensional.

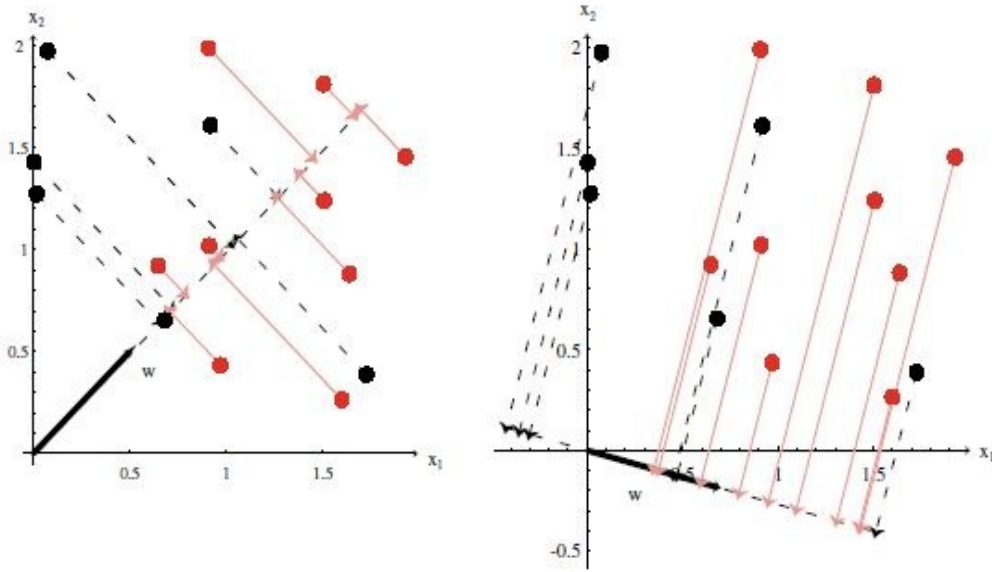


Figura 3.6: Projeção de um mesmo conjunto de amostras em duas retas diferentes na direção de \mathbf{w} [42].

Para escolher a melhor direção para \mathbf{w} , espera-se que seja permitida a classificação exata. Uma medida de separação entre os pontos projetados é a diferença da média amostral. Se \mathbf{m}_i é a média amostral d -dimensional obtida pela Equação (3.16)

$$\mathbf{m}_i = \frac{1}{n_i} \sum_{\mathbf{x} \in D_i} \mathbf{x}, \quad (3.16)$$

então a média amostral dos pontos projetados é encontrada utilizando-se a Equação (3.17)

$$\hat{m}_i = \frac{1}{n_i} \sum_{y \in Y_i} y = \frac{1}{n_i} \sum_{\mathbf{x} \in D_i} \mathbf{w}^T \mathbf{x} = \mathbf{w}^T \mathbf{m}_i \quad (3.17)$$

onde:

\hat{m}_i - projeção de \mathbf{m}_i .

A distância entre as médias projetadas é encontrada utilizando-se a Equação (3.18).

$$|\hat{m}_1 - \hat{m}_2| = |\mathbf{w}^T(\mathbf{m}_1 - \mathbf{m}_2)|. \quad (3.18)$$

Consegue-se fazer a diferença da Equação (3.18) tão grande quanto se deseja, simplesmente pelo escalonamento de \mathbf{w} . Para obter uma boa separação dos dados projetados a diferença entre as médias deve ser grande.

Define-se o espalhamento \tilde{s}_i^2 para as amostras ω_i pela Equação (3.19)

$$\tilde{s}_i^2 = \sum_{y \in Y_i} (y - \hat{m}_i)^2, \quad (3.19)$$

onde:

y - projeção de \mathbf{x} na direção de \mathbf{w} ; e

\hat{m}_i - projeção de \mathbf{m}_i .

Assim, $\frac{1}{n}(\tilde{s}_1^2 - \tilde{s}_2^2)$ é uma estimativa da variância dos dados reunidos e $\tilde{s}_1^2 - \tilde{s}_2^2$ é chamado de dispersão total dentro da classe das amostras projetadas.

O discriminante linear de Fisher utiliza a função linear $\mathbf{w}^T \mathbf{x}$ onde a função critério mostrada na Equação (3.20)

$$J(\mathbf{w}) = \frac{|\hat{m}_1 - \hat{m}_2|^2}{\tilde{s}_1^2 - \tilde{s}_2^2} \quad (3.20)$$

é máxima e independente de $\|\mathbf{w}\|$. Considerando \mathbf{w}^* o vetor que maximiza a Equação (3.20), tem-se assim o vetor que conduz a melhor separação entre os dois conjuntos projetados.

Para obter a Equação (3.20) como uma função explícita de \mathbf{w} , define-se a matriz espalhamento S_i e S_W pelas Equações (3.21) e (3.22)

$$S_i = \sum_{\mathbf{x} \in D_i} (\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^T \quad (3.21)$$

e

$$S_W = S_1 - S_2. \quad (3.22)$$

Então, das Equações (3.15) e (3.17) pode-se escrever a Equação (3.19) conforme a Equação (3.23)

$$\begin{aligned} \tilde{s}_i^2 &= \sum_{\mathbf{x} \in D_i} (\mathbf{w}^T \mathbf{x} - \mathbf{w}^T \mathbf{m}_i)^2 \\ &= \sum_{\mathbf{x} \in D_i} \mathbf{w}^T (\mathbf{x} - \mathbf{m}_i) (\mathbf{x} - \mathbf{m}_i)^T \mathbf{w} \\ &= \mathbf{w}^T S_i \mathbf{w}. \end{aligned} \quad (3.23)$$

Portanto,

$$\tilde{s}_1^2 - \tilde{s}_2^2 = \mathbf{w}^T S_W \mathbf{w}. \quad (3.24)$$

Analogamente, as separações das médias projetadas obedece a Equação (3.25)

$$\begin{aligned} (\hat{m}_1 - \hat{m}_2)^2 &= (\mathbf{w}^T \mathbf{m}_1 - \mathbf{w}^T \mathbf{m}_2)^2 \\ &= \mathbf{w}^T (\mathbf{m}_1 - \mathbf{m}_2) (\mathbf{m}_1 - \mathbf{m}_2)^T \mathbf{w} \\ &= \mathbf{w}^T S_B \mathbf{w}, \end{aligned} \quad (3.25)$$

onde:

$$S_B = (\mathbf{m}_1 - \mathbf{m}_2) (\mathbf{m}_1 - \mathbf{m}_2)^T. \quad (3.26)$$

Denomina-se S_W como *matriz espalhamento dentro da classe*. Ela é simétrica e semi-definida positiva, normalmente não singular se $n > d$. Da mesma forma, S_B é chamada de *matriz espalhamento entre as classes*. Ela também é simétrica e semidefinida positiva. Em particular, para qualquer \mathbf{w} , $S_B \mathbf{w}$ está na direção de $\mathbf{m}_1 - \mathbf{m}_2$ e S_B é absolutamente singular.

Em termos de S_B e S_W , a Equação (3.20) pode ser escrita conforme a Equação (3.27)

$$J(\mathbf{w}) = \frac{\mathbf{w}^T S_B \mathbf{w}}{\mathbf{w}^T S_W \mathbf{w}}. \quad (3.27)$$

A Equação (3.27) é bem conhecida na física matemática como o quociente de Rayleigh generalizado [42]. O vetor \mathbf{w}^* deve satisfazer a Equação (3.28)

$$S_B \mathbf{w} = \lambda S_W \mathbf{w} \quad (3.28)$$

para alguma constante λ , que é um problema de autovalor generalizado. Se S_W é não singular pode-se obter um problema de autovalor convencional escrevendo a Equação (3.29)

$$S_W^{-1} S_B \mathbf{w} = \lambda \mathbf{w}. \quad (3.29)$$

Na Equação (3.29) é desnecessário resolver para autovalores e autovetores de $S_W^{-1} S_B$ porque $S_B \mathbf{w}$ está sempre na direção de $\mathbf{m}_1 - \mathbf{m}_2$. Pode-se imediatamente escrever a solução \mathbf{w}^* que otimiza a Equação (3.20) pela Equação (3.30)

$$\mathbf{w}^* = S_W^{-1}(\mathbf{m}_1 - \mathbf{m}_2). \quad (3.30)$$

Assim, obtém-se \mathbf{w}^* para o discriminante linear de Fisher e a classificação é convertida de um problema d -dimensional para um com dimensão um [42].

3.6 Máquina de Vetores Suporte

Fundamentada na teoria da aprendizagem estatística, a máquina de vetores suporte, do inglês Support Vectors Machine (SVM), foi desenvolvida por [45] com o intuito de resolver problemas de classificação de padrões. Devido a sua eficiência em trabalhar com dados de alta dimensionalidade, ela é reportada na literatura como uma técnica altamente robusta, muitas vezes comparada as redes neurais [46, 47].

O objetivo da teoria do aprendizado estatístico é estabelecer condições matemáticas que permitam escolher um classificador, com bom desempenho, para um conjunto de dados disponíveis para treinamento e teste [48].

Como a Equação (3.31)

$$f(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + b \quad (3.31)$$

é um hiperplano, pode-se definir como margem ρ a menor distância entre os exemplos do conjunto de treinamento e o hiperplano utilizado para a separação das classes [49]. A margem determina quão bem duas classes podem ser separadas [50].

A margem é obtida pela distância entre o hiperplano e os vetores que estão mais próximos a ele. Esses vetores são denominados de vetores suporte. De acordo com [50], os vetores suporte são padrões críticos, que sozinhos determinam o hiperplano ótimo. Os outros padrões (não-críticos) são irrelevantes, podendo ser removidos do conjunto de treinamento sem afetar os resultados. Na Figura 3.7, os vetores suporte são realçados por círculos externos nos padrões.

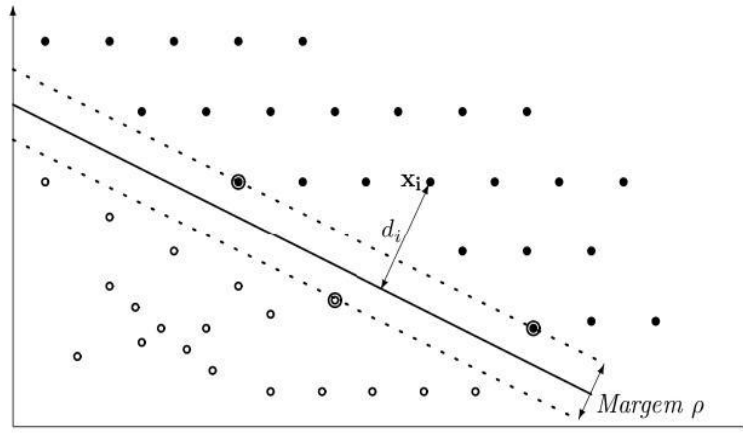


Figura 3.7: Identificação da margem ρ e dos vetores suporte sobre a linha pontilhada.

Uma classificação linear consiste em determinar uma função $f : X \subseteq \mathbb{R}^N \rightarrow \mathbb{R}$, que atribui um valor $(+1)$ se $f(\mathbf{x}) > 0$ e (-1) caso contrário. Uma função linear pode ser representada pela Equação (3.32).

$$\begin{aligned} f(\mathbf{x}) &= \mathbf{w} \cdot \mathbf{x} + b \\ &= \sum_{i=1}^N w_i x_i + b \end{aligned} \quad (3.32)$$

onde:

\mathbf{w} - vetor peso; e

b - bias.

Na Equação (3.32) os parâmetros $\mathbf{w} \in \mathbb{R}^N$ e $b \in \mathbb{R}$ são responsáveis pelo controle da função $f(\mathbf{x})$ e pela regra de decisão [51]. Os valores de \mathbf{w} e b são obtidos pelo processo de aprendizagem a partir dos dados de entrada. Eles podem ser interpretados geometricamente sobre um hiperplano. Um hiperplano é um subespaço afim, que divide um espaço em duas partes, correspondendo aos dados de duas classes distintas [51].

Um SVM linear busca encontrar um hiperplano que separa perfeitamente os dados de cada classe e cuja margem de separação é máxima. Esse hiperplano é denominado de hiperplano ótimo.

Assumindo que o conjunto de treinamento é linearmente separável, o hiperplano ótimo é definido pela Equação (3.33).

$$\mathbf{w} \cdot \mathbf{x} + b = 0. \quad (3.33)$$

Considerando a restrição imposta pela Equação (3.34)

$$\begin{aligned} \mathbf{w} \cdot \mathbf{x}_i + b &\geq 1, & \text{para } f(\mathbf{x}_i) = +1 \\ \mathbf{w} \cdot \mathbf{x}_i + b &\leq -1, & \text{para } f(\mathbf{x}_i) = -1 \end{aligned}, \quad (3.34)$$

os classificadores lineares que separam um conjunto de treinamento possuem margem positiva, ou seja, essa restrição afirma que não existe nenhum dado entre $\mathbf{w} \cdot \mathbf{x} + b = 0$ e $\mathbf{w} \cdot \mathbf{x} + b = \pm 1$, onde a margem sempre é maior do que a distância entre os hiperplanos $\mathbf{w} \cdot \mathbf{x} + b = 0$ e $|\mathbf{w} \cdot \mathbf{x} + b| = 1$. Devido a essas suposições as SVMs obtidas são normalmente chamadas de SVMs com margens rígidas (ou largas).

A Figura 3.8 mostra a distância entre os hiperplanos e os vetores suporte.

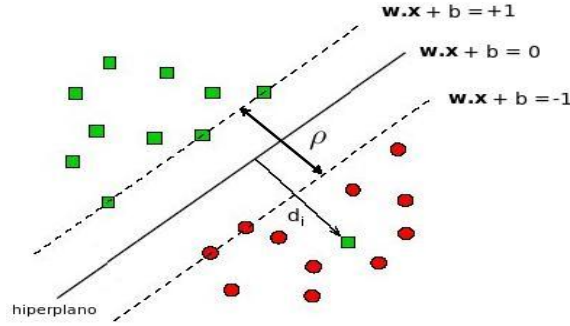


Figura 3.8: Distância entre os hiperplanos e os vetores suporte.

Um conjunto de dados é considerado ser não linearmente separável, caso não seja possível separar os dados com um hiperplano. A Figura 3.9 mostra um conjunto linearmente separável e outro não-linearmente separável.

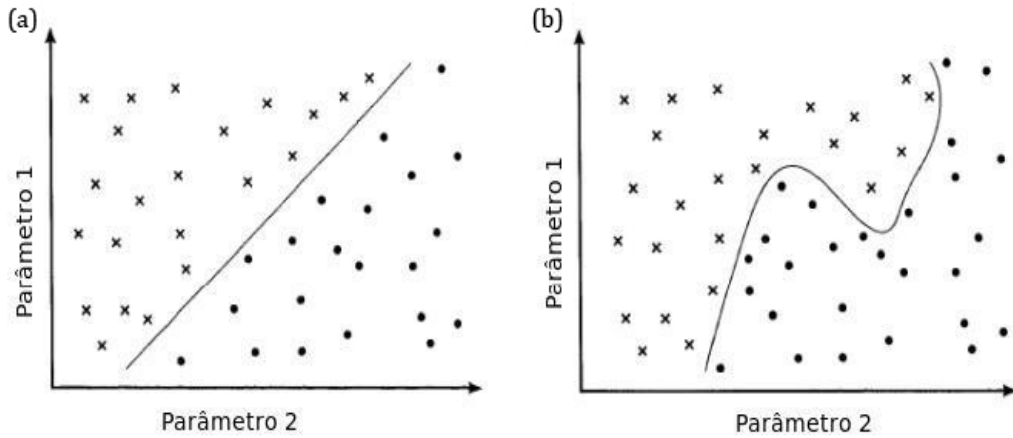


Figura 3.9: (a) Conjunto linearmente separável; (b) conjunto não-linearmente separável.

O teorema de Cover afirma que um problema não-linear tem maior probabilidade de ser linearmente separável, em um espaço de mais alta dimensionalidade [50]. Esse fato também é uma estratégia fundamental na qualidade do classificador polinomial que é apresentado no Capítulo 4 deste trabalho.

3.7 Árvores de Decisão

Uma árvore de decisão é constituída internamente por:

- a) nós de decisão que particionam o espaço de características por meio de superfícies de decisão; e
- b) nós terminais ou “folhas” que permitem associar uma classe a cada objeto.

Para características numéricas, utilizam-se normalmente funções lineares como modelos para as superfícies de decisão.

A Figura 3.10 ilustra uma árvore de decisão univariada do inglês *Ordinary Binary Classification Tree* (OBCT), suas regras de classificação e a respectiva representação da classificação dos dados em um espaço de características. Esse é o caso mais simples de árvore de decisão, chamada de univariada em função de suas regras estarem associadas somente a uma única característica [52].

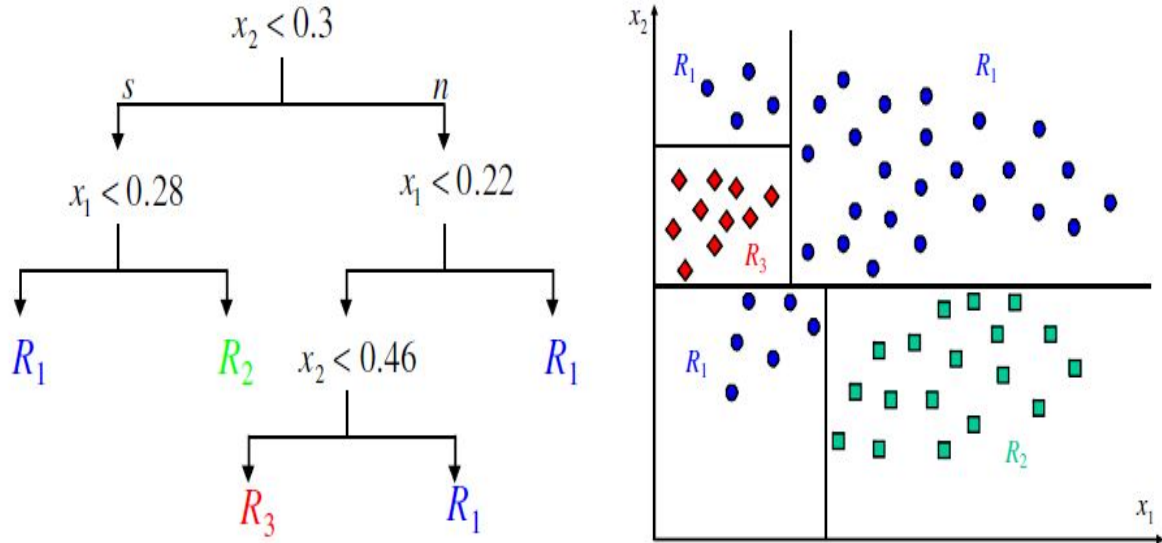


Figura 3.10: Exemplo de árvore de decisão univariada [42].

Em cada nó da OBCT, o hiperplano de decisão intercepta apenas um dos eixos coordenados e seu modelo, para um determinado atributo x_j é descrito pela Equação (3.35)

$$f(\mathbf{x}|j, w_0) = x_j + w_0 = 0 \quad (3.35)$$

onde:

\mathbf{x} - vetor de atributos do objeto,

e o hiperplano intercepta o eixo x_j em $-w_0$.

Na Figura 3.11 a árvore de decisão é chamada de oblíqua do inglês *Oblique Decision Tree* (ODT) ou multivariada linear [53], porque os testes consistem de hiperplanos com orientação oblíqua aos eixos dos atributos.

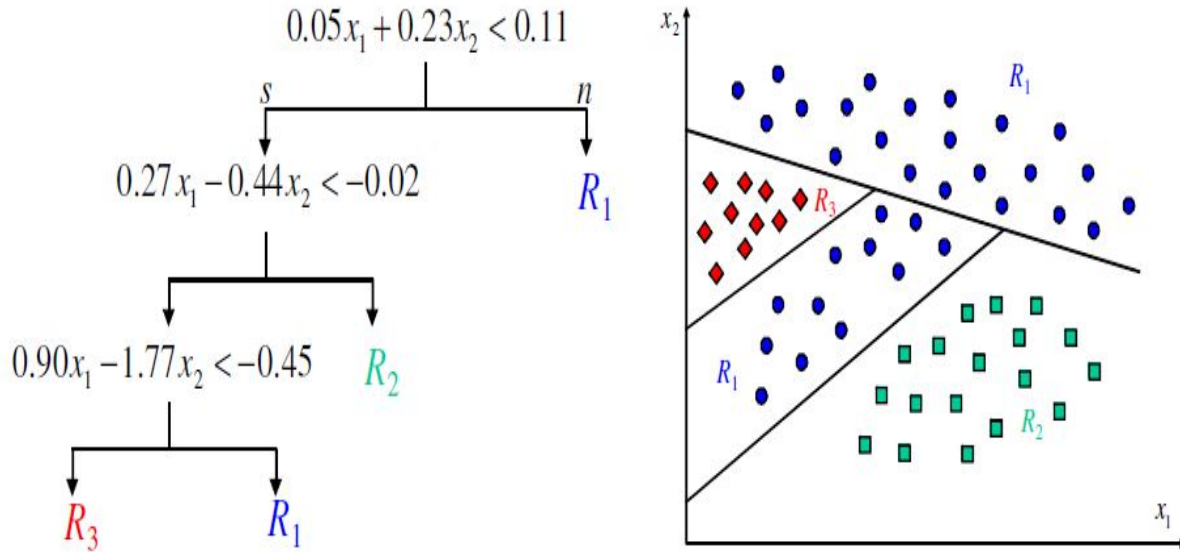


Figura 3.11: Exemplo de árvore de decisão oblíqua [42].

Na ODT, o modelo adotado para as superfícies de decisão é mostrado na Equação (3.36)

$$f_m(\mathbf{x}|\mathbf{w}_m, w_{m0}) = \mathbf{w}_m^T \mathbf{x} + w_{m0} \quad (3.36)$$

onde:

m - nó de decisão;

\mathbf{w}_m - vetor de pesos do hiperplano d-dimensional;

w_{m0} - valor limitante quando \mathbf{x} está projetado sobre \mathbf{w}_m ; e

T - vetor transposto.

Tanto na OBCT quanto na ODT, o teste a ser realizado em cada nó de decisão pode ser descrito como uma função booleana da forma: se $f_m(\mathbf{x}|\Phi) > 0$, então siga o galho da direita; caso contrário, siga o galho de esquerda [53].

A classificação de um padrão \mathbf{x} consiste em apresentá-lo ao primeiro nó de decisão da árvore (chamado de “raiz”), realizar o teste, seguir para o nó do galho selecionado e sucessivamente aplicar testes até atingir um nó folha onde sua classe será definida [52].

O aprendizado indutivo de árvores de decisão é geralmente dividido em aprendizado supervisionado e não-supervisionado, embora o aprendizado semi-supervisionado também tenha sido considerado ao longo dos últimos anos [54].

Conforme [55], a classificação pode ser utilizada para os seguintes propósitos: modelagem descritiva e modelagem preditiva. Na modelagem descritiva, um modelo de classificação é utilizado como um modelo para distinguir exemplos de diferentes classes. E na modelagem preditiva, um modelo é utilizado para classificar exemplos cujas classes são desconhecidas, ou seja, exemplos que não foram utilizados na construção do modelo. Uma das razões para que a árvore de decisão seja uma técnica muito poderosa e amplamente utilizada em problemas de classificação é o fato do conhecimento adquirido ser representado por regras que podem ser expressas em linguagem natural, facilitando o entendimento das pessoas.

Existem muitas maneiras de uma árvore de decisão ser estruturada a partir de um conjunto de características. Nesse sentido, algoritmos baseados em heurísticas têm sido desenvolvidos para a indução de árvores de decisão. Mesmo que eles não garantam uma solução ótima, apresentam resultados satisfatórios em tempo hábil.

O *Top-Down Induction of Decision Tree* (TDIDT) é um algoritmo bem conhecido e é utilizado como base para muitos algoritmos de indução de árvores de decisão, dentre eles os mais conhecidos são ID3 [56], C4.5 [57] e CART [58].

O TDIDT produz regras de decisão de forma implícita em uma árvore de decisão, a qual é construída por sucessivas divisões dos exemplos de acordo com os valores de seus atributos preditivos. De acordo com [59], esse processo é conhecido como particionamento recursivo.

O esqueleto do algoritmo TDIDT é baseado em três possibilidades sobre um conjunto de treinamento T contendo classes C_1, C_2, \dots, C_k :

1. T contém um ou mais objetos, onde todos são da classe C_j . Assim, a árvore de decisão para T é um nó folha que identifica a classe C_j ;
2. T não contém objetos. A árvore de decisão também é um nó folha, mas a classe associada deve ser determinada por uma informação externa; e
3. T contém exemplos pertencentes a mais de uma classe. Neste caso, a ideia é dividir T em sub-conjuntos que são, ou tendem a dirigir-se para, coleções de exemplos com classes únicas. Para isso é escolhido um atributo preditivo A que possui um ou mais possíveis resultados O_1, O_2, \dots, O_n . T é particionado em subconjuntos T_1, T_2, \dots, T_n , onde T_i contém todos os exemplos de T que têm resultado O_i para o atributo A . A árvore de decisão para T consiste de um nó de decisão identificando o teste sobre o atributo A e uma aresta para cada possível resultado, ou seja, n arestas. No lugar de um único atributo A , ele também pode ser considerado um subconjunto de atributos.

O mesmo algoritmo de indução de árvores de decisão (passos 1, 2 e 3) é aplicado recursivamente para cada sub-conjunto de exemplos T_i com i variando de 1 até n [60].

Uma das medidas baseadas em impureza é o ganho de informação, que usa a entropia como medida de impureza. O algoritmo ID3 [56], pioneiro em indução de árvores de decisão, utiliza essa medida. Para determinar o quão boa é uma condição de teste realizada é necessário comparar o grau de entropia do nó-pai (antes da divisão) com o grau de entropia dos nós-filhos (após a divisão). O atributo que gerar uma maior diferença é escolhido como condição de teste. O ganho de informação é obtido pela Equação (3.37)

$$ganho = entropia(pai) - \sum_{j=1}^n \left[\frac{N(v_j)}{N} entropia(v_j) \right] \quad (3.37)$$

onde:

n - número de valores do atributo;

N - número total de objetos do nó-pai; e

$N(v_j)$ - número de exemplos associados ao nó filho.

O grau de entropia é definido pela Equação (3.38)

$$entropia(nó) = - \sum_{i=1}^c p(i/nó) \log_2[p(i/nó)] \quad (3.38)$$

onde:

$p(i/nó)$ - fração dos registros pertencentes à classe i no nó;

c - número de classes.

O critério de ganho seleciona como atributo-teste aquele que maximiza o ganho de informação. O grande problema ao se utilizar esse ganho é que ele prefere atributos com muitos valores possíveis (número de arestas).

Para solucionar o problema do ganho de informação, foi proposto em [57] a razão de ganho (do inglês *gain ratio*), que nada mais é do que o ganho de informação relativo como critério de avaliação.

A razão de ganho é definida pela Equação (3.39)

$$razão\ de\ ganho(nó) = \frac{ganho}{entropia(nó)}. \quad (3.39)$$

Pela Equação (3.39) é possível perceber que a razão não é definida quando o denominador é igual a zero. Além disso, a razão de ganho favorece atributos cujo denominador, ou seja, a entropia possui valor pequeno. Em [61] é sugerido que a razão de ganho seja realizada em duas etapas. Na primeira etapa é calculado o ganho de informação para todos os atributos. Após isso, considera apenas aqueles atributos que obtiveram um ganho de informação acima da média e então escolhe aquele que apresentar a melhor razão de ganho. Em [61] mostra-se que a razão de ganho supera o ganho de informação tanto em termos de acurácia quanto em termos de complexidade das árvores de decisão geradas.

Outra medida bastante conhecida é o Gini, que emprega um índice de dispersão estatística proposto em 1912 pelo estatístico italiano Corrado Gini. Esse índice é muito utilizado

em análises econômicas e sociais, por exemplo, para quantificar a distribuição de renda em um certo país.

O Gini é utilizado no algoritmo CART. Para um problema de c classes, o $gini_{index}$ é definido pela Equação (3.40)

$$gini_{index}(\text{nó}) = 1 - \sum_{i=1}^c p(i/\text{nó}). \quad (3.40)$$

Assim, como no cálculo do ganho de informação, basta calcular a diferença entre o $gini_{index}$ antes e após a divisão. Essa diferença, Gini, é representada pela Equação (3.41)

$$\text{Gini} = gini_{index}(\text{pai}) - \sum_{j=1}^n \left[\frac{N(v_j)}{N} gini_{index}(v_j) \right] \quad (3.41)$$

onde:

n - número de valores do atributo;

N - número total de objetos do nó-pai; e

$N(v_j)$ - número de exemplos associados ao nó filho.

Assim, é selecionado o atributo que gerar um maior valor para Gini. Para outras medidas, consultar [62].

3.7.1 Métodos de poda

Quando as árvores de decisão são construídas, muitas arestas ou sub-árvores podem refletir ruídos ou erros. Isso é um problema conhecido como *sobreajuste*, significando um aprendizado muito específico do conjunto de treinamento não permitindo a generalização do modelo. Para detectar e excluir essas arestas e sub-árvores são utilizados métodos de poda (*pruning*) da árvore, cujo objetivo é melhorar a taxa de acerto do modelo para novos exemplos que não foram utilizados no conjunto de treinamento [63].

Consequentemente, a árvore podada torna-se mais simples facilitando a sua interpretação por parte do usuário. Existem diversas formas de realizar poda em uma árvore de decisão e todas elas são classificadas como pré-poda ou pós-poda [60].

O método pré-poda é realizado durante o processo de construção da árvore em que o processo pode simplesmente parar de dividir o conjunto de elementos e transformar o nó corrente em um nó folha da árvore. O pós-poda é realizado após a construção da árvore de decisão removendo ramos completos onde tudo que está abaixo de um nó interno é excluído e esse nó é transformado em folha, representando a classe mais frequente no ramo [60].

Embora a poda seja um método bastante utilizado e eficaz na solução do problema de sobreajuste deve-se ter cuidado para não podar demais a árvore. Quando isso ocorre, tem-se o problema conhecido como sub-ajuste, em que o modelo de classificação não aprendeu o suficiente sobre os dados de treinamento. Dentre os métodos de poda existentes, pode-se destacar: *Cost Complexity Pruning*, *Reduced Error Pruning*, *Minimum Error Pruning* (MEP), *Pessimistic Pruning*, *Error-Based Pruning* (EBP), *Minimum Description Length* (MDL) *Pruning*, *Minimum Message Length* (MML) *Pruning*, *Critical Value Pruning* (CVP), OPT e OPT-2 [60].

3.7.2 Algoritmos de indução de árvores de decisão

Serão apresentados sucintamente os três principais algoritmos para a indução de árvores de decisão. São eles: ID3 [56], C4.5 [57] e CART [58].

Na literatura, existem novos algoritmos para a indução de árvores de decisão, inclusive alguns que divergem do algoritmo básico TDIDT. Como exemplos, pode-se mencionar: NBTree [64], ADTree [65], LMT [66] e BFTree [67].

- O ID3 é o algoritmo pioneiro em indução de árvore de decisão. Ele é um algoritmo recursivo baseado em busca gulosa procurando sobre um conjunto de atributos que “melhor” dividem os exemplos, gerando sub-árvores. A principal limitação é que esse algoritmo só trabalha com os atributos categóricos não-ordinais, não sendo possível apresentar a ele conjuntos de dados com atributos contínuos. Outra limitação é que o ID3 não apresenta nenhuma forma para tratar valores desconhecidos, ou seja, todos os exemplos do conjunto de treinamento devem ter valores conhecidos para todos os seus atributos.

O ID3 utiliza o ganho de informação para selecionar a melhor divisão. No entanto, esse critério não considera o número de divisões (número de arestas) e isso pode resultar em árvores mais complexas. Esse algoritmo também não apresenta nenhum método de pós-poda o que poderia amenizar esse problema de árvores mais complexas.

- O algoritmo C4.5 representa uma evolução significativa em relação ao ID3. As principais melhorias em relação ao ID3 são: utilizar atributos categóricos (ordinais ou não-ordinais) ou atributos contínuos. Quando utiliza-se atributos contínuos, esse algoritmo define um limiar e então divide os exemplos de forma binária: aqueles cujo valor do atributo é maior do que o limiar e aqueles cujo valor do atributo é menor ou igual ao limiar. O algoritmo C4.5 permite que os valores desconhecidos para um determinado atributo sejam representados como “?”. Esse algoritmo trata esses valores de forma especial. Esses valores não são utilizados nos cálculos do ganho e da entropia [60].

O C4.5 utiliza a medida da razão do ganho para selecionar o atributo que melhor divide os exemplos. Esse algoritmo trabalha com problemas em que os atributos possuem custos diferenciados e apresenta um método de pós-poda das árvores geradas. O algoritmo C4.5 faz uma busca na árvore de baixo para cima e transforma em nós folha aqueles ramos que não apresentam nenhum ganho significativo.

O software Waikato Environment for Knowledge Analysis (WEKA) [8] disponibiliza a implementação do algoritmo C4.5, chamado de J48 nesse software.

O C4.5 é um dos algoritmos mais utilizados na literatura por ter mostrado ótimos resultados em problemas de classificação. O C4.5 é do tipo:

- “Guloso”: executa sempre o melhor passo avaliado localmente sem se preocupar se esse passo, junto à sequência completa dos passos, vai resultar na melhor solução ao final; e
- “Dividir para conquistar”: partindo da raiz criam-se as sub-árvores até chegar nas folhas o que implica em uma divisão hierárquica em múltiplos sub-problemas de

decisão, que tendem a ser mais simples do que o problema original.

- O algoritmo *Classification and Regression Trees* (CART) consiste de uma técnica não-paramétrica que induz tanto árvores de classificação quanto árvores de regressão, dependendo se o atributo é nominal (classificação) ou contínuo (regressão).

Dentre as principais vantagens do CART está a grande capacidade de pesquisa das relações entre os dados, mesmo quando elas não são evidentes, bem como a obtenção dos resultados sob a forma de árvores de decisão de grande simplicidade e legibilidade [68].

As árvores geradas pelo algoritmo CART são sempre binárias e elas podem ser percorridas da sua raiz até as folhas respondendo apenas as questões simples do tipo “sim” ou “não”. Nessa árvore utiliza-se a técnica de pesquisa exaustiva para definir os limiares a serem utilizados nos nós para dividir os atributos contínuos. Diferente das abordagens adotadas por outros algoritmos, que utilizam a pré-poda, o CART expande a árvore exaustivamente realizando a pós-poda por meio da redução do fator custo-complexidade [58].

3.8 Considerações Finais Deste Capítulo

Existem inúmeros outros classificadores com várias aplicações, entre eles pode-se citar as redes neurais que têm sido utilizadas para solucionar diversos problemas de engenharia como em processamento digital de imagens, robótica e processamento digital de sinais. As principais características no estudo e na utilização das redes neurais artificiais são: a capacidade das mesmas de aprenderem por meio da experiência, a capacidade de se adaptarem a situações adversas e a sua tolerância a ruídos.

Este capítulo apresentou os padrões e as classes de padrões. Depois foram descritos os classificadores: funções discriminantes, K vizinhos mais próximos (K-NN), discriminante linear de Fisher, máquina de vetores suporte e árvores de decisão.

O próximo capítulo descreve os classificadores polinomiais utilizados na classificação

3.8. CONSIDERAÇÕES FINAIS DESTE CAPÍTULO

de duas classes e de várias classes de padrões. Esse classificador é o método proposto neste trabalho.

Capítulo 4

Método Proposto

4.1 Introdução

As funções discriminantes lineares são consideradas o método de classificação mais simples entre os métodos de classificação supervisionado. Nesse método, uma sequência de vetores de características são linearmente mapeados em uma sequência de classes identificadas. Problemas de classificação multi-classes, ou seja, mais que duas classes, podem ser reduzidos, para múltiplos problemas de duas classes.

As funções discriminantes lineares são úteis com dados linearmente separáveis, portanto, elas são menos precisas com dados não linearmente separáveis. Como uma solução para esse problema, muitos métodos de classificação não linear foram introduzidos nas últimas décadas incluindo classificadores neurais e estatísticos [69].

O classificador polinomial é um método de classificação supervisionado capaz de separar perfeitamente tanto os dados linearmente separáveis como os não linearmente separáveis. Ele tem fornecido bons resultados em várias aplicações de reconhecimento de padrão incluindo reconhecimento de fala e separação de sinais biomédicos [70, 71, 72, 73].

Os métodos típicos de treinamento dos classificadores polinomiais são baseados em métodos estatísticos ou na minimização da função critério da soma dos quadrados dos erros. Essa minimização é solucionada por meio dos mínimos quadrados [42].

O classificador polinomial é uma aplicação não linear parametrizada que expande não linearmente uma sequência de vetores de entrada para um espaço de dimensão superior e identifica esses vetores para uma sequência de saída desejada. A proposta dessa expansão é melhorar a separação das diferentes classes no espaço vetorial expandido [69]. Os classificadores polinomiais são técnicas discutidas na literatura devido a sua qualidade na classificação para a resolução de problemas [74, 75]. Um dos principais motivos de utilizar os polinômios é porque ele é baseado no teorema da aproximação de Weierstrass [76]. Este teorema é a aproximação universal do classificador ótimo de Bayes [77].

O classificador polinomial é o método proposto neste trabalho. Este capítulo descreve esse classificador utilizado na classificação de duas ou mais classes de padrões.

4.2 Classificação Polinomial de Duas Classes

Como o princípio do classificador polinomial é expandir o espaço de características da entrada em um espaço de dimensão superior, foi mostrado na seção 3.5 do Capítulo 3 desta tese que quando a dimensão do espaço de características é reduzida, o discriminante linear de Fisher pode resultar em uma separação pobre entre as classes, principalmente quando elas não são linearmente separáveis.

A Figura 4.1 mostra um exemplo de uma distribuição de duas classes no espaço de características. Nessa figura as características das duas classes ω_1 e ω_2 são representadas por circunferências e quadrados, respectivamente.

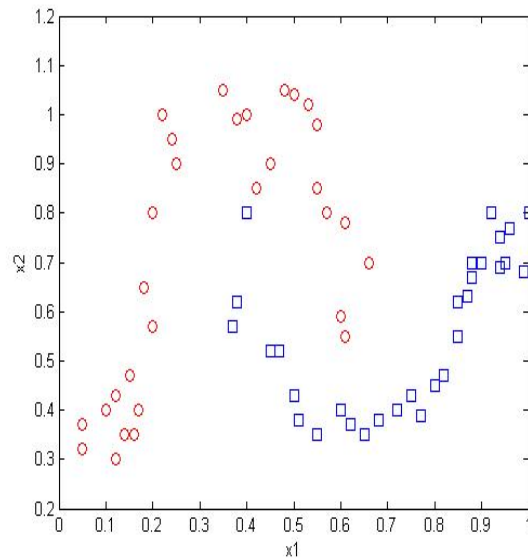


Figura 4.1: Distribuição de duas classes no espaço de características.

A Figura 4.2 ilustra a classificação das duas classes da Figura 4.1 utilizando o discriminante linear de Fisher.

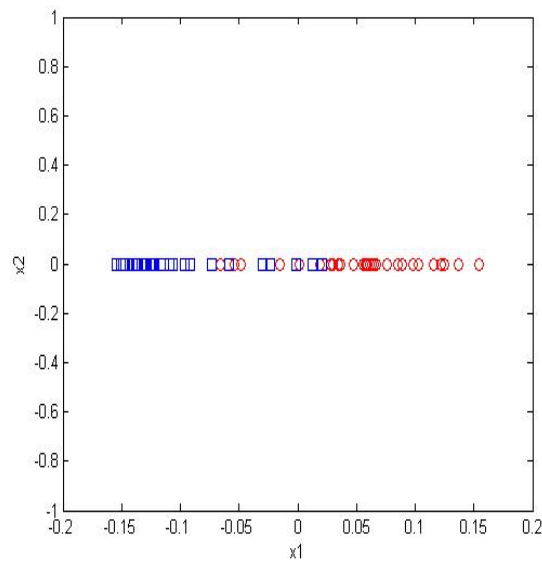


Figura 4.2: Classificação das duas classes da Figura 4.1 utilizando o discriminante linear de Fisher.

Pode-se observar na Figura 4.2 que a redução da dimensionalidade resultou em uma mistura entre as duas classes e não foi possível determinar o limiar que as separam totalmente.

A grande vantagem em expandir o espaço de características consiste na possibilidade

de buscar uma melhor separação entre as classes em um espaço de dimensão superior.

O classificador polinomial é um método de classificação supervisionado muito poderoso e seu desempenho na classificação tem mostrado resultados superiores aos classificadores SVM e árvore de decisão, principalmente quando os dados não são linearmente separáveis. Esses resultados são mostrados com mais detalhes no Capítulo 5 deste trabalho.

Após a explicação do método proposto de classificação polinomial, é mostrada a capacidade de separação desse método ao conseguir uma classificação exata das duas classes utilizadas no exemplo mostrado na Figura 4.1, não classificadas corretamente pelo classificador linear de Fisher.

Esta seção apresenta a classificação polinomial para apenas duas classes, a extensão para várias classes é apresentada na seção 4.3 deste capítulo.

Na seção 3.3 do Capítulo 3 deste trabalho, a função discriminante polinomial foi definida pela Equação (3.11). Existem muitas maneiras de representar as funções discriminantes, para não existir divergência de notação, no decorrer deste trabalho define-se a função polinomial pela Equação (4.1)

$$g(\mathbf{x}) = \mathbf{a}^T p_n(\mathbf{x}) \quad (4.1)$$

onde:

\mathbf{a} - vetor de coeficientes da função da base polinomial;

$p_n(\mathbf{x})$ - função da base polinomial; e

n - ordem ou grau da função polinomial.

A estrutura do classificador polinomial pode ser representada conforme o diagrama de blocos da Figura 4.3. Este diagrama está dividido em duas fases: treinamento e teste.

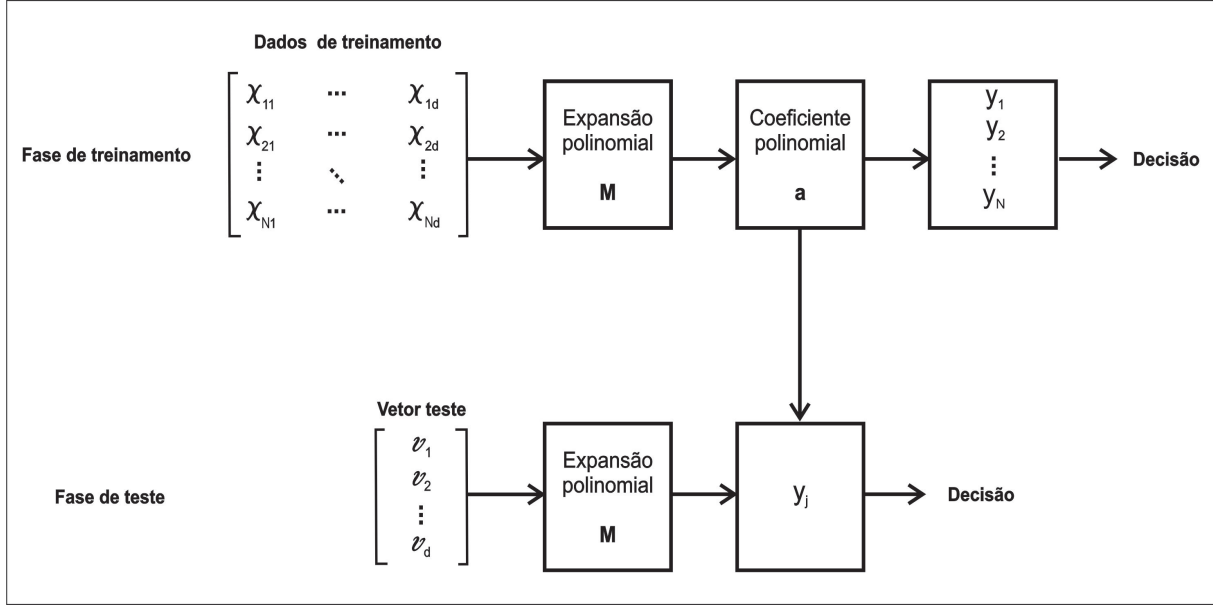


Figura 4.3: Diagrama de blocos do classificador polinomial.

Na fase de treinamento, após a entrada dos padrões no classificador, o primeiro passo consiste na transformação dos vetores de características d -dimensionais nos vetores da base polinomial L -dimensional. Por exemplo, considere um vetor de entrada de duas dimensões $\mathbf{x} = [x_1 \ x_2]^T$ e que os elementos de $p_2(\mathbf{x})$ são da forma mostrada na Equação (4.2) [70]

$$p_2(\mathbf{x}) = [1 \ x_1 \ x_2 \ x_1^2 \ x_1x_2 \ x_2^2]^T. \quad (4.2)$$

Após essa transformação, o objetivo é encontrar os coeficientes da função polinomial que minimizam a diferença entre a saída obtida e a saída desejada. Isto é possível aplicando o método dos mínimos quadrados, que é detalhado a seguir neste capítulo. Finalmente, a saída $y_i = g(\mathbf{x})$ é determinada após uma combinação linear dos termos expandidos $p_n(\mathbf{x})$ com o vetor de coeficientes polinomiais como mostrado na Equação (4.1).

De acordo com o sinal algébrico de $y_i = g(\mathbf{x})$ na Equação (4.1) é realizada a decisão de qual é a classe que o vetor de característica \mathbf{x} pertence.

Na fase de teste, o vetor teste, não presente na fase de treinamento é expandido em termos da base polinomial conforme o exemplo mostrado na Equação (4.2). Em seguida, é realizado um produto escalar desse vetor expandido com o vetor de coeficientes polinomiais obtido na fase de treinamento produzindo uma saída como mostrado na Equação (4.1). De

acordo com o sinal algébrico da saída é realizada a decisão sobre qual classe o vetor de teste pertence.

Para duas classes, considerando que o vetor de características é $\mathbf{x} = [x_1 \ x_2]^T$ e a função da base polinomial é $p_2(\mathbf{x})$ obtém-se a Equação (4.3)

$$g(\mathbf{x}) = a_1 + a_2x_1 + a_3x_2 + a_4x_1^2 + a_5x_1x_2 + a_6x_2^2. \quad (4.3)$$

Este trabalho mostra que a dimensionalidade do vetor expandido $p_n(\mathbf{x})$ pode ser expressa em termos da ordem polinomial n e da dimensionalidade d do vetor de entrada \mathbf{x} como mostrado na Equação (4.4)

$$L_n^d = L_n^{d-1} + L_{n-1}^d \quad p/ \ d > 2, \quad (4.4)$$

onde:

$$L_n^2 = \frac{n^2 + 3n + 2}{2}; \quad \text{e}$$

$$L_{-1}^d = 0.$$

Relembrando do Capítulo 3 deste trabalho que a função discriminante classifica o padrão conforme o seu sinal algébrico.

O objetivo do classificador polinomial quando existem apenas duas classes é o mesmo da regra de decisão mostrada na Equação (4.5)

$$\text{Decide} \begin{cases} \omega_1, & \text{se } g(\mathbf{x}) > 0 \\ \omega_2, & \text{se } g(\mathbf{x}) < 0 \end{cases}. \quad (4.5)$$

Na Equação (4.5) no lugar de utilizar as desigualdades $g(\mathbf{x}) > 0$ para os padrões pertencentes a classe ω_1 e $g(\mathbf{x}) < 0$ para os padrões pertencentes a classe ω_2 , é mais simples trabalhar com as igualdades $g(\mathbf{x}) = 1$ para os padrões pertencentes a classe ω_1 e $g(\mathbf{x}) = -1$ para os padrões pertencentes a classe ω_2 . Então, o problema de encontrar a solução para um conjunto de desigualdades pode ser transformado em um problema de encontrar a solução para um conjunto de equações.

A notação de equações simultâneas é mais simples quando utiliza-se a notação matricial mostrada na Equação (4.6)

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1d} \\ x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Nd} \end{bmatrix}. \quad (4.6)$$

Na Equação (4.6) cada linha representa um vetor de características de um determinado padrão identificado e N é o número de padrões utilizados na fase de treinamento.

Para duas classes, N_1 indica o número de amostras pertencentes a classe ω_1 , N_2 representa o número de amostras pertencentes a classe ω_2 e $N = N_1 + N_2$.

Após o pré-processamento dos dados de entrada no classificador é realizada a expansão da base polinomial para cada amostra de acordo com a ordem polinomial estabelecida. Colocando todas as expansões em linha define-se a matriz expansão polinomial \mathbf{M} de dimensão $N \times L$, como mostrado na Equação (4.7)

$$\mathbf{M} = \begin{bmatrix} p_n(\mathbf{x}_1)^T \\ p_n(\mathbf{x}_2)^T \\ \vdots \\ p_n(\mathbf{x}_N)^T \end{bmatrix}. \quad (4.7)$$

Como para um vetor de características da classes ω_1 e ω_2 as saídas desejadas são 1 e -1 , respectivamente, o problema proposto na Equação (4.5) é encontrar o vetor de coeficientes \mathbf{a} que satisfaça a Equação (4.8)

$$\begin{pmatrix} p_n(\mathbf{x}_1)^T \\ p_n(\mathbf{x}_2)^T \\ \vdots \\ p_n(\mathbf{x}_N)^T \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_L \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_N \end{pmatrix}, \quad (4.8)$$

ou seja,

$$\mathbf{Ma} = \mathbf{b}, \quad (4.9)$$

onde \mathbf{b} é um vetor cujos elementos são todos constantes arbitrárias de valor 1 ou -1 , de acordo com a classe do padrão de entrada, isto é, no caso de duas classes pode-se considerar os N_1 primeiros elementos como pertencentes a classe ω_1 e os outros N_2 elementos como pertencentes a classe ω_2 .

Se a matriz \mathbf{M} é não singular, pode-se escrever a Equação (4.10)

$$\mathbf{a} = \mathbf{M}^{-1}\mathbf{b} \quad (4.10)$$

e obter uma solução formal para o problema. Porém, geralmente \mathbf{M} é retangular, isto é, com mais linhas do que colunas. Quando existem mais equações do que incógnitas, não existe uma única solução, portanto pode-se procurar um vetor que minimiza o erro entre \mathbf{Ma} e \mathbf{b} .

Definindo o vetor erro pela Equação (4.11)

$$\mathbf{e} = \mathbf{Ma} - \mathbf{b}, \quad (4.11)$$

uma aproximação é tentar minimizar o comprimento ao quadrado do vetor erro que é equivalente a minimizar a soma dos quadrados dos erros como mostrado na Equação (4.12)

$$J_s(\mathbf{a}) = \|\mathbf{Ma} - \mathbf{b}\|^2 = \sum_{i=1}^N (\mathbf{a}^T p_n(x_i) - b_i)^2. \quad (4.12)$$

O problema de minimizar a Equação (4.12) é um problema clássico conhecido na literatura como mínimos quadrados. Esse problema pode ser resolvido encontrando o gradiente da função critério na Equação (4.12) e igualando-o a zero na Equação (4.13)

$$\nabla J_s = \sum_{i=1}^N 2(\mathbf{a}^T p_n(x_i) - b_i)p_n(x_i) = 2\mathbf{M}^T(\mathbf{Ma} - \mathbf{b}). \quad (4.13)$$

Então, o problema de resolver $\mathbf{Ma} = \mathbf{b}$ foi convertido em resolver a Equação (4.14)

$$\mathbf{M}^T\mathbf{Ma} = \mathbf{M}^T\mathbf{b}. \quad (4.14)$$

A vantagem da Equação (4.14) é que a matriz $\mathbf{M}^T\mathbf{M}$ é uma matriz quadrada de dimensão $L \times L$ e geralmente ela é não singular. Como ela é não singular, pode-se encontrar uma única solução \mathbf{a} conforme a Equação (4.15)

$$\mathbf{a} = (\mathbf{M}^T\mathbf{M})^{-1}\mathbf{M}^T\mathbf{b} = \mathbf{M}^\dagger\mathbf{b}. \quad (4.15)$$

onde a matriz \mathbf{M}^\dagger de dimensão $L \times N$ como mostrado na Equação (4.16)

$$\mathbf{M}^\dagger = (\mathbf{M}^T \mathbf{M})^{-1} \mathbf{M}^T \quad (4.16)$$

é chamada de *pseudoinversa* de \mathbf{M} [42].

Retornando ao exemplo mostrado na Figura 4.1 onde o classificador linear de Fisher não conseguiu uma perfeita classificação dos dados, pode-se observar na Figura 4.4 que o classificador polinomial é realmente um método de classificação bastante eficiente, principalmente para os dados não linearmente separáveis. Nessa figura o classificador polinomial de ordem 5 resultou em uma perfeita separação das classes.

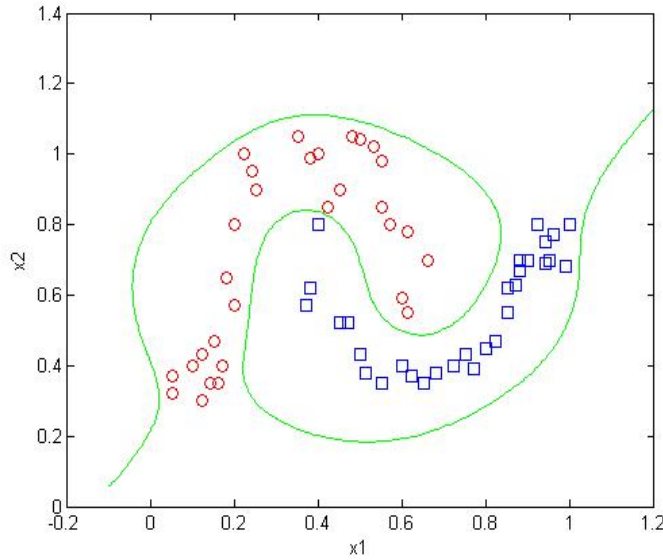


Figura 4.4: Classificação polinomial de duas classes utilizando uma base polinomial de ordem 5.

4.3 Classificação Polinomial em Várias Classes

Na seção 4.2 deste capítulo foi mostrado apenas a separação de duas classes. Nesta seção é mostrado como o classificador polinomial pode ser estendido na classificação em várias classes.

Não existe uma única maneira de estender todos os casos de duas classes para o caso de várias classes. Na seção 3.3 identificou-se um padrão pelo cálculo de c funções discriminantes

lineares como mostrado na Equação (4.17)

$$g_i(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + w_{i0} \quad i = 1, \dots, c, \quad (4.17)$$

e relacionou-se o vetor \mathbf{x} pertencente a classe correspondente ao maior valor de todas as funções discriminantes. Isso pode ser estendido e generalizado simplesmente para as funções polinomiais escrevendo a Equação (4.18)

$$g_i(\mathbf{x}) = \mathbf{a}_i^T p_n(\mathbf{x}) \quad i = 1, \dots, c, \quad (4.18)$$

onde \mathbf{x} é assumido como pertencente a classe ω_i se $g_i(\mathbf{x}) > g_j(\mathbf{x})$, para todo $j \neq i$.

A maneira mais simples de obter uma generalização natural dos procedimentos de mínimos quadrados para o caso de várias classes é considerar o problema como um conjunto de c problemas de duas classes. O i -ésimo problema é encontrar um vetor peso \mathbf{a}_i que é a solução do erro quadrático mínimo para as Equações (4.19)

$$\begin{aligned} \mathbf{a}_i^T p_n(\mathbf{x}) &= +1 & \forall \mathbf{x} \in \omega_i \\ \mathbf{a}_i^T p_n(\mathbf{x}) &= -1 & \forall \mathbf{x} \notin \omega_i. \end{aligned} \quad (4.19)$$

A solução pseudoinversa dos problemas dos mínimos quadrados para várias classes pode ser escrita de forma análoga ao caso de duas classes. Seja \mathbf{M} a matriz expansão polinomial de dimensão $N \times L$ das amostras de treinamento, que pode ser particionada como mostrado na Equação (4.20)

$$\mathbf{M} = \begin{bmatrix} \mathbf{P}_1 \\ \mathbf{P}_2 \\ \vdots \\ \mathbf{P}_c \end{bmatrix}, \quad (4.20)$$

com as amostras pertencentes a classe ω_i correspondendo as linhas de \mathbf{P}_i . Similarmente, a matriz \mathbf{A} de dimensão $L \times c$ dos vetores de coeficientes das funções da base polinomial, como mostrado na Equação (4.21)

$$\mathbf{A} = [\mathbf{a}_1 \ \mathbf{a}_2 \ \cdots \ \mathbf{a}_c] \quad (4.21)$$

e \mathbf{B} , de dimensão $N \times c$, como apresentado na Equação (4.22)

$$\mathbf{B} = [\mathbf{b}_1 \ \mathbf{b}_2 \ \cdots \ \mathbf{b}_c] \quad (4.22)$$

onde todos os elementos de \mathbf{b}_i são -1 exceto para aqueles na i -ésima linha, que são iguais a 1.

Então, a solução que minimiza o erro quadrático é obtida pela Equação (4.23)

$$\mathbf{A} = \mathbf{M}^\dagger \mathbf{B} \quad (4.23)$$

onde, como mencionado na seção 4.2 deste capítulo, \mathbf{M}^\dagger é a *pseudoinversa* de \mathbf{M} [42].

As Figuras 4.5 a 4.8 ilustram os exemplos de classificação polinomial para mais de duas classes. A Figura 4.5 mostra a classificação polinomial de três classes utilizando uma base polinomial de ordem 1.

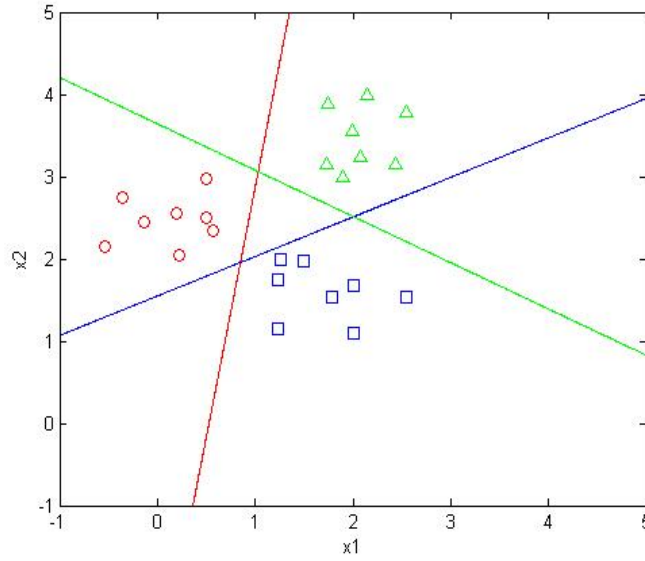


Figura 4.5: Classificação polinomial de três classes utilizando uma base polinomial de ordem 1.

Pode-se observar na Figura 4.5 que o classificador polinomial de grau 1 conseguiu classificar corretamente (100%) as três classes.

A Figura 4.6 ilustra a classificação polinomial de três classes utilizando uma base polinomial de ordem 2.

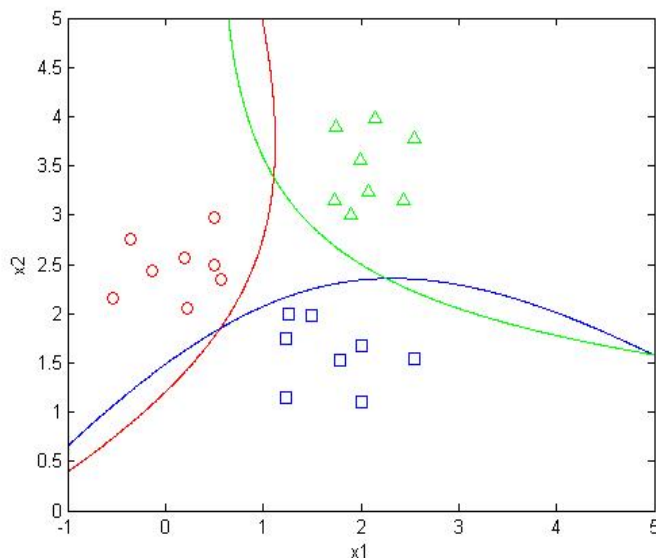


Figura 4.6: Classificação polinomial de três classes utilizando uma base polinomial de ordem 2.

Pode-se verificar na Figura 4.6, que além do classificador polinomial conseguir uma classificação exata para as três classes, ocorreu uma melhora na classificação em relação ao classificador polinomial de grau 1 utilizado na Figura 4.5, ou seja, ocorreu uma diminuição nas regiões de incerteza, isto é, nas regiões onde o classificador fica sujeito a uma decisão errada ou não consegue identificar a classe.

4.3. CLASSIFICAÇÃO POLINOMIAL EM VÁRIAS CLASSES

A Figura 4.7 mostra a classificação polinomial de quatro classes utilizando uma base polinomial de ordem 1.

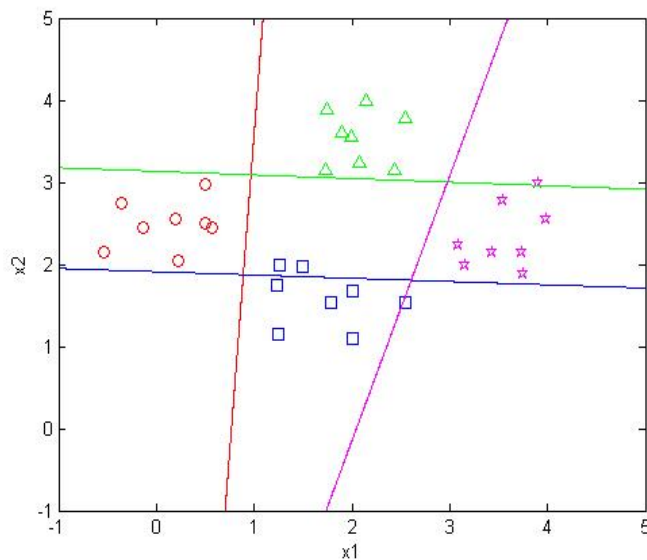


Figura 4.7: Classificação polinomial de quatro classes utilizando uma base polinomial de ordem 1.

Pode-se observar na Figura 4.7 que não foi possível uma classificação exata para as quatro classes, pois elas não são linearmente separáveis duas a duas.

A Figura 4.8 ilustra a classificação polinomial de quatro classes utilizando uma base polinomial de ordem 2.

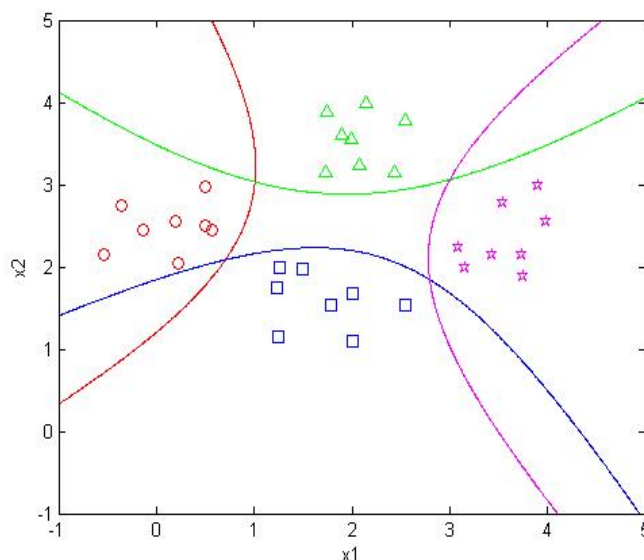


Figura 4.8: Classificação polinomial de quatro classes utilizando uma base polinomial de ordem 2.

Pode-se verificar na Figura 4.8 a importância do classificador polinomial, principalmente quando os dados não são linearmente separáveis, pois com uma base polinomial de ordem 2 foi possível encontrar uma classificação exata para as quatro classes.

4.4 Considerações Finais Deste Capítulo

Este capítulo descreveu o classificador polinomial de duas ou mais classes. Também foi realçada a importância desse método na classificação de dados, principalmente aqueles que são linearmente não separáveis.

O classificador polinomial apresenta um desempenho muito superior ao discriminante linear de Fisher, pois ele consegue mapear o espaço vetorial amostral onde seria impossível separá-las em um espaço de dimensão superior por meio da expansão polinomial das características dos padrões, na qual seja possível uma separação.

Deve-se valorizar também o desempenho desse classificador pela capacidade que ele tem de aumentar a ordem da base polinomial, podendo melhorar a superfície de separação

dos dados.

O próximo capítulo apresenta os resultados obtidos nos testes realizados com o classificador polinomial e a comparação desses resultados com os classificadores SVM e árvore de decisão.

Capítulo 5

Resultados Obtidos

5.1 Introdução

Este capítulo mostra os resultados obtidos nos testes realizados utilizando o classificador polinomial na classificação da base de dados IRIS [5] e na identificação de pixels parciais em ROIs de imagens mamomográficas. Com o objetivo de comparar os resultados obtidos com esse classificador, foram realizados também testes com os classificadores máquina de vetores suporte (SVM) e árvore de decisão utilizando o algoritmo J48.

5.2 Classificação Polinomial da Base de Dados Iris

Esta seção apresenta os resultados obtidos da aplicação dos classificadores polinomiais na base de dados IRIS, introduzida em 1936 por R. A. Fisher [5].

Os dados da base de dados IRIS contém quatro características de três espécies de flores Iris da família das Iridáceas. Essas características são:

- comprimento da sépala (x_1);
- largura da sépala (x_2);
- comprimento da pétala (x_3); e

- largura da pétala (x_4).

O classificador polinomial utiliza essas características para separar 3 tipos de flores: Iris-Virginica, Iris-Versicolor e Iris-Setosa. A base de dados da IRIS possui 50 flores de cada um desses tipos, portanto o número total de flores dessa base é igual a 150.

A Figura 5.1 mostra as imagens das flores setosa, versicolor e virginica da base de dados obtida por Fisher [5].



(a) Setosa

(b) Versicolor

(c) Virginica

Figura 5.1: Imagem das três flores Iris da base de dados obtida Fisher.

Um dos objetivos desta tese é mostrar como o classificador polinomial comporta-se ao variar o número de características e o grau da base polinomial.

A classificação de três classes foi realizada separando a classe de uma flor das outras duas flores. Por exemplo, no primeiro teste tem-se uma função polinomial separando a classe da flor setosa das flores versicolor e virginica. No segundo teste a classe da flor versicolor foi separada das classes das flores setosa e virginica. Finalmente, no terceiro teste a classe da flor virginica foi separada das classes das flores setosa e versicolor.

5.2.1 Classificação polinomial da base de dados Iris utilizando duas características

Neste primeiro momento são apresentados os resultados obtidos utilizando apenas duas características. Foram realizados todos os cruzamentos possíveis dois a dois, ou seja, utilizando a característica comprimento da sépala (x_1) versus largura da sépala (x_2), comprimento

da sépala (x_1) versus comprimento da pétala (x_3), comprimento da sépala (x_1) versus largura da pétala (x_4), largura da sépala (x_2) versus comprimento da pétala (x_3), largura da sépala (x_2) versus largura da pétala (x_4) e comprimento da pétala (x_3) versus largura da pétala (x_4).

A Figura 5.2 ilustra a distribuição do espaço de características do comprimento da sépala (x_1) versus a largura da sépala (x_2). Nessa figura os pontos vermelhos representam a classe da flor setosa, os pontos azuis a classe da flor versicolor e os pontos verdes representam a classe da flor virginica.

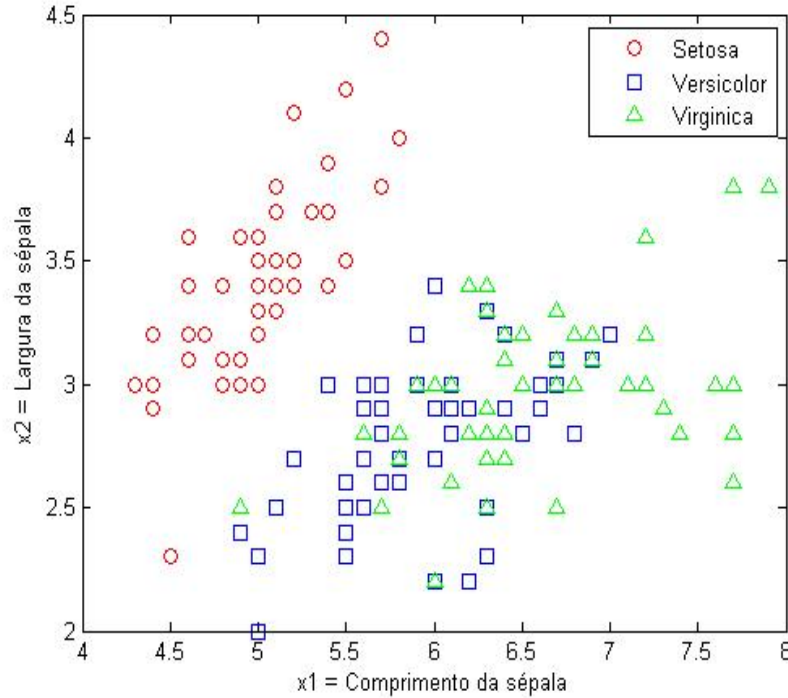


Figura 5.2: Distribuição das características das flores setosa, versicolor e virginica no espaço de características comprimento da sépala (x_1) versus largura da sépala (x_2).

Os resultados dos testes com as características comprimento da sépala (x_1) e largura da sépala (x_2) são mostrados na Tabela 5.1, variando o grau da base polinomial de 1 a 50. Foram omitidos alguns resultados por não apresentarem melhoria na separação das classes.

Tabela 5.1: Resultados obtidos da classificação polinomial utilizando as características comprimento da sépala (x_1) e largura da sépala (x_2).

Grau do polinômio	Setosa		Versicolor		Virginica	
	Erro	% Acerto	Erro	% Acerto	Erro	% Acerto
1	1	99,33	41	72,67	29	80,67
2	1	99,33	38	74,67	29	80,67
3	1	99,33	34	77,33	27	82,00
4	0	100,00	30	80,00	26	82,67
5	0	100,00	25	83,33	23	84,67
6	0	100,00	24	84,00	24	84,00
7	0	100,00	26	82,67	24	84,00
8	0	100,00	26	82,67	27	82,00
9	0	100,00	48	68,00	45	70,00
10	0	100,00	26	82,67	26	82,67
11	0	100,00	26	82,67	25	83,33
12	0	100,00	26	82,67	24	84,00
13	0	100,00	26	82,67	26	82,67
14	0	100,00	26	82,67	25	83,33
15	0	100,00	27	82,00	27	82,00
20	15	90,00	33	78,00	49	67,33
30	31	79,33	53	64,67	33	78,00
40	0	100,00	23	84,67	23	84,67
50	1	99,33	50	66,67	43	71,33

Pode-se observar na Tabela 5.1 que as 50 (100%) flores setosa foram separadas corretamente utilizando a base polinomial de grau 4 a 15 e 40. E as 27 (84,67%) flores de cada um dos tipos versicolor e virginica apresentaram a melhor classificação utilizando a base polinomial de grau 40.

É importante salientar na Tabela 5.1 que ao realizar a variação do grau da base polinomial ocorreu uma perda na separação da classe setosa nos graus 20 e 30. Isso ocorreu devido a oscilação da função polinomial no espaço de características.

A Figura 5.3 mostra a curva de nível do classificador polinomial de grau 40 que obteve a melhor separação entre as 3 classes utilizando as características comprimento da sépala (x_1) e largura da sépala (x_2). O desempenho desse classificador foi de 100% para a flor setosa e 84,67% para as flores versicolor e virginica.

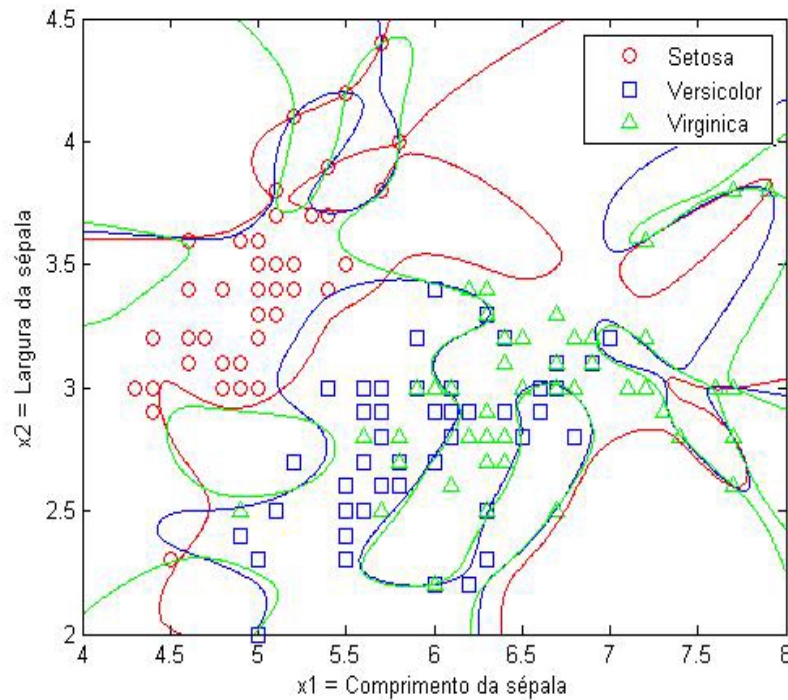


Figura 5.3: Curva de nível do classificador polinomial de grau 40 para as flores setosa, versicolor e virginica.

A Tabela 5.2 mostra os resultados obtidos com os classificadores SVM e árvore de decisão (J48) utilizando as características comprimento da sépala (x_1) e largura da sépala (x_2) da base de dados IRIS.

5.2. CLASSIFICAÇÃO POLINOMIAL DA BASE DE DADOS IRIS

Tabela 5.2: Resultados obtidos pelos classificadores SVM e árvore de decisão (J48) utilizando as características comprimento da sépala (x_1) e largura da sépala (x_2).

Classificador	Setosa		Versicolor		Virginica	
	Erro	% Acerto	Erro	% Acerto	Erro	% Acerto
SVM	1	99,33	12	92,00	17	88,67
J48	1	99,33	15	90,00	13	91,33

Pode-se verificar na Tabela 5.2 que o classificador árvore de decisão separou corretamente 49 (99,33%) flores setosa, 35 (90,00%) flores versicolor e 37 (91,33%) flores virginica, enquanto o classificador SVM separou corretamente 49 (99,33%) flores setosa, 38 (92,00%) flores versicolor e 33 (88,67%) flores virginica.

A Figura 5.4 ilustra um gráfico comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando as características comprimento da sépala (x_1) e largura da sépala (x_2).

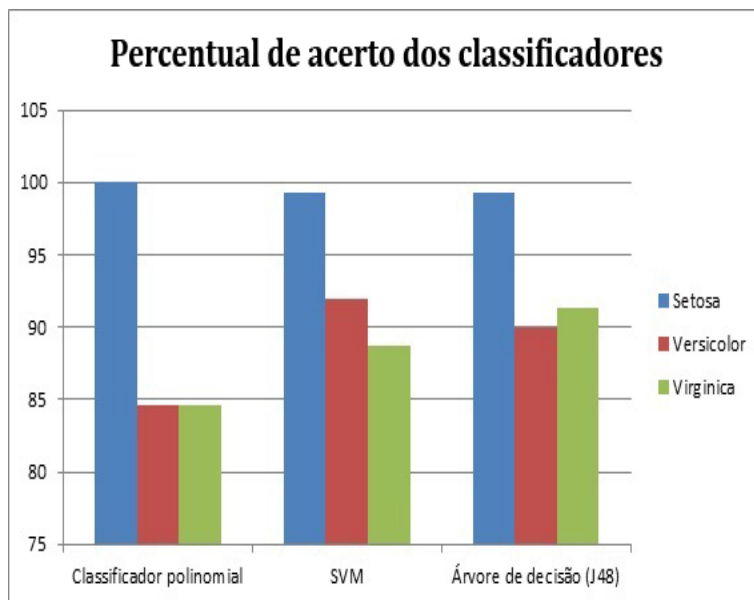


Figura 5.4: Comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando as características comprimento da sépala (x_1) e largura da sépala (x_2)

Ao comparar os resultados obtidos nas Tabelas 5.1 e 5.2 e na Figura 5.4, pode-se observar que o classificador árvore de decisão (J48) apresentou melhores resultados do que

os classificadores SVM e polinomial.

A Figura 5.5 ilustra a distribuição das flores setosa, versicolor e virginica no espaço de características comprimento da sépala (x_1) versus comprimento da pétala (x_3).

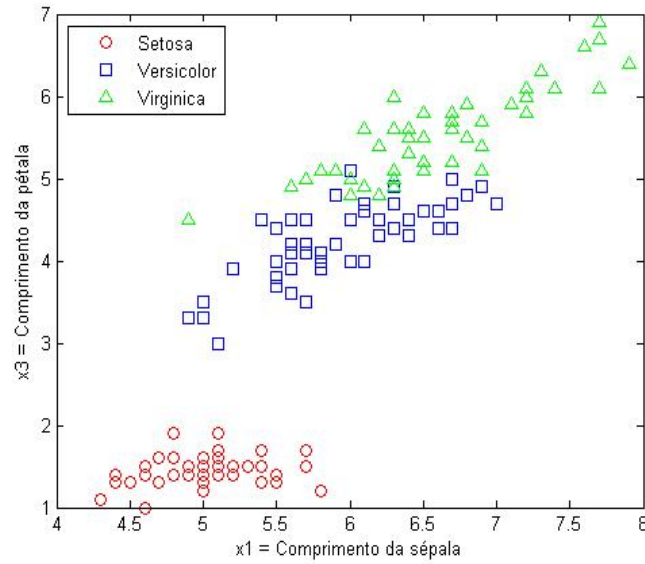


Figura 5.5: Distribuição das características das flores setosa, versicolor e virginica no espaço de características comprimento da sépala (x_1) versus comprimento da pétala (x_3).

A Tabela 5.3 mostra os resultados obtidos nos testes realizados com as características comprimento da sépala (x_1) e comprimento da pétala (x_3) variando o grau da base polinomial de 1 a 50.

5.2. CLASSIFICAÇÃO POLINOMIAL DA BASE DE DADOS IRIS

Tabela 5.3: Resultados obtidos da classificação polinomial utilizando as características comprimento da sépala (x_1) e comprimento da pétala (x_3).

Grau do polinômio	Setosa		Versicolor		Virginica	
	Erro	% Acerto	Erro	% Acerto	Erro	% Acerto
1	0	100,00	58	61,33	14	90,67
2	0	100,00	10	93,33	7	95,33
3	0	100,00	7	95,33	6	96,00
4	0	100,00	5	96,67	6	96,00
5	0	100,00	5	96,67	5	96,67
6	0	100,00	30	80,00	28	81,33
7	0	100,00	5	96,67	5	96,67
8	0	100,00	5	96,67	5	96,67
9	0	100,00	4	97,33	4	97,33
10	0	100,00	3	98,00	3	98,00
11	0	100,00	6	96,00	4	97,33
12	0	100,00	3	98,00	4	97,33
13	0	100,00	4	97,33	4	97,33
14	0	100,00	4	97,33	4	97,33
15	0	100,00	6	96,00	4	97,33
20	0	100,00	4	97,33	4	97,33
29	0	100,00	2	98,67	2	98,67
30	0	100,00	9	94,00	8	94,67
40	0	100,00	8	94,67	8	94,67
50	0	100,00	6	96,00	12	92,00

Pode-se observar na Tabela 5.3 que as 50 (100%) flores setosa foram separadas corretamente utilizando a base polinomial de grau 1 ou superior, mostrando que essa flor é linearmente separável das flores versicolor e virginica. E as 48 (98,67%) flores de cada um

dos tipos versicolor e virginica apresentaram a melhor classificação utilizando a base polinomial de grau 29.

A Figura 5.6 mostra a curva de nível do classificador polinomial de grau 29 que obteve a melhor separação entre as 3 classes utilizando as características comprimento da sépala (x_1) e comprimento da pétala (x_3). O desempenho desse classificador foi de 100% para a flor setosa e 98,67% para as flores versicolor e virginica.

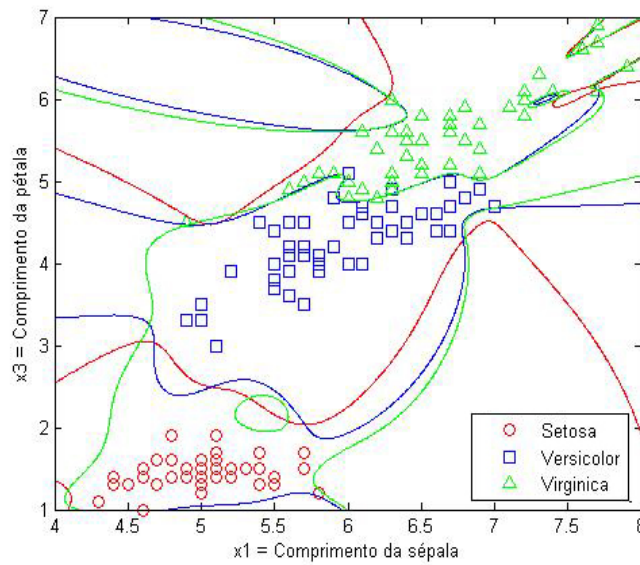


Figura 5.6: Curva de nível do classificador polinomial de grau 29 para as flores setosa, versicolor e virginica.

A Tabela 5.4 mostra os resultados obtidos com os classificadores SVM e árvore de decisão (J48) utilizando as características comprimento da sépala (x_1) e comprimento da pétala (x_3) da base de dados IRIS.

Tabela 5.4: Resultados obtidos pelos classificadores SVM e árvore de decisão (J48) utilizando as características comprimento da sépala (x_1) e comprimento da pétala (x_3).

Classificador	Setosa		Versicolor		Virginica	
	Erro	% Acerto	Erro	% Acerto	Erro	% Acerto
SVM	0	100,00	7	95,33	10	93,33
J48	0	100,00	6	96,00	1	99,33

Pode-se verificar na Tabela 5.4 que os classificadores SVM e árvore de decisão (J48) separaram corretamente as 50 (100%) flores setosas. O classificador SVM separou corretamente 43 (95,33%) flores versicolor e 40 (93,33%) flores virginica, enquanto o classificador árvore de decisão classificou corretamente 44 (96,00%) flores versicolor e 49 (99,33%) flores virginica.

A Figura 5.7 ilustra um gráfico comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando as características comprimento da sépala (x_1) e comprimento da pétala (x_3).

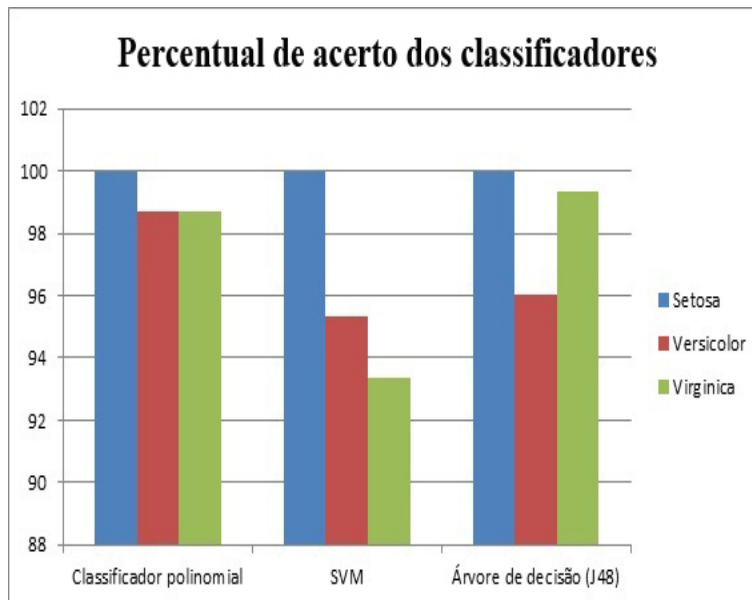


Figura 5.7: Comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando as características comprimento da sépala (x_1) e comprimento da pétala (x_3)

Ao comparar os resultados obtidos nas Tabelas 5.3 e 5.4 e na Figura 5.7, pode-se observar que o classificador polinomial apresentou melhores resultados do que os classificadores SVM e árvore de decisão (J48).

A Figura 5.8 ilustra a distribuição das flores setosa, versicolor e virginica no espaço de características comprimento da sépala (x_1) versus largura da pétala (x_4).

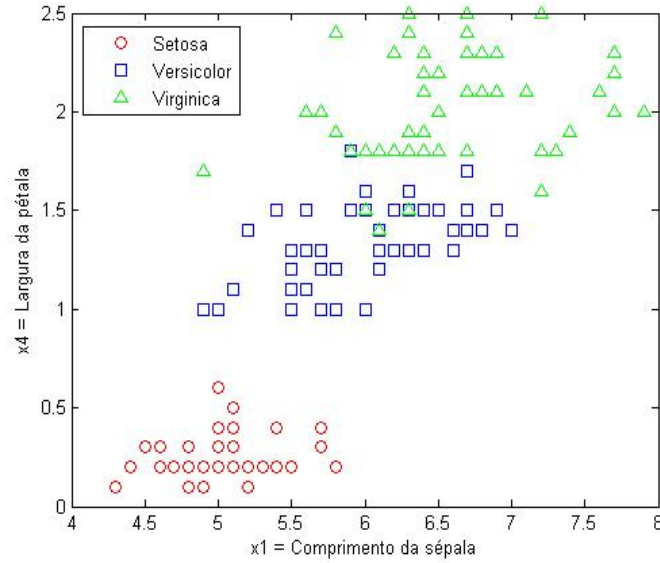


Figura 5.8: Distribuição das características das flores setosa, versicolor e virginica no espaço de características comprimento da sépala (x_1) versus largura da pétala (x_4).

A Tabela 5.5 mostra os resultados obtidos nos testes realizados com as características comprimento da sépala (x_1) e largura da pétala (x_4), variando o grau da base polinomial de 1 a 50.

5.2. CLASSIFICAÇÃO POLINOMIAL DA BASE DE DADOS IRIS

Tabela 5.5: Resultados obtidos da classificação polinomial utilizando as características comprimento da sépala (x_1) e largura da pétala (x_4).

Grau do polinômio	Setosa		Versicolor		Virginica	
	Erro	% Acerto	Erro	% Acerto	Erro	% Acerto
1	0	100,00	50	66,67	8	94,67
2	0	100,00	14	90,67	5	96,67
3	0	100,00	8	94,67	6	96,00
4	0	100,00	6	96,00	6	96,00
5	0	100,00	5	96,67	5	96,67
6	0	100,00	5	96,67	5	96,67
7	0	100,00	5	96,67	4	97,33
8	0	100,00	4	97,33	4	97,33
9	0	100,00	4	97,33	4	97,33
10	0	100,00	4	97,33	4	97,33
11	0	100,00	4	97,33	4	97,33
12	0	100,00	4	97,33	4	97,33
13	0	100,00	5	96,67	4	97,33
14	0	100,00	4	97,33	4	97,33
15	0	100,00	9	94,00	10	93,33
20	0	100,00	6	96,00	5	96,67
25	4	97,33	40	73,33	44	70,67
30	0	100,00	4	97,33	4	97,33
40	0	100,00	20	86,67	23	84,67
50	0	100,00	4	97,33	5	96,67

Pode-se observar na Tabela 5.5 que as 50 (100%) flores setosas foram separadas corretamente utilizando a base polinomial de grau 1 ou superior, exceto para o grau 25, mostrando que essa flor é linearmente separável das flores versicolor e virginica. E as 46 (97,33%) flores

de cada um dos tipos versicolor e virginica apresentaram o melhor resultado usando a base polinomial de grau 8. Para essas duas flores, outras bases polinomiais de graus superiores apresentaram resultados iguais, porém escolheu-se o polinômio de grau menor que apresentou o melhor resultado porque o tempo de processamento dele foi menor. Para exemplificar, com as bases polinomiais de grau 8 e 14, os tempos de processamento foram 0,0151 e 0,0202 segundos, respectivamente, utilizando neste trabalho os recursos computacionais conforme mostrado na Tabela 1.1 do Capítulo 1.

A Figura 5.9 mostra a curva de nível do classificador polinomial de grau 8 que obteve a melhor separação entre as 3 classes utilizando as características comprimento da sépala (x_1) e largura da pétala (x_4). O desempenho desse classificador com esse grau foi de 100% para a flor setosa e 97,33% para as flores versicolor e virginica.

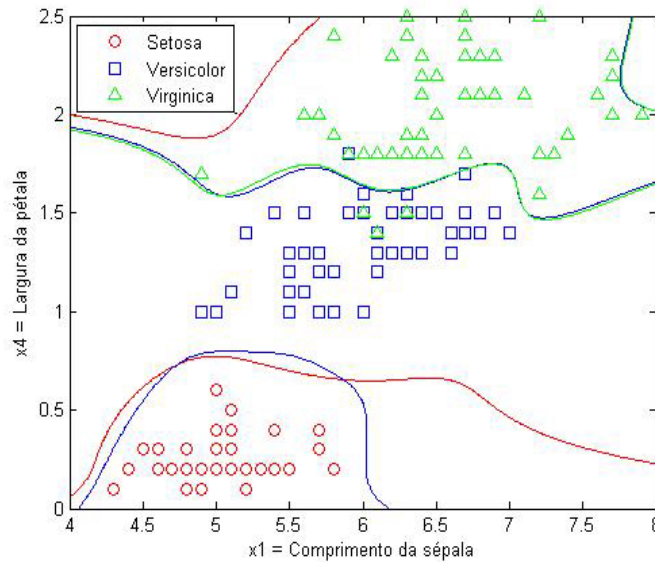


Figura 5.9: Curva de nível do classificador polinomial de grau 8 para as flores setosa, versicolor e virginica.

A Tabela 5.6 mostra os resultados obtidos com os classificadores SVM e árvore de decisão (J48) utilizando as características comprimento da sépala (x_1) e largura da pétala (x_4) da base de dados IRIS.

Tabela 5.6: Resultados obtidos pelos classificadores SVM e árvore de decisão (J48) utilizando as características comprimento da sépala (x_1) e largura da pétala (x_4).

Classificador	Setosa		Versicolor		Virginica	
	Erro	% Acerto	Erro	% Acerto	Erro	% Acerto
SVM	0	100,00	2	98,67	4	97,33
J48	0	100,00	1	99,33	5	96,67

Pode-se verificar na Tabela 5.6 que os classificadores SVM e árvore de decisão (J48) separaram corretamente as 50 (100%) flores setosas. O classificador SVM separou corretamente 48 (98,67%) flores versicolor e 46 (97,33%) flores virginica, enquanto o classificador árvore de decisão classificou corretamente 49 (99,33%) flores versicolor e 45 (96,67%) flores virginica.

A Figura 5.10 ilustra um gráfico comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando as características comprimento da sépala (x_1) e largura da pétala (x_4).

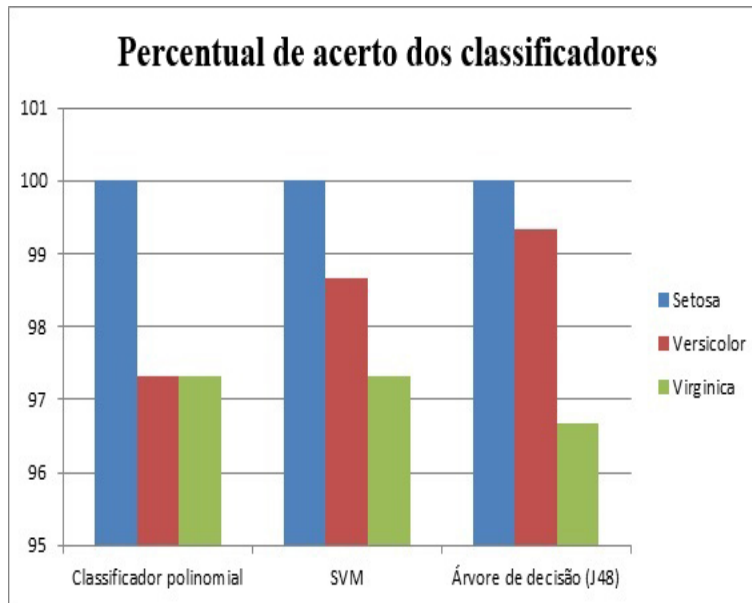


Figura 5.10: Comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando as características comprimento da sépala (x_1) e largura da pétala (x_4).

Ao comparar os resultados obtidos nas Tabelas 5.5 e 5.6 e na Figura 5.10, pode-se ob-

servar que os classificadores SVM e árvore de decisão (J48) apresentaram melhores resultados do que o classificador polinomial.

A Figura 5.11 ilustra a distribuição das flores setosa, versicolor e virginica no espaço de características largura da sépala (x_2) versus comprimento da pétala (x_3).

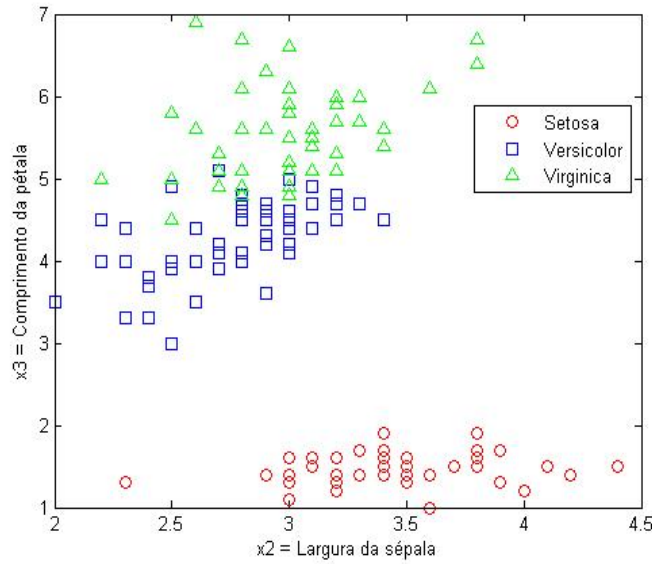


Figura 5.11: Distribuição das características das flores setosa, versicolor e virginica no espaço de características largura da sépala (x_2) versus comprimento da pétala (x_3).

A Tabela 5.7 mostra os resultados obtidos nos testes realizados com as características largura da sépala (x_2) e comprimento da pétala (x_3) variando o grau da base polinomial de 1 a 50.

5.2. CLASSIFICAÇÃO POLINOMIAL DA BASE DE DADOS IRIS

Tabela 5.7: Resultados obtidos da classificação polinomial com as características largura da sépala (x_2) e comprimento da pétala (x_3).

Grau do polinômio	Setosa		Versicolor		Virginica	
	Erro	% Acerto	Erro	% Acerto	Erro	% Acerto
1	0	100,00	41	72,67	17	88,67
2	0	100,00	10	93,33	7	95,33
3	0	100,00	10	93,33	6	96,00
4	0	100,00	7	95,33	7	95,33
5	0	100,00	7	95,33	6	96,00
6	0	100,00	6	96,00	6	96,00
7	0	100,00	7	95,33	7	95,33
8	0	100,00	6	96,00	6	96,00
9	0	100,00	6	96,00	6	96,00
10	0	100,00	6	96,00	6	96,00
11	0	100,00	6	96,00	6	96,00
12	0	100,00	6	96,00	6	96,00
13	0	100,00	7	95,33	7	95,33
14	0	100,00	6	96,00	6	96,00
15	0	100,00	7	95,33	7	95,33
17	0	100,00	5	96,67	5	96,67
20	0	100,00	9	94,00	9	94,00
30	0	100,00	5	96,67	5	96,67
40	0	100,00	17	88,67	23	84,67
50	0	100,00	23	84,67	21	86,00

Pode-se observar na Tabela 5.7 que as 50 (100%) flores setosas foram separadas corretamente utilizando a base polinomial de grau 1 ou superior, mostrando que essa flor é linearmente separável das flores versicolor e virginica. E as 45 (96,67%) flores de cada um

dos tipos versicolor e virginica apresentaram o melhor resultado usando a base polinomial de grau 17. Para essas duas flores, outras bases polinomiais de graus superiores apresentaram resultados iguais, porém escolheu-se o polinômio de grau menor que apresentou o melhor resultado porque o tempo de processamento dele foi menor.

A Figura 5.12 mostra a curva de nível do classificador polinomial de grau 17 que obteve a melhor separação entre as 3 classes utilizando as características largura da sépala (x_2) e comprimento da pétala (x_3). O desempenho do classificador polinomial de grau 17 foi de 100% para a flor setosa e 96,67% para as flores versicolor e virginica.

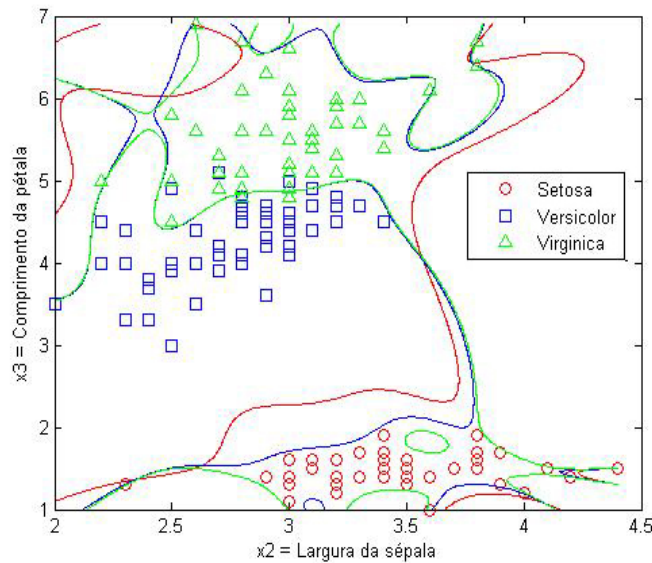


Figura 5.12: Curva de nível do classificador polinomial de grau 17 para as flores setosa, versicolor e virginica.

A Tabela 5.8 mostra os resultados obtidos com os classificadores SVM e árvore de decisão (J48) utilizando as características largura da sépala (x_2) e comprimento da pétala (x_3) da base de dados IRIS.

5.2. CLASSIFICAÇÃO POLINOMIAL DA BASE DE DADOS IRIS

Tabela 5.8: Resultados obtidos pelos classificadores SVM e árvore de decisão (J48) utilizando as características largura da sépala (x_2) e comprimento da pétala (x_3).

Classificador	Setosa		Versicolor		Virginica	
	Erro	% Acerto	Erro	% Acerto	Erro	% Acerto
SVM	0	100,00	3	98,00	8	94,67
J48	0	100,00	6	96,00	1	99,33

Pode-se verificar na Tabela 5.8 que os classificadores SVM e árvore de decisão (J48) separaram corretamente as 50 (100%) flores setosas. O classificador SVM separou corretamente 47 (98,00%) flores versicolor e 42 (94,67%) flores virginica, enquanto o classificador árvore de decisão classificou corretamente 44 (96,00%) flores versicolor e 49 (99,33%) flores virginica.

A Figura 5.13 ilustra um gráfico comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando as características largura da sépala (x_2) e comprimento da pétala (x_3).

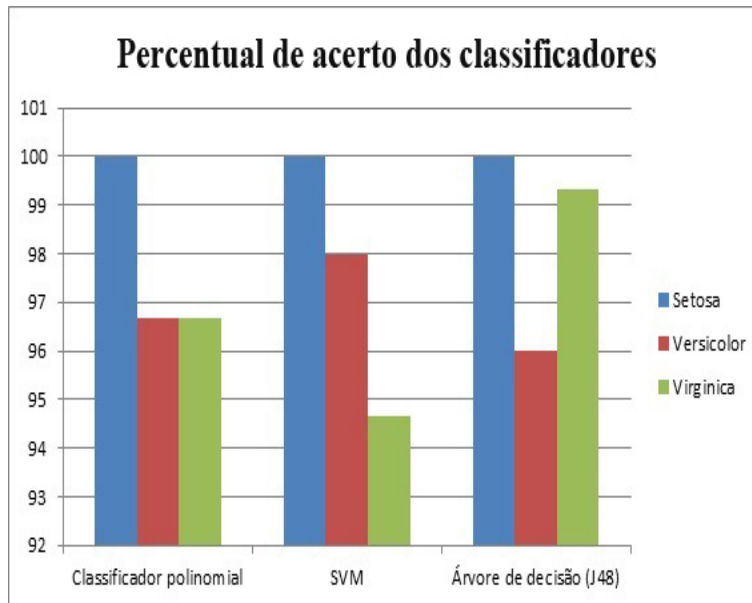


Figura 5.13: Comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando as características largura da sépala (x_2) e comprimento da pétala (x_3).

Ao comparar os resultados obtidos nas Tabelas 5.7 e 5.8 e na Figura 5.13, pode-se

observar que o classificador árvore de decisão (J48) apresentou um resultado um pouco melhor do que os classificadores SVM e polinomial.

A Figura 5.14 ilustra a distribuição das flores setosa, versicolor e virginica no espaço de características largura da sépala (x_2) versus largura da pétala (x_4).

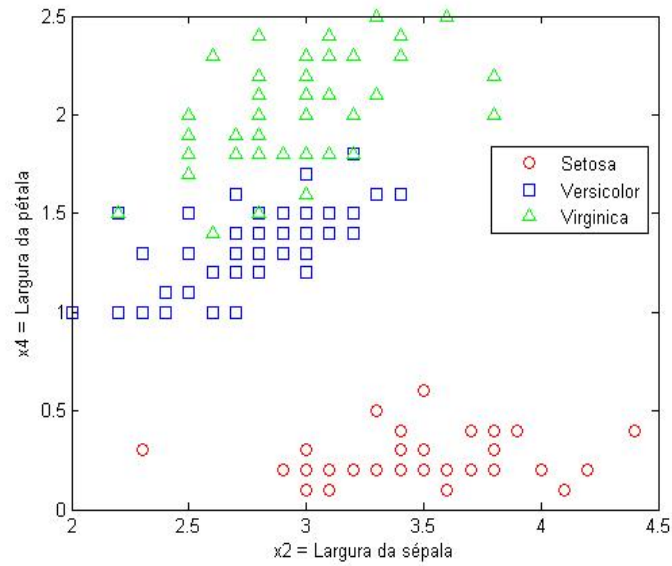


Figura 5.14: Distribuição das características das flores setosa, versicolor e virginica no espaço de características largura da sépala (x_2) versus largura da pétala (x_4).

A Tabela 5.9 mostra os resultados obtidos nos testes realizados com as características largura da sépala (x_2) e largura da pétala (x_4) variando o grau da base polinomial de 1 a 50.

5.2. CLASSIFICAÇÃO POLINOMIAL DA BASE DE DADOS IRIS

Tabela 5.9: Resultados obtidos da classificação polinomial utilizando as características largura da sépala (x_2) e largura da pétala (x_4).

Grau do polinômio	Setosa		Versicolor		Virginica	
	Erro	% Acerto	Erro	% Acerto	Erro	% Acerto
1	1	99,33	38	74,67	9	94,00
2	0	100,00	8	94,67	6	96,00
3	0	100,00	6	96,00	6	96,00
4	0	100,00	6	96,00	7	95,33
5	0	100,00	7	95,33	7	95,33
6	0	100,00	7	95,33	7	95,33
7	0	100,00	7	95,33	7	95,33
8	0	100,00	7	95,33	7	95,33
9	0	100,00	7	95,33	7	95,33
10	0	100,00	7	95,33	7	95,33
11	0	100,00	7	95,33	7	95,33
12	0	100,00	7	95,33	7	95,33
13	0	100,00	7	95,33	7	95,33
14	0	100,00	6	96,00	6	96,00
15	0	100,00	7	95,33	7	95,33
20	0	100,00	6	96,00	6	96,00
30	0	100,00	6	96,00	6	96,00
35	1	99,33	40	73,33	37	75,33
40	0	100,00	13	91,33	14	90,67
50	0	100,00	6	96,00	6	96,00

Pode-se observar na Tabela 5.9 que as 50 (100%) flores setosas foram separadas corretamente utilizando a base polinomial de grau 2 ou superior, exceto para o grau 35. E as 44 (96,00%) flores de cada um dos tipos versicolor e virginica apresentaram o melhor resultado usando a base polinomial de grau 3. Para essas duas flores, outras bases polinomiais de graus

superiores apresentaram resultados iguais, porém escolheu-se o polinômio de grau menor que apresentou o melhor resultado porque o tempo de processamento dele foi menor.

A Figura 5.15 mostra a curva de nível do classificador polinomial de grau 3 que obteve a melhor separação entre as 3 classes utilizando as características largura da sépala (x_2) e largura da pétala (x_4). O desempenho desse classificador com esse grau foi de 100% para a flor setosa e 96% para as flores versicolor e virginica.

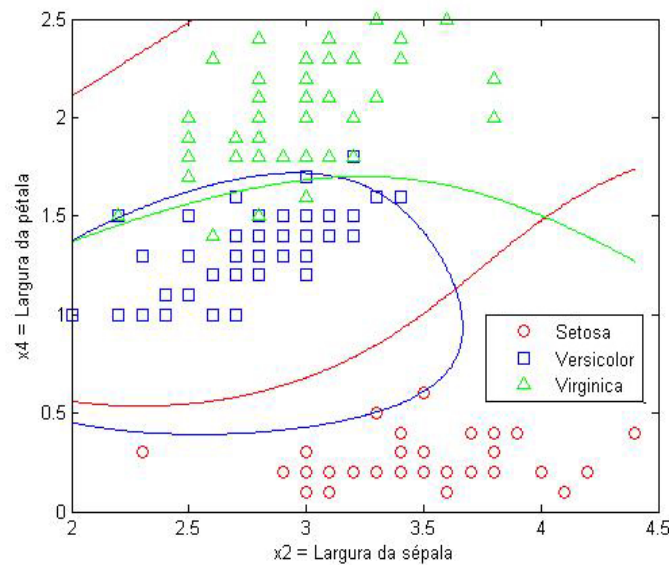


Figura 5.15: Curva de nível do classificador polinomial de grau 3 para as flores setosa, versicolor e virginica.

A Tabela 5.10 mostra os resultados obtidos com os classificadores SVM e árvore de decisão (J48) utilizando as características largura da sépala (x_2) e largura da pétala (x_4) da base de dados IRIS.

Tabela 5.10: Resultados obtidos pelos classificadores SVM e árvore de decisão (J48) utilizando as características largura da sépala (x_2) e largura da pétala (x_4).

Classificador	Setosa		Versicolor		Virginica	
	Erro	% Acerto	Erro	% Acerto	Erro	% Acerto
SVM	0	100,00	1	99,33	5	96,67
J48	0	100,00	1	99,33	5	96,67

Pode-se verificar na Tabela 5.10 que os classificadores SVM e árvore de decisão (J48) separaram corretamente as 50 (100%) flores setosas. Esse classificadores separaram corretamente 49 (99,33%) flores versicolor e 45 (96,67%) flores virginica.

A Figura 5.16 ilustra um gráfico comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando as características largura da sépala (x_2) e largura da pétala (x_4).

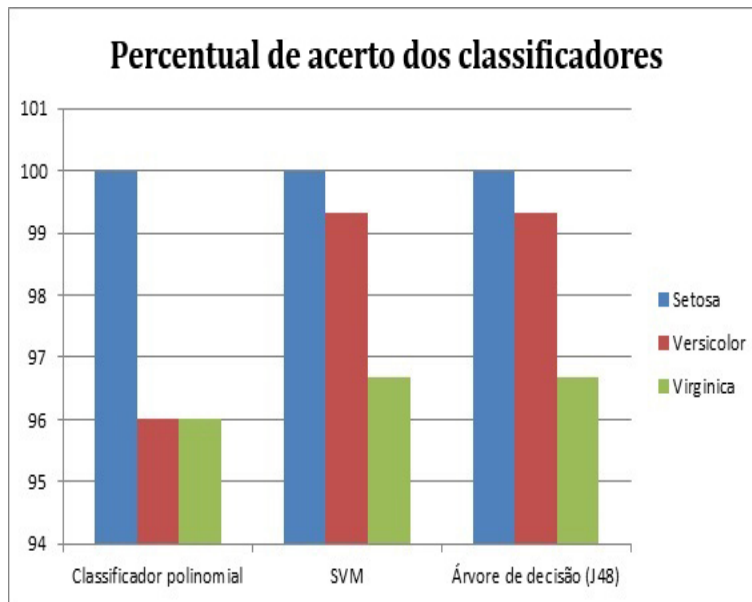


Figura 5.16: Comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando as características largura da sépala (x_2) e largura da pétala (x_4) .

Ao comparar os resultados obtidos nas Tabelas 5.9 e 5.10 e na Figura 5.16, pode-se observar que os classificadores SVM e árvore de decisão (J48) apresentaram um resultado um pouco melhor do que o classificador polinomial.

A Figura 5.17 ilustra a distribuição das flores setosa, versicolor e virginica no espaço de características comprimento da pétala (x_3) versus largura da pétala (x_4).

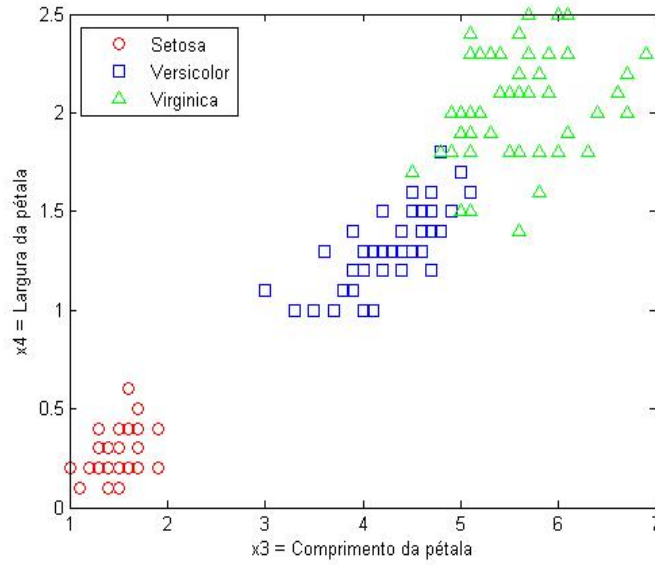


Figura 5.17: Distribuição das características das flores setosa, versicolor e virginica no espaço de características comprimento da pétala (x_3) versus largura da pétala (x_4).

A Tabela 5.11 mostra os resultados obtidos nos testes realizados com as características comprimento da pétala (x_3) e largura da pétala (x_4) variando o grau da base polinomial de 1 a 50.

5.2. CLASSIFICAÇÃO POLINOMIAL DA BASE DE DADOS IRIS

Tabela 5.11: Resultados obtidos da classificação polinomial com as características comprimento da pétala (x_3) e largura da pétala (x_4).

Grau do polinômio	Setosa		Versicolor		Virginica	
	Erro	% Acerto	Erro	% Acerto	Erro	% Acerto
1	1	99,33	54	64,00	8	94,67
2	0	100,00	8	94,67	7	95,33
3	0	100,00	4	97,33	3	98,00
4	0	100,00	4	97,33	4	97,33
5	0	100,00	3	98,00	3	98,00
6	0	100,00	3	98,00	3	98,00
7	0	100,00	2	98,67	3	98,00
8	0	100,00	2	98,67	2	98,67
9	0	100,00	7	95,33	8	94,67
10	0	100,00	5	96,67	6	96,00
11	0	100,00	2	98,67	2	98,67
12	0	100,00	2	98,67	2	98,67
13	0	100,00	2	98,67	2	98,67
14	0	100,00	2	98,67	2	98,67
15	0	100,00	15	90,00	9	94,00
20	0	100,00	2	98,67	3	98,00
30	0	100,00	25	83,33	17	88,67
31	1	99,33	27	82,00	32	78,67
34	1	99,33	41	72,67	25	83,33
40	0	100,00	47	68,67	61	59,33
44	1	99,33	23	84,67	24	84,00
50	0	100,00	42	72,00	15	90,00

Pode-se observar na Tabela 5.11 que as 50 (100%) flores setosas foram separadas corretamente utilizando a base polinomial de grau 2 ou superior, exceto nos graus 31, 34 e 44. E as 48 (98,67%) flores de cada um dos tipos versicolor e virginica apresentaram o melhor resultado usando a base polinomial de grau 8. Para essas duas flores, outras bases polinomiais de graus superiores apresentaram resultados iguais, porém escolheu-se o polinômio de grau menor que apresentou o melhor resultado porque o tempo de processamento foi menor.

A Figura 5.18 mostra a curva de nível do classificador polinomial de grau 8 que obteve a melhor separação entre as 3 classes utilizando as características comprimento da pétala (x_3) e largura da pétala (x_4). O desempenho do classificador polinomial de grau 3 foi de 100% para a flor setosa e 98,67% para as flores versicolor e virginica.

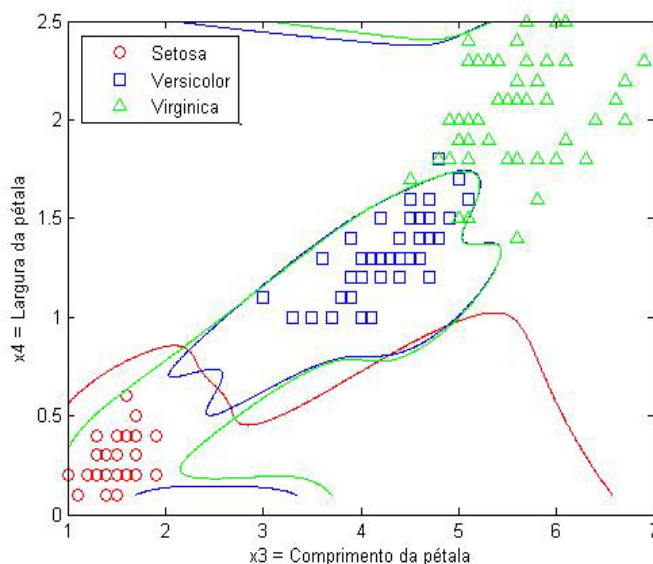


Figura 5.18: Curva de nível do classificador polinomial de grau 8 para as flores setosa, versicolor e virginica.

A Tabela 5.12 mostra os resultados obtidos com os classificadores SVM e árvore de decisão (J48) utilizando as características comprimento da pétala (x_3) e largura da pétala (x_4) da base de dados IRIS.

5.2. CLASSIFICAÇÃO POLINOMIAL DA BASE DE DADOS IRIS

Tabela 5.12: Resultados obtidos pelos classificadores SVM e árvore de decisão (J48) utilizando as características comprimento da pétala (x_3) e largura da pétala (x_4).

Classificador	Setosa		Versicolor		Virginica	
	Erro	% Acerto	Erro	% Acerto	Erro	% Acerto
SVM	0	100,00	2	98,67	4	97,33
J48	0	100,00	1	99,33	2	98,67

Pode-se verificar na Tabela 5.12 que os classificadores SVM e árvore de decisão (J48) separaram corretamente as 50 (100%) flores setosas. O classificador SVM separou corretamente 48 (98,67%) flores versicolor e 46 (97,33%) flores virginica. Enquanto o classificador árvore de decisão (J48) separou corretamente 49 (99,33%) flores versicolor e 48 (98,67%) flores virginica.

A Figura 5.19 ilustra um gráfico comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando as características comprimento da pétala (x_3) e largura da pétala (x_4).

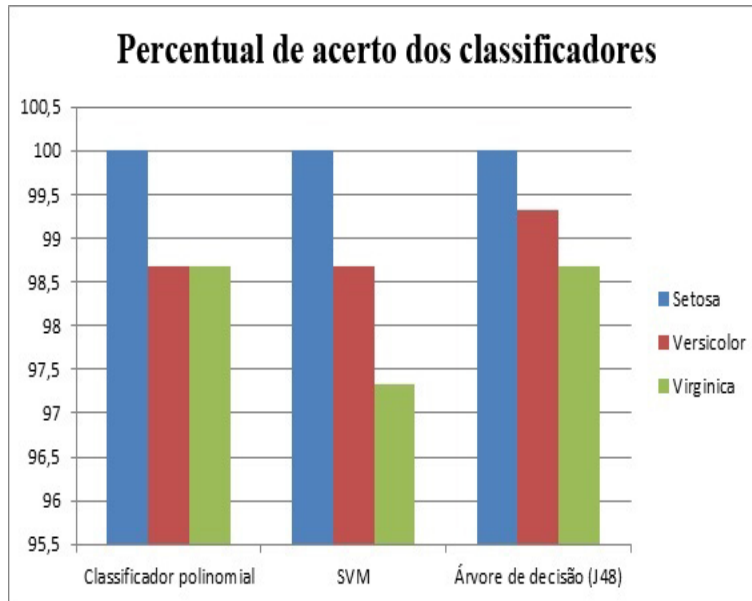


Figura 5.19: Comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando as características comprimento da pétala (x_3) e largura da pétala (x_4).

Ao comparar os resultados obtidos nas Tabelas 5.11 e 5.12 e na Figura 5.19, pode-

se observar que o classificador árvore de decisão (J48) apresentou um resultado um pouco melhor do que o classificador polinomial.

A Tabela 5.13 mostra o resumo dos resultados obtidos da classificação polinomial utilizando duas características.

Tabela 5.13: Resumo dos resultados obtidos da classificação polinomial utilizando duas características: comprimento da sépala (x_1), largura da sépala (x_2), comprimento da pétala (x_3) e largura da pétala (x_4).

Características	Grau	Setosa		Versicolor		Virginica	
		Erro	% Acerto	Erro	% Acerto	Erro	% Acerto
x_1 versus x_2	40	0	100,00	23	84,67	23	84,67
x_1 versus x_3	29	0	100,00	2	98,67	2	98,67
x_1 versus x_4	8	0	100,00	4	97,33	4	97,33
x_2 versus x_3	17	0	100,00	5	96,67	5	96,67
x_2 versus x_4	3	0	100,00	6	96,00	6	96,00
x_3 versus x_4	8	0	100,00	2	98,67	2	98,67

Pode-se observar na Tabela 5.13 que ao utilizar apenas duas características o classificador polinomial separou corretamente as 50 (100%) flores setosas e não conseguiu classificar todas as flores versicolor e virginica, mas, mesmo assim, para essas flores esse classificador apresentou bons resultados.

5.2.2 Classificação polinomial da base de dados Iris utilizando três características

A seguir são mostrados os resultados obtidos nos testes realizados utilizando-se três características da base de dados IRIS.

A Tabela 5.14 mostra os resultados obtidos nos testes realizados com as características comprimento da sépala (x_1), largura da sépala (x_2) e comprimento da pétala (x_3) variando o grau da base polinomial de 1 a 20.

5.2. CLASSIFICAÇÃO POLINOMIAL DA BASE DE DADOS IRIS

Tabela 5.14: Resultados obtidos da classificação polinomial utilizando as características comprimento da sépala (x_1), largura da sépala (x_2) e comprimento da pétala (x_3).

Grau do polinômio	Setosa		Versicolor		Virginica	
	Erro	% Acerto	Erro	% Acerto	Erro	% Acerto
1	0	100,00	43	71,33	15	90,00
2	0	100,00	8	94,67	5	96,67
3	0	100,00	6	96,00	5	96,67
4	0	100,00	5	96,67	5	96,67
5	0	100,00	4	97,33	4	97,33
6	0	100,00	4	97,33	4	97,33
7	0	100,00	4	97,33	4	97,33
8	0	100,00	2	98,67	2	98,67
9	0	100,00	3	98,00	2	98,67
10	0	100,00	5	96,67	8	94,67
11	0	100,00	1	99,33	1	99,33
12	0	100,00	12	92,00	11	92,67
13	0	100,00	35	76,67	34	77,33
14	0	100,00	34	77,33	33	78,00
15	0	100,00	0	100,00	0	100,00
16	0	100,00	1	99,33	1	99,33
17	0	100,00	0	100,00	0	100,00
18	0	100,00	0	100,00	0	100,00
19	0	100,00	11	92,67	12	92,00
20	0	100,00	18	88,00	18	88,00

Pode-se observar na Tabela 5.14 que as 50 (100%) flores setosas foram separadas corretamente utilizando a base polinomial de grau 1 ou superior, mostrando que essa flor é linearmente separável das flores versicolor e virginica. E também a mesma quantidade des-

sas duas últimas flores (50, ou seja, 100%) foram separadas corretamente usando as bases polinomiais de graus 15, 17 e 18.

A Tabela 5.15 mostra os resultados obtidos utilizando os classificadores SVM e árvore de decisão (J48) na classificação das características comprimento da sépala (x_1), largura da sépala (x_2) e comprimento da pétala (x_3) da base de dados IRIS.

Tabela 5.15: Resultados obtidos pelos classificadores SVM e árvore de decisão (J48) utilizando as características comprimento da sépala (x_1), largura da sépala (x_2) e comprimento da pétala (x_3).

Classificador	Setosa		Versicolor		Virginica	
	Erro	% Acerto	Erro	% Acerto	Erro	% Acerto
SVM	0	100,00	6	96,00	11	92,67
J48	0	100,00	6	96,00	1	99,33

Pode-se verificar na Tabela 5.15 que os classificadores SVM e árvore de decisão separaram corretamente 50 (100%) flores setosa. O classificador SVM separou corretamente 44 (96,00%) flores versicolor e 39 (92,67%) flores virginica, enquanto o classificador árvore de decisão separou corretamente 44 (96,00%) flores versicolor e 49 (99,33%) flores virginica.

A Figura 5.20 ilustra um gráfico comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando as características comprimento da sépala (x_1), largura da sépala (x_2) e comprimento da pétala (x_3).

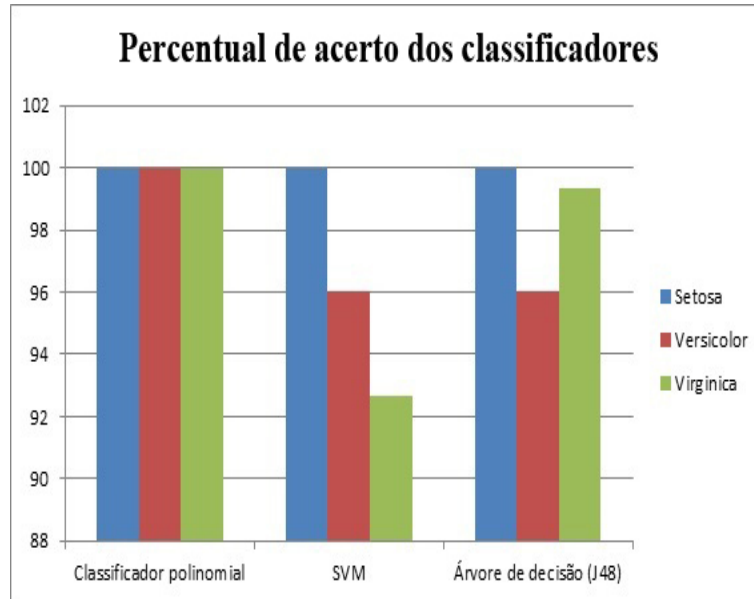


Figura 5.20: Comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando as características comprimento da sépala (x_1), largura da sépala (x_2) e comprimento da pétala (x_3).

Ao comparar os resultados obtidos nas Tabelas 5.14 e 5.15 e na Figura 5.20, pode-se observar que o classificador polinomial conseguiu separar corretamente as 50 flores (100%) de cada uma das classes setosa, versicolor e virginica utilizando as bases polinomiais de graus 15, 17 e 18. Enquanto os classificadores SVM e árvore de decisão não conseguiram esse mesmo resultado na separação dessas três classes.

A Tabela 5.16 mostra os resultados obtidos nos testes realizados com as características comprimento da sépala (x_1), largura da sépala (x_2) e largura da pétala (x_4) variando o grau da base polinomial de 1 a 20.

5.2. CLASSIFICAÇÃO POLINOMIAL DA BASE DE DADOS IRIS

Tabela 5.16: Resultados obtidos da classificação polinomial utilizando as características comprimento da sépala (x_1), largura da sépala (x_2) e largura da pétala (x_4).

Grau do polinômio	Setosa		Versicolor		Virginica	
	Erro	% Acerto	Erro	% Acerto	Erro	% Acerto
1	0	100,00	38	74,67	10	93,33
2	0	100,00	8	94,67	4	97,33
3	0	100,00	7	95,33	5	96,67
4	0	100,00	4	97,33	4	97,33
5	0	100,00	4	97,33	5	96,67
6	0	100,00	3	98,00	3	98,00
7	0	100,00	29	80,67	26	82,67
8	0	100,00	2	98,67	2	98,67
9	0	100,00	31	79,33	27	82,00
10	0	100,00	1	99,33	1	99,33
11	0	100,00	1	99,33	1	99,33
12	0	100,00	1	99,33	1	99,33
13	0	100,00	4	97,33	5	96,67
14	0	100,00	0	100,00	0	100,00
15	0	100,00	0	100,00	0	100,00
16	0	100,00	0	100,00	0	100,00
17	0	100,00	0	100,00	0	100,00
18	0	100,00	3	98,00	4	97,33
19	0	100,00	0	100,00	2	98,67
20	0	100,00	0	100,00	0	100,00

Pode-se observar na Tabela 5.16 que as 50 (100%) flores setosas foram separadas corretamente utilizando a base polinomial de grau 1 ou superior mostrando que essa flor é linearmente separável das flores versicolor e virginica. E também a mesma quantidade dessas duas últimas flores (50, ou seja, 100%) foram separadas corretamente usando a base

polinomial de grau 14 a 17 e 20.

A Tabela 5.17 mostra os resultados obtidos pelos classificadores SVM e árvore de decisão (J48) utilizando as características comprimento da sépala (x_1), largura da sépala (x_2) e largura da pétala (x_4) da base de dados IRIS.

Tabela 5.17: Resultados obtidos pelos classificadores SVM e árvore de decisão (J48) utilizando as características comprimento da sépala (x_1), largura da sépala (x_2) e largura da pétala (x_4).

Classificador	Setosa		Versicolor		Virginica	
	Erro	% Acerto	Erro	% Acerto	Erro	% Acerto
SVM	0	100,00	1	99,33	4	97,33
J48	0	100,00	1	99,33	5	96,67

Pode-se verificar na Tabela 5.17 que os classificadores SVM e árvore de decisão separaram corretamente 50 (100%) flores setosa. O classificador SVM separou corretamente 49 (99,33%) flores versicolor e 46 (97,33%) flores virginica. Enquanto o classificador árvore de decisão classificou corretamente 49 (99,33%) flores versicolor e 45 (96,67%) flores virginica.

A Figura 5.21 ilustra um gráfico comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando as características comprimento da sépala (x_1), largura da sépala (x_2) e largura da pétala (x_4).

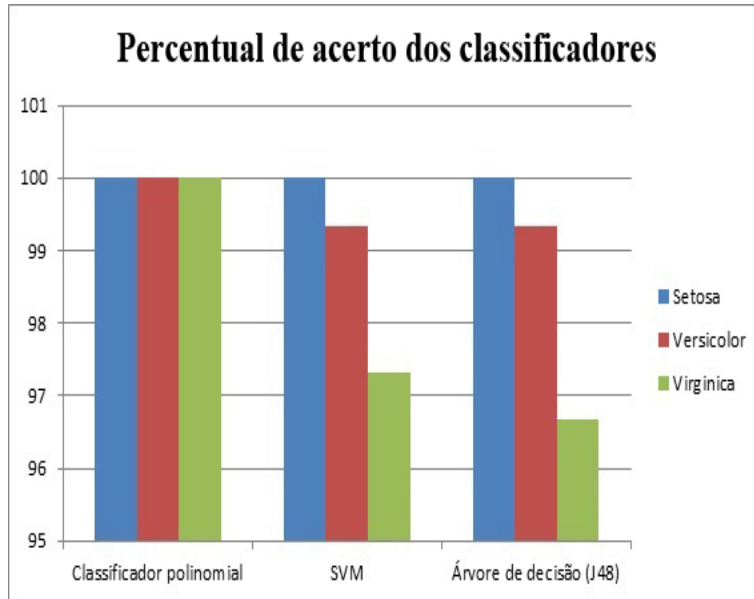


Figura 5.21: Comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando as características comprimento da sépala (x_1), largura da sépala (x_2) e largura da pétala (x_4).

Ao comparar os resultados obtidos nas Tabelas 5.16 e 5.17 e na Figura 5.21, pode-se observar que o classificador polinomial conseguiu separar corretamente as 50 flores (100%) de cada uma das classes setosa, versicolor e virginica utilizando as bases polinomiais de graus 14 a 17 e 20. Enquanto os classificadores SVM e árvore de decisão não conseguiram esse mesmo resultado na separação dessas três classes.

A Tabela 5.18 mostra os resultados obtidos nos testes realizados com as características comprimento da sépala (x_1), comprimento da pétala (x_3) e largura da pétala (x_4) variando o grau da base polinomial de 1 a 20.

Tabela 5.18: Resultados obtidos da classificação polinomial utilizando as características comprimento da sépala (x_1), comprimento da pétala (x_3) e largura da pétala (x_4).

Grau do polinômio	Setosa		Versicolor		Virginica	
	Erro	% Acerto	Erro	% Acerto	Erro	% Acerto
1	0	100,00	44	70,67	8	94,67
2	0	100,00	9	94,00	6	96,00
3	0	100,00	4	97,33	4	97,33
4	0	100,00	4	97,33	4	97,33
5	0	100,00	2	98,67	2	98,67
6	0	100,00	2	98,67	2	98,67
7	0	100,00	0	100,00	1	99,33
8	0	100,00	0	100,00	0	100,00
9	0	100,00	0	100,00	0	100,00
10	0	100,00	1	99,33	1	99,33
11	0	100,00	0	100,00	0	100,00
12	0	100,00	0	100,00	0	100,00
13	0	100,00	5	96,67	5	96,67
14	0	100,00	0	100,00	0	100,00
15	0	100,00	0	100,00	0	100,00
16	0	100,00	0	100,00	0	100,00
17	0	100,00	0	100,00	0	100,00
18	0	100,00	0	100,00	0	100,00
19	0	100,00	1	99,33	3	98,00
20	0	100,00	0	100,00	0	100,00

Pode-se observar na Tabela 5.18 que as 50 (100%) flores setosas foram separadas corretamente utilizando a base polinomial de grau 1 ou superior mostrando que essa flor é linearmente separável das flores versicolor e virginica. E também a mesma quantidade dessas duas últimas flores (50, ou seja, 100%) foram separadas corretamente usando a base

polinomial de grau 8, 9, 11, 12, 14 a 18 e 20.

A Tabela 5.19 mostra os resultados obtidos pelos classificadores SVM e árvore de decisão (J48) utilizando as características comprimento da sépala (x_1), comprimento da pétala (x_3) e largura da pétala (x_4) da base de dados IRIS.

Tabela 5.19: Resultados obtidos pelos classificadores SVM e árvore de decisão (J48) utilizando as características comprimento da sépala (x_1), comprimento da pétala (x_3) e largura da pétala (x_4).

Classificador	Setosa		Versicolor		Virginica	
	Erro	% Acerto	Erro	% Acerto	Erro	% Acerto
SVM	0	100,00	2	98,67	4	97,33
J48	0	100,00	1	99,33	2	98,67

Pode-se verificar na Tabela 5.19 que os classificadores SVM e árvore de decisão separaram corretamente 50 (100%) flores setosa. O classificador SVM separou corretamente 48 (98,67%) flores versicolor e 46 (97,33%) flores virginica. Enquanto o classificador árvore de decisão classificou corretamente 49 (99,33%) flores versicolor e 48 (98,67%) flores virginica.

A Figura 5.22 ilustra um gráfico comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando as características comprimento da sépala (x_1), comprimento da pétala (x_3) e largura da pétala (x_4).

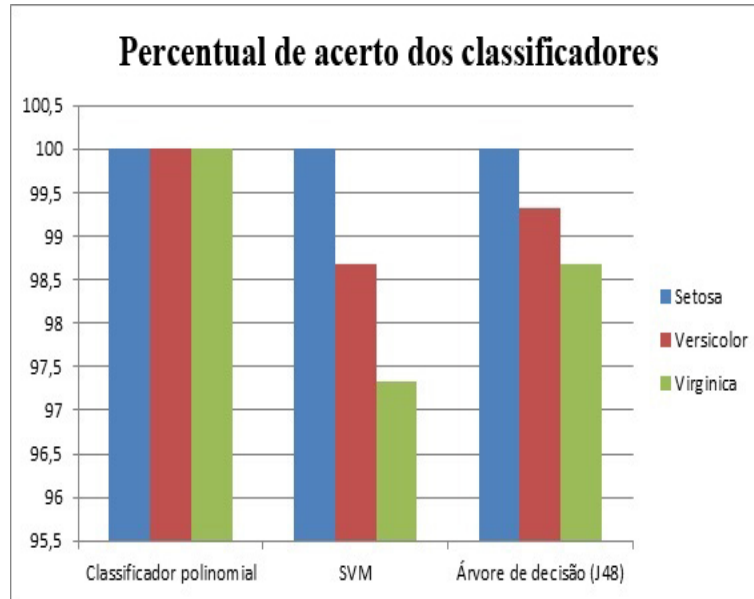


Figura 5.22: Comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando as características comprimento da sépala (x_1), comprimento da pétala (x_3) e largura da pétala (x_4).

Ao comparar os resultados obtidos nas Tabelas 5.18 e 5.19 e na Figura 5.22, pode-se observar que o classificador polinomial conseguiu separar corretamente as 50 flores (100%) de cada uma das classes setosa, versicolor e virginica utilizando as bases polinomiais de graus 8, 9, 11, 12, 14 a 18 e 20. Enquanto os classificadores SVM e árvore de decisão não conseguiram esse mesmo resultado na separação dessas três classes.

A Tabela 5.20 mostra os resultados obtidos nos testes realizados com as características largura da sépala (x_2), comprimento da pétala (x_3) e largura da pétala (x_4) variando o grau da base polinomial de 1 a 20.

5.2. CLASSIFICAÇÃO POLINOMIAL DA BASE DE DADOS IRIS

Tabela 5.20: Resultados obtidos da classificação polinomial utilizando as características largura da sépala (x_2), comprimento da pétala (x_3) e largura da pétala (x_4).

Grau do polinômio	Setosa		Versicolor		Virginica	
	Erro	% Acerto	Erro	% Acerto	Erro	% Acerto
1	0	100,00	39	74,00	9	94,00
2	0	100,00	6	96,00	4	97,33
3	0	100,00	3	98,00	3	98,00
4	0	100,00	1	99,33	1	99,33
5	0	100,00	1	99,33	1	99,33
6	0	100,00	1	99,33	1	99,33
7	0	100,00	2	98,67	2	98,67
8	0	100,00	0	100,00	0	100,00
9	0	100,00	0	100,00	0	100,00
10	0	100,00	0	100,00	0	100,00
11	0	100,00	0	100,00	0	100,00
12	0	100,00	0	100,00	0	100,00
13	0	100,00	0	100,00	0	100,00
14	0	100,00	0	100,00	0	100,00
15	0	100,00	5	96,67	5	96,67
16	0	100,00	0	100,00	0	100,00
17	0	100,00	47	68,67	48	68,00
18	0	100,00	0	100,00	0	100,00
19	0	100,00	0	100,00	0	100,00
20	0	100,00	0	100,00	0	100,00

Pode-se observar na Tabela 5.20 que as 50 (100%) flores setosas foram separadas corretamente utilizando a base polinomial de grau 1 ou superior mostrando que essa flor é linearmente separável das flores versicolor e virginica. E também a mesma quantidade dessas duas últimas flores (50, ou seja, 100%) foram separadas corretamente usando a base

polinomial de grau 8 a 14, 16 e 18 a 20.

A Tabela 5.21 mostra os resultados obtidos pelos classificadores SVM e árvore de decisão (J48) utilizando as características largura da sépala (x_2), comprimento da pétala (x_3) e largura da pétala (x_4) da base de dados IRIS.

Tabela 5.21: Resultados obtidos pelos classificadores SVM e árvore de decisão (J48) utilizando as características largura da sépala (x_2), comprimento da pétala (x_3) e largura da pétala (x_4).

Classificador	Setosa		Versicolor		Virginica	
	Erro	% Acerto	Erro	% Acerto	Erro	% Acerto
SVM	0	100,00	2	98,67	4	97,33
J48	0	100,00	1	99,33	2	98,67

Pode-se verificar na Tabela 5.21 que os classificadores SVM e árvore de decisão separaram corretamente 50 (100%) flores setosa. O classificador SVM separou corretamente 48 (98,67%) flores versicolor e 46 (97,33%) flores virginica. Enquanto o classificador árvore de decisão classificou corretamente 49 (99,33%) flores versicolor e 48 (98,67%) flores virginica.

A Figura 5.23 ilustra um gráfico comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando as características largura da sépala (x_2), comprimento da pétala (x_3) e largura da pétala (x_4).

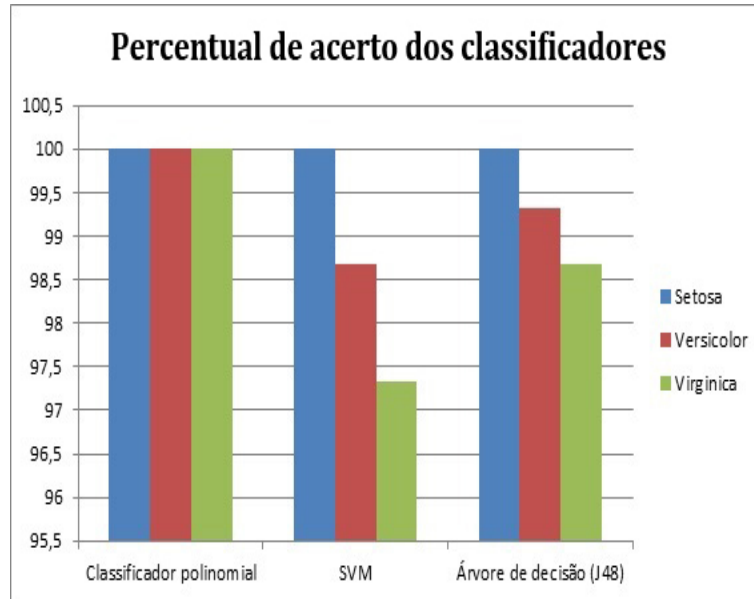


Figura 5.23: Comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando as características largura da sépala (x_2), comprimento da pétala (x_3) e largura da pétala (x_4).

Ao comparar os resultados obtidos nas Tabelas 5.20 e 5.21 e na Figura 5.23, pode-se observar que o classificador polinomial conseguiu separar corretamente as 50 flores (100%) de cada uma das classes setosa, versicolor e virginica utilizando as bases polinomiais de graus 8 a 14, 16 e 18 a 20. Enquanto os classificadores SVM e árvore de decisão não conseguiram esse mesmo resultado na separação dessas três classes.

Pode-se observar nas Tabelas de 5.14 a 5.20 e nas Figuras de 5.20 a 5.23 que o classificador polinomial mostrou uma perfeita separação das classes setosa, versicolor e virginica utilizando três características e essa separação não foi possível usando os classificadores SVM e árvore de decisão.

5.2.3 Classificação polinomial da base de dados Iris utilizando quatro características

A Tabela 5.22 mostra os resultados obtidos nos testes realizados com as características comprimento da sépala (x_1), largura da sépala (x_2), comprimento da pétala (x_3) e largura da pétala (x_4) variando o grau da base polinomial de 1 a 15.

Tabela 5.22: Resultados obtidos da classificação polinomial utilizando as características comprimento da sépala (x_1), largura da sépala (x_2), comprimento da pétala (x_3) e largura da pétala (x_4).

Grau do polinômio	Setosa		Versicolor		Virginica	
	Erro	% Acerto	Erro	% Acerto	Erro	% Acerto
1	0	100,00	40	73,33	11	92,67
2	0	100,00	5	96,67	2	98,67
3	0	100,00	2	98,67	2	98,67
4	0	100,00	2	98,67	2	98,67
5	0	100,00	0	100,00	0	100,00
6	0	100,00	0	100,00	0	100,00
7	0	100,00	0	100,00	0	100,00
8	0	100,00	0	100,00	0	100,00
9	0	100,00	0	100,00	0	100,00
10	0	100,00	0	100,00	0	100,00
11	0	100,00	0	100,00	0	100,00
12	0	100,00	0	100,00	0	100,00
13	0	100,00	0	100,00	0	100,00
14	0	100,00	0	100,00	0	100,00
15	0	100,00	0	100,00	0	100,00

Pode-se observar na Tabela 5.22 que as 50 (100%) flores setosas foram separadas corretamente utilizando a base polinomial de grau 1 ou superior mostrando que essa flor é linearmente separável das flores versicolor e virginica. E também a mesma quantidade des-

As duas últimas flores (50, ou seja, 100%) foram separadas corretamente usando a base polinomial de grau 5 ou superior.

A Tabela 5.23 mostra os resultados obtidos pelos classificadores SVM e árvore de decisão (J48) utilizando as características comprimento da sépala (x_1), largura da sépala (x_2), comprimento pétala (x_3) e largura da pétala (x_4) da base de dados IRIS.

Tabela 5.23: Resultados obtidos pelos classificadores SVM e árvore de decisão (J48) utilizando as características comprimento da sépala (x_1), largura da sépala (x_2), comprimento da pétala (x_3) e largura da pétala (x_4).

Classificador	Setosa		Versicolor		Virginica	
	Erro	% Acerto	Erro	% Acerto	Erro	% Acerto
SVM	0	100,00	1	99,33	4	97,33
J48	0	100,00	1	99,33	2	98,67

Pode-se verificar na Tabela 5.23 que os classificadores SVM e árvore de decisão separaram corretamente 50 (100%) flores setosa. O classificador SVM separou corretamente 49 (99,33%) flores versicolor e 46 (97,33%) flores virginica. Enquanto o classificador árvore de decisão classificou corretamente 49 (99,33%) flores versicolor e 48 (98,67%) flores virginica.

A Figura 5.24 ilustra um gráfico comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando as características comprimento da sépala (x_1), largura da sépala (x_2), comprimento da pétala (x_3) e largura da pétala (x_4).

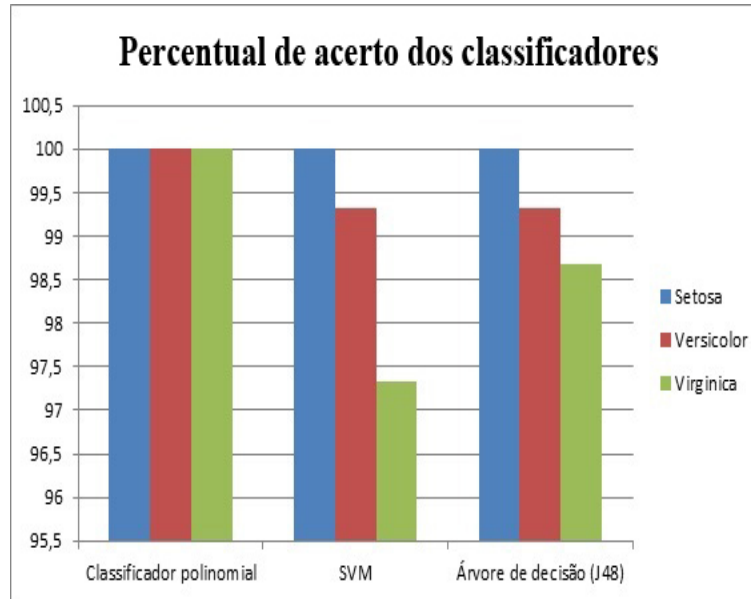


Figura 5.24: Comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando as características comprimento da sépala (x_1), largura da sépala (x_2), comprimento da pétala (x_3) e largura da pétala (x_4).

Ao comparar os resultados obtidos nas Tabelas 5.22 e 5.23 e na Figura 5.24, pode-se observar que o classificador polinomial conseguiu separar corretamente as 50 flores, (100%) de cada uma das classes setosa, versicolor e virginica utilizando as bases polinomiais de graus 5 ou superior. Enquanto os classificadores SVM e árvore de decisão não conseguiram esse mesmo resultado na separação dessas três classes.

5.2.4 Classificação polinomial da base de dados Iris utilizando o método de validação cruzada

Esta seção apresenta os resultados obtidos no teste realizado com as características comprimento da sépala (x_1), largura da sépala (x_2), comprimento da pétala (x_3) e largura da pétala (x_4) da base de dados IRIS utilizando o método de validação cruzada (*cross-validation*). O objetivo desse método é avaliar a precisão e a capacidade de generalização do método de classificação.

O método de validação cruzada foi utilizado neste trabalho para avaliar e comparar o

classificador polinomial com os classificadores SVM e árvore de decisão aplicando um padrão de treinamento e validação (teste) iguais para todos os classificadores. Na validação cruzada os exemplos foram particionados dentro de n grupos de forma que o número de amostras e a distribuição das classes fossem o mais uniforme possível. O conjunto de treinamento foi constituído utilizando $n - 1$ grupos. O grupo restante foi utilizado para a validação do classificador [78].

Nesta tese as 150 flores foram divididas em 10 grupos disjuntos. Cada grupo continha 5 flores setosa, 5 flores versicolor e 5 flores virginica, totalizando 15 flores. Para o treinamento foi selecionado 9 grupos totalizando 135 (90%) e um grupo para validação contendo 15 flores (10%). Esse processo foi repetido 10 vezes alterando o conjunto de validação com um grupo do conjunto de treinamento. Com essa metodologia todos os dados foram utilizados no treinamento e na validação. A Figura 5.25 ilustra os grupos de amostras de treinamento e validação utilizados no método de validação cruzada.

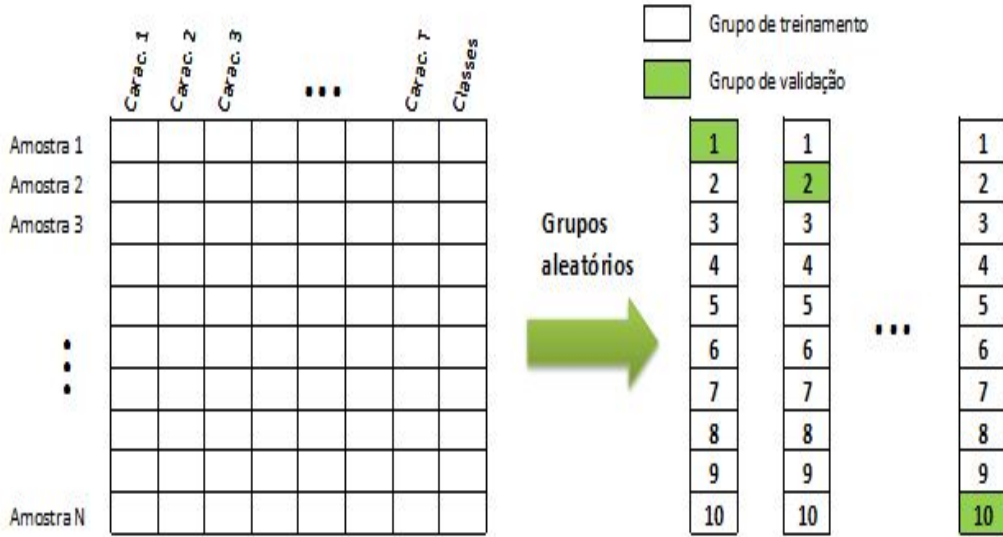


Figura 5.25: Método de validação cruzada.

A Tabela 5.24 mostra os resultados obtidos nos testes realizados com as características comprimento da sépala (x_1), largura da sépala (x_2), comprimento da pétala (x_3) e largura da pétala (x_4) da base de dados IRIS utilizando o método de validação cruzada.

5.3. CLASSIFICAÇÃO POLINOMIAL PARA IDENTIFICAÇÃO DE PIXELS PARCIAIS EM REGIÕES DE INTERESSE DE IMAGENS MAMOGRÁFICAS

Tabela 5.24: Resultados obtidos nos testes realizados com as características comprimento da sépala (x_1), largura da sépala (x_2), comprimento da pétala (x_3) e largura da pétala (x_4) utilizando o método de validação cruzada.

Classificador	Setosa		Versicolor		Virginica	
	Erro	% Acerto	Erro	% Acerto	Erro	% Acerto
Classificador Polinomial	0	100,00	2	96,00	1	98,00
SVM	0	100,00	1	98,00	5	90,00
J48	1	98,00	3	94,00	2	96,00

Pode-se verificar na Tabela 5.24 que as médias de acerto do classificador polinomial para as flores setosa, versicolor e virginica foram 100%, 96% e 98%, respectivamente. Para o classificador SVM essas médias para as flores setosa, versicolor e virginica foram 100%, 98% e 90%, respectivamente, enquanto para o classificador árvore de decisão as médias de acerto para as flores setosa, versicolor e virginica foram 98%, 94% e 96%, respectivamente.

5.3 Classificação Polinomial para Identificação de Pixels Parciais em Regiões de Interesse de Imagens Mamográficas

Esta seção apresenta os resultados obtidos da aplicação do classificador polinomial para a identificação de pixels parciais em regiões de interesse de imagens mamográficas.

A Figura 5.26 ilustra o diagrama de blocos das etapas utilizadas no processamento das imagens mamográficas.

5.3. CLASSIFICAÇÃO POLINOMIAL PARA IDENTIFICAÇÃO DE PIXELS PARCIAIS EM REGIÕES DE INTERESSE DE IMAGENS MAMOGRÁFICAS

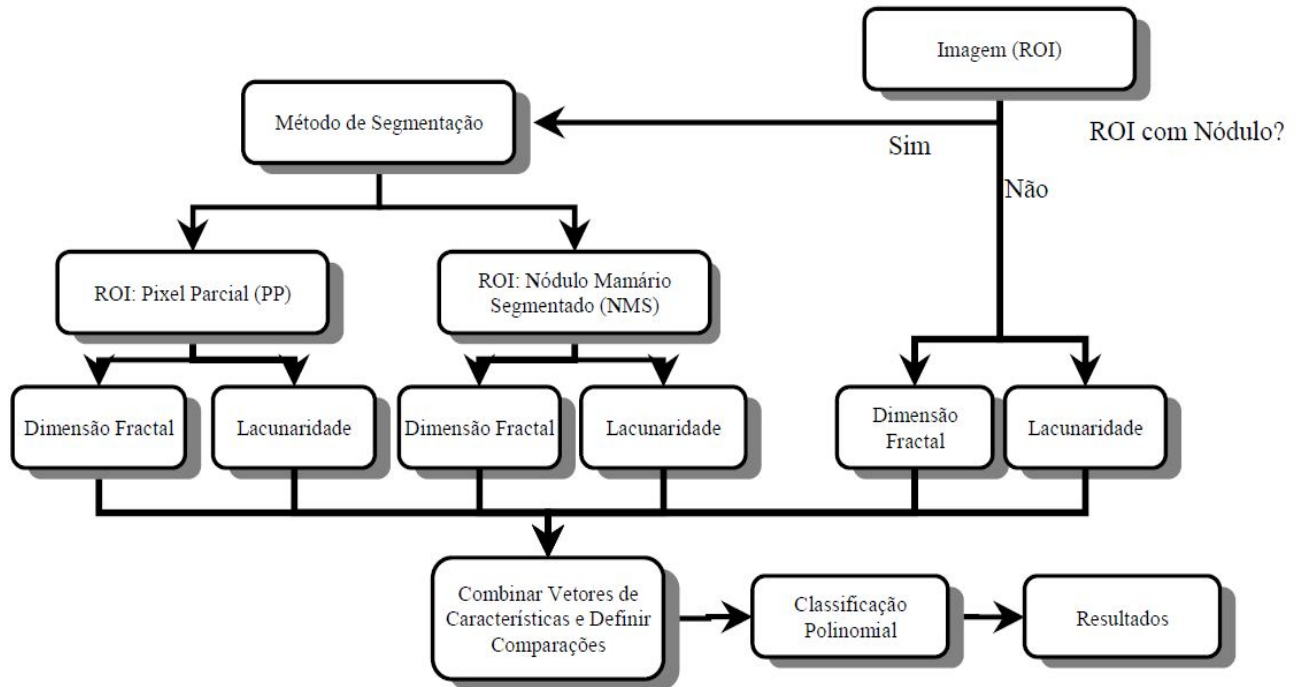


Figura 5.26: Diagrama de blocos das etapas utilizadas no processamento das imagens mamográficas: segmentação, quantificação, análise e classificação.

Os atributos fractais dimensão fractal (D) e lacunaridade (L) foram calculados a partir das imagens obtidas da base Digital Database for Screening Mammography (DDSM) da Universidade do Sul da Flórida, EUA [13]. A base DDSM é constituída por aproximadamente 2.500 casos, obtidos entre os anos de 1988 e 1999, por meio de projetos desenvolvidos por pesquisadores de quatro instituições dos Estados Unidos: Massachusetts General Hospital (MGH), Wake Forest University School of Medicine (WFUSM), Sacred Heart Hospital (SHH) e Washington University (WU).

As imagens da base de dados DDSM foram digitalizadas em diferentes equipamentos, tais como Howtek (na universidade WU e no hospital MGH), Lumisys (na universidade WFUSM e no hospital SHH) e DBA (no hospital MGH). Essas imagens são das projeções médio-lateral oblíquo (MLO) e crânio-caudal (CC), representadas em níveis de cinza com quantização de 12 ou 16 bits.

5.3. CLASSIFICAÇÃO POLINOMIAL PARA IDENTIFICAÇÃO DE PIXELS PARCIAIS EM REGIÕES DE INTERESSE DE IMAGENS MAMOGRÁFICAS

Para este trabalho foram selecionadas 166 imagens nas visões CC e MLO distribuídas em dois conjuntos: nódulos mamários malignos (NM) com 58 casos (29 CC e 29 MLO) e as 108 imagens restantes são das estruturas sadias (ES) com (54 CC e 54 MLO). Por questões de padronização de resolução, apenas as imagens obtidas pelo digitalizador LUMYSIS foram utilizadas nesta tese. A resolução desse equipamento é $43,5 \mu\text{m}$ e 12 bits de quantização.

Para quantificar as estruturas sadias e doentes, os grupos nódulos mamários (NM) e estruturas sadias (ES) foram redefinidos para conter apenas regiões de interesses do inglês *Regions of Interest* (ROIs). O estudo das lesões por meio de ROIs é um processo comumente utilizado na literatura [21, 79, 80]. O tamanho da janela utilizada neste trabalho foi de 65 x 65 pixels. Este tamanho é suficiente para conter regiões nodulares. As ROIs do grupo NM foram obtidas a partir do arquivo *.ics*, que contém o código de cadeia (*code-chain*) da posição da lesão na base de dados DDSM. Para o grupo ES, as ROIs foram extraídas de imagens sadias (considerando diferentes localizações). Essas extrações foram realizadas manualmente por um especialista. A Figura 5.27 ilustra as ROIs utilizadas para o processamento de uma imagem denominada C-0100 da base de dados DDSM.

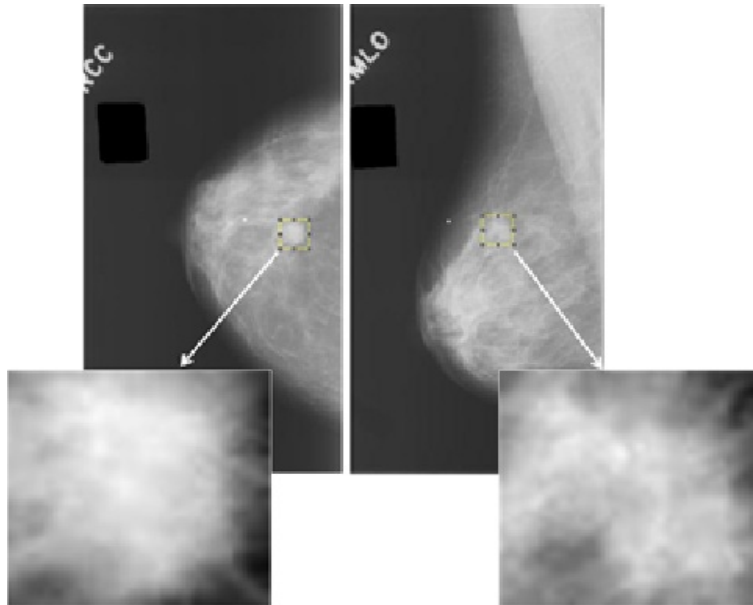


Figura 5.27: Imagens mamográficas das visões CC e MLO do caso C-0100.

Definir ROIs doentes a partir de janelas quadradas ou retangulares pode incluir regiões de fundo e influenciar as quantificações e as distinções dos padrões desejados. As regiões

5.3. CLASSIFICAÇÃO POLINOMIAL PARA IDENTIFICAÇÃO DE PIXELS PARCIAIS EM REGIÕES DE INTERESSE DE IMAGENS MAMOGRÁFICAS

de fundo são conhecidas como pixels parciais (PP) que são os tecidos entre os estados sadio e doente. Os pixels parciais foram quantificados separadamente para mostrar que essas imagens têm padrões diferentes, colaborando com o estudo apresentado em [21]. Essa tarefa foi realizada aplicando um algoritmo de segmentação nas ROIs do grupo NM proporcionando dois novos grupos: um constituído somente com os nódulos malignos segmentados (NMS) e o outro com as estruturas remanescentes, os pixels parciais (PP).

As imagens que constituem o grupo NM foram segmentadas com o método entropia máxima [81]. A Figura 5.28 mostra o resultado obtido com as ROIs nas visões CC e MLO do grupo NM.

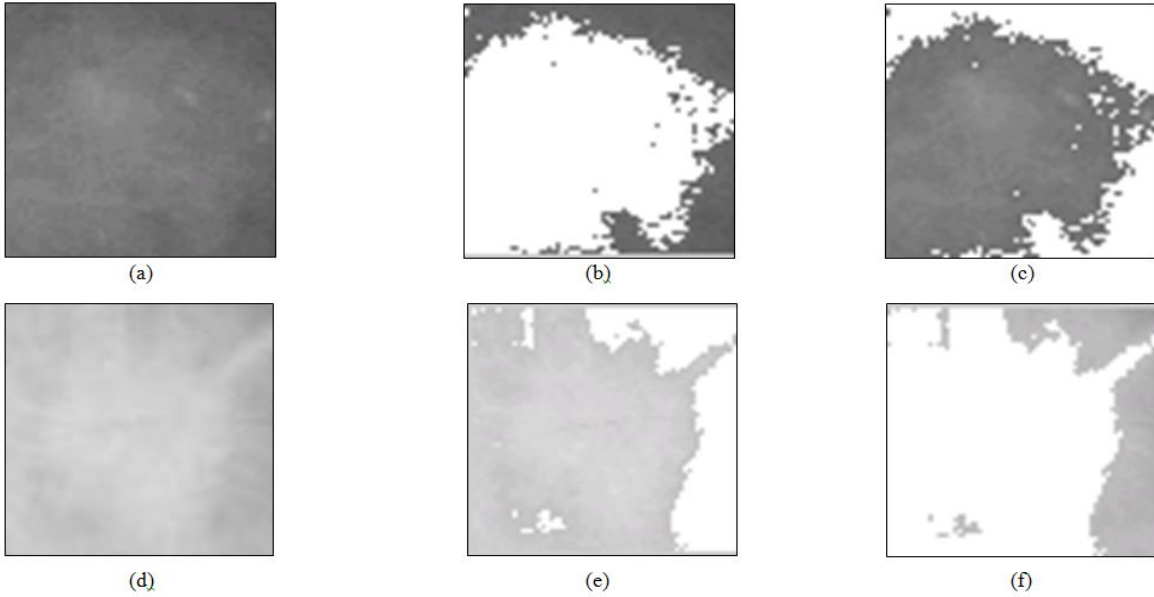


Figura 5.28: Visualização das ROIs utilizadas para a quantificação dos comportamentos das texturas. (a) e (d) ROIs do grupo NM, nas visões CC e MLO, respectivamente; (b) e (e) estruturas segmentadas do grupo NMS; (c) e (f) estruturas segmentadas do grupo PP.

As Figuras 5.29 e 5.30 ilustram as assinaturas das texturas dos grupos NMS versus PP evidenciando as aproximações das retas obtidas por meio da regressão (gráfico log-log) do número de caixas ocupadas na imagem ($Nr(I)$) por tamanho do lado dessa caixa (r). As aproximações das retas mostram os comportamentos diferentes entre os grupos NMS versus PP. Os valores da dimensão fractal e da lacunaridade que representam cada ROI foram obtidos calculando-se os coeficientes angulares das retas apresentadas. Os vetores

5.3. CLASSIFICAÇÃO POLINOMIAL PARA IDENTIFICAÇÃO DE PIXELS PARCIAIS EM REGIÕES DE INTERESSE DE IMAGENS MAMOGRÁFICAS

das características foram obtidos extrapolando essa abordagem para todas as ROIs que constituem cada grupo.

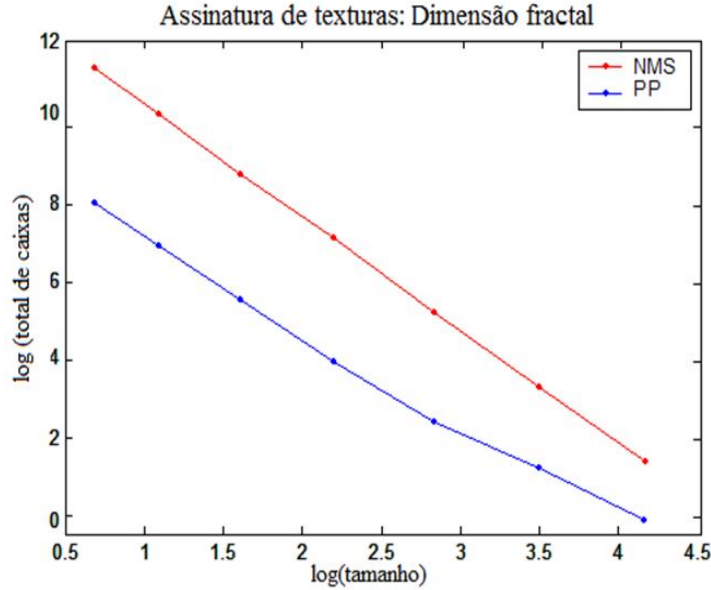


Figura 5.29: Assinaturas das texturas das ROIs representadas em gráficos log-log das dimensões fractais.

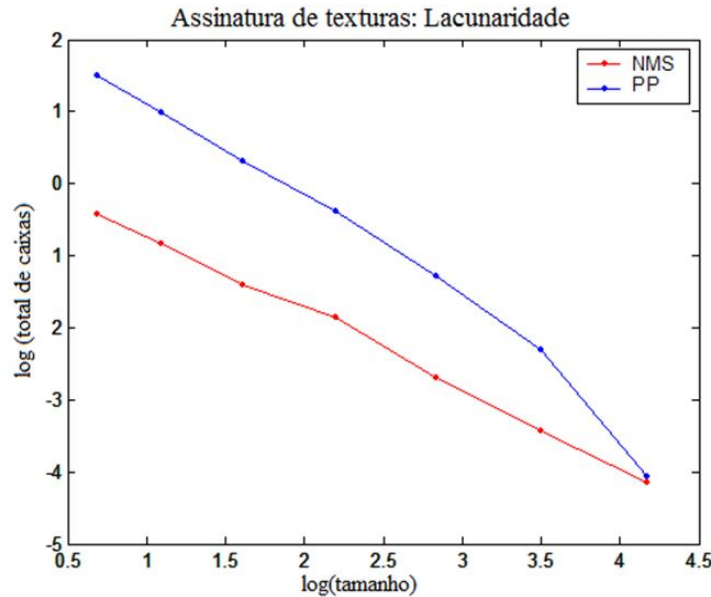


Figura 5.30: Assinaturas das texturas das ROIs representadas em gráficos log-log das lacunaridades.

Nas visões CC e MLO foram realizadas as comparações entre os grupos NMS versus PP, NMS versus ES e PP versus ES, considerando as características:

5.3. CLASSIFICAÇÃO POLINOMIAL PARA IDENTIFICAÇÃO DE PIXELS PARCIAIS EM REGIÕES DE INTERESSE DE IMAGENS MAMOGRÁFICAS

1. Dimensão fractal CC;
2. Dimensão fractal MLO;
3. Dimensão fractal CC e MLO;
4. Lacunaridade CC;
5. Lacunaridade MLO;
6. Lacunaridade CC e MLO;
7. Dimensão fractal e lacunaridade CC;
8. Dimensão fractal e lacunaridade MLO; e
9. Dimensão fractal e lacunaridade CC e MLO.

Essas características descrevem o conjunto selecionado das imagens utilizadas neste trabalho com os atributos discretizados, possibilitando 27 combinações diferentes.

Obtidos os vetores de características, foram realizadas as comparações entre os grupos NMS versus PP, NMS versus ES e PP versus ES utilizando o classificador polinomial e os classificadores SVM e árvore de decisão.

A Tabela 5.25 mostra as taxas de acertos obtidas pelos classificadores polinomial, SVM e árvore de decisão utilizando a característica dimensão fractal.

Tabela 5.25: Resultados obtidos pelos classificadores polinomial (CP), SVM e árvore de decisão (J48) utilizando a característica dimensão fractal.

Classificador	NMS × ES			NMS × PP			PP × ES		
	CC	MLO	CC/MLO	CC	MLO	CC/MLO	CC	MLO	CC/MLO
CP	87,95%	87,95%	87,35%	93,10%	89,66%	85,34%	92,77%	89,16%	89,16%
SVM	65,06%	65,06%	65,06%	77,59%	75,86%	76,72%	84,34%	83,13%	83,73%
J48	85,54%	90,36%	88,55%	86,21%	79,31%	81,03%	87,95%	86,75%	87,35%

5.3. CLASSIFICAÇÃO POLINOMIAL PARA IDENTIFICAÇÃO DE PIXELS PARCIAIS EM REGIÕES DE INTERESSE DE IMAGENS MAMOGRÁFICAS

Pode-se observar na Tabela 5.25 que para o grupo NMS versus ES o classificador polinomial separou corretamente 73 imagens (87,95%) nas visões CC e MLO e 145 imagens (87,35%) na visão CC/MLO. Para o grupo NMS versus PP esse classificador separou corretamente nas visões CC, MLO e CC/MLO 54 imagens (93,10%), 52 imagens (89,66%) e 99 imagens (85,34%), respectivamente. E para o grupo PP versus ES o classificador polinomial separou corretamente para as visões CC, MLO e CC/MLO 77 imagens (92,77%), 74 imagens (89,16%) e 148 imagens (89,16%), respectivamente. O classificador SVM separou corretamente para o grupo NMS versus ES 54 imagens (65,06%) nas visões CC e MLO e 108 imagens (65,06%) na visão CC/MLO. Para o grupo NMS versus PP esse classificador separou corretamente para as visões CC, MLO e CC/MLO 45 imagens (77,59%), 44 imagens (75,86%) e 89 imagens (76,72%), respectivamente. E para o grupo PP versus ES o classificador SVM separou corretamente para as visões CC, MLO e CC/MLO 70 imagens (84,34%), 69 imagens (83,13%) e 139 imagens (83,73%), respectivamente. No classificador árvore de decisão (algoritmo J48) para o grupo NMS versus ES foram separados corretamente para as visões CC, MLO e CC/MLO 71 imagens (85,54%), 75 imagens (90,36%) e 147 imagens (88,55%), respectivamente. Para o grupo NMS versus PP esse classificador separou corretamente para as visões CC, MLO e CC/MLO 50 imagens (86,21%), 46 imagens (79,31%) e 94 imagens (81,03%), respectivamente. E para o grupo PP versus ES o classificador árvore de decisão separou corretamente para as visões CC, MLO e CC/MLO 73 imagens (87,95%), 72 imagens (86,75%) e 145 imagens (87,35%), respectivamente.

A Figura 5.31 ilustra o gráfico comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando a característica dimensão fractal para o grupo NMS versus ES.

5.3. CLASSIFICAÇÃO POLINOMIAL PARA IDENTIFICAÇÃO DE PIXELS PARCIAIS EM REGIÕES DE INTERESSE DE IMAGENS MAMOGRÁFICAS

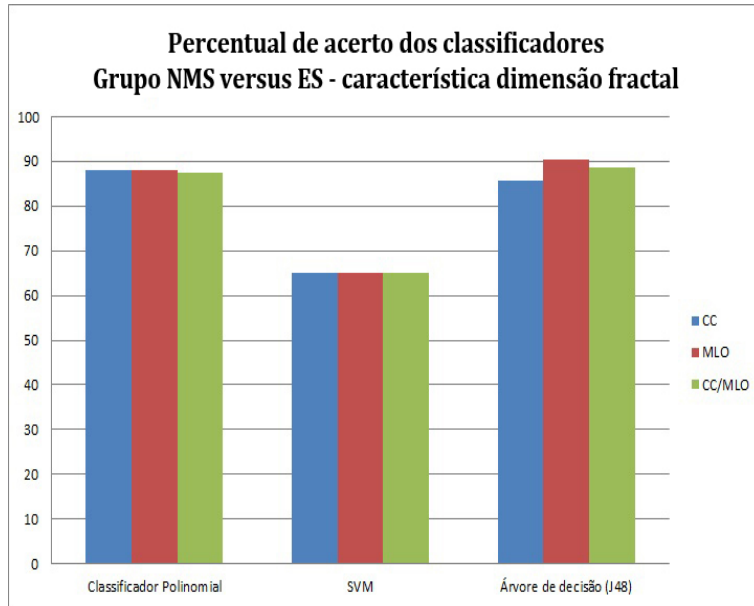


Figura 5.31: Comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando a característica dimensão fractal para o grupo NMS versus ES.

A Figura 5.32 mostra o gráfico comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando a característica dimensão fractal para o grupo NMS versus PP.

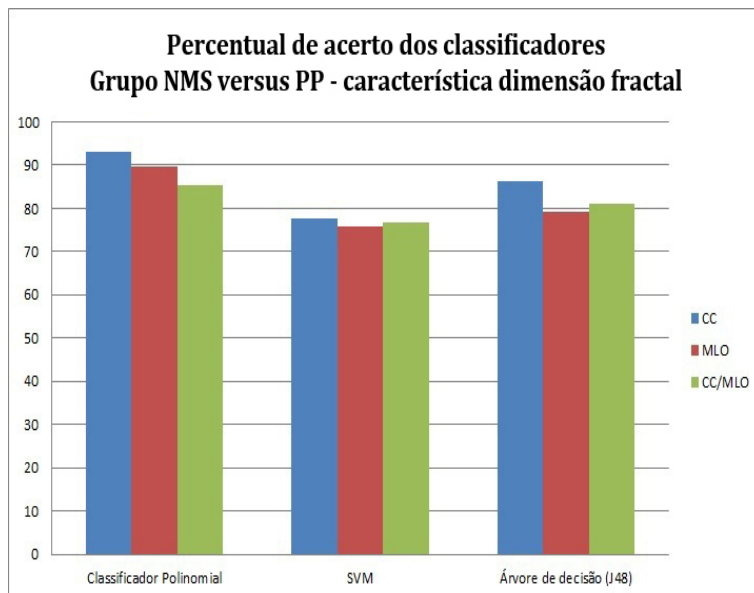


Figura 5.32: Comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando a característica dimensão fractal para o grupo NMS versus PP.

5.3. CLASSIFICAÇÃO POLINOMIAL PARA IDENTIFICAÇÃO DE PIXELS PARCIAIS EM REGIÕES DE INTERESSE DE IMAGENS MAMOGRÁFICAS

A Figura 5.33 ilustra o gráfico comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando a característica dimensão fractal para o grupo PP versus ES.

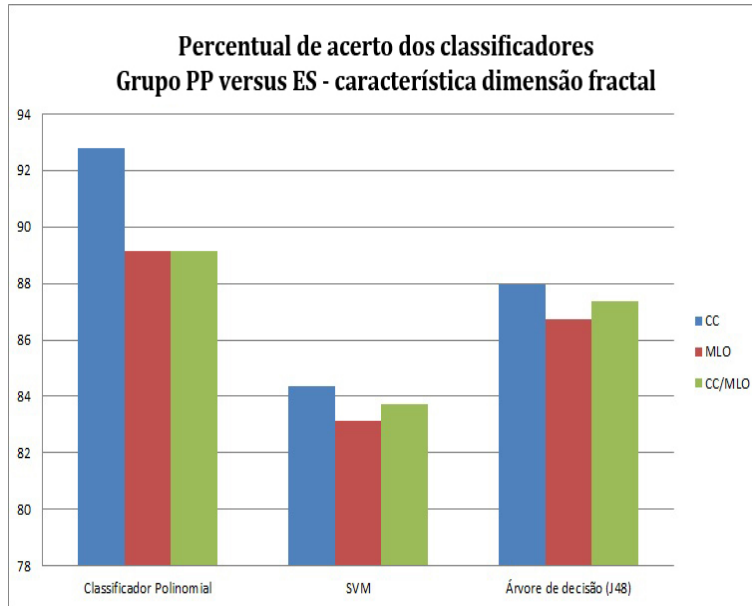


Figura 5.33: Comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando a característica dimensão fractal para o grupo PP versus ES.

Ao comparar os resultados obtidos na Tabela 5.25 e nas Figuras 5.31 a 5.33 pode-se verificar que o classificador polinomial obteve melhores resultados do que o classificador SVM em todos os grupos, para todas as projeções. Comparando com o classificador árvore de decisão, o classificador polinomial só não apresentou melhores resultados no grupo NMS versus ES nas projeções MLO e CC/MLO. Para os grupos NMS versus PP e PP versus ES em todos os casos os resultados obtidos pelo classificador polinomial foram melhores do que os resultados do classificador árvore de decisão.

A Tabela 5.26 ilustra as taxas de acertos obtidas pelos classificadores polinomial, SVM e árvore de decisão utilizando a característica lacunaridade.

5.3. CLASSIFICAÇÃO POLINOMIAL PARA IDENTIFICAÇÃO DE PIXELS PARCIAIS EM REGIÕES DE INTERESSE DE IMAGENS MAMOGRÁFICAS

Tabela 5.26: Resultados obtidos pelos classificadores polinomial (CP), SVM e árvore de decisão (J48) utilizando a característica lacunaridade.

Classificador	NMS \times ES			NMS \times PP			PP \times ES		
	CC	MLO	CC/MLO	CC	MLO	CC/MLO	CC	MLO	CC/MLO
CP	87,95%	87,95%	86,75%	91,38%	82,76%	86,21%	92,77%	90,36%	89,16%
SVM	65,06%	65,06%	65,06%	79,31%	75,86%	77,59%	85,54%	83,13%	84,34%
J48	85,54%	90,36%	88,55%	87,93%	79,31%	81,90%	91,57%	87,95%	90,36%

Pode-se observar na Tabela 5.26 que para o grupo NMS versus ES o classificador polinomial separou corretamente 73 imagens (87,95%) nas visões CC e MLO e 144 imagens (86,75%) na visão CC/MLO. Para o grupo NMS versus PP esse classificador separou corretamente nas visões CC, MLO e CC/MLO 53 imagens (91,38%), 48 imagens (82,76%) e 100 imagens (86,21%), respectivamente. E para o grupo PP versus ES o classificador polinomial separou corretamente para as visões CC, MLO e CC/MLO 77 imagens (92,77%), 75 imagens (90,36%) e 148 imagens (89,16%), respectivamente. O classificador SVM separou corretamente para o grupo NMS versus ES 54 imagens (65,06%) nas visões CC e MLO e 108 imagens (65,06%) na visão CC/MLO. Para o grupo NMS versus PP esse classificador separou corretamente para as visões CC, MLO e CC/MLO 46 imagens (79,31%), 44 imagens (75,86%) e 90 imagens (77,59%), respectivamente. E para o grupo PP versus ES o classificador SVM separou corretamente para as visões CC, MLO e CC/MLO 71 imagens (85,54%), 69 imagens (83,13%) e 140 imagens (84,34%), respectivamente. No classificador árvore de decisão (algoritmo J48) para o grupo NMS versus ES foram separados corretamente para as visões CC, MLO e CC/MLO 71 imagens (85,54%), 75 imagens (90,36%) e 147 imagens (88,55%), respectivamente. Para o grupo NMS versus PP esse classificador separou corretamente para as visões CC, MLO e CC/MLO 51 imagens (87,93%), 46 imagens (79,31%) e 95 imagens (81,90%), respectivamente. E para o grupo PP versus ES o classificador árvore de decisão separou corretamente para as visões CC, MLO e CC/MLO 76 imagens (91,57%), 73 imagens (87,95%) e 150 imagens (90,36%), respectivamente.

A Figura 5.34 ilustra o gráfico comparativo do percentual de acerto dos classificadores

5.3. CLASSIFICAÇÃO POLINOMIAL PARA IDENTIFICAÇÃO DE PIXELS PARCIAIS EM REGIÕES DE INTERESSE DE IMAGENS MAMOGRÁFICAS

polinomial, SVM e árvore de decisão utilizando a característica lacunaridade para o grupo NMS versus ES.

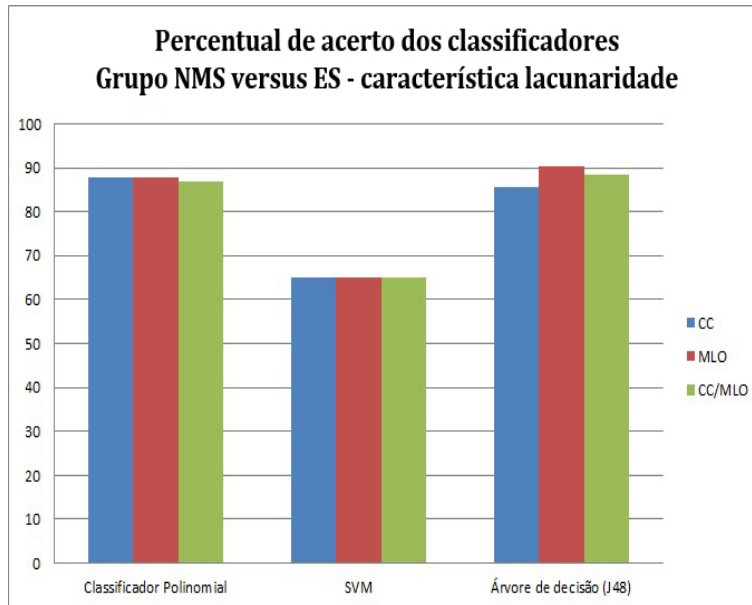


Figura 5.34: Comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando a característica lacunaridade para o grupo NMS versus ES.

A Figura 5.35 mostra o gráfico comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando a característica lacunaridade para o grupo NMS versus PP.

5.3. CLASSIFICAÇÃO POLINOMIAL PARA IDENTIFICAÇÃO DE PIXELS PARCIAIS EM REGIÕES DE INTERESSE DE IMAGENS MAMOGRÁFICAS

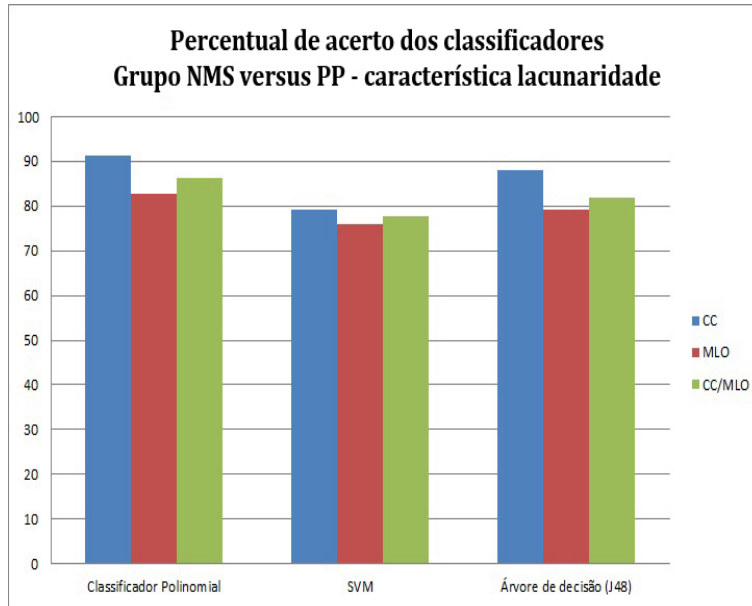


Figura 5.35: Comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando a característica lacunaridade para o grupo NMS versus PP.

A Figura 5.36 ilustra o gráfico comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando a característica lacunaridade para o grupo PP versus ES.

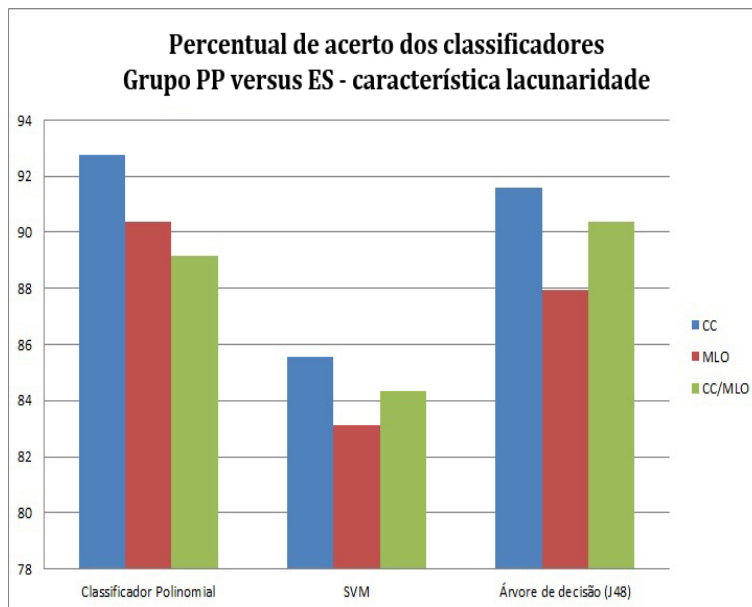


Figura 5.36: Comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando a característica lacunaridade para o grupo PP versus ES.

5.3. CLASSIFICAÇÃO POLINOMIAL PARA IDENTIFICAÇÃO DE PIXELS PARCIAIS EM REGIÕES DE INTERESSE DE IMAGENS MAMOGRÁFICAS

Ao comparar os resultados obtidos na Tabela 5.26 e nas Figuras 5.34 a 5.36 pode-se verificar que o classificador polinomial obteve melhores resultados do que o classificador SVM em todos os grupos, para todas as projeções. Comparando com o classificador árvore de decisão, o classificador polinomial só não apresentou melhores resultados no grupo NMS versus ES nas projeções MLO e CC/MLO e no grupo PP versus ES na projeção CC/MLO.

A Tabela 5.27 mostra as taxas de acertos obtidas pelos classificadores polinomial, SVM e árvore de decisão utilizando as características dimensão fractal e lacunaridade.

Tabela 5.27: Resultados obtidos pelos classificadores polinomial (CP), SVM e árvore de decisão (J48) utilizando as características dimensão fractal e lacunaridade.

Classificador	NMS × ES			NMS × PP			PP × ES		
	CC	MLO	CC/MLO	CC	MLO	CC/MLO	CC	MLO	CC/MLO
CP	89,16%	87,95%	86,75%	93,10%	93,10%	87,93%	92,77%	91,57%	90,36%
SVM	65,06%	65,06%	65,06%	77,59%	75,86%	76,72%	84,34%	83,13%	83,73%
J48	85,54%	90,36%	88,55%	87,93%	79,31%	81,90%	91,57%	86,75%	87,35%

Pode-se verificar na Tabela 5.27 que para o grupo NMS versus ES o classificador polinomial separou corretamente 74 imagens (89,16%) na visão CC, 73 imagens (87,95%) na visão MLO e 144 imagens (86,75%) na visão CC/MLO. Para o grupo NMS versus PP esse classificador separou corretamente nas visões CC, MLO e CC/MLO 54 imagens (93,10%), 54 imagens (93,10%) e 102 imagens (87,93%), respectivamente. E para o grupo PP versus ES o classificador polinomial separou corretamente para as visões CC, MLO e CC/MLO 77 imagens (92,77%), 76 imagens (91,57%) e 150 imagens (90,36%), respectivamente. O classificador SVM separou corretamente para o grupo NMS versus ES 54 imagens (65,06%) nas visões CC e MLO e 108 imagens (65,06%) na visão CC/MLO. Para o grupo NMS versus PP esse classificador separou corretamente para as visões CC, MLO e CC/MLO 45 imagens (77,59%), 44 imagens (75,86%) e 89 imagens (76,72%), respectivamente. E para o grupo PP versus ES o classificador SVM separou corretamente para as visões CC, MLO e CC/MLO 70 imagens (84,34%), 69 imagens (83,13%) e 139 imagens (83,73%), respectivamente. No classificador árvore de decisão (algoritmo J48) para o grupo NMS versus ES foram separa-

5.3. CLASSIFICAÇÃO POLINOMIAL PARA IDENTIFICAÇÃO DE PIXELS PARCIAIS EM REGIÕES DE INTERESSE DE IMAGENS MAMOGRÁFICAS

dos corretamente para as visões CC, MLO e CC/MLO 71 imagens (85,54%), 75 imagens (90,36%) e 147 imagens (88,55%), respectivamente. Para o grupo NMS versus PP esse classificador separou corretamente para as visões CC, MLO e CC/MLO 51 imagens (87,93%), 46 imagens (79,31%) e 95 imagens (81,90%), respectivamente. E para o grupo PP versus ES o classificador árvore de decisão separou corretamente para as visões CC, MLO e CC/MLO 76 imagens (91,57%), 72 imagens (86,75%) e 145 imagens (87,35%), respectivamente.

A Figura 5.37 ilustra o gráfico comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando as características dimensão fractal e lacunaridade para o grupo NMS versus ES.

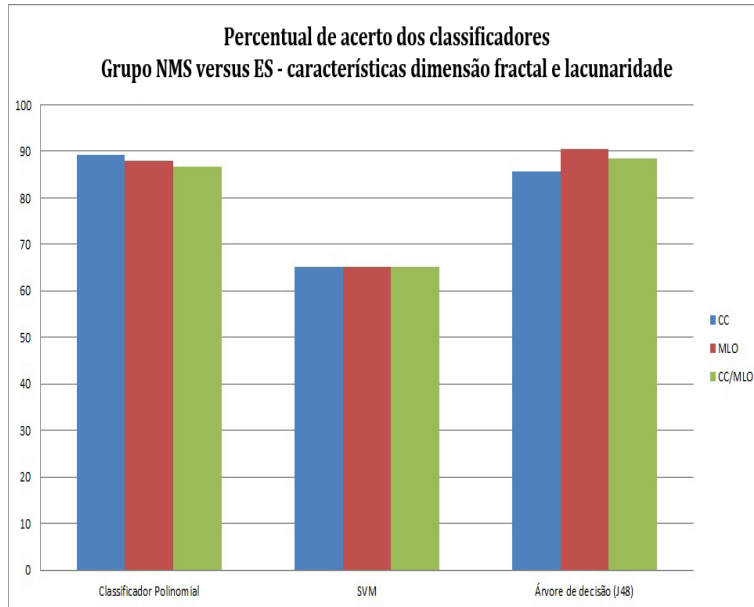


Figura 5.37: Comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando as características dimensão fractal e lacunaridade para o grupo NMS versus ES.

A Figura 5.38 mostra o gráfico comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando as características dimensão fractal e lacunaridade para o grupo NMS versus PP.

5.3. CLASSIFICAÇÃO POLINOMIAL PARA IDENTIFICAÇÃO DE PIXELS PARCIAIS EM REGIÕES DE INTERESSE DE IMAGENS MAMOGRÁFICAS

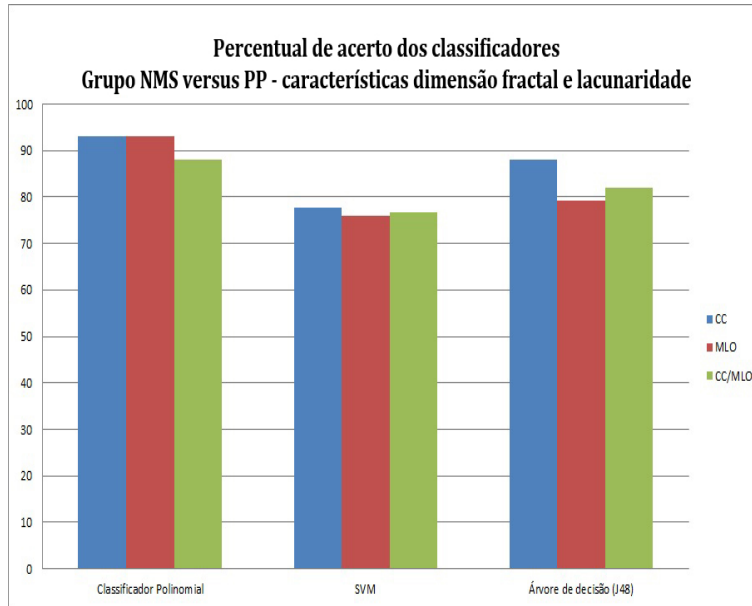


Figura 5.38: Comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando as características dimensão fractal e lacunaridade para o grupo NMS versus PP.

A Figura 5.39 ilustra o gráfico comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando as características dimensão fractal e lacunaridade para o grupo PP versus ES.

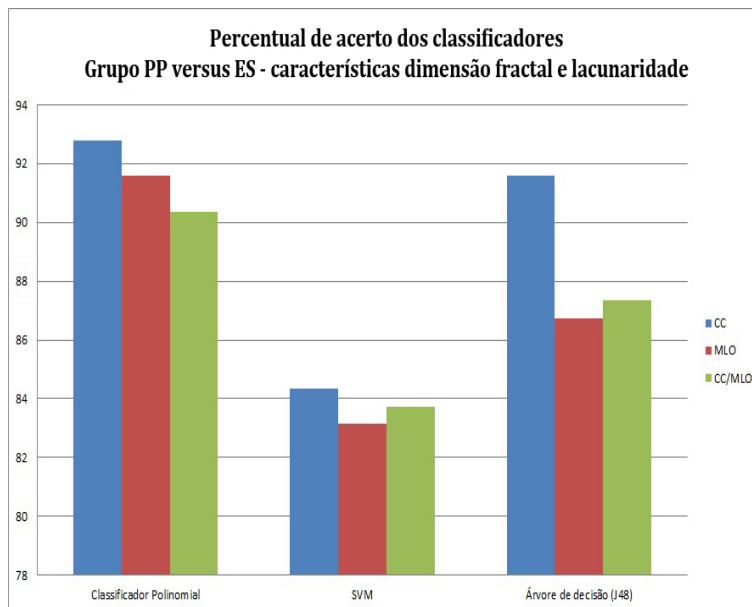


Figura 5.39: Comparativo do percentual de acerto dos classificadores polinomial, SVM e árvore de decisão utilizando as características dimensão fractal e lacunaridade para o grupo PP versus ES.

Ao comparar os resultados obtidos na Tabela 5.27 e nas Figuras 5.37 a 5.39 pode-se observar que o classificador polinomial obteve melhores resultados do que o classificador SVM em todos os grupos, para todas as projeções. Comparando com o classificador árvore de decisão, o classificador polinomial só não apresentou melhores resultados no grupo NMS versus ES nas projeções MLO e CC/MLO. Para os grupos NMS versus PP e PP versus ES, os valores obtidos indicam diferentes comportamentos. Nas projeções CC e MLO, as características dimensão fractal e lacunaridade, utilizando o classificador polinomial, permitiram uma taxa de acerto maior do que 90%.

O menor número de imagens que o classificador polinomial separou corretamente foi 48 imagens (82,76%) no grupo NMS versus PP considerando a característica lacunaridade na visão MLO. Mesmo assim esse classificador apresentou melhores resultados do que os classificadores SVM e árvore de decisão como mostrado na Tabela 5.26. No entanto, utilizando como entrada as características dimensão fractal e lacunaridade esse número subiu para 54 imagens (93,10%) nesse mesmo grupo como ilustrado na Tabela 5.27.

Os valores apresentados nas Tabelas 5.25 a 5.27 permitem mostrar que existem diferenças entre ES, PP e NMS, justificando a necessidade de considerar PP como uma classe intermediária que pode influenciar na identificação das estruturas sadias e doentes.

5.4 Conclusões Deste Capítulo

Este capítulo apresentou os resultados obtidos nos testes realizados utilizando o classificador polinomial na classificação da base de dados IRIS e na identificação de pixels parciais em ROIs de imagens mamográficas. Com o objetivo de comparar os resultados obtidos com esse classificador foram realizados também testes com os classificadores máquina de vetores suporte (SVM) e árvore de decisão (J48).

Nos testes realizados utilizando a base de dados IRIS foram utilizadas duas, três e quatro características dessa base. Pôde-se verificar que o classificador polinomial classificou corretamente as classes das flores setosa, versicolor e virginica da base de dados IRIS ao utilizar três ou quatro características.

O classificador polinomial separou corretamente todas as flores de cada uma das três classes utilizando a base polinomial, porém o grau utilizado depende de quais características foram utilizadas, enquanto os classificadores SVM e árvore de decisão não conseguiram esse mesmo resultado na separação dessas três classes.

Uma grande vantagem do classificador polinomial é a possibilidade de utilizar menos características sem perder a qualidade de classificação. Isto é fundamental quando pretende-se selecionar as melhores características de um conjunto diminuindo o custo computacional do processo. Além disso, esse classificador consegue identificar as melhores características para realizar a separação entre as classes.

Também foi apresentado um estudo para quantificar os comportamentos das ROIs de imagens mamográficas sadias e com câncer utilizando segmentação baseada em entropia, técnicas fractais e o classificador polinomial.

O uso da dimensão fractal, da lacunaridade multiescalas e das assinaturas fractais correspondentes permitiram quantificar as estruturas mamárias sadias, doentes e parciais. A utilização combinada do classificador polinomial e das características fractais forneceram diferenças significativas principalmente na projeção CC, nos grupos NMS versus PP e PP versus ES. Essa abordagem permitiu comprovar que os pixels parciais têm comportamentos diferentes, úteis para possibilitar propostas de modelos ou técnicas computacionais de apoio ao diagnóstico considerando ROIs mamográficas.

Os resultados obtidos neste capítulo mostraram a importância do classificador polinomial na classificação propriamente dita, na seleção das características e também no processo de otimização.

O próximo capítulo apresenta as conclusões e as contribuições deste trabalho, os trabalhos futuros que poderão ser desenvolvidos a partir desta tese e a publicação deste trabalho.

Capítulo 6

Conclusões, Contribuições Deste Trabalho e Trabalhos Futuros

6.1 Introdução

Este capítulo apresenta as conclusões e as contribuições deste trabalho, os trabalhos futuros que poderão ser desenvolvidos a partir desta tese e a publicação deste trabalho.

6.2 Conclusões

O objetivo deste trabalho é desenvolver um classificador polinomial de grau n , $n \geq 1$, que expande o vetor de características projetado no espaço \mathbb{R}^d para um espaço de dimensão superior onde é possível a classificação de padrões de duas ou mais classes. Este classificador mapeia um conjunto de amostras de treinamento e classifica as amostras previamente identificadas.

Na fase de treinamento após a entrada dos padrões no classificador polinomial, o próximo passo consiste no mapeamento dos vetores de características d -dimensionais em vetores da base polinomial L -dimensional. Essa expansão no espaço de características mostrou que ao variar o grau da base polinomial desse classificador consegue separar os dados que eram inseparáveis em separáveis.

A implementação do classificador polinomial é simples pois as funções da base polinomial utilizam fórmulas matemáticas simples, como soma e produto. O grande problema na utilização desse classificador é a ordem, ou seja, o grau da base polinomial aplicada, como mostrado na Equação (4.4) do Capítulo 4 deste trabalho. É importante ressaltar que para um grau muito alto o tempo de processamento pode ser elevado e mesmo assim apresentar resultados insatisfatórios.

O aumento ou a diminuição do número de características é outro fator importante utilizado na classificação polinomial. Como mostrado no Capítulo 5 desta tese, esse classificador não conseguiu classificar 100% as flores setosa, versicolor e virginica da base de dados IRIS de Fisher utilizando apenas duas características. Entretanto, ele conseguiu isso com três e quatro características. Os classificadores SVM e árvore de decisão não obtiveram esses mesmos resultados para essa base de dados.

A visualização do espaço de separação mostra que o classificador polinomial procura uma melhor separação entre as classes deixando as regiões limitadas pelas curvas de níveis polinomiais.

Foi mostrado também no Capítulo 5 uma aplicação do classificador polinomial na identificação de pixels parciais em regiões de interesse de imagens mamográficas. Seus resultados foram comparados com os classificadores SVM e árvore de decisão.

Utilizando a característica dimensão fractal foi mostrado na Tabela 5.25 e nas Figuras 5.31 a 5.33 do Capítulo 5 deste trabalho que o classificador polinomial obteve melhores resultados do que o classificador SVM em todos os grupos, para todas as projeções. Comparando com o classificador árvore de decisão, o classificador polinomial só não apresentou melhores resultados no grupo NMS versus ES nas projeções MLO e CC/MLO.

Na Tabela 5.26 e nas Figuras 5.34 a 5.36 o classificador polinomial também conseguiu melhores resultados do que o classificador SVM em todos os grupos, para todas as projeções. O classificador árvore de decisão apresentou melhores resultados no grupo NMS versus ES nas projeções MLO e CC/MLO e no grupo PP versus ES na projeção CC/MLO.

Na Tabela 5.27 e nas Figuras 5.37 a 5.39 o classificador polinomial obteve melhores resultados do que o classificador SVM em todos os grupos, para todas as projeções. Com-

parando com o classificador árvore de decisão, o classificador polinomial só não apresentou melhores resultados no grupo NMS versus ES nas projeções MLO e CC/MLO.

Para os grupos NMS versus PP e PP versus ES, os valores obtidos indicam diferentes comportamentos. Os valores apresentados nas Tabelas 5.25 a 5.27 permitem mostrar que existem diferenças entre ES, PP e NMS, justificando a necessidade de considerar os PP como uma classe intermediária que pode influenciar na identificação das estruturas sadias e doentes.

6.3 Contribuições Deste Trabalho

Este trabalho apresenta o classificador polinomial como um método de classificação eficiente, pois os resultados obtidos mostraram qualidade na separação das classes de dados linearmente separáveis e não linearmente separáveis.

O classificador polinomial conseguiu separar 100% as flores setosa, versicolor e virginica da base de dados IRIS de Fisher utilizando três ou quatro características. Esse resultado não foi alcançado pelos classificadores SVM e árvore de decisão.

A variação no grau do classificador polinomial é um importante resultado deste trabalho, pois mostra que é possível melhorar a classificação de padrões utilizando esse classificador.

Um conjunto de imagens mamográficas da base de dados *Digital Database for Screening Mammography* (DDSM) foi utilizado nos testes do classificador polinomial. O comportamento das estruturas sadias e doentes foi quantificado pelo cálculo da dimensão fractal e lacunaridade multiescalas. E a separação dos pixels parciais foi obtida com a aplicação de um método de segmentação multinível baseado na entropia máxima. Os resultados obtidos pelo classificador polinomial conseguiram identificar que existem diferenças entre os grupos ES, PP e NMS, justificando a necessidade de considerar os PP como uma classe intermediária.

O desempenho do classificador polinomial foi comparado com os classificadores SVM e árvore de decisão mostrando que esse primeiro classificador apresenta melhores resultados.

6.4 Trabalhos Futuros

Este trabalho possibilita pesquisas em classificadores polinomiais, onde novas funções bases além da utilizada nesta tese como mostrado na Equação (4.2) podem ser combinadas de maneiras diferentes com o objetivo de obter novos classificadores.

O classificador polinomial é um método importante de classificação e seu desempenho pode ser comparado com outros classificadores, tais como: redes neurais, algoritmos genéticos, K vizinhos mais próximos, etc.

O classificador polinomial é um classificador simples de fácil implementação e seu estudo pode ser aplicado em outros tipos de imagens médicas.

A aplicabilidade desse classificador é um fator que pode propiciar estudos na identificação de patologias em sinais de voz.

Como o classificador polinomial separa dados não linearmente separáveis, isso possibilita novos estudos considerando diferentes descritores, tais como: transformada wavelets, transformada discreta do cosseno, momentos de zernike, etc. Esse classificador é importante na seleção das melhores características para obter a melhor separação das classes.

6.5 Publicação Deste Trabalho

- A. S. Martins, L. A. Neves, M. Z. Nascimento, M. F. Godoy, E. L. Flôres and G. A. Carrijo. Multiscale Fractal Descriptors and Polynomial Classifier for Partial Pixels Identification in Regions of Interest of Mammographic Images, IEEE Latin America Transactions, vol. 10, n. 4, p. 1999-2005, 2012.

6.6 Considerações Finais Deste Capítulo

O reconhecimento de padrões utilizando o método de classificação polinomial mostrou que ele é um importante método de classificação que pode ser utilizado para classificar dados não linearmente separáveis. A área de aplicação desse classificador é ampla podendo

contribuir para muitas pesquisas. Uma importante aplicação foi a utilização do classificador polinomial nas identificações de pixels parciais em regiões de interesse de imagens mamográficas.

Este capítulo apresentou as conclusões e as contribuições deste trabalho, os trabalhos futuros que poderão ser desenvolvidos a partir desta tese e a publicação deste trabalho.

Referências Bibliográficas

- [1] FRANCO, C. R. de. Novos Métodos de Classificação Nebulosa e Validação de Categorias e suas Aplicações a Problemas de Reconhecimento de Padrões. 2002. 133 f.. Dissertação (Mestrado) - Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2002.
- [2] ROSS, T. J. Fuzzy logic with engineering applications. New York: McGraw-Hill; 1995.
- [3] HUANG, T. M.; KECMAN, V.; KOPRIVA, I. Kernel Based Algorithms for Mining Huge Data Sets: Supervised, Semi-supervised, and Unsupervised Learning (Studies in Computational Intelligence). New York: Springer-Verlag; 2006.
- [4] PAPA, J. P. Classificação supervisionada de padrões utilizando floresta de caminhos ótimos. 2008. Tese (Doutorado). UNICAMP. Campinas, 2008.
- [5] BLAKE, C. L.; MERZ, C. J. UCI Repository of Machine Learning Databases, Univ. of California, Irvine, Dept. of Information and Computer Science, <http://www.ics.uci.edu/mlearn/MLRepository.html>, 1998.
- [6] BERRY, M. J. A.; LINOFF, G. S. Data mining techniques: for marketing, sales, and customer relationship management. New York: John Wiley; 2004.
- [7] ZHANG, Y.; TOMURO, N.; FURST, J.; RAICU, D. S. Building an ensemble system for diagnosing masses in mammograms. International Journal of Computer Assisted Radiology and Surgery. 2012; 7(2): 323-329.
- [8] FRANK, E. HALL, M.; TRIGG, L. Weka 3-Data Mining with Open Source Machine Learning Software in Java. Nova Zelândia: University of Waikato; 2000.

- [9] TUCERYAN, M.; JAIN, A. K. Handbook of Pattern Recognition and Computer Vision (2nd Edition), by C. H. Chen, L. F. Pau, P. S. P. Wang (eds.), pp. 207-248, World Scientific Publishing Co, 1998.
- [10] DANTAS, R. D. Explorando Informações Texturais das Visões Crânio-Caudal e Médio-Lateral Oblíqua para Extração e Classificação Fuzzy de Imagens Mamográficas. 2010. Dissertação (Mestrado). Universidade Federal do ABC, Programa de Pós-graduação em Engenharia da Informação, Santo André, 2010.
- [11] PEREIRA, R. R.; AZEVEDO, M. P.; HONDA, M.; KINOSHITA, S.; ELGELMANN, R.; MURAMATSU, C.; DOI, K. Usefulness of Texture Analysis for Computerized Classification of Breast Lesions on Mammograms. *Journal of Digital Imaging*. 2007; v. 20: 248-255.
- [12] KINOSHITA, S.; AZEVEDO, M. P.; PEREIRA, R. R.; RODRIGUES, J.; RANAYAN, R. Content-based retrieval of mammograms using visual features related to breast density patterns. *Journal of Digital Imaging*. 2007; v. 20: 172-190.
- [13] JIANG, J.; YAO, B.; WASON, A. M. A Genetic Algorithm Design for Microcalcification Detection and Classification in Digital Mammograms. *Computerized Medical Imaging and Graphics*. 2007; v. 31: 49-61.
- [14] DOI, K.; MACMAHON, H.; KATSURAGAMA, S.; NISHIKAWA, R.; JIANG, Y. Computer-Aided Diagnosis in Radiology: Potential and Pitfalls. *European Journal of Radiology*. 1999; v. 31: 97-109.
- [15] TANG, J.; RANGAYAN, R.; XU, J.; EL NAGA, I.; YANG, Y. Computer-Aided Detection and Diagnosis of Breast Cancer With Mammography: Recent Advances. *IEEE Transactions on Information Technology in Biomedicine*. 2009; v. 13: 236-251.
- [16] NISHIKAWA, R. Current status and future directions of computer-aided diagnosis in mammography, *Computerized Medical Imaging and Graphics*. 2007; v. 31: 224-235.

- [17] SCLOWITZ, M. L.; MENEZES, A. M. B.; GIGANTE, D. P.; TESSARO, S. Conduitas na prevenção secundária do câncer de mama e fatores associados, *Revista de Saúde Publica*. 2005; 39(3): 340-349.
- [18] MONTAG, A.; KUMAR, V. O sistema genital feminino e a Mama, Em: A. K. Abbas, V. Kumar, N. Fausto, R. N. Mitchell, editors. *Robbins Patologia Básica*, 8 ed. Rio de Janeiro: Saunders Elsevier; pp.775-818, 2008 .
- [19] ROVEDA JUNIOR, D.; PIATO, S.; OLIVEIRA, V. M. de; RINALDI, J. F.; FERREIRA, C. A. P.; FLEURY, E. de C. F. Valores preditivos das categorias 3, 4 e 5 do sistema BI-RADS em lesões mamárias nodulares não-palpáveis avaliadas por mamografia, ultra-sonografia e ressonância magnética. 2007. *Radiol Bras*; 40(2): 93-98.
- [20] COSTARIDOU, L.; SAKELLAROPOULOS, P.; STEFANOYIANNIS, A. P.; UNGUREANU, E.; PANAYIOYAKIS, G. Quantifying image quality at breast periphery vs mammary gland in mammography using wavelet analysis. *The British Journal of Radiology*. 2001; v. 74: 913-919.
- [21] MENCATTINI, A.; SALMERI, M.; LOJACONO, R.; FRIGERIO, M.; CASELLI, F. Mammographic images enhancement and denoising for breast cancer detection using dyadic wavelet processing. *IEEE Trans. on Instrumentation and Measurement*. 2008; 57(7): 1422-1430.
- [22] MARQUES, P. M. A. Diagnóstico Auxiliado por Computador na Radiologia. *Radiologia Brasileira*. 2001; v. 34(5): 285-293.
- [23] GONZALES, R. C.; WOODS, R. E. *Processamento de Imagens Digitais*. Tradução: Roberto Marcondes Cesar Junior, Luciano da Fontoura Costa, Edgard Blücher, São Paulo, 2000.
- [24] SILVA, L. A. Categorização de imagens médicas baseada em transformada wavelet e mapas auto-organizáveis, 2009. Tese (Doutorado). Escola Politécnica da Universidade de São Paulo. São Paulo, 2009.

- [25] MÜLLER, H.; MICHOUX, N.; BANDON, D.; GEISSBUHLER, A. A review of content-based image retrieval systems in medical applications-clinical benefits and future directions. *International Journal of Medical Informatic*. 2004; v. 73: 1-23.
- [26] SILVA, L. A. da; DEL-MORAL-HERNANDEZ, E.; RANGAYYAN, R. M. Classification of breast masses using a committes machine of artificial neural networks. *Journal of Eletronic Imaging*. 2008; v. 17, n. 1: 13-17.
- [27] CHAUDHURI, B. B.; SARKAR, N. Texture Segmentation Using Fractal Dimension. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1995; n.17: 72-76.
- [28] BRUNO, O. M.; PLOTZE, R. O.; FALVO, M.; CASTRO, M. Fractal dimension applied to plant identification. *Information Sciences*. 2008; 178(12): 2722-2733.
- [29] CARLIN, M. Measuring the complexity of non-fractal shapes by a fractal method. *Pattern Recognition Letters*. 2000; 21: 1013-1017.
- [30] KENKEL, N. C.; WALTER, D. J. Fractals and ecology. *Abst. Bot*. 1993; 17: 53-70.
- [31] SUGIHARA, G.; MAY, R. M. Applications of fractals in ecology. *Trends Ecol*. 1990; 5: 79-86.
- [32] GULICK, D. *Encounters with Chaos*. New York: McGraw-Hill International Editions - Mathematics and Statistics Series; 1992.
- [33] CESAR, R. M.; COSTA, L. F. *Shape Analysis and Classification: Theory and Practice*. Hardcover: CRC Press; 2000.
- [34] FALCONER, K. J. *Fractal Geometry: mathematical and applications*. New York: John Wiley; 1990.
- [35] TRICOT, C. *Curves and Fractal Dimension*. New York: Springer-Verlag; 1995.
- [36] CORNFORTH, D.; JELINEK, H.; PEICHL, L. Fractop: a tool for automated biological image classification. In: 6th Australasia-Japan Joint Workshop, ANU, Canberra, Australia 30 November - 1 December 2002.

- [37] MANDELBROT, B. B. The Fractal Geometry of Nature. 19 ed. New York: W. H. Freeman & Company; 2000.
- [38] COELHO, R. C.; COSTA, L. da F. The Box-Counting Fractal. Dimension: Does it provide an Accurate Subsidy for Experimental Shape Characterization? If So, How to Use It?. In: Anais do Sibgrapi 95, 1995. 183-191.
- [39] ALLAIN, C.; CLOITRE, M. Characterizing the lacunarity of random and deterministic fractals sets, Phys. Rev. A. 1991; 44(6): 3552-3558.
- [40] PLOTNICK, R. E.; GRADNER, R. H.; HARGROVE, W. W.; PRESTEGAARD, K.; PERLMUTTER, M. Lacunarity analysis: a general technique for the analysis of spatial patterns. Physical Review E, v. 53, n. 5. 1996, 5461-5468.
- [41] BACKES, A. R.; BRUNO, O. M. Segmentação de texturas por análise de complexidade, INFOCOMP Journal of Computer Science. 2006; v. 5: 87-95.
- [42] DUDA, R. O.; HART, P. E.; STORK, D. G. Pattern Classification. 2. ed. New York: Wiley-Interscience; 2000.
- [43] BEZERRA, M. E. R. Métodos Baseados na regra do vizinho mais próximo para reconhecimento de imagens. (Trabalho de Conclusão de Curso). Recife: Escola Politécnica de Pernambuco, Engenharia da Computação; 2006.
- [44] RAICHAROEN, T.; LURSINSAP, C. A Divide-And-Conquer Approach To The Pairwise Opposite Class-Nearest Neighbor (POC-NN) Algorithm. Pattern Recognition Letters, 2005; v. 26, n. 10: 1554-1567.
- [45] VAPNIK, V. N. The nature of statistical learning theory. New York: Springer-Verlag; 1995.
- [46] SUNG, A. H.; MUKKAMALA, S. Identifying important features for intrusion detection using support vector machines and neural networks. Proceedings of International Symposium on Applications and the Internet; 2003; Orlando, Florida. USA: 209-216.

- [47] DING, C. H.; DUBCHAK, I. Multi-class protein fold recognition using support vector machines and neural networks. 2001. *Bioinformatics*; v. 17, n. 4: 349-358.
- [48] LORENA, A. C.; CARVALHO, A. C. P. L. de. *Introdução aos Classificadores de Margens Largas*. São Carlos: 2003. ICMC, 52 p. (Relatório Técnico n. 195).
- [49] LORENA, A. C.; CARVALHO, A. C. P. L. de. *Introdução às Máquinas de Vetores Suporte*. São Carlos:2003. ICMC, 66 p. (Relatório Técnico n. 192).
- [50] SMOLA, A. J. et al. Introduction to large margin classifiers. In: *Advances in Large Margin Classifiers*. Morgan-Kaufman, London, England: 1999. cap. 1, 1-28.
- [51] LIMA, A. R. G. *Máquinas de Vetores Suporte na Classificação de Impressões Digitais*. 2002. Dissertação (Mestrado). Universidade Federal do Ceará. Fortaleza, 2002.
- [52] MEDEIROS, I. P.; CASTRO FILHO, C. A. P.; ERTHAL, G. J.; DUTRA, L. V. Classificação de imagens pelo método de Árvore de Decisão Oblíqua. In: *SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO*, 15. (SBSR), 2011, Curitiba. Anais... São José dos Campos: INPE, 2011. p. 4255-4262. DVD, Internet. ISBN 978-85-17-00056-0 (Internet), 978-85-17-00057-7 (DVD). Disponível em: <<http://urlib.net/3ERPFQRTRW/39UFKJE>>. Acesso em: 13 out. 2012.
- [53] YILDIZ, O. T.; ALPAYDIN, E. Linear discriminant trees. *International Journal of Pattern Recognition and Artificial Intelligence*. v. 19, n. 3. 2005. 763-782.
- [54] CHPELLE, O.; SCHOLKOFF, B.; ZIEN, A. *Semi-supervised learning*. Cambridge: MIT Press; 2006.
- [55] TAN, P.-N.; STEINBACH, M.; KUMAR, V. *Introduction to Data Mining*. Boston: Addison-Wesley Longman Publishing; 2005.
- [56] QUINLAN, J. R. Induction of decision trees. *Machine Learning*, 2005; 1(1): 81-106.
- [57] QUINLAN, J. R. *C4.5: programs for machine learning*. San Francisco: Morgan Kaufmann Publishers Inc.; 1993.

- [58] BREIMAN, L.; FRIEDMAN, J. H.; OLSHEN, R. A.; STONE, C. J. Classification and Regression Trees. Monterey, Calif.: Wadsworth; 1984.
- [59] BRAMER, M. Principles of data mining. London: Springer; 2007.
- [60] BASGALUPP, M. P. LEGAL-Tree: Um algoritmo genético multi-objetivo lexicográfico para indução de árvores de decisão. 2010. Tese (Doutorado), ICMC-USP, São Carlos, 2010.
- [61] QUINLAN, J. R. Decision trees and multivalued attributes. Machine Intelligence, 1988; 11: 305-318.
- [62] ROKACH, L.; MAIMON, O. Data mining with decision trees: Theory and applications. Singapore: World Scientific Publishing; 2008.
- [63] HAN, J. Data Mining: Concepts and Techniques. San Francisco: Morgan Kaufmann Publishers Inc.; 2001.
- [64] KOHANI, R. Scaling up the accuracy of Naive-Bayes classifiers: a decision-tree hybrid. In: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, 1996. 202-207.
- [65] FREUND, Y.; MASON, L. The alternating decision tree learning algorithm. In Proc. 16th International Conf. on Machine Learning, Morgan Kaufmann, San Francisco: 1999; 124-133.
- [66] LANDWEHR, N.; HALL, M.; FRANK, E. Logistic model trees. Machine Learning, 2005; 59(1-2): 161-205.
- [67] SHI, H. Best-first decision tree learning. Master's thesis, University of Waikato, Hamilton, NZ. COMP594. 2007.
- [68] FONSECA, J. M. M. R. da. Indução de árvores de decisão. 1994. Dissertação (Mestrado). Universidade Nova de Lisboa. Lisboa, 1994.

- [69] SHANABLEH, T.; ASSALEH, K. T., Feature modeling using polynomial classifiers and stepwise regression. *Neurocomputing*. 2010; v. 73: 1752-1759.
- [70] CAMPBELL, W. M.; ASSALEH, K. T.; BROUN, C. C. Speaker Recognition With Polynomial Classifiers. *IEEE Transactions on Speech and Audio Processing*. 2002; 10(4): 205-212.
- [71] ASSALEH, K. T.; CAMPEBELL, W. M. Speaker identification using a polynomial-based classifier, In: *Proceeding of the Fourth International Symposium on Signal Processing and its Applications ISSPA 99*, Brisbane, Australia, August 1999.
- [72] CAMPBELL, W. M.; ASSALEH, K. T. Low-complexity small-vocabulary speech recognition for portable devices. In: *Proceeding of the Fourth International Symposium on Signal Processing and its Applications ISSPA 99*, Brisbane, Australia, August 1999.
- [73] ASSALEH, K.; AL-NASHASH, H. A novel technique for the extraction of fetal ECG using polynomial networks. *IEEE Transactions on Biomedical Engineering*. 2005; 52(6): 1148-1152.
- [74] FUKUNAGA, K. *Introduction to Statistical Pattern Recognition*. New York: Academic Press; 1990.
- [75] SPECHT, D. F. Generation of polynomial discriminant functions for pattern recognition. *IEEE Transactions on Electronic Computers*. 1967; v. EC-16: 308-319.
- [76] SANTOS, C. J.; PIANTELLA, A. C. O Teorema da Aproximação de Weierstrass: uma demonstração probabilística. *XXXIII Congresso Nacional de Matemática Aplicada e Computacional*. 2010; v. 3: 1161-1162.
- [77] DEVROYE, L.; GYORFI, L.; LUGOSI, G. *A Probabilistic Theory of Pattern Recognition*. New York: Springer-Verlag; 1996.

- [78] GARCIA, S. C. Uso de Arvore de Decisão na Descoberta de Conhecimento na Área da Saúde. 2003. Dissertação (Mestrado). Porto Alegre: Instituto de Informática, Universidade Federal do Rio Grande do Sul, 2003.
- [79] GUPTA, S.; MARKEY, M. K. Correspondence in texture features between two mammographic views. *Medical Physics*, 2005; 32: 1598-1600.
- [80] MENCATTINI, A.; SALMERI, M.; RABOTTINO, G.; SALICONE, S. Metrological characterization of a cadx system for the classification of breast masses in mammograms. *IEEE Trans. on Instrumentation and Measurement*. 2010; 59(11): 2792-2799.
- [81] NEVES, L. A.; OLIVEIRA, F. R.; PERES, F. A.; MOREIRA, R. D.; MORIEL, A. R.; GODOY, M. F. de; MURTA JUNIOR, L. O. Maximum entropy, fractal dimension and lacunarity in quantification of cellular rejection in myocardial biopsy of patients submitted to heart transplantation. *Journal of Physics: Conference Series*, 285, 2011; 012-032.