

ANDRÉIA MARINI

**ANÁLISE DE ESTRATÉGIAS DE
REJEIÇÃO PARA PROBLEMAS COM
MÚLTIPLAS CLASSES UTILIZANDO
CURVAS ROC**

Dissertação apresentada ao Programa de Pós-Graduação em Informática da Pontifícia Universidade Católica do Paraná como requisito parcial para obtenção do título de Mestre em Informática.

Curitiba
2007

ANDRÉIA MARINI

ANÁLISE DE ESTRATÉGIAS
DE REJEIÇÃO PARA
PROBLEMAS COM
MÚLTIPLAS CLASSES
UTILIZANDO CURVAS ROC

Dissertação apresentada ao Programa de Pós-Graduação em Informática da Pontifícia Universidade Católica do Paraná como requisito parcial para obtenção do título de Mestre em Informática.

Área de Concentração: Ciência da Computação

Orientador: Prof. Dr. Alessandro L. Koerich

Curitiba
2007

Marini, Andréia

ANÁLISE DE ESTRATÉGIAS DE REJEIÇÃO PARA PROBLEMAS COM MÚLTIPLAS CLASSES UTILIZANDO CURVAS ROC. Curitiba, 2007.

Dissertação - Pontifícia Universidade Católica do Paraná. Programa de Pós-Graduação em Informática.

1. Estratégias de Rejeição 2. Curvas ROC 3. Problemas com múltiplas classes I.Pontifícia Universidade Católica do Paraná. Centro de Ciências Exatas e Tecnologia. Programa de Pós-Graduação em Informática II - t

Agradecimentos

Agradeço ao Jerri que teve que suportar minha ausência em tantos momentos que poderíamos estar juntos. Aos meus familiares e meus amigos que sempre me incentivam nos momentos que preciso ter força para seguir a diante. Ao professor Alessandro L. Koerich pela orientação, paciência e disponibilidade, durante a elaboração deste trabalho. Também, ao professor Luiz Eduardo S. Oliveira, pelos questionamentos e contribuições.

Sumário

Agradecimentos	i
Sumário	ii
Lista de Figuras	v
Lista de Tabelas	ix
Lista de Símbolos	x
Lista de Algoritmos	xi
Lista de Abreviações	xii
Resumo	xiii
Abstract	xiv
Capítulo 1	
Introdução	1
1.1 Descrição do Problema	2
1.2 Objetivos	3
1.3 Justificativas	3
1.4 Contribuições	4
1.5 Estrutura do Trabalho	5
Capítulo 2	
Revisão Bibliográfica	6
2.1 Reconhecimento de Padrões	6
2.2 Redes Neurais Artificiais	8
2.3 Estratégias de Rejeição	9
2.4 Fundamentação Estatística	9
2.5 Análise ROC	11
2.5.1 Definições para ROC	11
2.5.2 Curvas ROC	12

2.5.3	Interpretação do Gráfico ROC	13
2.6	Trabalhos Relacionados	14
2.7	Análise Crítica	20

Capítulo 3

Metodologia Proposta		21
3.1	Definição do Problema	22
3.2	Definição da Base de Dados	22
3.3	Classificação	23
3.3.1	Classificação para duas classes	24
3.3.2	Classificação para múltiplas classes	25
3.4	Estratégias de Rejeição	27
3.4.1	Metodologia aplicada aos experimentos	27
3.4.2	Rejeição com um limiar(Chow)	30
3.4.3	Rejeição com múltiplos limiares (Fumera)	30
3.4.4	Rejeição com múltiplos limiares (FumeraMod)	31
3.4.5	Rejeição utilizando a diferença entre os valores de confiança (DIF)	31
3.4.6	Rejeição utilizando Média das Classes (<i>Class Average</i>)	32
3.4.7	Rejeição utilizando Média da Classes e Desvio Padrão (<i>Class Average and Standard Deviation</i>)	34
3.5	Avaliação de Desempenho	34
3.6	Interpretação dos Resultados	36

Capítulo 4

Experimentos e Resultados		37
4.1	Experimentos em Problemas com Duas Classes	37
4.1.1	Dados Balanceados e Desbalanceados	38
4.1.1.1	Análise do Impacto do Balanceamento	38
4.1.2	Separação dos Dados	40
4.1.2.1	Análise do Impacto da Separação dos Dados	40
4.1.3	Distribuição de Probabilidade	42
4.1.3.1	Análise do impacto da Distribuição de Probabilidade	43
4.1.4	Resumo dos Resultados	43
4.2	Experimentos em problemas com múltiplas classes	65
4.2.1	Base NIST Caracteres Maiúsculos - (<i>Upper</i>)	65
4.2.2	Base NIST Caracteres Minúsculos - (<i>Lower</i>)	67

4.2.3	Base NIST Caracteres Maiúsculos e Minúsculos - (<i>UpperLower</i>) . . .	69
4.2.4	Análise dos resultados utilizando múltiplas classes	71
Capítulo 5		
Conclusão		73
5.1	Conclusões	73
5.2	Trabalhos Futuros	75
Referências Bibliográficas		76
Apêndice A		
Experimentos com duas classes		79
A.1	Distribuição Normal	79
A.2	Distribuição Normal Multivariável	88
A.3	Distribuição <i>Chi-square</i>	89

Lista de Figuras

Figura 2.1	O teste de hipóteses levando em consideração a variação do limiar. Adaptada de (MASSAD et al., 2004) ilustrando distribuição de doentes e não doentes com um valor de limiar para corte.	10
Figura 2.2	Exemplo de um gráfico ROC relacionando FAR e FRR para diferentes estratégias de rejeição.	13
Figura 2.3	Aplicação da regra de Chow para as probabilidades <i>a posteriori</i> “verdadeiras” e “estimadas” (FUMERA; ROLI; GIACINTO, 2000).	16
Figura 2.4	Utilização de dois limiares de rejeição diferentes T1 e T2 para a tarefa de classificação da Figura 2.3 (FUMERA; ROLI; GIACINTO, 2000). . .	17
Figura 3.1	Diagrama de procedimentos aplicados à avaliação das estratégias de rejeição.	21
Figura 3.2	Diversidade de exemplos na base de dados NIST.	26
Figura 3.3	Detalhamento da saída da RNA para um conjunto qualquer com duas classes e nove instâncias. Apresentando a probabilidade <i>a posteriori</i> atribuída para cada uma das classes (níveis de confiança). Representação normalizadas expressas por D e R.	28
Figura 3.4	Obtendo os limiares através do conjunto de validação.	29
Figura 3.5	Obtendo os limiares de forma automática através do conjunto de validação.	29
Figura 3.6	Testando os limiares através do conjunto de testes.	30
Figura 3.7	Exemplo de uma matriz de confusão para um problema de 26 classes.	35
Figura 4.1	Conjunto de dados NBS - Gráfico Erro/Rejeição e Gráfico ROC para as seis estratégias implementadas.	44
Figura 4.2	Conjunto de dados NBPS - Gráfico Erro/Rejeição e Gráfico ROC para as seis estratégias implementadas.	45

Figura 4.3	Conjunto de dados NBSob - Gráfico Erro/Rejeição e Gráfico ROC para as seis estratégias implementadas.	46
Figura 4.4	Conjunto de dados NDS - Gráfico Erro/Rejeição e Gráfico ROC para as seis estratégias implementadas.	47
Figura 4.5	Conjunto de dados NDPS - Gráfico Erro/Rejeição e Gráfico ROC para as seis estratégias implementadas.	48
Figura 4.6	Conjunto de dados NDSob - Gráfico Erro/Rejeição e Gráfico ROC para as seis estratégias implementadas.	49
Figura 4.7	Conjunto de dados NMBS - Gráfico Erro/Rejeição e Gráfico ROC para as seis estratégias implementadas.	50
Figura 4.8	Conjunto de dados NMBPS - Gráfico Erro/Rejeição e Gráfico ROC para as seis estratégias implementadas.	51
Figura 4.9	Conjunto de dados NMBSob - Gráfico Erro/Rejeição e Gráfico ROC para as seis estratégias implementadas.	52
Figura 4.10	Conjunto de dados NMDS - Gráfico Erro/Rejeição e Gráfico ROC para as seis estratégias implementadas.	53
Figura 4.11	Conjunto de dados NMDPS - Gráfico Erro/Rejeição e Gráfico ROC para as seis estratégias implementadas.	54
Figura 4.12	Conjunto de dados NMDSob - Gráfico Erro/Rejeição e Gráfico ROC para as seis estratégias implementadas.	55
Figura 4.13	Conjunto de dados CBS - Gráfico Erro/Rejeição e Gráfico ROC para as seis estratégias implementadas.	56
Figura 4.14	Conjunto de dados CBPS - Gráfico Erro/Rejeição e Gráfico ROC para as seis estratégias implementadas.	57
Figura 4.15	Conjunto de dados CBSob - Gráfico Erro/Rejeição e Gráfico ROC para as seis estratégias implementadas.	58
Figura 4.16	Conjunto de dados CDS - Gráfico Erro/Rejeição e Gráfico ROC para as seis estratégias implementadas.	59
Figura 4.17	Conjunto de dados CDPS - Gráfico Erro/Rejeição e Gráfico ROC para as seis estratégias implementadas.	60
Figura 4.18	Conjunto de dados CDSob - Gráfico Erro/Rejeição e Gráfico ROC para as seis estratégias implementadas.	61
Figura 4.19	Avaliação das estratégias de rejeição utilizando o compromisso erro/rejeição para o conjunto Upper.	66
Figura 4.20	Avaliação das estratégias de rejeição utilizando Curva ROC para o conjunto Upper.	66

Figura 4.21 Avaliação das estratégias de rejeição utilizando o compromisso erro/rejeição para Base Lower.	68
Figura 4.22 Avaliação das estratégias de rejeição utilizando Curva ROC para Base Lower.	68
Figura 4.23 Avaliação das estratégias de rejeição utilizando o compromisso erro/rejeição para Base UpperLower.	70
Figura 4.24 Avaliação das estratégias de rejeição utilizando Curva ROC para Base UpperLower.	70
Figura A.1 Separação dos dados para uma Distribuição Normal - NBS.	81
Figura A.2 Separação dos dados para o conjunto de teste e respectiva superfície de separação traçada pela RNA-MLP para dados NBS.	81
Figura A.3 Separação dos dados para uma Distribuição Normal - NBPS.	82
Figura A.4 Separação dos dados para o conjunto de teste e respectiva superfície de separação traçada pela RNA-MLP para dados NBPS.	82
Figura A.5 Separação dos dados para uma Distribuição Normal - NBSob.	83
Figura A.6 Separação dos dados para o conjunto de teste e respectiva superfície de separação traçada pela RNA-MLP para dados NBSob.	83
Figura A.7 Separação dos dados para uma Distribuição Normal - Desbalanceados - Separados - NDS.	84
Figura A.8 Separação dos dados para o conjunto de validação e respectiva superfície de separação traçada pela RNA-MLP para o experimento NDS.	84
Figura A.9 Separação de dados para o conjunto de teste e respectiva superfície de separação traçada pela RNA-MLP para o experimento NDS.	85
Figura A.10 Separação de dados para uma Distribuição Normal - Desbalanceados - Parcialmente Sobrepostos - NDPS.	85
Figura A.11 Separação de dados para o conjunto de teste e respectiva superfície de separação traçada pela RNA-MLP para o experimento NDPS.	86
Figura A.12 Separação de dados para uma Distribuição Normal - Desbalanceados - Sobrepostos - NDSob.	86
Figura A.13 Separação de dados para o conjunto de teste e respectiva superfície de separação traçada pela RNA-MLP para o experimento NDSob.	87
Figura A.14 Separação de dados em uma distribuição Normal Multivariável - Dados Desbalanceados - Sobrepostos - NMDSOB.	88
Figura A.15 Separação de dados para o conjunto de teste e respectiva superfície de separação traçada pela RNA-MLP para o experimento NMDSOB.	88

Figura A.16 Separação de dados em uma Distribuição <i>Chi-square</i> - Balanceados	
- Dados Separados - CDS.	89
Figura A.17 Separação de dados para o conjunto de teste e respectiva superfície	
de separação traçada pela RNA-MLP para o experimento CDS.	89

Lista de Tabelas

Tabela 2.1	Resumo de ações para teste de hipóteses	11
Tabela 3.1	Matriz de confusão clássica.	34
Tabela 3.2	Matriz de confusão para problemas com múltiplas classes	35
Tabela 4.1	Resumo do comportamento das estratégias de rejeição em relação a problemas com duas classes gerados através de uma Distribuição Normal.	62
Tabela 4.2	Resumo do comportamento das estratégias de rejeição em relação a problemas com duas classes gerados através de uma Distribuição Normal Multivariável.	63
Tabela 4.3	Resumo do comportamento das estratégias de rejeição em relação a problemas com duas classes gerados através de uma Distribuição <i>Chi-square</i> .	64
Tabela 4.4	Resumo do comportamento das estratégias de rejeição em relação a problemas multiclases.	72
Tabela A.1	Os conjuntos são construídos alterando os parâmetros de média e variância na Distribuição Normal e Normal Multivariável e graus de liberdade na distribuição <i>Chi-square</i> pré-definindo suas características de apresentação.	80

Lista de Símbolos

w_i	classe do problema
T_i	limiar de rejeição
N	número de classes do problema
x	padrão
C	classificador
c_i	confiança atribuída pelo classificador a um padrão
$P(.)$	probabilidade
$\hat{P}(.)$	probabilidade estimada
μ	média
σ	desvio padrão
D	conjunto de níveis de confiança
R	conjunto de rótulos
N	número de linhas de D
c	número de classes de D

Lista de Algoritmos

1	Algoritmo aplicado na obtenção dos limiares através do conjunto de validação para a estratégia FumeraMod	31
2	Algoritmo aplicado na obtenção dos limiares através do conjunto de validação para a estratégia DIF.	32
3	Algoritmo aplicado na obtenção dos limiares através do conjunto de validação para a estratégia Média das Classes	33
4	Algoritmo aplicado na obtenção dos limiares através do conjunto de validação para a estratégia Média/Desvio	33

Lista de Abreviações

CA	<i>Correta Aceitação</i>
CBPS	<i>Distribuição Chi-square, balanceada e parcialmente sobreposta</i>
CBS	<i>Distribuição Chi-square, balanceada e separada</i>
CBSob	<i>Distribuição Chi-square, balanceada e sobreposta</i>
CDPS	<i>Distribuição Chi-square, desbalanceada e parcialmente sobreposta</i>
CDS	<i>Distribuição Chi-square, desbalanceada e separada</i>
CDSob	<i>Distribuição Chi-square, desbalanceada e sobreposta</i>
CR	<i>Correta Rejeição</i>
CRT	<i>Class - Related Thresholds (Limiar relacionado a classe)</i>
FA	<i>Falsa Aceitação</i>
FAR	<i>Taxa de Falsa Aceitação</i>
FR	<i>Falsa Rejeição</i>
FRR	<i>Taxa de Falsa Rejeição</i>
MLP	<i>Multi - Layer Perceptron (Perceptron multicamadas)</i>
NBPS	<i>Distribuição Normal, balanceada e parcialmente sobreposta</i>
NBS	<i>Distribuição Normal, balanceada e separada</i>
NBSob	<i>Distribuição Normal, balanceada e sobreposta</i>
NDPS	<i>Distribuição Normal, desbalanceada e parcialmente sobreposta</i>
NDS	<i>Distribuição Normal, desbalanceada e e separada</i>
NDSob	<i>Distribuição Normal, desbalanceada e sobreposta</i>
NIST	<i>National Institute of Standards and Technology</i>
NMBPS	<i>Distribuição Normal Multivariável, balanceada e parcialmente sobreposta</i>
NMBS	<i>Distribuição Normal Multivariável, balanceada e separada</i>
NMBSob	<i>Distribuição Normal Multivariável, balanceada e sobreposta</i>
NMDPS	<i>Distribuição Normal Multivariável, desbalanceada e parcialmente sobreposta</i>
NMDS	<i>Distribuição Normal Multivariável, desbalanceada e separada</i>
NMDSob	<i>Distribuição Normal Multivariável, desbalanceada e sobreposta</i>
RNA	<i>Redes Neurais Artificiais</i>
ROC	<i>Receiver Operating Characteristic (Característica de operação do receptor)</i>

Resumo

Estratégias de rejeição são utilizadas para melhorar o desempenho de sistemas de reconhecimento de padrões. Entretanto, as opções descritas na literatura referem-se a problemas bem definidos em conjuntos de dados e objetivos específicos, em geral considerando apenas duas classes. Realizamos um estudo comparativo para caracterizar o desempenho de algumas estratégias de rejeição em problemas com duas classes e posteriormente múltiplas classes, sob diferentes condições de balanceamento e separação dos dados. Os resultados em relação as estratégias implementadas sugerem que as características dos conjuntos de dados utilizados podem influenciar o comportamento dos mecanismos de rejeição e que métodos clássicos da literatura podem apresentar resultados similares a estratégias heurísticas. Dessa maneira, a contribuição do presente trabalho é uma análise de alguns métodos de rejeição utilizando curvas ROC e o compromisso erro/rejeição, destacando sua necessidade e importância na construção de sistemas classificadores confiáveis.

Palavras-chave: Estratégias de rejeição, Problemas com múltiplas classes, Curvas ROC, Compromisso Erro/Rejeição

Abstract

Rejection strategies have been employed to improve the performance of pattern recognition systems. However most of the rejection strategies described in literature are related to well-conditioned data and a limited number of classes, usually only two. We present a comparative study that evaluates several rejection strategies on two-class and multi-class problems but taking into account ill-conditioned data with different balancing and overlapping conditions. The experimental results achieved by the rejection strategies suggest that the characteristics of the data may have an influence on the performance of the rejection strategies, and that classical rejection strategies described in the literature as optimal under certain constraints may be surpassed by heuristics strategies depending on the complexity of the problem. The main contribution of this work is a critical analysis of several rejection methods through ROC and error-rejection curves, highlighting their importance and relevance in building reliable intelligent systems.

Keywords: Rejection strategies, multiclass problems, ROC curves, trade-off error/reject

Capítulo 1

Introdução

A aprendizagem computacional é um conjunto de técnicas envolvendo várias áreas de conhecimento. Um sistema de aprendizagem é um programa de computador que toma decisões baseadas na solução de problemas anteriores. A aprendizagem de máquina está relacionada com programas computacionais que melhoram seu desempenho através da experiência, por exemplo, a medida de desempenho de um programa capaz de jogar xadrez pode ser observada em relação a sua habilidade em vencer, utilizando como experiência jogos de xadrez contra si próprio. A aprendizagem de máquina concentra-se em como modelar um sistema de predição com base em observações prévias, que segundo Mitchell (1997) pode ser definido como:

“Um programa de computador aprende a partir de uma experiência “E” em relação a alguma classe de tarefas “T”, e uma medida de desempenho “P” nas tarefas “T” que aumenta com a experiência “E” adquirida”.

Nos últimos anos, várias técnicas têm sido desenvolvidas e testadas visando encontrar melhores resultados para sistemas inteligentes. A busca é direcionada em atribuir à máquina, capacidade de aprendizagem e generalização similar a de um ser humano, sendo a classificação uma das principais tarefas. A classificação é uma tarefa básica e essencial de aprendizagem em alguns sistemas inteligentes que reproduzem o comportamento humano. De acordo com Mitchell (1997), o desempenho de um classificador deve melhorar com o treinamento. Logo “aprender” consiste em classificar corretamente os dados com um desempenho cada vez melhor. Em qualquer processo de aprendizagem o aprendiz deve utilizar os conhecimentos que possui para obter novos conhecimentos.

Visando incrementar o processo de classificação, o conceito de rejeição admite que um sistema de reconhecimento aplique uma decisão global de aceitar ou recusar uma hipótese se o classificador não estiver certo o suficiente. Uma evidência sobre a certeza é dada pelo valor da probabilidade atribuída às hipóteses fornecidas pelo classificador. A

recusa de uma hipótese pode acontecer quando um padrão é ambíguo não podendo ser associado a uma única classe com grande certeza, ou quando este padrão pode ser associado com baixa confiança à diversas classes. Em um sistema inteligente, uma estratégia de rejeição busca aumentar a confiabilidade através da minimização do efeito dos erros cometidos pelo classificador para um dado nível de rejeição.

1.1 Descrição do Problema

O desempenho de um sistema inteligente é um aspecto importante, pois, pode ser influenciado pela variação da distribuição dos dados, o tamanho da amostra de dados utilizada no treinamento, dimensionalidade dos dados, a habilidade do analista e a aplicação de uma estratégia de rejeição. O ponto chave é que os métodos tradicionalmente utilizados para avaliar o desempenho baseados nas taxas de erros e reconhecimento, podem não levar em conta todas as informações relevantes para o cálculo dessas taxas (PROVOST; FAWCETT, 1997). Como exemplo disso, podemos considerar o fato de que em aplicações reais geralmente há uma desproporção no número de exemplos pertencentes às classes do sistema de reconhecimento, sendo comum o classificador não lidar bem com essa variação. Outra situação que se torna implícita é que os custos de uma classificação correta ou incorreta não são levados em conta, ou seja, o resultado da classificação é indiferente, não provocando conseqüências à solução do problema. Entretanto, é difícil imaginar um domínio de aplicação em que um sistema inteligente não precise preocupar-se com os custos dos erros de classificação. Considerando aplicações reais, o resultado proposto pelo sistema sugere ações que podem vir a manifestar graves conseqüências em ambientes onde equívocos são raramente permitidos (PROVOST; FAWCETT, 1997).

Para ilustrar a relevância desta situação, considere o seguinte exemplo apresentado em (DUDA; HART; STORK, 2000): uma indústria recebe dois tipos de peixes, salmão e robalo. O processo de classificação é manual, porém a indústria gostaria de automatizar tal processo. Para tanto, seriam avaliadas imagens da esteira onde os peixes são recebidos utilizando parâmetros para separar de maneira confiável os dois tipos de peixes. Podemos aceitar a hipótese de que um pedaço de salmão possa ser embalado junto com robalo, mas o contrário jamais pode acontecer. Isso sugere que associado à decisão, surgirá um custo adicional aos rendimentos da empresa. A tarefa de um sistema inteligente consiste em encontrar uma regra de decisão que minimize esse custo, podendo tratar tal questão como um problema de otimização.

Buscamos ao longo do trabalho analisar o comportamento das estratégias de rejeição em diferentes problemas de classificação considerando c classes, onde $c \geq 2$ tentando

responder, como estabelecer uma regra de rejeição eficiente para minimizar os erros de classificação? Neste trabalho estudamos problemas de classificação utilizando uma abordagem tradicional, verificando em seguida seu desempenho utilizando métodos que levam em conta a possibilidade de rejeitar um exemplo, caso o classificador não tenha certeza em associar uma instância a uma classe.

1.2 Objetivos

O principal objetivo deste trabalho é avaliar diferentes estratégias de rejeição através de curvas ROC (*Receiver Operating Characteristics*) em diferentes problemas de classificação, utilizando inicialmente dados sintéticos com parâmetros controlados e em seguida dados reais sobre os quais não temos controle sobre os parâmetros.

Analisamos o impacto da distribuição de dados, quantidade de classes, número de amostras por classe, além de outros parâmetros do desempenho dos métodos de rejeição, realizando um estudo comparativo entre os mesmos. Toda atenção estará voltada a implementação de estratégias de rejeição para identificar os padrões problemáticos que poderiam ter sido reconhecidos de forma incorreta pelo classificador. A originalidade do trabalho esta fundamentada na avaliação de como as características do problema podem impactar nos resultados apresentados pelas diferentes estratégia de rejeição.

1.3 Justificativas

Uma estratégia de rejeição é uma ferramenta útil para melhorar a confiabilidade no processo de classificação. Em algumas aplicações, o custo de rejeitar certos padrões e controlar os procedimentos necessários após sua implementação, realizando uma classificação manual pode ser menor que o custo de uma classificação incorreta. Mesmo trabalhando com taxas de rejeição elevadas essa opção pode se tornar viável em tarefas que são realizadas, tradicionalmente, de forma manual. Dentre as possíveis aplicações para este contexto destacamos a manipulação de vários tipos de documentos, como: formulários, recibos, cheques bancários e envelopes postais. Outro exemplo, caso considerarmos problemas que apresentam custos relacionados a classificação incorreta descrito em (PROVOST; FAWCETT, 1997) é a possibilidade de considerar um paciente doente diagnosticado como saudável. Isso pode ser um erro fatal, fazendo com que o paciente chegue à morte, enquanto que um paciente saudável classificado como doente pode ser considerado como um erro menos sério, uma vez que o erro pode ser corrigido em exames futuros.

Muitos trabalhos encontrados na literatura avaliam estratégias de rejeição em problemas específicos. Não encontramos nenhum estudo mais abrangente que compare várias estratégias de rejeição baseadas em heurísticas e estratégias utilizando formalismo estatístico, como é o caso das estratégias propostas por Chow (1970) e Fumera, Roli e Giacinto (2000). Ainda, outro fator importante é que os trabalhos encontrados na literatura fazem referência a um único problema e procuram encontrar uma solução ótima para tal problema. Neste trabalho estudamos diferentes estratégias de rejeição, pois, a aplicação de uma estratégia em problema real pode variar de acordo com as exigências atribuídas ao sistema de reconhecimento. Para tanto estudamos algumas estratégias de rejeição considerando problemas de classificação que possuem características distintas em conjuntos de dados diferentes.

Para avaliar a eficiência das estratégias, além da redução da taxa de erro propiciada pelo mecanismo de rejeição, analisamos outras estatísticas, tais como, a taxa de falsa aceitação e a taxa de falsa Rejeição. A utilização da análise ROC possui como principais características realizar uma análise independente de certos fatores, como por exemplo: um limiar fixo de classificação; a possibilidade ajustar modelos de classificação; a facilidade de realizar uma análise visual da relação entre as instâncias corretas ou incorretamente classificadas verificando o compromisso entre *falsa aceitação* e *falsa rejeição*. Outro ponto importante é que os pontos no gráfico são independentes das distribuições das instâncias na classe e pesos associados aos erros não sendo influenciada por classes desbalanceadas e diferentes tipos de dados. Concluindo, esta ferramenta apresenta-se de forma adequada para medir e especificar problemas de desempenho provendo uma avaliação mais rica do que simplesmente avaliar um modelo a partir de uma única medida.

1.4 Contribuições

A contribuição científica do presente trabalho é uma análise crítica dos métodos de rejeição, destacando sua necessidade e importância na construção de sistemas classificadores confiáveis, realizando um estudo comparativo para caracterizar o desempenho de alguns métodos de rejeição sob diferentes condições, buscando maximizar o desempenho de um sistema inteligente.

Outra contribuição importante é direcionada a comunidade em geral através de aplicações comerciais e industriais. Um mecanismo de rejeição ótimo pode ser aplicado para resolver diversos problemas onde se faz necessário sistemas confiáveis. Nos dias atuais esta preocupação vai desde evitar fraudes e falsificação em documentos, implantação de sistemas de identificação pessoal baseados em características biométricas, até melhorar a

precisão de sistemas de produção automatizados que buscam realizar seus processos de forma mais rápida e segura.

1.5 Estrutura do Trabalho

Este trabalho desenvolve-se ao longo de cinco capítulos. Após uma breve introdução, o segundo capítulo apresenta uma perspectiva geral sobre o estado da arte relacionada aos métodos de rejeição e à análise ROC. No terceiro capítulo, é apresentada em detalhes a metodologia de desenvolvimento deste trabalho. No capítulo 4 são apresentados os resultados obtidos através dos experimentos realizados. Por fim, o último capítulo apresenta as conclusões e perspectivas de trabalhos futuros.

Capítulo 2

Revisão Bibliográfica

Neste capítulo são apresentados trabalhos relacionados ao tema central da dissertação, a análise ROC e estratégias de rejeição. Embora, o relacionamento de alguns trabalhos com os temas aqui propostos, não seja direto, todos são focados e contribuem muito para o entendimento de tais conceitos.

2.1 Reconhecimento de Padrões

Diversos paradigmas de aprendizagem computacional têm sido explorados para a construção de sistemas inteligentes. Algumas áreas relacionadas a esses tipos de sistemas são: reconhecimento de padrões, aprendizagem de máquina e visão computacional. O termo “padrão” é uma palavra de nosso vocabulário que expressa alguma regularidade, algo capaz de servir como modelo, ou algo representando uma idéia do que foi observado (SCHURMANN, 1996). O reconhecimento de padrões no escopo computacional é descrito como “área relacionada ao reconhecimento de regularidades significativas em ambientes ruidosos ou complexos através de máquinas” (DUDA; HART; STORK, 2000). O reconhecimento de padrões utiliza o processo de classificar objetos por categorias (classes). Um padrão caracteriza uma dada classe, logo, reconhecer um padrão é identificar a classe a qual um dado objeto pertence. Esses objetos podem ser, por exemplo, imagens ou medidas. A maioria das coisas que nos cercam podem ser definidas como padrões. Algumas áreas que utilizam reconhecimento de padrões são: biologia, psicologia, medicina, marketing, visão computacional e engenharia (JAIN; DUIN; MAO, 2000).

A construção de um sistema de aprendizagem envolve diversas atividades: a definição da base de dados, a escolha das características, escolha do algoritmo de treinamento e análise da evolução do classificador. Formalmente, três aspectos são apontados por Jain, Duin e Mao (2000):

1. Aquisição de dados e pré-processamento;
2. Representação de dados;
3. E a decisão do que fazer com esses dados.

Diversos paradigmas de aprendizagem de máquina já foram propostos e vem sendo largamente utilizados. O principais métodos para reconhecimento de padrões são:

- Comparação de modelos (*template matching*): Uma das abordagens para reconhecer padrões é a técnica de comparação de modelos. Trata-se de uma operação genérica que determina a similaridade entre duas entidades do mesmo tipo. Normalmente se apresenta como uma forma 2D ou um protótipo. O padrão a ser reconhecido é comparado, observando todas as variações possíveis em termos de: translação, rotação e mudanças de escalas, com o armazenamento de todos os modelos, tratando-se de um método computacionalmente caro.
- Método estatístico: Utilizando a abordagem estatística, um padrão é representado por um vetor de características com d dimensões. Os conceitos da teoria de decisão estatística são utilizados para estabelecer fronteiras de decisão entre classes e padrões, podendo ser manipulado de duas maneiras: treinamento (aprendizagem) e classificação (teste). Classificadores estatísticos são poderosos em muitos casos, pois, possuem um bom método de tratamento para padrões distorcidos ou com ruídos.
- Método sintático: Esse método freqüentemente resolve algumas limitações do método estatístico, tais como a necessidade de um grande número de exemplos para o projeto de um classificador. Em muitos casos de reconhecimentos de padrões complexos trata-se do método mais adequado.
- Redes Neurais Artificiais: A principal diferença deste método para os anteriores é sua capacidade de aprender relacionamentos complexos não lineares entre dados de entrada e saída através do processo de treinamento. Os modelos de redes neurais utilizam alguns princípios organizacionais como: aprendizado, generalização, adaptabilidade, tolerância à falhas e computação distribuída.

Neste trabalho nos concentramos em aprendizagem supervisionada utilizando como classificador redes neurais artificiais para sua implementação. Detalhes sobre este tópico são apresentados na próxima seção.

2.2 Redes Neurais Artificiais

As redes neurais artificiais (RNA) são sistemas baseados no funcionamento do cérebro humano e caracterizam-se pela união de uma grande quantidade de células de processamentos interligadas por um grande número de conexões, que processam a informação de forma paralela. As pesquisas em RNA foram desenvolvidas, originalmente, na década de 40, pelo neurofisiologista Warren McCulloch, e pelo matemático Walter Pitts, os quais fizeram uma analogia entre as células nervosas vivas e o processo eletrônico em um trabalho publicado em 1943 (MITCHELL, 1997). Retomadas enfaticamente a partir da década de 80, diversos modelos de RNA têm surgido visando aperfeiçoar e aplicar este método. Sua maior vantagem provém de sua capacidade de aprendizado, ou seja, a capacidade de se auto ajustar na tentativa de reconhecer padrões a partir das informações dadas. A capacidade das redes para aprender e generalizar tais relacionamentos as torna menos sensíveis ao ruído que outros sistemas. A capacidade de representar relacionamentos não lineares as torna adequadas para resolver inúmeros problemas de classificação.

As RNA emergiram como uma ferramenta importante para classificação nos últimos tempos. Podemos utilizá-las para resolver inúmeros problemas e combinando-as podemos mudar a arquitetura conforme a necessidade da aplicação. As recentes atividades de pesquisa são vastas em classificação neural, estabelecendo-as como uma alternativa para vários métodos de classificação convencionais (ZHANG, 2000).

No presente trabalho, a classificação é realizada com a utilização de uma RNA do tipo *Multi-Layer Perceptron* (MLP), ou perceptron multicamadas. Tal rede possui uma camada de neurônios artificiais escondidos entre duas camadas (entrada e saída). O perceptron multicamadas consiste de uma rede de neurônios artificiais, organizados em camadas e cada neurônio possui um peso e uma função de ativação. Esses pesos são ajustados na etapa de treinamento através de um algoritmo chamado *backpropagation*. Basicamente, os dados de entrada (treinamento) são submetidos à rede, que irá produzir uma saída. Esta saída é comparada com a saída esperada e os pesos são atualizados de maneira a diminuir o erro médio quadrático (diferença entre saída desejada e saída obtida). A variação do erro em função dos pesos pode ser representada pela descida de gradiente, possibilitando a avaliação da taxa de aprendizagem do algoritmo (MITCHELL, 1997). A escolha das redes neurais como classificador de base se justifica pelo fato de que elas provêm na saída a probabilidade *a posteriori* (RICHARD; LIPPMANN, 1991). As redes neurais podem calcular probabilidade *a posteriori* que serve como base para estabelecer regras de classificação, análises estatísticas e também possibilita a aplicação de uma estratégia de rejeição.

2.3 Estratégias de Rejeição

Segundo Schurmann (1996) a classificação de padrões é uma decisão baseada na observação. Um padrão, para o nosso contexto, é um par de valores $[v, k]$ dos quais somente v está presente no sistema de classificação. O classificador de padrões tem a decisão sobre a classe k para cada padrão apresentado. O conceito de rejeição admite a recusa de uma hipótese caso o classificador não esteja seguro de sua resposta. Tal recusa pode ser fundamentada em duas razões:

1. As evidências não são suficientes para a chegar a uma única decisão, pois, nenhuma das hipóteses parece adequada, ou muitas podem parecer adequadas.
2. O classificador admite nunca ter visto tal caso v , obviamente ocasionando dificuldade para uma classificação correta.

Quando um sistema de classificação supervisionada aplica uma decisão global que possibilita a escolha de aceitar ou rejeitar um resultado, estamos adotando uma estratégia de rejeição. Seu principal objetivo concentra-se em minimizar o número de erros para um dado número de rejeições. Dizemos que um erro de classificação ocorre quando um padrão é associado a uma classe diferente daquela a que ele realmente pertence.

O interesse por estratégias de rejeição para otimização de classificadores é encontrado em muitos trabalhos. A rejeição é cada vez mais necessária para diferentes contextos de classificação principalmente em algumas tarefas que necessitam de uma precisão muito alta ou em qualquer sistema de reconhecimento que não aceite atribuir padrões a classes diferentes das que eles realmente pertencem (MOUCHERE; ANQUETIL, 2006).

Neste trabalho buscamos uma estratégia que mais se aproxime do ideal, possibilitando a rejeição de todos os exemplos classificados incorretamente pelo classificador e aceitação de todos os exemplos classificados corretamente em uma determinada base de dados. A idéia central deste trabalho é avaliar qual estratégia de rejeição mais se aproxima da situação descrita acima, buscando que esta se adapte a mais de uma aplicação e especialmente a problemas com múltiplas classes. A metodologia proposta se concentra em avaliar métodos de rejeição descritos na literatura e métodos heurísticos.

2.4 Fundamentação Estatística

O problema da rejeição pode ser definido como um teste de hipóteses ou tomada de decisão estatística. A teoria estatística sugere rejeitar uma hipótese se a probabilidade desta for menor que certo limiar. Assim, a tarefa de uma estratégia simples de rejeição

deve apenas estabelecer qual é este limiar. Por hipótese estatística entendemos qualquer consideração, pergunta ou suposição, feita acerca de um parâmetro relacionado. Um teste de hipóteses sempre compara duas hipóteses definidas como (MASSAD et al., 2004):

- Hipótese nula H_0 . É a hipótese da não diferença. Supõe que a diferença observada é atribuída somente ao acaso. Em geral, queremos afastá-la e provar que não é válida.
- Hipótese alternativa H_1 . É a hipótese alternativa à hipótese nula. Propõe que haja uma diferença real e não atribuída ao acaso. Em geral, queremos comprová-la.

Um exemplo da utilização do teste de hipóteses é mapear cada exemplo de uma base de dados como um problema com dois estados: a hipótese H_0 e a hipótese H_1 . A primeira assume o incorreto reconhecimento enquanto a segunda o correto. A Figura 2.1 apresenta no eixo x as medidas de confiança atribuídas pelo classificador e no eixo y a densidade probabilidade. O critério de decisão é aceitar um exemplo ou rejeitar um exemplo de acordo com limiar. A aceitação é apresentada a direita do limiar enquanto a rejeição é vista a esquerda do limiar. Diferentes valores de limiar podem produzir diferentes valores de taxa de falsa aceitação e taxa de falsa rejeição. Como ambas são alteradas de acordo com o limiar é claro o compromisso entre as taxas.

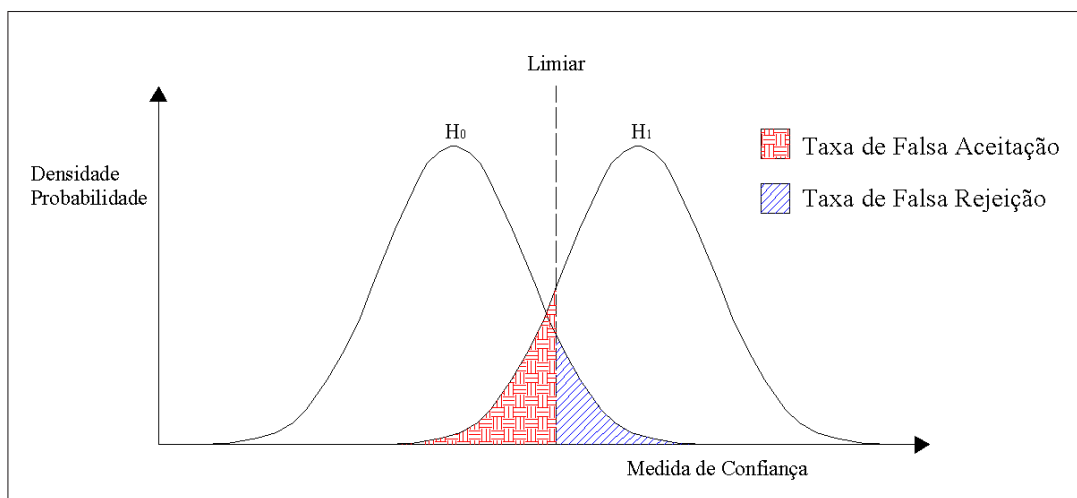


Figura 2.1: O teste de hipóteses levando em consideração a variação do limiar. Adaptada de (MASSAD et al., 2004) ilustrando distribuição de doentes e não doentes com um valor de limiar para corte.

Podemos encontrar dois tipos de erros envolvidos no procedimento de testes de hipóteses: rejeitar a hipótese nula sendo ela verdadeira, conhecido como Erro do Tipo I ou erro α , ou a aceitarmos sendo falsa, ocorrendo neste caso um Erro do Tipo II ou erro β . A Tabela 2.1 resume as ações possíveis ao testar hipóteses estatísticas, segundo os

princípios gerais que obedecem as regras desenvolvidas por Neyman e Pearson (MASSAD et al., 2004).

Tabela 2.1: Resumo de ações para teste de hipóteses

Conclusão do teste	H_0 Verdadeira	H_0 Falsa
Não rejeita H_0	Correto	Erro do Tipo II (ou erro β)
Rejeita H_0	Erro do Tipo I (ou erro α)	Incorreto

2.5 Análise ROC

Quando nos deparamos com a necessidade de avaliação e análise, precisamos escolher um método eficaz. A análise ROC (*Receiver Operating Characteristic*) é uma ferramenta poderosa para medir e especificar problemas no desempenho. A análise ROC teve a sua origem na teoria de detecção de sinais, para avaliar a qualidade de transmissão de sinal em um canal com ruído. Hoje é muito utilizada na área médica, para avaliação de diagnósticos e para analisar a qualidade de um teste clínico (exames), e a algum tempo vêm sendo adotada em técnicas de aprendizagem de máquina e mineração de dados como uma ferramenta para avaliação de modelos (FAWCETT, 2006). É particularmente útil em casos onde a desproporção de classes é grande e quando se faz necessário levar em conta diferentes considerações de custos para diferentes erros ou acertos de um sistema de classificação.

2.5.1 Definições para ROC

Para avaliar a eficiência das estratégias de rejeição usamos a terminologia convencional utilizada para reconhecimento de padrões que refere-se à falsa aceitação e falsa rejeição, tratando o correto e o incorreto reconhecimento, verificados através da implementação de rotinas de pós-processamento. Sendo estas taxas calculadas através de valores fundamentados na saída de uma análise de pós-processamento em relação a aceitação ou rejeição de cada exemplo da base de dados comparados, com os resultados fornecidos pelo classificador aplicado. A saída do pós-processamento pode ser definida de quatro formas:

- **CA:** *Correta Aceitação* - Instâncias corretamente reconhecida pelo classificador e após aplicação de uma estratégia de rejeição ela é novamente aceita como correta.
- **CR:** *Correta Rejeição* - Instâncias incorretamente reconhecida pelo classificador, e rejeitada pela estratégia de rejeição.

- **FA:** *Falsa Aceitação* - O classificador classifica incorretamente uma instância, mas a estratégia de rejeição aceita.
- **FR:** *Falsa Rejeição* - O classificador classifica corretamente uma instância, mas a estratégia de rejeição rejeita esta instância.

Tais estatísticas, são utilizadas para avaliação do desempenho de uma tarefa de pós-processamento ou simplesmente, das estratégias de rejeição. Podemos medir a precisão de tais rotinas observando (PITRELLI; PERRONE, 2002):

- **Taxa de Falsa Aceitação** - FAR, ou *Erro do tipo I*: Aceitar que um exemplo foi classificado corretamente quando na realidade foi classificado incorretamente. A frequência de ocorrências deste tipo é chamada de FAR calculada pela Equação 2.1:

$$FAR = \frac{FA}{FA + CR} \quad (2.1)$$

- **Taxa de Falsa Rejeição** - FRR, ou *Erro do tipo II*: Aceitar que um exemplo foi classificado incorretamente quando na realidade foi classificado corretamente. A frequência de ocorrências deste tipo é chamada de FRR calculada pela Equação 2.2:

$$FRR = \frac{FR}{FR + CA} \quad (2.2)$$

2.5.2 Curvas ROC

Geometricamente, a curva é um gráfico de pares “X” e “Y” em um plano onde as coordenadas representam medidas de probabilidade, e por esse motivo variam entre zero e um. Um modelo de classificação é representado por um ponto no espaço ROC. O ponto inferior esquerdo (0,0) representa a estratégia de sempre prever positivo. O ponto superior direito (1,1) representa a estratégia de sempre prever negativo. Os melhores desempenhos situam-se no canto inferior esquerdo do gráfico, isto é, representam menor Erro do Tipo I associado ao menor Erro do Tipo II.

A Figura 2.2 representa um gráfico ROC típico. Observando o gráfico podemos realizar uma análise visual do comportamento das curvas para um dado problema. Para a construção do gráfico e suas respectivas curvas plotamos valores de FAR no eixo das ordenadas (eixo X) e valores de FRR no eixo das abscissas (eixo Y).

Uma vantagem na utilização de curvas ROC está na avaliação ordenada de exemplos. Nesse contexto o sistema de aprendizagem, ou a estratégia de rejeição não prediz uma classe e sim um valor que pode ser contínuo ou ordinal. Assim, para a avaliação de

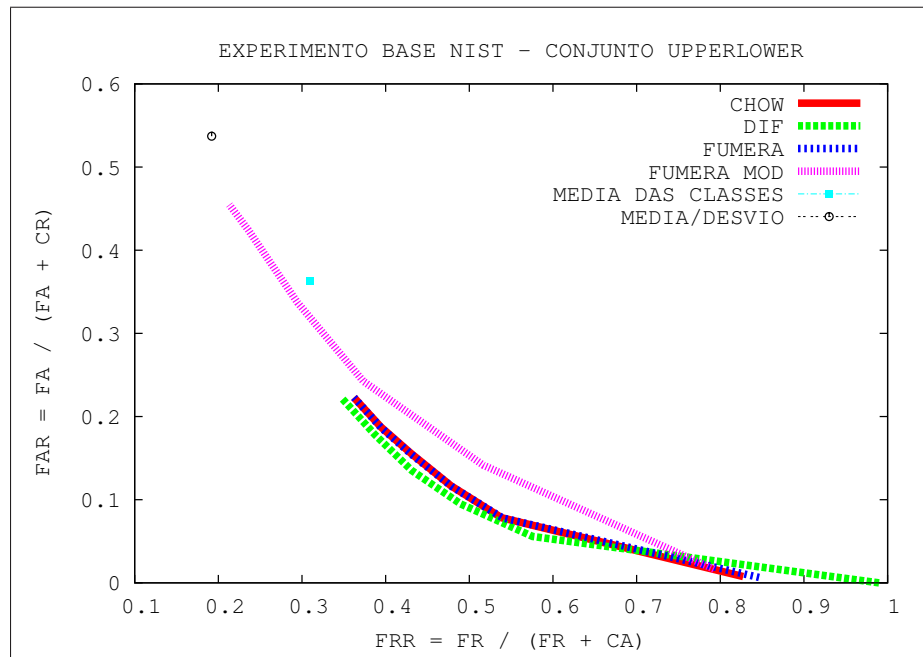


Figura 2.2: Exemplo de um gráfico ROC relacionando FAR e FRR para diferentes estratégias de rejeição.

um modelo podemos utilizar um limiar. Desta forma, ao invés de escolhermos um limiar arbitrário para representar o desempenho de um sistema com apenas um único ponto no espaço ROC, podemos simular o comportamento do sistema para vários limiares e o desempenho pode ser avaliado por uma curva no espaço ROC sendo independente da escolha de um determinado limiar observando o compromisso entre erros e acertos específicos de cada um dos limiares.

2.5.3 Interpretação do Gráfico ROC

Os gráficos ROC conseguem trabalhar muito bem com problemas de duas classes, mas uma de suas desvantagens é a dificuldade de trabalhar com problemas com mais classes. Apesar dos princípios básicos serem os mesmos, o número de curvas cresce exponencialmente ao número de classes dificultando a análise visual. No entanto, algumas soluções são viáveis. Um trabalho que discute esse problema envolvendo mais de duas classes foi proposto por Hand e Till (2001). A idéia geral é transformar um problema multiclases em duas classes comparando duas classes de cada vez. Observando na Figura 2.2 a comparação de várias curvas, caso não exista nenhuma intersecção ou sobreposição, ou existam mas mesmo assim é possível identificar através de uma análise visual, a curva que mais se aproxima do ponto (0,0) é a de melhor desempenho. Neste caso a estratégia que apresenta o melhor compromisso entre a falsa rejeição e falsa aceitação é a DIF e o

pior é apresentado por Média/Desvio.

2.6 Trabalhos Relacionados

O objetivo de um mecanismo de rejeição é minimizar o número de erros em um sistema de reconhecimento para um dado número de rejeições. Um erro ocorre quando um padrão é identificado em uma classe diferente a que ele realmente pertence. A rejeição ocorre quando um padrão é deixado de lado para uma posterior classificação. Em uma aplicação de classificação conforme aumentamos o número de exemplos rejeitados, igualmente, diminuimos o número de exemplos rotulados como erros. Conseqüentemente é evidente a existência de um compromisso entre os valores utilizados para descrever o desempenho de um sistema inteligente cujo objetivo principal seja classificação. Tal evidência (compromisso erro/rejeição) é definido pela estratégia de rejeição aplicada para um dado número de rejeições.

O trabalho de Chow (1970), busca otimizar o procedimento de rejeição para um classificador Bayesiano considerando também problemas com múltiplas classes. A idéia central desta estratégia é rejeitar uma instância que possua uma probabilidade global menor do que um determinado limiar.

No estudo de Fumera, Roli e Giacinto (2000) uma nova proposta de estratégia de rejeição é apresentada. Um classificador de N classes é utilizado para subdividir o espaço de características em regiões de decisão D_i , onde $i=1, \dots, N$, tais que os padrões x das classes ω_i pertençam à região D_i . De acordo com a teoria estatística de reconhecimento de padrões, tais regiões de decisão são definidas para maximizar a probabilidade de reconhecimento correto, chamada de precisão do classificador, conforme a Equação 2.3, onde $P(\omega_i)$ é uma probabilidade *a priori* e $p(x | \omega_i)$ é a probabilidade condicional:

$$precisão=P(correto) = \sum_{i=1}^N \int_{D_i} p(x | \omega_i) P(\omega_i) dx \quad (2.3)$$

E, conseqüentemente, para minimizar a probabilidade de erro do classificador, conforme a Equação 2.4:

$$P(erro) = \sum_{i=1}^N \int_{D_i} \sum_{j=1, j \neq i}^N p(x | \omega_j) P(\omega_j) dx \quad (2.4)$$

Para este fim, a então denominada regra de decisão de Bayes associa cada padrão x à classe cuja a probabilidade *a posteriori* $P(\omega_i|x)$ é máxima.

Utilizando uma estratégia de rejeição, os padrões que são mais propensos a serem

classificados erroneamente são rejeitados, ou seja, não classificados. Uma estratégia apresentando o compromisso entre erro e rejeição foi proposta por Chow (1970). De acordo com a regra de Chow um padrão é rejeitado se:

$$\max_{k=1, \dots, N} P(\omega_k | x) = P(\omega_i | x) < T \quad (2.5)$$

onde $T \in [0,1]$. Por outro lado, o padrão x é aceito e associado a classe ω_i se:

$$\max_{k=1, \dots, N} P(\omega_k | x) = P(\omega_i | x) \geq T \quad (2.6)$$

O espaço de características é subdividido em $N + 1$ regiões. A região de rejeição D_0 é definida de acordo com a Equação 2.5, enquanto as regiões de decisão $D_1 \dots D_n$ são definidas de acordo com a Equação 2.6. Então a probabilidade de um padrão ser rejeitado pode ser computada conforme a Equação 2.7:

$$P(\text{rejeição}) = \int_{D_0} P(x) dx \quad (2.7)$$

A função densidade de probabilidade é representada por $P(x)$. Em contraste, a precisão do classificador é definida como a probabilidade condicional que um padrão classificado corretamente, dado que ele tenha sido aceito conforme a Equação 2.8:

$$\text{precisão} = P(\text{correto} | \text{aceito}) = \frac{P(\text{correto})}{P(\text{correto}) + P(\text{erro})} \quad (2.8)$$

De acordo com Fumera, Roli e Giacinto (2000) uma análise do trabalho de Chow (1970) permite apontar que a regra Chow provê uma fronteira ótima de erro-rejeição, somente se as probabilidades *a posteriori* são exatamente conhecidas. No mesmo trabalho eles propõem a estratégia CRT (*Class-Related Thresholds*) que utiliza limiares particulares para cada classe do problema, permitindo obter regiões de decisão conforme a Figura 2.3. O uso de múltiplos limiares de rejeição é indicado para problemas multiclases visando obter as regiões ótimas de decisão e rejeição, mesmo se as probabilidades *a posteriori* forem afetadas por erros.

Quanto à suposição de que as probabilidades *a posteriori* podem ser afetadas por erros, os autores mostram em seus experimentos que, para qualquer taxa de rejeição R , existem valores dos CRTs $T_0 \dots T_n$ que correspondem a precisão de um classificador $A_{T_0 \dots T_n}$ ser igual ou superior à precisão $A(T)$ provida pela regra de Chow, dada pela Equação 2.9:

$$\forall R \exists T_1, T_2 \dots T_n : A(T_1, T_2 \dots T_n) \geq A(T) \quad (2.9)$$

Portanto, utilizando o mecanismo CRT para uma tarefa de classificação com N classes de dados caracterizadas por probabilidades *a posteriori* estimadas $\hat{P}(\omega_i | x)$, $i = 1, \dots, N$ um padrão x é rejeitado se:

$$\max_{k=1, \dots, N} \hat{P}(\omega_k | x) = \hat{P}(\omega_i | x) < T_i \quad (2.10)$$

Enquanto um padrão x é aceito e associado à classe ω_i , se:

$$\max_{k=1, \dots, N} \hat{P}(\omega_k | x) = \hat{P}(\omega_i | x) \geq T_i \quad (2.11)$$

Para exemplificar, consideremos uma tarefa simples de classificação unidimensional com duas classes ω_1 e ω_2 caracterizadas por distribuições Gaussianas, como mostradas na Figura 2.3. Os termos $P(\omega_i | x)$ e $\hat{P}(\omega_i | x)$, $i = 1, 2$ indicam as probabilidades *a posteriori* “verdadeiras” e “estimadas”, respectivamente. Considerando a hipótese de que erros significantes afetam as probabilidades estimadas nas variações dos valores das características nas quais duas classes estão “sobrepostas”, as regiões ótimas de decisão e rejeição providas pela regra de Chow aplicada às probabilidades “verdadeiras” são indicadas pelos termos D_1 , D_2 e D_0 respectivamente. O termo T indica um limiar de rejeição de Chow. Analogamente, os termos \hat{D}_1 , \hat{D}_2 e \hat{D}_0 indicam regiões de decisão e rejeição providas pela regra de Chow aplicada para probabilidades estimadas. A Figura 2.3 mostra que as regiões estimadas diferem das ótimas nos intervalos $(\hat{D}_1 - D_1)$ e $(\hat{D}_2 - \hat{D}_2)$.

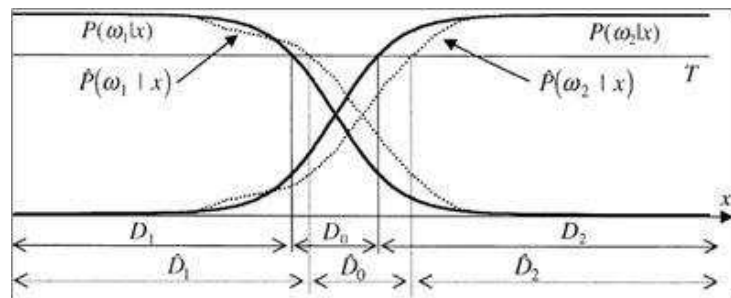


Figura 2.3: Aplicação da regra de Chow para as probabilidades *a posteriori* “verdadeiras” e “estimadas” (FUMERA; ROLI; GIACINTO, 2000).

Podemos constatar que a regra de Chow aceita os padrões pertencentes ao intervalo $(\hat{D}_1 - D_1)$, visto que a probabilidade *a posteriori* $\hat{P}(\omega_i | x)$ contém valores superiores aos de “T” neste intervalo. Sendo que o correto seria que tais valores fossem rejeitados utilizando um valor de limiar $T_1 \geq T$. Da mesma forma, os padrões pertencentes a $(D_2 - \hat{D}_2)$ são erroneamente rejeitados, pois a probabilidade *a posteriori* $\hat{P}(\omega_2 | x)$ contém valores inferiores aos “verdadeiros” dentro deste intervalo. Tais padrões deveriam ser corretamente

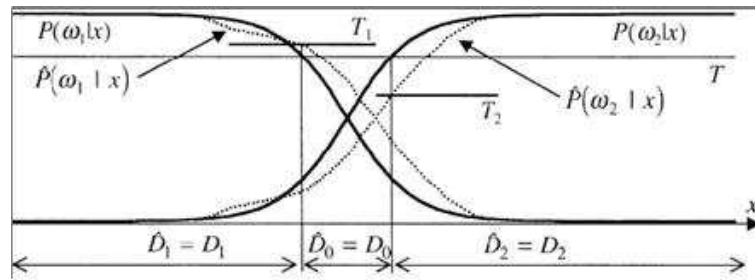


Figura 2.4: Utilização de dois limiares de rejeição diferentes T_1 e T_2 para a tarefa de classificação da Figura 2.3 (FUMERA; ROLI; GIACINTO, 2000).

aceitos utilizando um valor de limiar $T_2 \geq T$. Desta forma validando a idéia de utilizar um limiar para cada classe do problema. A Figura 2.4 mostra o uso de dois limiares de rejeição diferentes T_1 e T_2 aplicadas para as probabilidades *a posteriori* estimadas da tarefa de classificação da Figura 2.3.

Uma alteração da regra de Chow foi proposta por Ha (1996), apresentado uma nova regra de decisão. No trabalho de Chow, um padrão é rejeitado se a maior probabilidade *a posteriori* é menor que um limiar, desconsiderando a distribuição de probabilidade das demais classes. Na nova regra, os padrões não são rejeitados de todas as classes, mas, somente daquelas mais improváveis que este possa pertencer. Ao invés de simplesmente proceder com o mecanismo de rejeição, este é ignorado temporariamente, por exemplo, quando ocorrer uma saída de um padrão que não seja confiável associar a uma classe ou a nenhuma das múltiplas classes este primeiramente utiliza a *class-selective*. Em outras palavras, o espaço é dividido, apresentado padrões correspondentes a um subconjunto de classes. Desde que haja subconjuntos em um conjunto de N elementos, obtemos $2N - 1$ regiões, em um problema de N classes. O ponto chave da estratégia é a escolha do subconjunto a um dado padrão x que unicamente é especificada pela probabilidade *a priori* sendo representadas por $P_i(x) \ i=1....N$.

O trabalho de Mouchere e Anquetil (2006) descreve detalhadamente os tipos comuns de aplicação de mecanismos de rejeição para projetar um sistema de aprendizagem genérico e automático. Enfatiza que a maioria de trabalhos que aplicam mecanismos de rejeição não focalizam o tipo de rejeição, mas, otimização do compromisso erro-rejeição, como exemplo cita dois trabalhos descritos anteriormente (CHOW, 1970) e (FUMERA; ROLI; GIACINTO, 2000). Os dois tipos de rejeição mais comumente utilizados são: a rejeição através da observação da confusão e a rejeição através da observação da distância. O objetivo da rejeição através da observação da confusão é melhorar a precisão do reconhecimento quando a taxa de classificação incorreta é alta. Estes erros são próximos dos limites de decisão porque os valores relativos as classes são próximos. Então é necessário

definir uma zona de rejeição em cada lateral dos limites de decisão. Se um exemplo está dentro de uma destas zonas e a média de confiança é baixa, esse exemplo deve ser rejeitado. Estas zonas de rejeição são definidas comparando os valores das duas melhores classes. Caso sejam muito próximas o exemplo é rejeitado. Como usamos contagem de exemplos por classes este tipo de rejeição é possível para a maioria dos tipos de classificador apenas observando a matriz de confusão. Durante a classificação de um conjunto de dados completo N_{tot} ele pode dividir-se em três tipos de dados: N_{corr} no caso de exemplos corretamente classificados, N_{err} quando classificados incorretamente, e N_{rej} para os exemplos rejeitados. Assim uma rigorosa rejeição proporciona baixo desempenho e a precisão tende a aumentar caso os exemplos rejeitados sejam erros, então, ambos os valores devem ser maximizados. As Equações 2.12 e 2.13 definem a maneira de calcular o desempenho e a precisão:

$$desempenho = \frac{N_{corr}}{N_{tot}} \quad (2.12)$$

$$precisão = \frac{N_{corr}}{N_{corr} + N_{err}} \quad (2.13)$$

O mecanismo que leva em conta a distância permite delimitar o conhecimento do classificador usado e rejeitar exemplos que não pertençam a classes aprendidas. Conseqüentemente, se um exemplo é muito distante deve ser rejeitado. Assim pode ser usada para a descoberta de *outlier* e de melhores estratégias de rejeição. Para avaliar rejeição distância é utilizado o compromisso entre falsa rejeição e falsa aceitação. O objetivo aqui é minimizar ambas.

Para os dois mecanismos de rejeição apresentados utiliza-se a notação de confiança para formalizar a rejeição. A rejeição observando a matriz de confusão permite aumentar a precisão da informação obtida. Considerando o outro modelo obtemos informação intrínseca que permite delimitar o conhecimento do sistema de reconhecimento aumentando a robustez para a falsa rejeição. Os autores definem um formalismo que permite abstrair o mecanismo de rejeição e o classificador utilizados através de um algoritmo que aprende todo o mecanismo de rejeição automaticamente, porém ambos se apresentam de forma bem genérica para permitir novas estratégias de rejeição e aproximações de aprendizagem.

O interesse em otimizar o processo de classificação é mostrado muitas vezes em diversas aplicações e de diferentes formas. Algumas dessas formas são aqui destacadas e consideradas relevantes fontes de informação durante a realização deste trabalho. Primeiramente relatamos a proposta de construir um sistema de reconhecimento de padrões, com múltiplos estágios e opção rejeição apresentada por Pudil et al. (1992), onde os exemplos

rejeitados pelo primeiro estágio podem ser abordados por um segundo, utilizando novos critérios. Assim, os exemplos rejeitados por este estágio são abordados pelo estágio seguinte de forma que nenhum exemplo rejeitado permaneça aguardando nova classificação. Este método é interessante para aplicações de reconhecimento de padrões nas quais, apenas a apresentação dos parâmetros rejeitados não é aceita como um resultado final. Posteriormente, Gorski (1997) apresenta um método buscando encontrar um compromisso de erro-rejeição satisfatório. Uma estratégia de rejeição medindo a confiança na saída de uma rede neural pode, através de um *score*, aceitar ou rejeitar os exemplos e atribuí-los a uma lista de bons e maus candidatos. A implementação do método em um sistema de reconhecimento de cheques bancários se apresentou flexível e de fácil entendimento.

Os estudos de Marukatat et al. (2002) são direcionados à medidas de confiança para um sistema de reconhecimento em uma base de dados formada por caracteres *on-line* (base de dados UNIPEN). A decisão de aceitar ou rejeitar um exemplo para as quatro estratégias implementadas é realizada comparando o valor de confiança com um limiar e as regras de decisão aplicadas a entrada de uma sequência de observações $\mathcal{O}_T^1 = (o_1, o_2, \dots, o_T)$ reconhecidas na saída pela hipótese de uma palavra W . A decisão de rejeitar ou aceitar consiste: *medida de confiança* $\mathcal{O}_T^1 W < threshold$, a saída do sistema de reconhecimento é rejeitada. Caso a *medida de confiança* $\mathcal{O}_T^1 W \geq threshold$, a saída do sistema de reconhecimento é aceita. Uma estratégia de rejeição é projetada para rejeitar os exemplos que possivelmente fossem incorretamente classificados. Propõe a comparação de diferentes medidas de confiança obtendo precisão de 80% a 95% rejeitando 30% dos exemplos. O mecanismo de rejeição é implementado como uma tarefa de pós-processamento e as estatísticas de desempenho baseadas na matriz de confusão.

O trabalho de Pitrelli e Perrone (2002) compara várias opções de *scores* em oito estratégias de rejeição diferentes utilizando bases de dados formadas por um grande vocabulário de palavras *on-line* e posteriormente dígitos isolados. A análise ROC é utilizada para avaliar o desempenho das estratégias de rejeição. Algumas estratégias apresentam excelentes resultados sendo capazes de rejeitar 90% das palavras que seriam classificadas incorretamente, enquanto rejeitaria apenas 33% das palavras que seriam classificadas corretamente. Os resultados para as bases de dígitos isolados também apresentam resultados interessantes proporcionando uma correta aceitação de 90% dos exemplos e rejeição abaixo de 13%.

Em trabalhos similares, Koerich (2004), Zimmermann, Bertolami e Bunke (2004) investigam diferentes estratégias de rejeição aplicadas a sistemas de reconhecimento *off-line* de palavras implementadas em uma tarefa de pós-processamento através de *Hidden Markov Model*, ambos diminuem a taxa de erro e melhoram a confiança no processo de

classificação. O primeiro consegue reduzir a taxa de erro em até 10% para vocabulários maiores que 40.000 palavras (imagens de palavras), rejeitando 20% dos exemplos. Os resultados mostram que a melhor das três estratégias implementadas consegue fornecer 78% a 94% de taxa de reconhecimento, enquanto rejeita 30%. O segundo, em termos de estatísticas de erro e rejeição, obtém para 8825 palavras (imagens) no conjunto de treinamento, 0% de rejeição e uma taxa de erro de 19.9%. Para alcançar 5% de taxa de erro, precisa-se rejeitar 29% dos exemplos, para obter uma taxa de erro abaixo de 2% é necessário rejeitar 49% das palavras.

Finalmente, o trabalho Provost e Fawcett (1997) questiona a validade da utilização da precisão para avaliação de algoritmos de aprendizagem quando os exemplos nas classes são desbalanceados e por considerar os diferentes erros de classificação como igualmente importantes, pois em aplicações reais o resultado da classificação sugere ações que podem vir a manifestar graves conseqüências. Assim, um método combinando técnicas de análise ROC é empregado para verificação de tais problemas. O trabalho também enfatiza que a análise ROC adapta-se as particularidades da avaliação da aprendizagem dos classificadores, sendo incremental e minimizando a manipulação de dados para avaliar o desempenho do classificador; conseguindo inserir facilmente novos classificadores e possuindo clareza visual para a comparação de desempenho.

2.7 Análise Crítica

Concluindo, a contextualização de alguns temas chaves e os trabalhos citados neste capítulo, contribuem na elaboração deste trabalho, e principalmente ajudam a entender a complexidade do projeto de uma estratégia de rejeição e a avaliação de um problema de múltiplas classes. Observamos que os resultados apresentados por estes trabalhos são relativos a problemas específicos, a maioria deles voltados ao comportamento das estratégias de rejeição em bases de dados de palavras, *on-line* e *of-line*, considerando na maioria das vezes métodos estatísticos e não considerando que parâmetros ou características do problema podem impactar nos resultados apresentados pelas estratégias de rejeição. No capítulo seguinte, apresentamos detalhadamente a metodologia proposta para atingir os objetivos anteriormente descritos.

Capítulo 3

Metodologia Proposta

Neste capítulo é descrita a metodologia utilizada no desenvolvimento deste trabalho. A Figura 3.1 apresenta uma visão geral dos procedimentos a serem realizados e a seguir cada uma das etapas envolvidas é detalhada.

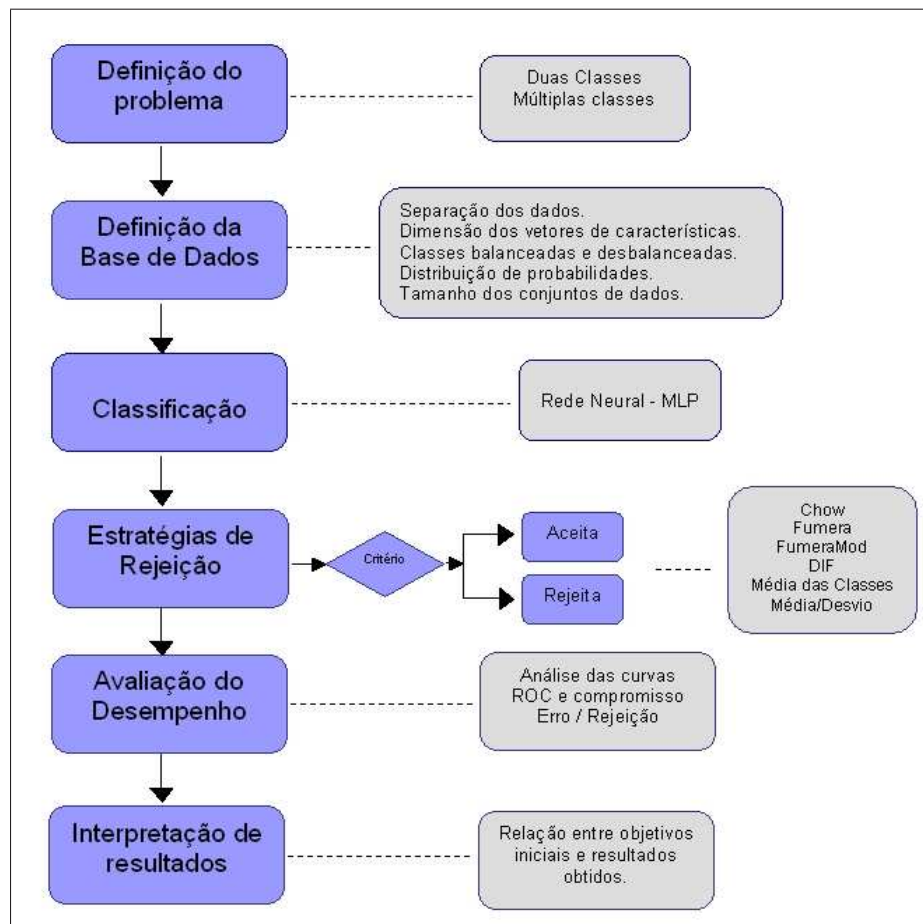


Figura 3.1: Diagrama de procedimentos aplicados à avaliação das estratégias de rejeição.

3.1 Definição do Problema

A primeira tarefa é a definição dos problemas nos quais as estratégias de rejeição serão avaliadas. Nesta fase, é essencial um estudo do problema abordado, visando conhecer o impacto que uma classificação incorreta pode ocasionar. Inicialmente serão considerados problemas de duas classes, construídos através de dados sintéticos, cujas distribuições são conhecidas e os parâmetros controlados. Posteriormente serão considerados problemas de múltiplas classes com dados reais e distribuições desconhecidas. As seções a seguir apresentam detalhes destas escolhas.

3.2 Definição da Base de Dados

A tarefa seguinte é a escolha da base de dados. Trata-se de uma escolha de extrema importância, pois desejamos avaliar se os parâmetros escolhidos podem influenciar os resultados obtidos pelas diferentes estratégias de rejeição. Os parâmetros definidos para escolhas das bases são:

- **Quantidade de classes no problema:** Tarefa simples de classificação com duas classes, onde um exemplo pertence a uma ou outra classe. Em seguida tarefas de classificação mais complexas, com múltiplas classes, onde uma determinada instância pode pertencer a qualquer uma das classes.
- **Tipos de dados estudados:** Dados sintéticos gerados através de simulação computacional envolvendo a geração de variáveis aleatórias com distribuições pré-definidas para problema de duas classes. Posteriormente dados reais, utilizando as bases NIST caracteres manuscritos maiúsculos e minúsculos.
- **Separação dos dados:** Conhecer a separação dos dados na base de dados sintéticas: dados sobrepostos, levemente sobrepostos ou separados para avaliar o comportamento das estratégias de rejeição.
- **Definição das distribuições de probabilidades dos dados:** Neste caso serão avaliadas três distribuições, podendo ser controladas na geração dos dados sintéticos: Distribuição Normal, *Chi-square* e Normal Multivariável. Para os experimentos em bases reais essa informação é desconhecida.
- **Tamanho da base de dados utilizada:** As bases de dados sintéticos para problemas de duas classes são implementadas através de combinações de 20.000 exemplos de forma balanceada controlada, com 10.000 exemplos em w_1 e 10.000 exemplos em

w_2 e desbalanceada com 16.000 em w_1 e 4.000 em w_2 . Em bases reais a quantidade de exemplos são significativamente maior. O método utilizado é o “*holdout validation*” com 70% da base para treinamento e validação e 30% para testes.

- **Dimensão dos vetores de características para o problema:** Inicialmente duas classes com duas características sendo representadas por vetores com valores controlados em relação a média e variância ou graus de liberdade, com dois atributos. Para problemas reais cada exemplo da base de dados é representado por um vetor com 108 atributos. Estes atributos representam características estruturais dos caracteres, como por exemplo, perfil horizontal e vertical, projeção inferior, superior, laterais e histograma direcional do contorno em 6 direções, estando distribuídas da seguinte forma: 20 características para o histograma de projeção; 40 características para o perfil de projeção; 48 características para a direção do contorno, (KOERICH, 2003).

3.3 Classificação

A terceira tarefa concentra-se na aplicação de um classificador nas bases de dados. A classificação é o processo realizado quando uma instância necessita ser atribuído a um determinado grupo ou classe baseada em um número de atributos relativos à instância em questão. O classificador escolhido é uma RNA-MLP com características específicas para cada tipo de problema. Essa escolha dever-se ao fato de que elas conseguem prover na saída probabilidades *a posteriori* o que possibilita a aplicação de uma estratégia de rejeição como atividade de pós-processamento.

Uma habilidade importante em relação a RNA é que elas são capazes de melhorar seu desempenho através do treinamento em um processo iterativo de ajustes aplicados a seus pesos. O processo de aprendizagem pode implicar na seguinte sequência de eventos:

1. A RNA é estimulada por um ambiente (conjunto de treinamento);
2. A RNA sofre modificações em nos seus parâmetros (pesos sinápticos e bias) como resultado desta estimulação;
3. A RNA responde de uma maneira nova ao ambiente, devido às modificações ocorridas na sua estrutura interna, caso ela seja submetida sempre ao mesmo conjunto dizemos que ela vai “decorrar” as amostras deste conjunto. A cada ciclo de treinamento, o erro médio quadrático, é avaliado, pois, durante o treinamento busca-se minimizar este erro.

4. Para todos os conjuntos, criamos uma subdivisão do conjunto de treinamento, estabelecendo um conjunto de validação, utilizado para verificar a eficiência da rede quanto a sua capacidade de generalização durante o treinamento, podendo ser empregado como critério de parada do treinamento.
5. Finalmente, a RNA é aplicada a um conjunto de dados desconhecidos (conjunto de teste) para verificar seu real desempenho.

3.3.1 Classificação para duas classes

Para a realização dos experimentos com dados sintéticos utilizamos o MATLAB, um software de alto desempenho para computação técnica que pode ser aplicado em diversas áreas, incluindo o reconhecimento de padrões. Este software conta com diversas *toolboxes* específicas para vários tipos de aplicações. Neste trabalho utilizamos a *toolbox* PRTools desenvolvida pelo grupo de reconhecimento de padrões da universidade de Delft na Holanda (DUIN et al., 2004). A versão atual (4.0) PRTools implementa aproximadamente 200 rotinas abrangendo uma grande parte da área de reconhecimento estatísticos de padrões. Para este tipo de problema utilizamos uma RNA com a seguinte arquitetura: 2 neurônios na camada de entrada (características controladas) uma única camada escondida com 3 neurônios, 2 neurônios na camada de saída (quantidade de classes). O algoritmo utilizado para o treinamento desta rede foi o *backpropagation*. Os experimentos para investigar o comportamento das estratégias de rejeição em dados com parâmetros conhecidos foram organizados em 18 conjuntos diferentes com balanceamento e desbalanceamento de exemplos entre as classes e dados separados, parcialmente separados ou sobrepostos. Esses conjuntos são separados de acordo com a distribuição de probabilidades e denominados com a seguinte nomenclatura:

- **Distribuição Normal:** Para esta distribuição foram criados seis diferentes conjuntos:
 1. Distribuição Normal, Balanceada e Separada (NBS)
 2. Distribuição Normal, Balanceada e Parcialmente Sobreposta (NBPS)
 3. Distribuição Normal, Balanceada e Sobreposta (NBSob)
 4. Distribuição Normal, Desbalanceada e Separada (NDS)
 5. Distribuição Normal, Desbalanceada e Parcialmente Sobreposta (NDPS)
 6. Distribuição Normal, Desbalanceada e Sobreposta (NDSob)

- **Distribuição Normal Multivariável:** Esta distribuição considera também a variação dos parâmetros anteriores:
 1. Distribuição Normal Multivariável, Balanceada e Separada (NMBS)
 2. Distribuição Normal Multivariável, Balanceada e Parcialmente Sobreposta (NMBPS)
 3. Distribuição Normal Multivariável, Balanceada e Sobreposta (NMBSob)
 4. Distribuição Normal Multivariável, Desbalanceada e Separada (NMDS)
 5. Distribuição Normal Multivariável, Desbalanceada e Parcialmente Sobreposta (NMDPS)
 6. Distribuição Normal Multivariável, Desbalanceada e Sobreposta (NMDSob)

- **Distribuição *Chi-square*:** Segue também os parâmetros das duas anteriores:
 1. Distribuição *Chi-square*, Balanceada e Separada (CBS)
 2. Distribuição *Chi-square*, Balanceada e Parcialmente Sobreposta (CBPS)
 3. Distribuição *Chi-square*, Balanceada e Sobreposta (CBSob)
 4. Distribuição *Chi-square*, Desbalanceada e Separada (CDS)
 5. Distribuição *Chi-square*, Desbalanceada e Parcialmente Sobreposta (CDPS)
 6. Distribuição *Chi-square*, Desbalanceada e Sobreposta(CDSob)

O Apêndice A apresenta detalhes de como os experimentos foram realizados ilustrando através de figuras a distribuição dos dados para classes balanceadas e desbalanceadas a separação dos dados em cada distribuição de probabilidade; os valores utilizados para gerar os conjuntos de dados; e a separação das classes efetuada pelo classificador no respectivo conjunto de teste.

3.3.2 Classificação para múltiplas classes

A base de dados utilizada para o reconhecimento de caracteres foi a NIST (*National Institute of Standards and Technology*) *NCharacter - SD19* contendo imagens no formato TIFF das 52 letras do alfabeto sendo cada letra uma classe. Foram considerados três conjuntos de dados:

- **Conjunto Upper** - Utilizamos 26 classes diferentes (“A - Z”) de caracteres maiúsculos totalizando 61.473 exemplos, divididos em três conjuntos: treinamento (37.440), validação (12.092) e teste (11.941). Os resultados do classificador MLP sem aplicação

de mecanismos de rejeição, ou seja, taxa de rejeição 0% apresentam taxa de reconhecimento de 97,87% para o conjunto treinamento, 93,60% para o conjunto de validação e 92,49 % para o conjunto de teste.

- **Conjunto Lower** - Utilizamos 26 classes diferentes de caracteres minúsculos (“a - z”) com 61.018 exemplos, divididos em treinamento (37.440), validação (11.578) e teste (12.000). Os resultados do classificador apresentam taxa de reconhecimento de 95,82% para o conjunto treinamento, 90,05% para o conjunto de validação e 86,73% para o conjunto de teste.
- **Conjunto UpperLower** - A combinação das bases anteriores em 52 classes, caracteres maiúsculos e minúsculos, onde “A” e “a” pertencem a mesma base e são classes distintas. Os resultados do classificador MLP apresentam taxa de reconhecimento de 80,93% para o conjunto treinamento, 73,59% para o conjunto de validação e 69,08 % para o conjunto de teste.








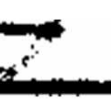
Letra A	Letra L	Letra O	Letra Z
			
			

Figura 3.2: Diversidade de exemplos na base de dados NIST.

O classificador escolhido é uma rede neural do tipo MLP utilizada por Koerich (2003). A seguinte arquitetura foi definida para esse problema: uma única camada oculta com 100 neurônios, 108 unidades de entrada (características extraídas das imagens) e 26 unidades de saída (quantidade de classes) sendo esta arquitetura adotada para os conjuntos Upper e Lower. A arquitetura adotada para o Conjunto UpperLower é composta por uma única camada oculta com 150 neurônios, com 108 unidade de entrada (características extraídas da imagens) e 52 unidades de saída (quantidade de classes). A Figura 3.2 apresenta exemplos da base de dados NIST Upper. Na linha superior exemplos de caracteres fáceis de identificar, na linha inferior exemplos mais difíceis dos mesmos caracteres. Podemos perceber em um breve exemplo tal dificuldade, considerando os estilos maiúsculo e minúsculo de um conjunto de caracteres: quando escrevemos a letra “V” ou “v” ambos

os estilos são similares, mas, quando escrevemos a letra “D” ou “d” não são nada similares e ainda “D” é muito similar a “O”, sendo caracteres pertencentes a diferentes classes, mas com formas e características muito semelhantes.

3.4 Estratégias de Rejeição

Para formalizar a rejeição usamos a notação de confiança associada a uma hipótese. Consideramos um problema de classificação atribuído a um classificador C que fornece na saída uma medida de confiança c_i para cada uma das “ c ” classes pertencentes ao problema onde, $\sum_{i=1}^c c_i = 1$. Quando um exemplo é apresentado ao classificador, este deve atribuir uma medida de confiança a cada uma das c classes as quais este exemplo possa pertencer. No caso do classificador utilizado ser uma RNA, este valor refere-se a probabilidade atribuída pelo classificador sempre representado por valores reais entre 0 e 1. Sendo que quanto mais próximo de 1 maior é o grau de confiança do exemplo pertencer a classe indicada. Idealmente espera-se que para uma dada instância i teremos, $c_j = 1$ e $\sum_{i=1}^{c-1} c_i = 0$ para todo o $i \neq j$.

Posteriormente, conhecidas as medidas de confiança $(c_1, c_2 \dots c_c)$ fornecidas pelo classificador, uma estratégia de rejeição pode ser aplicada. Com base nestas medidas, uma determinada estratégia de rejeição pode decidir se concorda com o classificador sobre a classificação de um determinado exemplo, ou se a rejeita e deixa tal exemplo para um procedimento mais sofisticado, uma classificação manual ou a utilização de outro classificador mais especializado, por exemplo. A rejeição ocorre quando um padrão ambíguo, propenso a ser incorretamente classificado, é deixado de lado para uma posterior classificação. A decisão de aceitar ou rejeitar um exemplo é controlada por um limiar t . Quando um exemplo possui medidas de confiança $\hat{c} \geq t$ onde $\hat{c} = \max c_j, 1 \leq j \leq c$ ele é aceito, enquanto exemplos com $\hat{c}_j < t$ são rejeitados. O objetivo de aplicar uma estratégia de rejeição em um dado problema é encontrar um valor ótimo para T . Quando este valor é encontrado uma estratégia de rejeição consegue rejeitar todos os exemplos incorretamente classificados pelo classificador e também aceitar todos os exemplos corretamente classificados.

3.4.1 Metodologia aplicada aos experimentos

A metodologia detalhada a seguir é aplicada tanto aos conjuntos sintéticos com duas classes quanto aos conjuntos com múltiplas classes. A Figura 3.4 apresenta a seqüência de passos para obter os limiares no conjunto de validação para as estratégias Chow,

Fumera, FumeraMod e DIF. O processo é iniciado com a divisão das bases de dados utilizando o método “*holdout validation*”, conforme as seções 3.2. O classificador é aplicado observando as arquiteturas descritas na seção 3.3.2. Depois que o classificador esteja devidamente ajustado podemos preparar os dados de forma normalizada para a aplicação da estratégia de rejeição como atividade de pós-processamento. Essa normalização é composta pelos níveis de confiança ($c_1 \dots c_c$) atribuída pelo classificador de um exemplo pertencer a cada uma das classes. Também apresenta R , sendo o rótulo real associado a classe, L sendo, o rótulo indicado pelo classificador, conforme a Figura 3.3. O processo de geração dos limiares para as estratégias Chow, Fumera, FumeraMod e DIF ocorre da mesma forma. Inicialmente, é atribuído ao limiar maior nível de confiança (0.999999), iterações vão sendo realizadas e um valor pré-definido (0.000001) é decrementado a cada iteração até chegar a zero mapeando o comportamento de diferentes limiares no conjunto de dados. Para todas as estratégias são resgatados os limiares verificando as taxas de erro proporcionada. Utilizamos os limiares conseguimos prover as taxas de erro de 0%, 1%, 2%, 3%, 4% e 5% para cada uma das estratégias. A escolha destes valores é fixada de acordo com o critério de erro máximo que o sistema de reconhecimento pode obter. Em um primeiro momento nos parece estranho buscar valores baixos de taxa de erro em conjuntos de dados que possuem taxas de erro elevadas no classificador de base, porém, em problemas reais de uma forma geral buscamos baixas taxas de erro e para diversas aplicações uma taxa de erro igual ou inferior a 5% é aceitável.

	Classe 0	Classe 1	Rótulos	
<i>Exemplo 1</i>	0.686295	0.313705	0	0
<i>Exemplo 2</i>	0.825520	0.174480	0	0
<i>Exemplo 3</i>	0.174096	0.825904	0	1
<i>Exemplo 4</i>	0.669631	0.330369	0	0
<i>Exemplo 5</i>	0.289311	0.710689	0	1
<i>Exemplo 6</i>	0.851577	0.148423	0	0
<i>Exemplo 7</i>	0.117586	0.882414	0	1
<i>Exemplo 8</i>	0.777224	0.222776	0	0
<i>Exemplo 9</i>	0.748834	0.251166	0	0
	Nível de confiança Classe 0	Nível de confiança Classe 1	Rótulo Real	Rótulo Classificador

$D[1..N, 1..C], R[1..N, 1..2]$

Figura 3.3: Detalhamento da saída da RNA para um conjunto qualquer com duas classes e nove instâncias. Apresentando a probabilidade *a posteriori* atribuída para cada uma das classes (níveis de confiança). Representação normalizadas expressas por D e R.

Para as estratégias Média das Classes e Média/Desvio a geração dos limiares é

obtida de forma automática. Para a primeira observamos o conjunto de validação e agrupamos para cada classe os exemplos corretamente classificados. Depois, calculamos a média de um exemplo ser corretamente classificado para cada classe do problema. O resultado obtido para cada classe é o valor utilizado como limiar para implementar o mecanismo de rejeição. Para a segunda, além de verificarmos a média buscamos também o desvio padrão de um exemplo ser corretamente classificado, para cada classe do problema. O limiar é obtido subtraindo o valor da média do valor do desvio padrão. A Figura 3.5 apresenta o processo de obtenção dos limiares para estas estratégias.

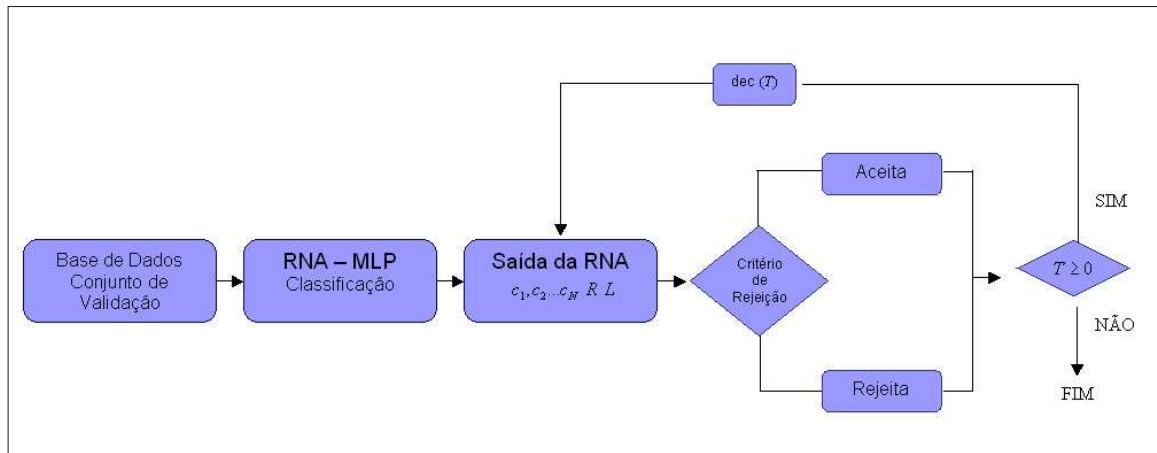


Figura 3.4: Obtendo os limiares através do conjunto de validação.

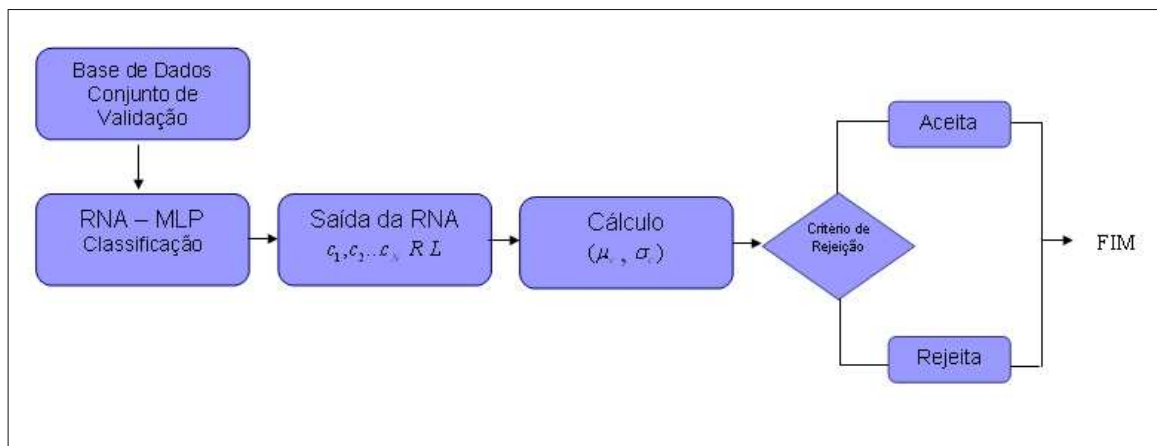


Figura 3.5: Obtendo os limiares de forma automática através do conjunto de validação.

A Figura 3.6 apresenta o procedimento realizado com os limiares obtidos no conjunto de validação. Esses limiares agora são testados sobre outro conjunto com dados completamente desconhecidos pelo classificador, neste caso utilizaremos o conjunto de testes. A saída do classificador para o conjunto de teste é normalizada e os seis limiares resgatados anteriormente são aplicados aos critérios de rejeição específicos de cada

estratégia. Os resultados obtidos sobre este conjunto provêm as taxas de erro, taxa de rejeição, taxa de falsa aceitação e taxa de falsa de rejeição utilizadas para avaliar o desempenho e o comportamento de cada uma das estratégias.

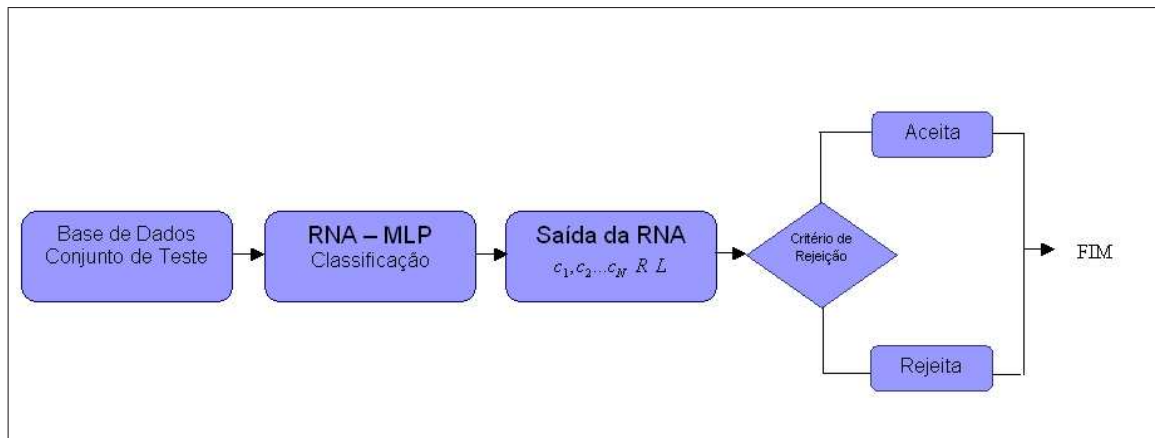


Figura 3.6: Testando os limiares através do conjunto de testes.

Neste trabalho adotamos métodos de rejeição estatísticos e também baseados em heurísticas. Os métodos estatísticos utilizados são descritos na literatura para encontrar o melhor compromisso erro/rejeição representados pelas estratégias Chow e Fumera. Dois novos métodos heurísticos (Médias das Classes e Média/Desvio) e um estatístico (FumeraMod) implementados afim de serem comparados com dois estatísticos já citados. Comparamos ainda o método heurístico DIF também descrito na literatura.

3.4.2 Rejeição com um limiar(Chow)

Este mecanismo é proporcionado pela regra de Chow (CHOW, 1970) em que apenas um limiar T é utilizado para todas classes do problema. Neste trabalho, T é encontrado através da aplicação da metodologia apresentada na Figura 3.4. De acordo com a regra de Chow uma instância é rejeitada caso obedeça a Equação 2.5, por outro lado, uma instância x é aceita e associada classe caso obedeça a Equação 2.6, descritas no segundo capítulo.

3.4.3 Rejeição com múltiplos limiares (Fumera)

A reprodução desta estratégia é baseada em (FUMERA; ROLI; GIACINTO, 2000) descrito na Seção 2.2.1 em que são utilizados múltiplos limiares, um para cada classe do problema. A escolha dos limiares utilizados nesta estratégia também é realizada no conjunto de validação, conforme a Figura 3.4. Para esta estratégia uma instância x é

rejeitada caso obedeça a Equação 2.10. Enquanto, uma instância x é aceita e associada à classe ω_i , caso obedeça a Equação 2.11.

3.4.4 Rejeição com múltiplos limiares (FumeraMod)

Esta estratégia é uma alteração a regra CRT, é baseada na idéia da estratégia gulosa (*Greedy*), ou seja, encontramos uma solução ótima para cada subproblema (Classe) pretendendo chegar a uma solução ótima global para o problema da rejeição. Neste caso, buscamos os limiares ótimos locais considerando uma classe de cada vez, para obter os limiares é necessário dividir o conjunto de validação em vários subconjuntos, sendo um para cada classe. Cada subconjunto é submetido a metodologia apresentada na Figura 3.4. A principal diferença em relação ao CRT é a escolha do limiar local e não global. A aplicação do mecanismo de rejeição da mesma forma que as estratégias anteriores. O Algoritmo 1 descreve a obtenção de $T_{FumeraMod}$ limiares ótimos locais para esta estratégia.

Algoritmo 1 Algoritmo aplicado na obtenção dos limiares através do conjunto de validação para a estratégia FumeraMod

```

1: Entrada: Conjunto de dados  $D[1...N, 1...c]$ ,  $R[1...N, 1...2]$  onde  $N$  é o número de instâncias,
    $c$  é o número de classes; Conjunto de rótulos  $R$ , onde  $R(.,1)$  contém o rótulo real da instância
   e  $R(.,2)$  contém o rótulo atribuído pelo classificador ( $L$ , figuras anteriores).
2: Saída: limiar =  $T_{FumeraMod}$  para cada classe
3: Informação Adicional: Critério de parada: Taxa de Erro = 0%, 1%, 2%, 3%, 4%, 5%
4:  $T_{FumeraMod} = 0$ 
5: Taxa de Erro = 0
6:  $T = 0.999999$ 
7:  $dec = 0.000001$ 
8: while  $i \leq N$  do
9:   for  $j = 1$  to  $c$  do
10:     calcule Taxa de Erro ( $T$ ) /* Conforme a Equação 3.1 */
11:     if Taxa de Erro == Critério de parada then
12:       Escreva : (Classe ( $j$ ,  $T_{FumeraMod}$ ))
13:     end if
14:      $T = T - dec$ 
15:   end for
16:    $i = i + 1$ 
17: end while

```

3.4.5 Rejeição utilizando a diferença entre os valores de confiança (DIF)

Esta estratégia é baseada em (PITRELLI; PERRONE, 2002) calculando a diferença entre os dois maiores níveis de confiança obtidos na saída da RNA. O Algoritmo 2 descreve o procedimento e a estruturação necessária ao conjunto de dados para sua implementação

e obtenção de limiares T_{DIF} .

Algoritmo 2 Algoritmo aplicado na obtenção dos limiares através do conjunto de validação para a estratégia DIF.

```

1: Entrada: Conjunto de dados  $D[1...N, 1...c]$ ,  $R[1...N, 1...2]$  onde  $N$  é o número de instâncias,
    $c$  é o número de classes; Conjunto de rótulos  $R$ , onde  $R(.,1)$  contém o rótulo real da instância
   e  $R(.,2)$  contém o rótulo atribuído pelo classificador ( $L$ , figuras anteriores).
2: Informação Adicional: Critério de parada: Taxa de Erro = 0%, 1%, 2%, 3%, 4%, 5%
3: Saída: limiar =  $T_{DIF}$ 
4:  $T_{dif} = 0$ 
5: Taxa de Erro = 0
6:  $T = 0.999999$ 
7:  $dec = 0.000001$ 
8: while  $i \leq N$  do
9:   for  $j = 1$  to  $c$  do
10:     encontre os dois valores máximos em  $D[i, j]$ 
11:      $Top_1 =$  primeiro valor máximo em  $D[i, j]$ 
12:      $Top_2 =$  segundo valor máximo em  $D[i, j]$ 
13:   end for
14:   Escreva =  $D_{dif} [Top_1 - Top_2] R [1...2]$ 
15:    $i = i + 1$ 
16: end while
   /* Fim da reorganização do conjunto */
17: while  $k \leq N$  do
18:   calcule Taxa de Erro(  $T$ ) para  $D_{dif}$  /* Conforme a Equação 3.1 */
19:   if Taxa de Erro == Critério de parada then
20:     Escreva:  $T_{DIF}$ 
21:   end if
22:    $T = T - dec$ 
23:    $k = k + 1$ 
24: end while

```

3.4.6 Rejeição utilizando Média das Classes (Class Average)

Esta estratégia é um método heurístico para calcular um suposto limiar ótimo a ser adotado como critério de rejeição. Aqui, o limiar é obtido de forma automática em relação ao conjunto de dados através da fórmula matemática para encontrar a média de um conjunto de valores. Neste caso, extraímos a média de um subconjunto (classe selecionada) considerando somente os exemplos corretamente classificados de uma determinada classe.

O Algoritmo 3 descreve o processo de obtenção dos limiares $T_{Média das Classes}$.

Algoritmo 3 Algoritmo aplicado na obtenção dos limiares através do conjunto de validação para a estratégia Média das Classes

```

1: Entrada: Conjunto de dados  $D[1...N, 1...c]$ ,  $R[1...N, 1...2]$  onde  $N$  é o número de instâncias,
    $c$  é o número de classes; Conjunto de rótulos  $R$ , onde  $R(.,1)$  contém o rótulo real da instância
   e  $R(.,2)$  contém o rótulo atribuído pelo classificador ( $L$ , figuras anteriores).
2: Saída:  $\text{limiar} = T_{\text{Média das classes}}$  Para cada Classe
3:  $T_{\text{Média das classes}} = 0$ 
4: while  $i \leq N$  do
5:   for  $j = 1$  to  $c$  do
6:     if  $R[i,1] == R[i,2]$  then
7:        $\text{Média} = \text{calcular a Média de } c$  /* Média Aritmética */
8:        $T_{\text{Média das classes}} = \text{Média de } c$ 
9:        $\text{Escreva} : (\text{Classe } (j, T_{\text{Média das classes}}))$ 
10:    end if
11:   end for
12:    $i = i + 1$ 
13: end while

```

Algoritmo 4 Algoritmo aplicado na obtenção dos limiares através do conjunto de validação para a estratégia Média/Desvio

```

1: Entrada: Conjunto de dados  $D[1...N, 1...c]$ ,  $R[1...N, 1...2]$  onde  $N$  é o número de instâncias,
    $c$  é o número de classes; Conjunto de rótulos  $R$ , onde  $R(.,1)$  contém o rótulo real do exemplo
   e  $R(.,2)$  contém o rótulo atribuído pelo classificador ( $L$ , figuras anteriores).
2: Saída:  $\text{limiar} = T_{\text{Média/Desvio}}$  Para cada Classe
3:  $T_{\text{Média/Desvio}} = 0$ 
4: while  $i \leq N$  do
5:   for  $j = 1$  to  $c$  do
6:     if  $R[i,1] == R[i,2]$  then
7:        $\text{Média} = \text{calcular Média de } c$  /* Média Aritmética */
8:        $\text{Desvio Padrão} = \text{calcular Desvio Padrão de } c$  /* Raiz quadrada da variância */
9:        $T_{\text{Média/Desvio}} = (\text{Média de } c - \text{Desvio Padrão de } c)$ 
10:       $\text{Escreva} : (\text{Classe } (j, T_{\text{Média/Desvio}}))$ 
11:    end if
12:   end for
13:    $i = i + 1$ 
14: end while

```

3.4.7 Rejeição utilizando Média da Classes e Desvio Padrão (Class Average and Standard Deviation)

Esta estratégia é derivada da anterior. Neste caso além de calcular a média calculamos também o desvio padrão para os dados corretamente classificados. O limiar é obtido através da diferença desses dois valores. O Algoritmo 4 descreve a obtenção dos limiares $T_{Média/Desvio}$.

3.5 Avaliação de Desempenho

Avaliação é um aspecto importante, pois o desempenho de um classificador pode ser diretamente influenciado por algumas características da base de dados. Baseados na matriz de confusão, diferentes métodos quantitativos podem ser derivados para a avaliação de um classificador. A denominação clássica de uma matriz de confusão refere-se a instâncias classificadas de maneira *booleana* em uma ou outra classe, sendo construída para mostrar a proporção de acertos na atribuição de classes às instâncias pelo classificador. Nessa matriz, pode ser vista a distribuição entre as classes e o relacionamento entre a primeira e a segunda linha. Assim, qualquer medida de desempenho que utilize valores de ambas as colunas será necessariamente sensível à desproporção entre as classes.

Considerando problemas de classificação de duas classes, cada exemplo x pode ser mapeado para um elemento do conjunto $\{correto, incorreto\}$. Para distinguir a classe atual e a classe prevista, podemos usar $\{aceitação, rejeição\}$ nas classes produzidas por um modelo. Dado um classificador e suas instâncias, há quatro resultados possíveis em uma matriz de confusão, conforme descritos na Seção 2.5.1 quando detalhada a maneira de traçar uma curva ROC. A Tabela 3.1 apresenta uma matriz de confusão clássica:

Tabela 3.1: Matriz de confusão clássica.

	Predição Correta	Predição Incorreta
Aceitação	CA - Correta Aceitação	FA - Falsa Aceitação
Rejeição	FR - Falsa Rejeição	CR - Correta Rejeição

A abordagem de problemas com mais de duas classes pode ser mais complexa e difícil de administrar, pois, as instâncias podem ser corretamente ou incorretamente classificadas em relação a qualquer classe. Com c classes, a matriz de confusão se torna uma matriz contendo os “ y ” resultados do classificador. Apesar desta advertência, este método na prática é viável. A Tabela 3.2 apresenta a matriz de confusão para problemas de múltiplas classes. A Figura 4.1 apresenta uma matriz de confusão para um problema com 26 classes em um sistema de reconhecimento de caracteres, similar ao conjunto Lower

adotado neste trabalho. Por exemplo, os valores relativos a correta classificação para a classe “a” é descrita n_{11} . A relação da classe “a” para cada uma das demais classes pertencentes ao problema é representada por n_{1c} , sendo c igual ao número de classes pertencentes a matriz. Neste caso n_{cc} representa os valores para a classe “z”.

Tabela 3.2: Matriz de confusão para problemas com múltiplas classes

	Classe 1	Classe 2	Classe c
Classe 1	n_{11}	n_{12}	n_{1c}
Classe 2	n_{21}	n_{22}	n_{2c}
Classe ...	$n_{...}$	$n_{...}$	$n_{...}$
Classe c	n_{1c}	n_{2c}	n_{cc}

Classe	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
a	2747	3	2	11	1	2	12	10	0	2	2	1	3	8	9	5	37	5	1	2	4	1	7	0	1	4
b	4	2811	0	18	2	1	12	8	0	1	2	5	0	1	1	2	3	3	1	0	0	0	0	1	1	3
c	3	0	2798	1	41	3	12	0	1	0	0	6	0	1	2	0	0	3	5	1	0	1	1	0	0	1
d	4	14	1	2787	1	0	0	1	2	4	1	2	0	7	38	2	7	0	1	0	4	0	3	1	0	0
e	4	7	14	3	2802	9	7	0	3	1	0	5	0	0	0	1	4	7	3	1	1	0	0	0	1	7
f	5	1	2	4	6	2798	7	1	3	1	2	1	0	0	0	13	3	10	4	15	0	0	0	0	1	3
g	18	23	5	12	7	8	2582	1	1	9	0	0	1	4	1	4	166	5	14	3	1	0	1	1	9	4
h	12	3	1	4	0	1	0	2746	0	0	13	17	8	40	0	2	0	4	0	2	7	0	7	4	7	2
i	2	4	0	1	2	1	2	2	2506	40	1	287	1	2	0	1	3	6	2	3	1	0	0	0	4	9
j	1	2	0	9	1	2	10	1	72	2740	0	6	0	0	0	0	0	0	12	12	4	2	0	1	4	1
k	4	9	0	5	0	1	0	15	3	0	2816	3	0	3	0	2	1	9	0	2	1	2	1	2	1	0
l	1	3	5	3	1	1	3	8	384	2	2	2447	0	0	0	2	4	1	2	2	0	2	3	0	4	
m	3	0	0	1	0	0	0	5	0	0	0	0	2837	21	1	1	0	1	0	0	2	1	1	1	3	2
n	3	2	0	0	0	0	11	1	3	2	0	24	2796	1	0	0	10	0	2	8	1	12	3	1	0	
o	6	6	3	45	3	0	5	0	1	1	0	0	1	2	2774	6	14	2	0	2	4	5	0	0	0	
p	2	2	0	14	1	6	3	0	0	1	0	2	0	1	1	2827	2	13	0	4	0	0	0	0	1	
q	42	5	2	5	5	2	84	3	1	3	2	1	0	1	16	11	2677	1	1	4	1	0	2	2	7	2
r	18	1	0	2	7	6	1	3	1	0	9	0	1	9	0	5	1	2796	1	2	1	8	1	3	3	1
s	4	4	1	12	2	1	8	0	0	10	1	3	0	4	2	0	0	3	2810	3	1	1	7	0	1	2
t	1	4	2	3	5	3	0	1	2	3	1	1	1	0	1	2	0	9	1	2829	0	2	0	1	5	3
u	4	2	1	0	0	0	1	2	1	4	2	3	1	3	4	0	1	0	0	2	2736	81	14	1	13	4
v	0	0	0	0	1	0	0	0	3	3	1	1	2	4	2	0	0	19	1	0	54	2766	3	1	18	1
w	3	0	0	2	0	0	2	0	2	1	0	2	24	0	0	0	1	0	0	5	8	2828	0	1	1	
x	4	1	0	2	1	3	2	6	0	5	15	0	2	3	0	4	1	6	0	1	4	0	2798	15	6	
y	1	2	0	3	0	3	11	6	1	12	0	1	3	0	1	1	2	10	2	4	12	27	3	10	2765	0
z	2	2	0	1	4	0	5	0	2	2	0	0	1	0	0	0	4	1	2	1	0	1	0	1	1	2850

Figura 3.7: Exemplo de uma matriz de confusão para um problema de 26 classes.

A forma de avaliação comumente utilizada é apresentada nas Equações 3.2 e 3.1 possibilitando a construção de um gráfico considerando Erro e Rejeição para identificar o melhor compromisso, ou seja, quantos exemplos são rejeitados para obter determinada taxa de erro, em uma estratégia de rejeição específica. Nos experimentos realizados no presente trabalho, tal como (FUMERA; ROLI; GIACINTO, 2000) considera-se a habitual exigência erro/rejeição de aplicações de reconhecimento de padrões em problemas reais, isto é, obter a mais alta taxa de reconhecimento associada a uma taxa de rejeição abaixo de um dado valor, neste caso taxa de rejeição que levem a erros entre 0% e 5%. Definimos o melhor compromisso erro/rejeição como sendo a menor taxa de rejeição associada a menor taxa de erro.

$$\text{Taxa de Erro} = \frac{FA - FR}{CA + FA + CR + FR} \quad (3.1)$$

$$\text{Taxa de Rejeição} = \frac{CR + FR}{CA + FA + CR + FR} \quad (3.2)$$

A utilização das curvas ROC foi detalhada nas Seções 2.5.1, 2.5.2 e 2.5.3 do Capítulo 2 quando revisados os temas centrais deste trabalho. Entretanto, as Equações 2.1, 2.2, relembram como essas taxas devem ser calculadas, sendo FAR a decisão que um exemplo foi classificado corretamente quando na realidade ele foi incorretamente classificado e FRR a decisão de que um exemplo foi classificado incorretamente quando na realidade foi classificado corretamente. Esse tipo de avaliação é mais apropriada quando necessitamos comparar mais de um modelo, sendo que a adição de um novo modelo pode ser facilmente realizada.

Para avaliar as estratégias de rejeição utilizaremos curvas ROC. Quando comparadas duas ou mais curvas, caso não existam nenhuma intersecção ou sobreposição, a curva que mais se aproxima do ponto (0,0) é a de melhor desempenho. Caso exista intersecção ou sobreposição, aumenta consideravelmente a complexidade da avaliação tendo que considerar outros detalhes, como por exemplo, a análise do valor obtido por FAR e FRR independente da curva, o compromisso erro/rejeição, os limiares e respectivas taxas obtidas nos conjuntos de validação e testes. A avaliação das estratégias de rejeição utilizadas neste trabalho será conduzida da seguinte forma:

1. Primeiramente, realizamos uma análise visual em relação aos gráficos erro/rejeição e as curvas ROC, afim de identificar a estratégia que apresenta os melhores compromissos para ambos os casos. Para identificar o compromisso analisamos a distância das curvas em relação ao ponto (0,0) do gráfico.
2. Em seguida, caso existam intersecção, sobreposição ou ajuste nas escalas, observamos os valores de erro/rejeição e também das taxas de FAR e FRR para os seis limiares aplicados ao conjunto de teste.
3. Para finalizar, caso seja necessário, verificamos os valores de CA, FA, CR e FR proporcionados pela estratégia em relação ao conjunto de dados utilizado.

3.6 Interpretação dos Resultados

A principal meta desta fase é compreensão dos resultados para verificar a relação entre os objetivos iniciais e os resultados obtidos. Nesta fase verificamos como as estratégias de rejeição podem ser influenciadas pelas características de um dado problema. O capítulo 4 descreve detalhadamente esta fase.

Capítulo 4

Experimentos e Resultados

Neste capítulo são apresentados os experimentos realizados e os resultados obtidos aplicando a metodologia proposta, investigando sua eficiência em relação as estratégias de rejeição em dois tipos de conjuntos de dados: sintéticos e reais. O primeiro conjunto é gerado através de parâmetros pré-definidos com apenas duas classes. Tratando-se de conjuntos especialmente desenvolvido para testar o comportamento das estratégias de rejeição. Sua importância está relacionada ao fato de que permitem variar suas características de acordo com nosso domínio de interesse, permitindo avaliar o desempenho de cada uma das estratégias para cada ambiente simulado. O segundo é formado por 26 classes de caracteres maiúsculos, minúsculos, posteriormente, combinados formando 52 classes, neste caso não se conhece o processo de geração dos dados.

4.1 Experimentos em Problemas com Duas Classes

Neste experimento todas as estratégias de rejeição e parâmetros detalhados na metodologia do Capítulo 3 são aplicadas em problemas de duas classes, buscando verificar o comportamento das estratégias de rejeição. Para sua realização foram gerados 18 conjuntos de dados sintéticos. Esses conjuntos são compostos por agrupamentos com parâmetros controlados já detalhados na seção 3.3.1. A avaliação de desempenho em relação as seis estratégias de rejeição implementadas para os 18 conjuntos de dados com duas classes são apresentados nesta seção. É importante destacar que todos os gráficos e a relação ordenada do desempenho das seis estratégias utilizadas nas seções seguintes serão apresentadas na seção 4.1.4. Outras informações sobre os valores adotados e demais detalhes referentes aos conjuntos de dados utilizados na realização dos experimentos podem ser encontrados no Apêndice A.

4.1.1 Dados Balanceados e Desbalanceados

A hipótese de que o desbalanceamento das classes pode influenciar o desempenho de um modelo de classificação em sistema de aprendizado supervisionado e que os algoritmos de aprendizado supervisionado podem encontrar dificuldades na indução nestes casos, é discutida em muitos trabalhos. Por exemplo, esta situação é abordada por Fawcett e Provost (1997) pesquisando o problema de detecção de fraudes em chamadas telefônicas no qual utilizam um sistema baseado em regras para extrair indicadores que podem ser usados na identificação de fraudes através do monitoramento de ligações.

4.1.1.1 Análise do Impacto do Balanceamento

Nesta seção a análise é conduzida pela seguinte questão: A utilização de um conjunto de dados balanceado ou desbalanceado pode interferir no desempenho de uma estratégia de rejeição? Para confirmar ou não essa hipótese realizamos várias comparações considerando conjuntos balanceados e desbalanceados.

A primeira comparação para verificar se existe essa influência nas estratégias de rejeição é realizada utilizando os conjuntos NBS e NDS. As Figuras 4.1 e 4.4 apresentam os gráficos citados. Observando os gráficos erro/rejeição percebemos que a estratégia mais adequada é Chow para ambos os conjuntos. Os piores desempenhos são apresentados pelas estratégias Média das Classes no primeiro e FumeraMod no segundo. A avaliação do gráfico ROC aponta DIF como a estratégia mais adequada e Média/Desvio e Média das Classes com desempenhos similares. Chow, Fumera e FumeraMod apresentam taxas de FAR elevadas descartando essas estratégias para aplicação neste problema. Para o segundo conjunto Chow apresenta-se como a melhor opção proporcionando taxas de FAR e FRR igual 0% e DIF, Fumera e FumeraMod novamente proporcionam taxas de FAR de 100%. Na avaliação destes conjuntos as estratégias que se destacaram foram Chow com o melhor desempenho e FumeraMod com o pior desempenho na maioria dos casos.

A segunda comparação é realizada em dois conjuntos gerados através de Distribuição Normal com dados parcialmente sobrepostos: NBPS e NDPS, representados pelas Figuras 4.2 e 4.5 mostram os gráficos citados durante a segunda avaliação. Para ambos os conjuntos percebe-se que o comportamento das estratégias Chow, Fumera e DIF é idêntico. A estratégia FumeraMod não apresenta resultados satisfatórios para NDPS e apresenta-se como estratégia mais adequada para NBPS. A análise do gráfico ROC confirma a equivalência das estratégias Chow, Fumera e DIF e aponta uma variação muito pequena em relação as demais estratégias. Indica também que o desempenho insatis-

fatório da estratégia FumeraMod com taxa de FAR de 66,7% e FRR de 90,0% para o conjunto NDPS. As taxas de FAR e FRR para a estratégia de melhor de desempenho Média/Desvio apontam 25,1% e 15,1%.

A terceira avaliação é aplicada em um conjunto de dados gerados através de uma Distribuição *Chi-square*, são utilizados os conjuntos CBS e CDS, representados pelas Figuras 4.13 e 4.16. Verificando os gráficos erro/rejeição, percebemos que as estratégias Chow, Fumera e DIF apresentam comportamento idêntico e uma pequena variação em relação a FumeraMod. As estratégias heurísticas Média das Classes e Média/Desvio apresentam desempenho ruim para o primeiro conjunto. Para o segundo a similaridade entre Chow e Fumera permanece. Entretanto, FumeraMod caracteriza o pior desempenho. Verificando o gráfico ROC fica evidente o péssimo desempenho da estratégia FumeraMod e percebe-se que as estratégias heurísticas Média das Classes e Média/Desvio são as mais indicadas. Para esses conjuntos as taxas de FAR e FRR são elevadas para todas as estratégias aumentando significativamente em relação aos demais casos.

A última avaliação observa os experimentos utilizando os conjuntos CBSob e CD-Sob caracterizando a maior dificuldade do classificador em atribuir rotulo correto a classe. Analisando o desempenho das estratégias em relação aos gráficos erro/rejeição constatamos novamente a equivalência das estratégias Chow, Fumera e DIF. A estratégia Média/Desvio apresenta-se como mais adequada no primeiro conjunto, porém, com uma diferença muito pequena em relação às demais. No segundo conjunto a aplicação de qualquer uma das estratégias é indiferente nenhuma consegue de destacar para o problema. Verificando o gráfico ROC para primeiro conjunto percebemos que a estratégia Média das Classes se destaca para esse tipo de problema apresentando taxas de FAR e FRR baixas. Neste caso fica evidente que as estratégias Chow, Fumera e DIF possuem equivalência, pois, seus resultados são similares quando seus desempenhos são ruins no primeiro conjunto e continuam similares quando seus resultados são os melhores no segundo conjunto. A dificuldade de ajustar o mecanismo de rejeição neste conjunto de dados é consideravelmente maior que nos demais, de forma que fica evidente que a aplicação de um mecanismo de rejeição, assim como a aplicação de um sistema de classificação, é influenciada pelas características do conjunto de dados. As Figuras 4.15 e 4.20 apresentam os gráficos utilizados nesta avaliação. É interessante observar a diferença deles em relação aos demais, expressando toda a dificuldade de estabelecer uma estratégia de rejeição em destaque.

Considerando todos os conjuntos avaliados constatamos que o desbalanceamento ou não das classes pode influenciar no desempenho do mecanismo de rejeição e dificultar a escolha de uma estratégia mais indicada ou menos indicada, principalmente para os conjuntos de dados desbalanceados. Entretanto, a observação dos conjuntos avaliados sugere

que essa característica apenas agrava a situação e que podem existir outros fatores que também podem influenciar ainda mais o comportamento do mecanismo. Por exemplo, considerando os dois últimos conjuntos avaliados percebemos que a dificuldade em encontrar estratégias de destaque é agravada trabalhando com dados sobrepostos, sugerindo que a influência nas estratégias de rejeição é dada mais pela separação dos dados do que pelo balanceamento das instâncias em uma classe. As estratégias que merecem destaque para os conjuntos avaliados são Chow, Fumera e DIF pela sua equivalência em praticamente todos os casos mesmo quando destacando-se pelo melhor ou pior desempenho. Com relação ao desempenho percebemos que as estratégias que conseguem trabalhar melhor com essa característica do conjunto são as heurísticas baseadas na média da correta classificação. Outro destaque é com relação a estratégia FumeraMod que em geral seu desempenho pode ser considerado insatisfatório.

4.1.2 Separação dos Dados

Outra análise importante é separação dos dados. Acredita-se que além do desbalanceamento entre as classes a separação dos dados pode influenciar os resultados apresentados pelo classificador. Realizamos uma série de experimentos para testar a hipótese de que as diferentes sobreposições entre as classes é fator complicante tanto para dados balanceados quanto para dados desbalanceados. O Apêndice A apresenta detalhes sobre estes experimentos. Verificando as Figuras A.5, A.6 e A.7, A.8, podemos visualizar como este problema pode ser encontrado tanto em dados balanceados quanto em dados desbalanceados e a dificuldade do classificador em manipular esse tipo de dados. Percebemos, também que a complexidade do problema é inicialmente pequena quando utilizados dados separados, sendo aumentada quando referenciados dados parcialmente sobrepostos e significativamente elevada quando utilizados dados sobrepostos.

4.1.2.1 Análise do Impacto da Separação dos Dados

Nesta análise verificamos experimentalmente a influência da separação dos dados em relação ao desempenho das estratégias de rejeição diferentes conjuntos de dados. As questões que orientam nossa análise nesta seção são: Qual o impacto da utilização de dados separados, parcialmente sobrepostos ou sobrepostos? Como as estratégias de rejeição se comportam nestas condições?

Primeiramente, avaliamos a separação dos dados em relação a conjuntos gerados a partir de uma distribuição normal. Serão avaliados os conjuntos NBS, NBPS e NBSob

representados pelos gráficos das Figuras 4.1, 4.2 e 4.3. No primeiro conjunto as estratégias Chow, Fumera, DIF e FumeraMod apresentam resultados similares sendo que todas possuem baixas taxas de FAR e FRR. As estratégias heurísticas baseadas em médias são as que apresentam resultados insatisfatórios. Para o segundo e o terceiro conjunto a estratégia FumeraMod apresenta resultados significativamente melhores que as demais, principalmente em relação ao terceiro. A avaliação do gráfico ROC aponta que com os conjuntos de dados separados possuem taxas de FAR e FRR baixas e que nos conjuntos parcialmente sobrepostos essas taxas são aumentadas. Quando considerados os dados sobrepostos essas taxas são significativamente elevadas em relação ao primeiro conjunto. Desconsiderando o melhor desempenho indicado pela estratégia FumeraMod em nenhum caso existe equivalência em relação a análise ROC e análise erro/rejeição. A análise destes conjuntos indicam fortes evidências de que as estratégias de rejeição são influenciadas pela separação de dados, principalmente porque essa situação se repete para as demais combinações dos conjuntos utilizados para os problemas com duas classes. Por exemplo, podemos considerar os conjuntos: NDS, NDPS e NDSob representados pelas Figuras 4.4, 4.5 e 4.6. Os conjuntos NMBS, NMBPS e NMBSob representados pelas Figuras 4.7, 4.8 e 4.9. Finalmente, consideramos também os conjuntos nos quais o classificador apresenta maior dificuldade em atribuir um exemplo a uma determinada classe. Os conjuntos: CDS, CDPS e CDSob representados pelas Figuras 4.16, 4.17 e 4.20. Sua avaliação sustenta a hipótese de que a separação dos dados influencia no comportamento de um mecanismo de rejeição.

Considerando todos os conjuntos avaliados podemos verificar de forma mais precisa que a separação dos dados influencia o desempenho do mecanismo de rejeição. Um argumento forte para validar essa idéia é a dificuldade em identificar estratégias de destaque para identificar o melhor ou pior desempenho. A equivalência entre Chow, Fumera e DIF é observada na maioria dos conjuntos, quando Fumera é melhor que Chow ou melhor que DIF essa diferença é muito pequena sendo praticamente imperceptível através da visualização dos gráficos ROC ou erro/rejeição, para tanto precisamos refinar a análise para encontrar o melhor desempenho. As estratégias Chow e DIF e Média/Desvio apresentam os melhores resultados quando utilizados conjuntos de dados separados ou parcialmente sobrepostos. Quando manipulados os conjuntos com dados sobrepostos as estratégias FumeraMod e Fumera apresentam os melhores desempenhos, acredita-se que isso deve-se ao fato destas estratégias manipularem um limiar específico para cada classe do problema, uma característica importante quando os dados não seguem nenhuma tendência.

4.1.3 Distribuição de Probabilidade

Para a verificação do impacto da distribuição de probabilidade em relação a uma estratégia de rejeição, inicialmente utilizamos os conjuntos com dados balanceados e separados: NBS, NMBS e CBS, representados pelas Figuras 4.1, 4.7 e 4.13. Observamos que a estratégia Média/Desvio possui um desempenho similar em relação aos três conjuntos. A estratégia Média das Classes também apresenta comportamento similar nos três casos. As estratégias Chow, Fumera, DIF e FumeraMod apresentaram ótimo desempenho analisando o compromisso erro/rejeição, neste caso a aplicação de qualquer uma é indiferente. Entretanto, o desempenho apontado pelas curvas ROC é consideravelmente variável, não apontando resultados de destaque para sugerir escolhas de estratégias adequadas ou inadequadas.

A comparação seguinte é relacionada aos conjuntos NBPS, NMBPS e CBPS. As Figuras 4.2, 4.8 e 4.14 apresentam os gráficos utilizados nesta avaliação. Novamente as estratégias Média/Desvio e Média das Classes apresentam um comportamento padrão em todos os casos que utilizam conjuntos com dados balanceados e parcialmente separados. Neste conjunto o desempenho da estratégia Média/Desvio é melhor ou muito parecido com a estratégia Média das Classes. As estratégias Chow, Fumera, DIF e FumeraMod embora em alguns casos apresentam valores melhores que os anteriores no geral os desempenhos são inconstantes.

Avaliando os conjuntos: NBSob, NMBSob e CBSob, representado pelas Figuras 4.4, 4.10 e 4.16, verificamos que nesses conjuntos a complexidade na análise é consideravelmente alterada. Da mesma forma que nos conjuntos anteriores as estratégias Média/Desvio e Média Classes possuem desempenho similares. Observamos que a estratégia FumeraMod se destaca em relação as demais apresentando resultados melhores ou taxas próximas aos melhores resultados. Para avaliar o impacto do desbalanceamento utilizamos os conjuntos: NDS, NMDS e CDS. Na maioria dos casos a estratégia Chow apresenta os melhores resultados. Os demais casos apresentam variações consideráveis, sendo indiferente a escolha da estratégia.

Para os conjuntos NDPS, NMDPS e CDPS, representados pelas Figuras 4.5, 4.11 e 4.17 a avaliação erro/rejeição indica as estratégias Chow, Fumera e DIF apresentando os melhores resultados sendo ambas equivalentes. Para a análise ROC a estratégia Média/Desvio apresenta o melhor desempenho e Chow, Fumera e DIF apresentam resultados ruins. Finalmente, os conjuntos: NDSob, NMDSob e CDSob apontam que Chow e Dif apresentam os melhores resultados, porém, Fumera apresenta resultado muito próximo. O mesmo padrão é seguido pelas estratégias Média/Desvio e Média das Classes, porém

Média/Desvio possui desempenho melhor. Neste caso, a estratégia FumeraMod não apresentou bons resultados.

4.1.3.1 Análise do impacto da Distribuição de Probabilidade

Em relação aos conjuntos balanceados, concluímos, em relação a avaliação erro/rejeição que existe uma equivalência entre as estratégias Chow, Fumera, DIF e FumeraMod, porém, quando os dados apresentam-se de forma sobrepostas temos alguns resultados bons com a estratégia FumeraMod, mas o destaque neste caso é para a estratégia Média/Desvio. Nos conjuntos desbalanceados, avaliando erro/rejeição, os melhores desempenhos são das estratégias Chow, Fumera e DIF. Entretanto, Fumera e DIF apresentam resultados ruins em alguns casos. Nos conjuntos desbalanceados utilizando a análise ROC a estratégia Média/Desvio apresenta um destaque com dados separados ou parcialmente sobrepostos. Com o conjunto de dados sobrepostos não é possível identificar qual estratégia se destaca. Análise geral em relação ao compromisso erro/rejeição aponta a estratégia Chow apresentando os melhores desempenhos, porém com os conjuntos com dados sobrepostos seu desempenho não merece destaque. Em geral, a estratégia Média das Classes apresenta resultados inferiores às demais estratégias. Análise geral ROC apresenta uma grande variação entre o desempenho das estratégias com destaque para a estratégia Média/Desvio em conjuntos desbalanceados e não sobrepostos.

Uma consideração importante em relação a essa verificação do comportamento do mecanismo de rejeição em diferentes distribuições de probabilidades é que em geral os experimentos realizados utilizando rejeição consideram somente distribuições normais, como o caso das estratégias Chow e Fumera. Percebemos experimentalmente, que as diferentes distribuições de probabilidade interferem nos resultados apresentados pelo mecanismos de rejeição. Entretanto, em um problema real na maioria dos casos a distribuição é desconhecida ou estimada possibilitando que essa característica influencie diretamente nos resultados do mecanismo de rejeição aplicado.

4.1.4 Resumo dos Resultados

Os gráficos utilizados para avaliar o desempenho das seis estratégias são apresentados nesta Seção. A relação dos desempenhos obtidos pelas estratégias de rejeição ordenadas com relação a menor distância do ponto (0,0) e respectivos dezoito conjuntos de dados são apresentados nas Tabelas 4.1, 4.2, 4.3. Essas tabelas são construídas utilizando a os valores apontados pela análise ROC e valores proporcionados pelo sistema de

reconhecimento em termos de taxa de erro e taxa de rejeição. As discussões dos resultados ilustrados foram apresentadas na seção anterior.

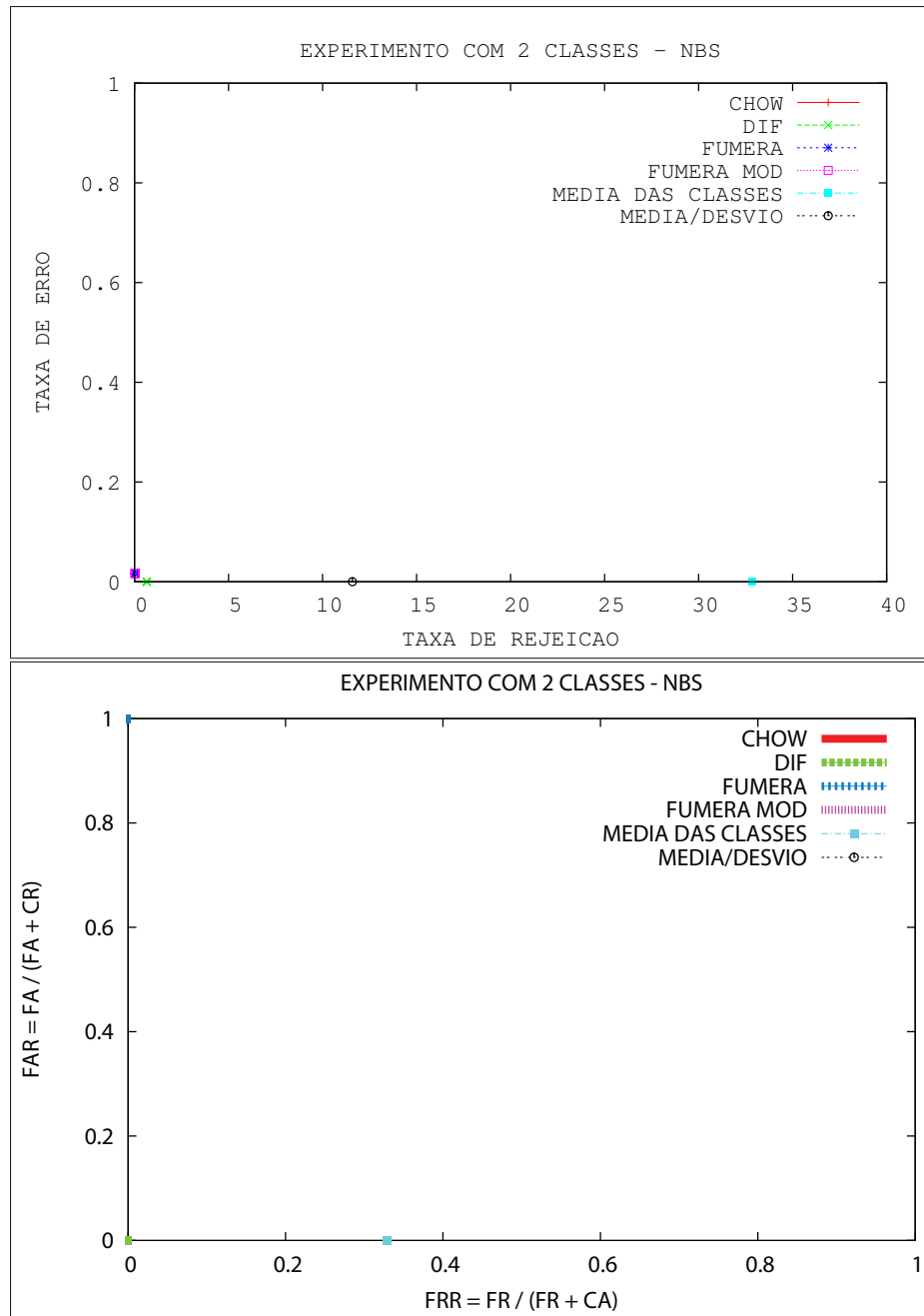


Figura 4.1: Conjunto de dados NBS - Gráfico Erro/Rejeição e Gráfico ROC para as seis estratégias implementadas.

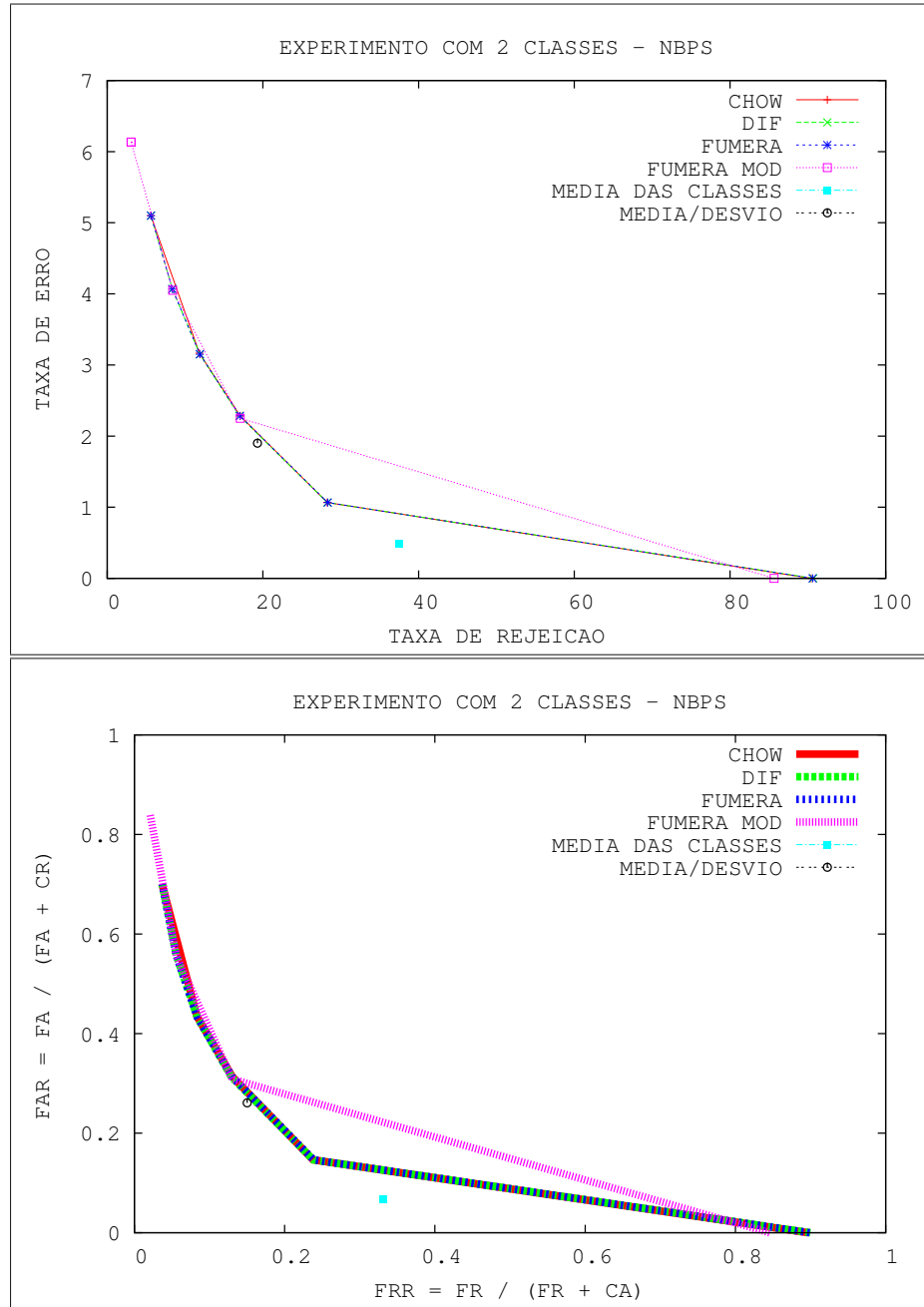


Figura 4.2: Conjunto de dados NBPS - Gráfico Erro/Rejeição e Gráfico ROC para as seis estratégias implementadas.

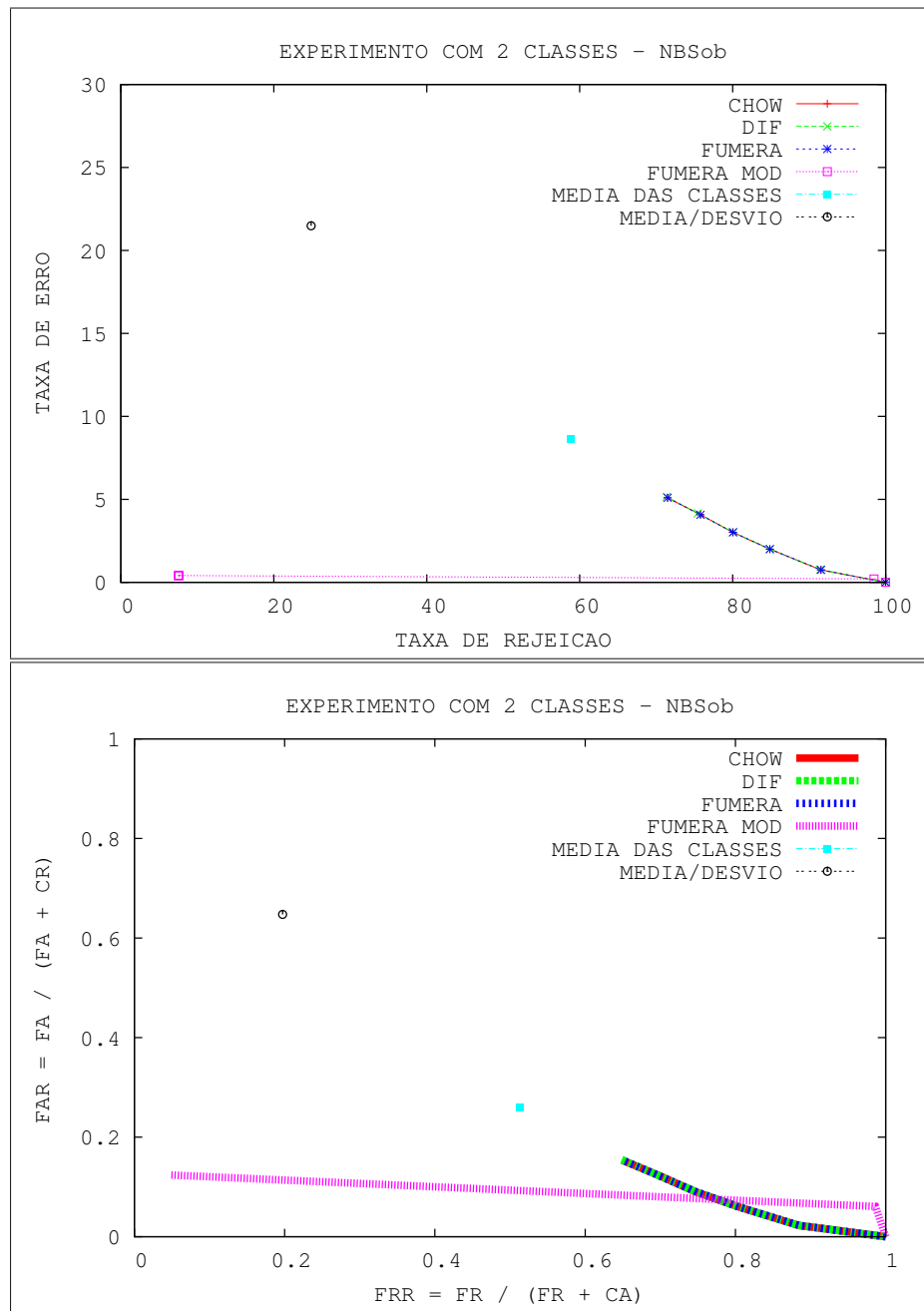


Figura 4.3: Conjunto de dados NBSob - Gráfico Erro/Rejeição e Gráfico ROC para as seis estratégias implementadas.

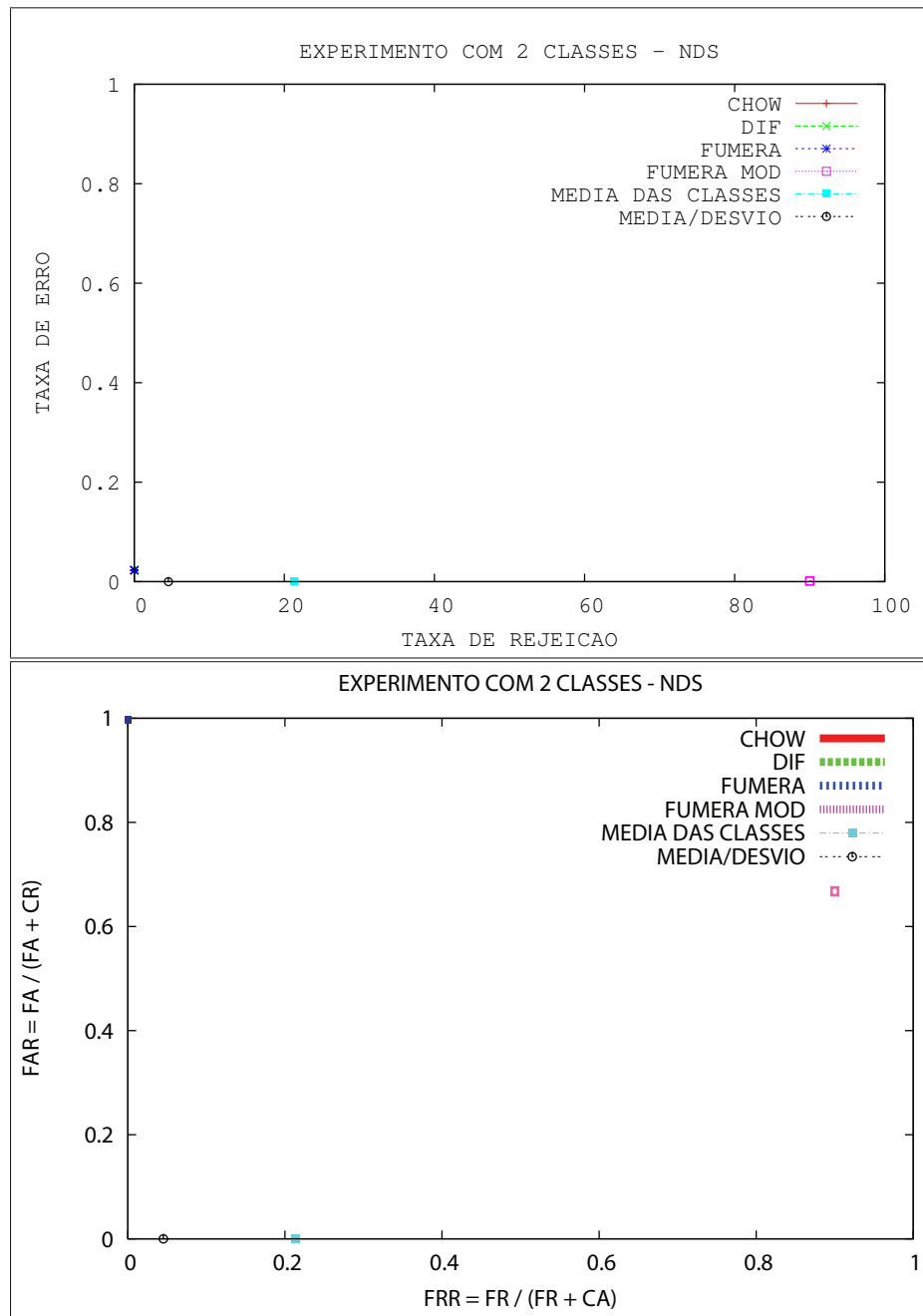


Figura 4.4: Conjunto de dados NDS - Gráfico Erro/Rejeição e Gráfico ROC para as seis estratégias implementadas.

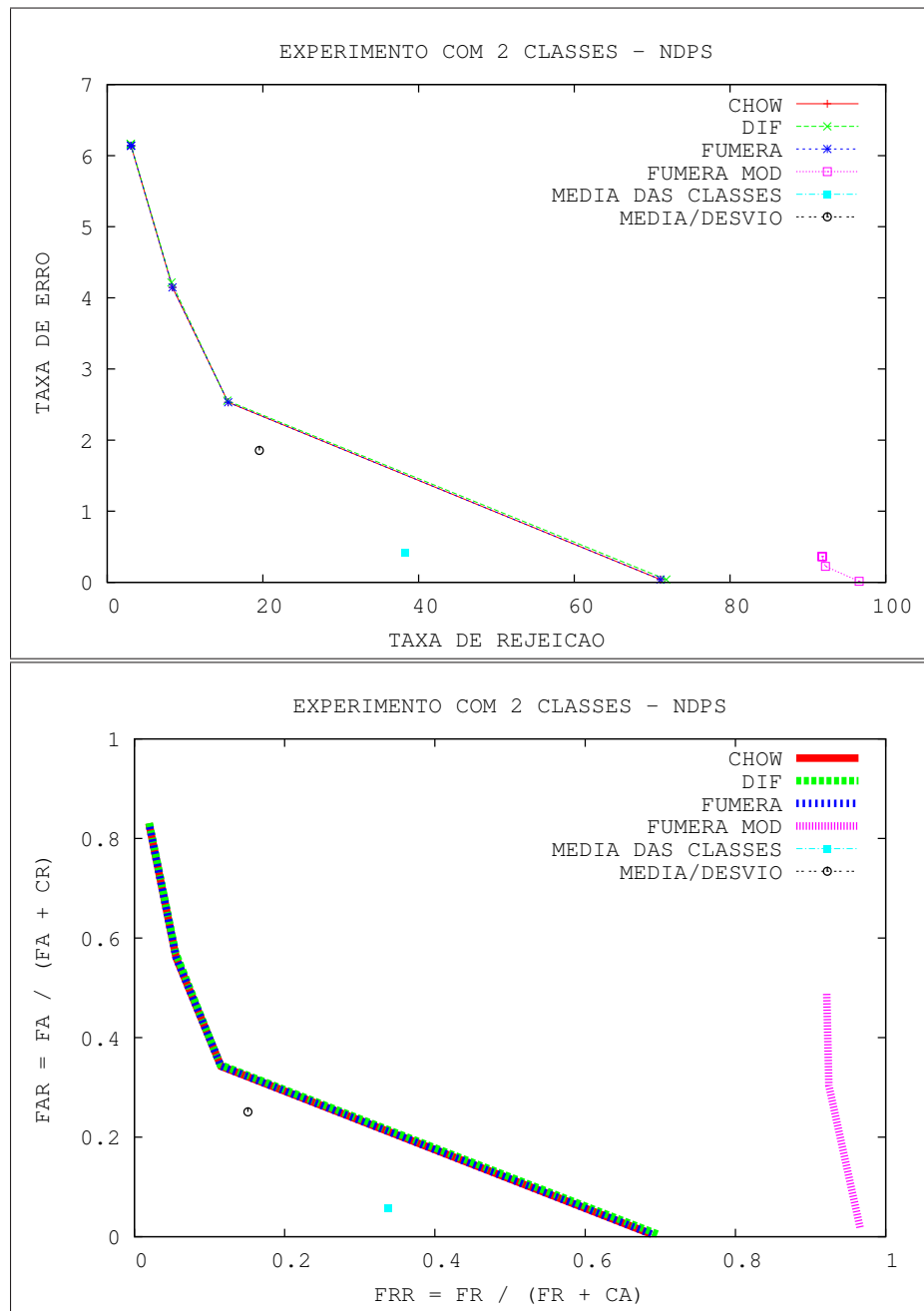


Figura 4.5: Conjunto de dados NDPS - Gráfico Erro/Rejeição e Gráfico ROC para as seis estratégias implementadas.

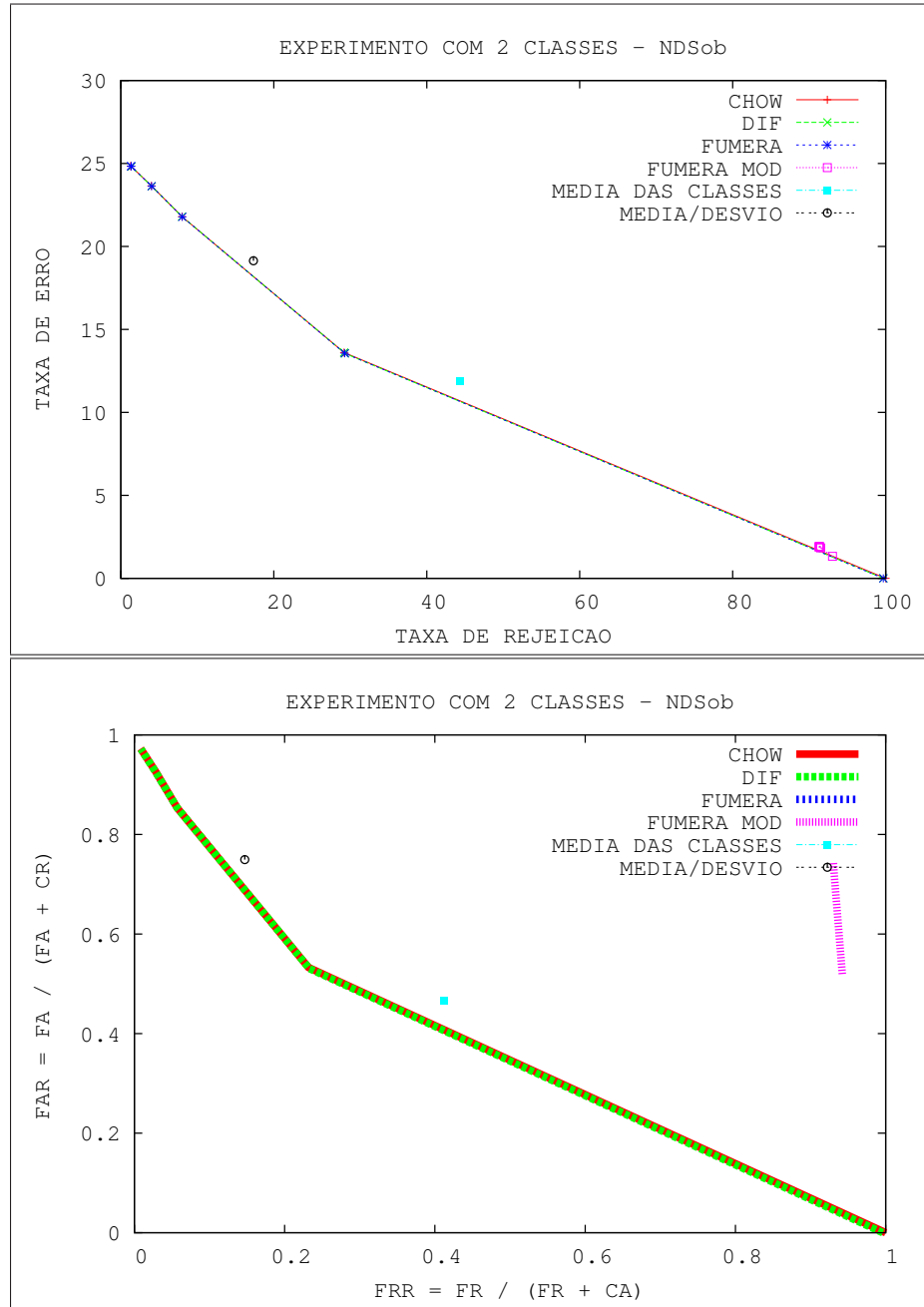


Figura 4.6: Conjunto de dados NDSob - Gráfico Erro/Rejeição e Gráfico ROC para as seis estratégias implementadas.

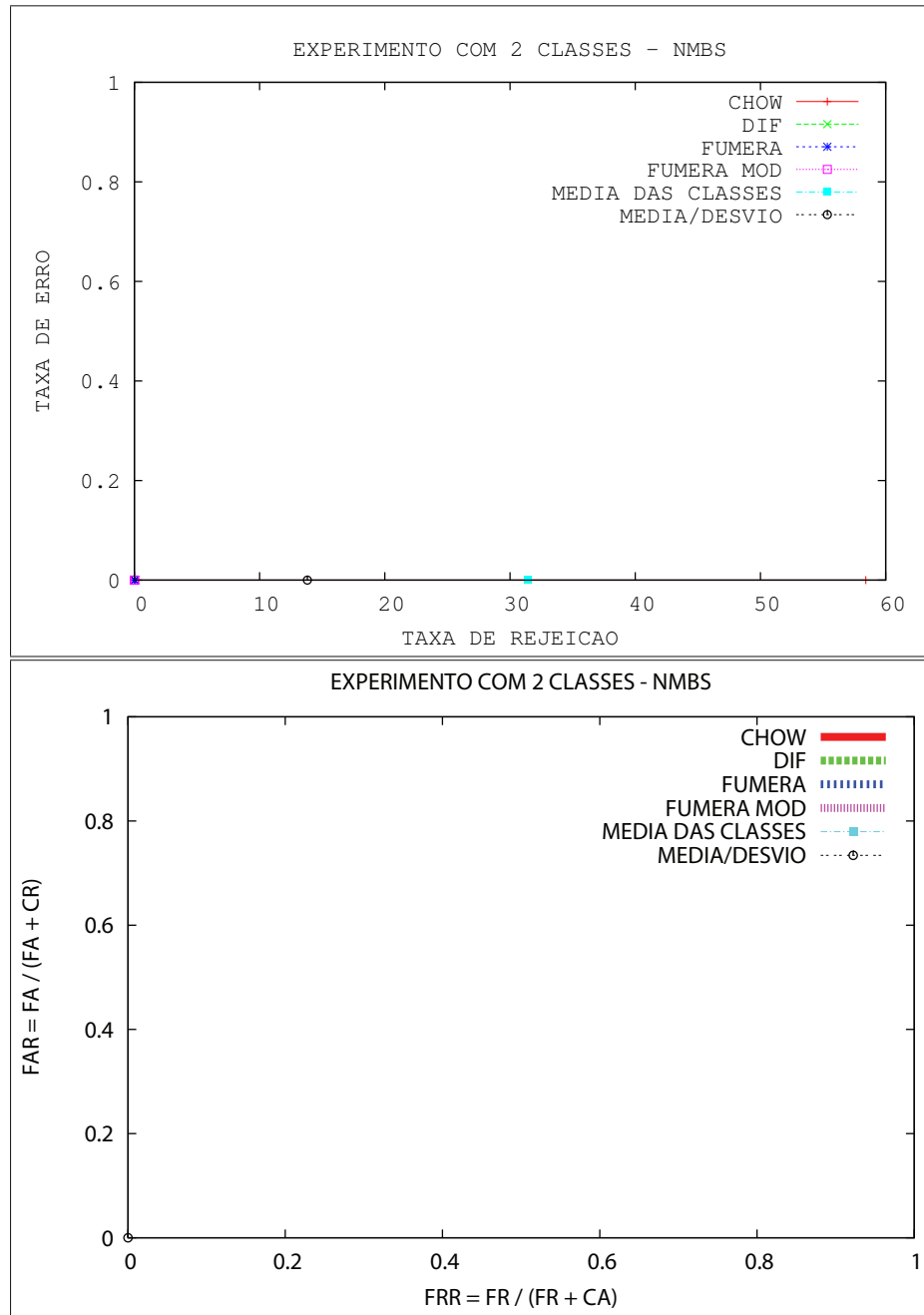


Figura 4.7: Conjunto de dados NMBS - Gráfico Erro/Rejeição e Gráfico ROC para as seis estratégias implementadas.

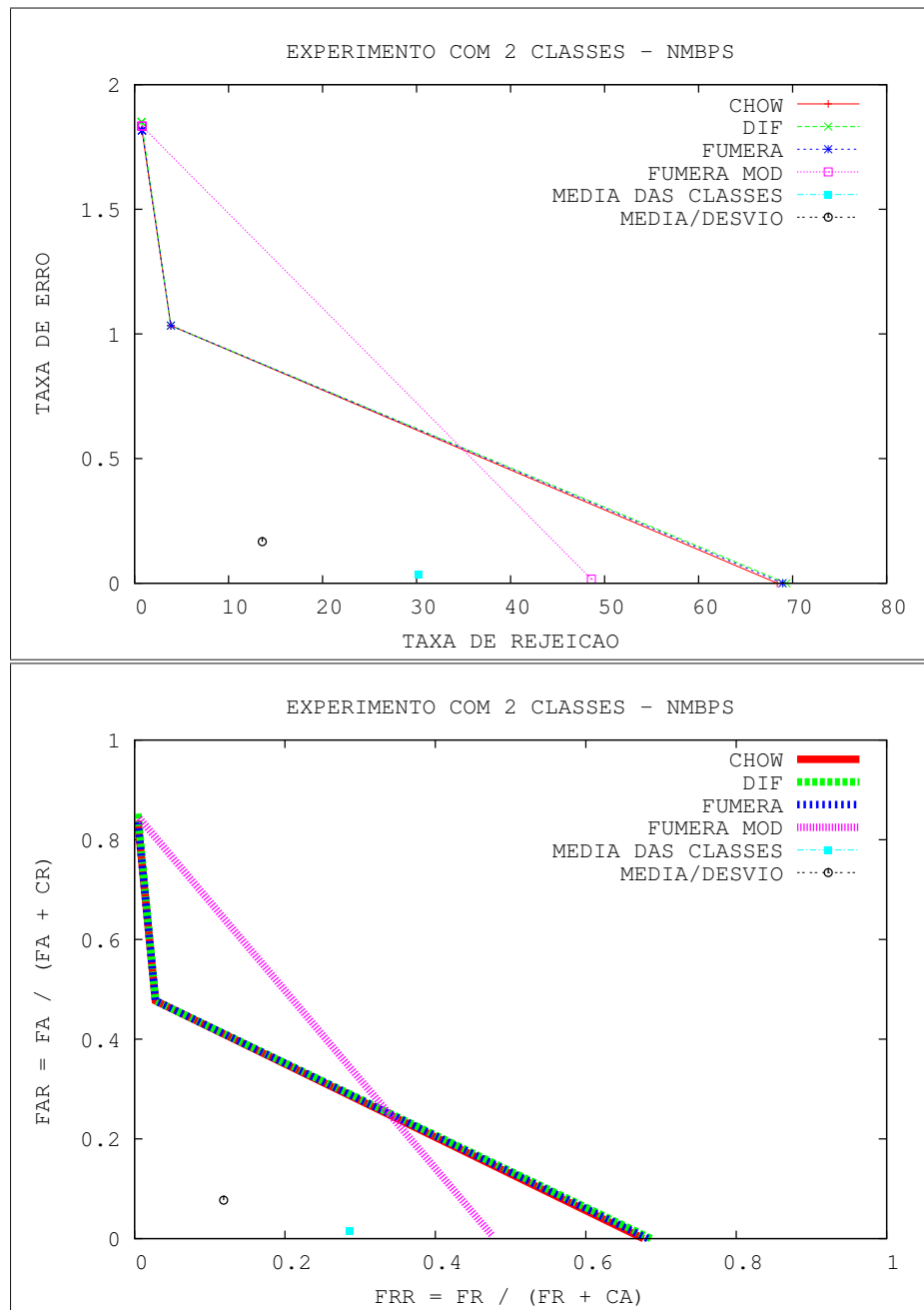


Figura 4.8: Conjunto de dados NMBPS - Gráfico Erro/Rejeição e Gráfico ROC para as seis estratégias implementadas.

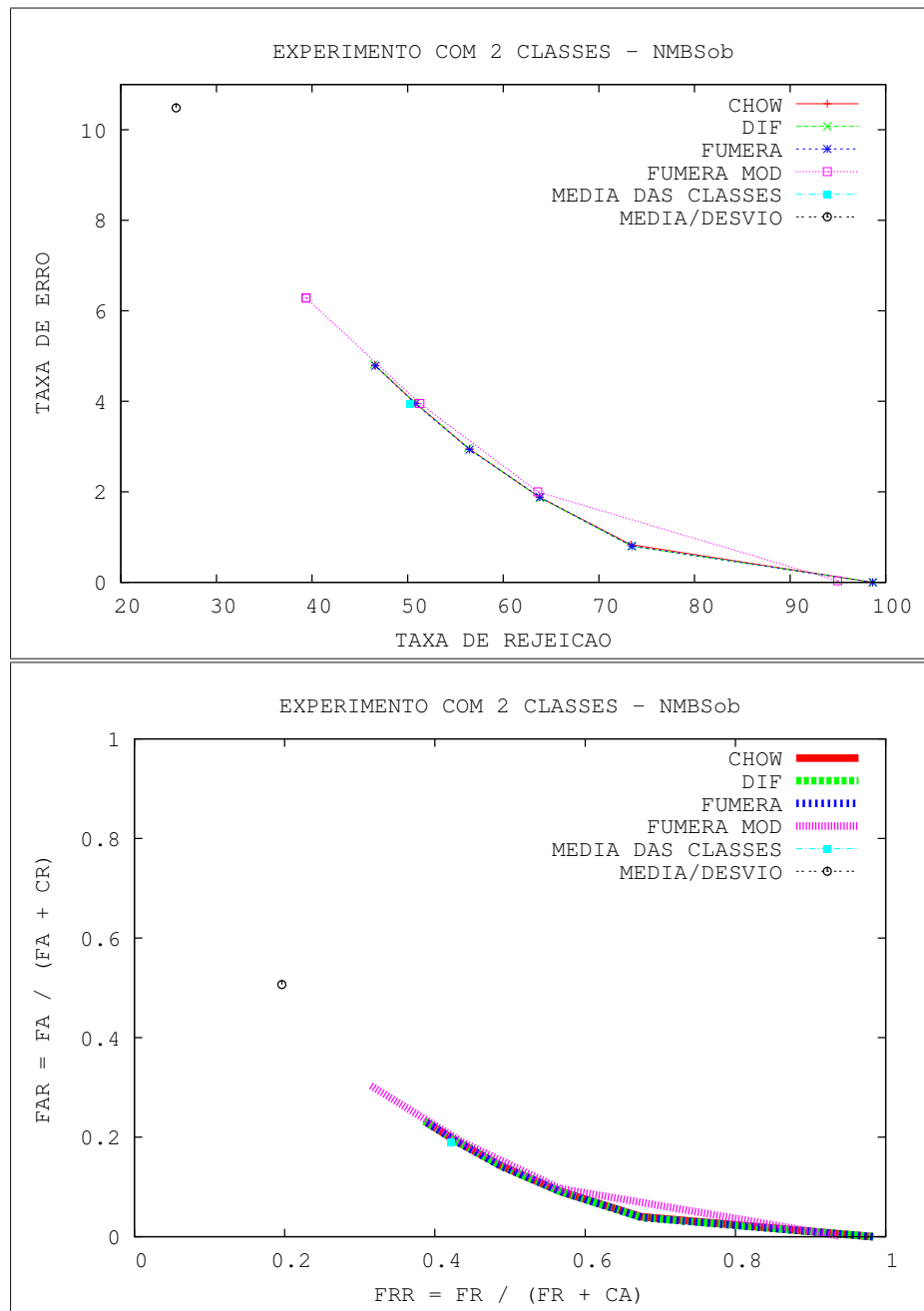


Figura 4.9: Conjunto de dados NMBSob - Gráfico Erro/Rejeição e Gráfico ROC para as seis estratégias implementadas.

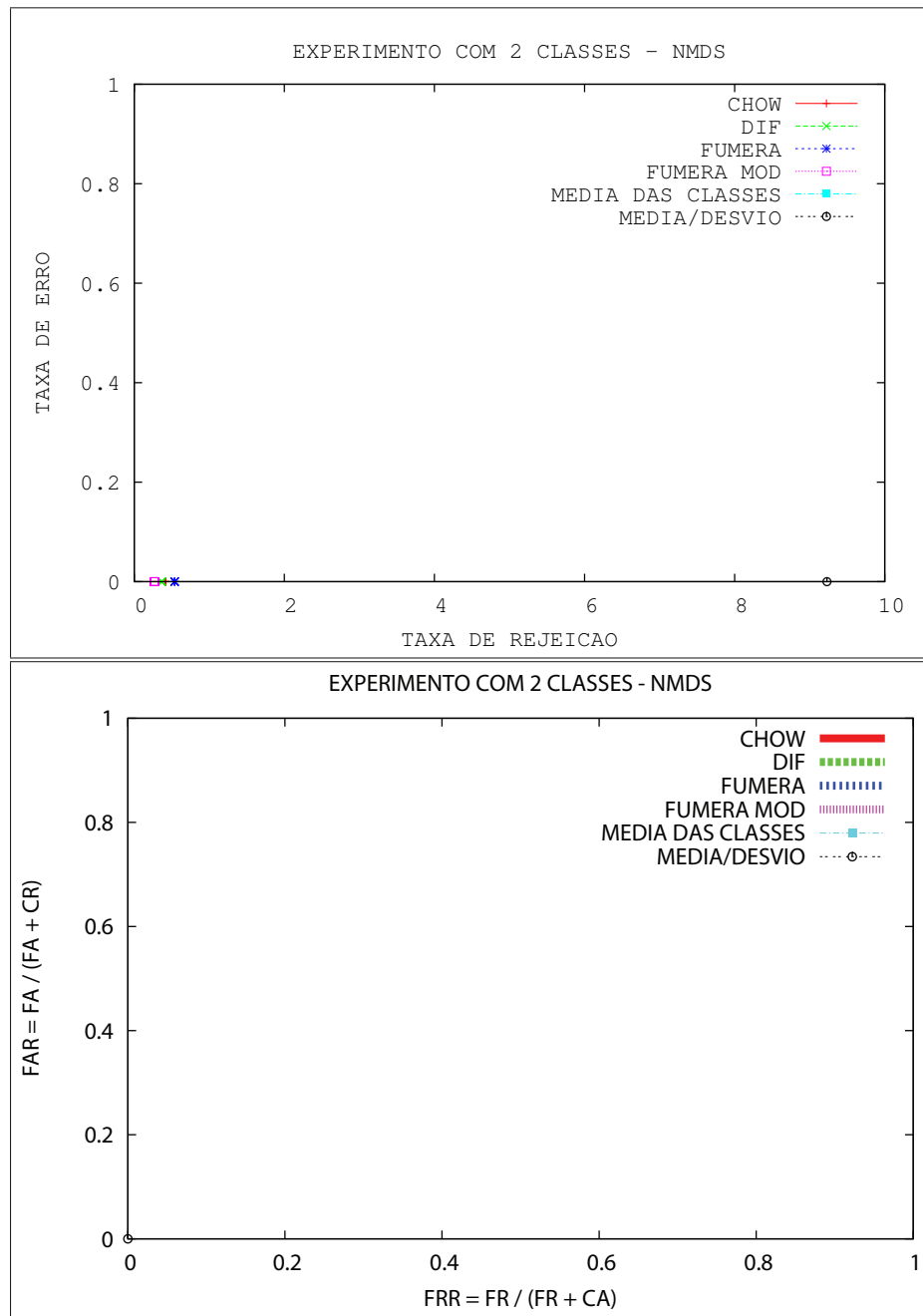


Figura 4.10: Conjunto de dados NMDS - Gráfico Erro/Rejeição e Gráfico ROC para as seis estratégias implementadas.

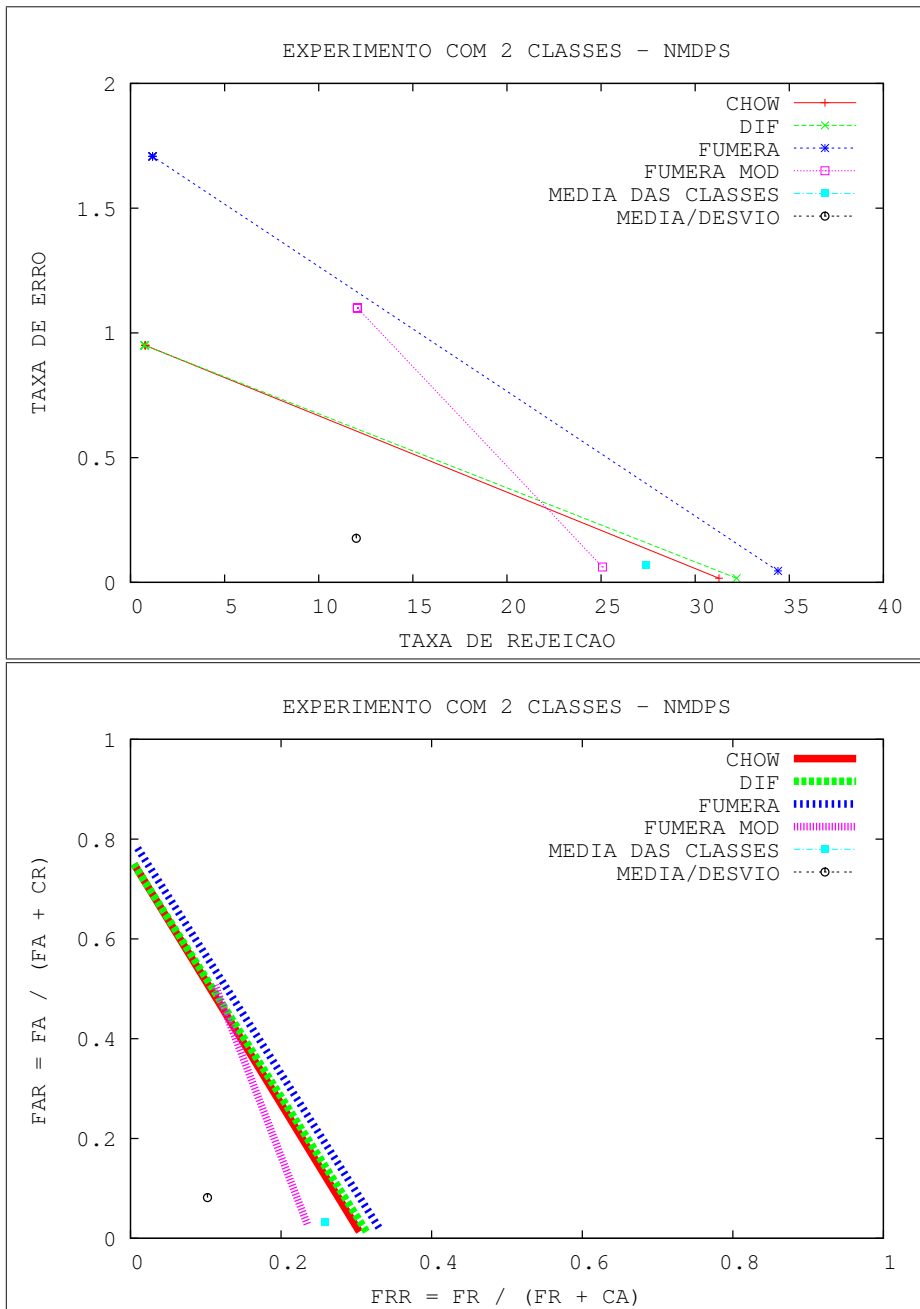


Figura 4.11: Conjunto de dados NMDPS - Gráfico Erro/Rejeição e Gráfico ROC para as seis estratégias implementadas.

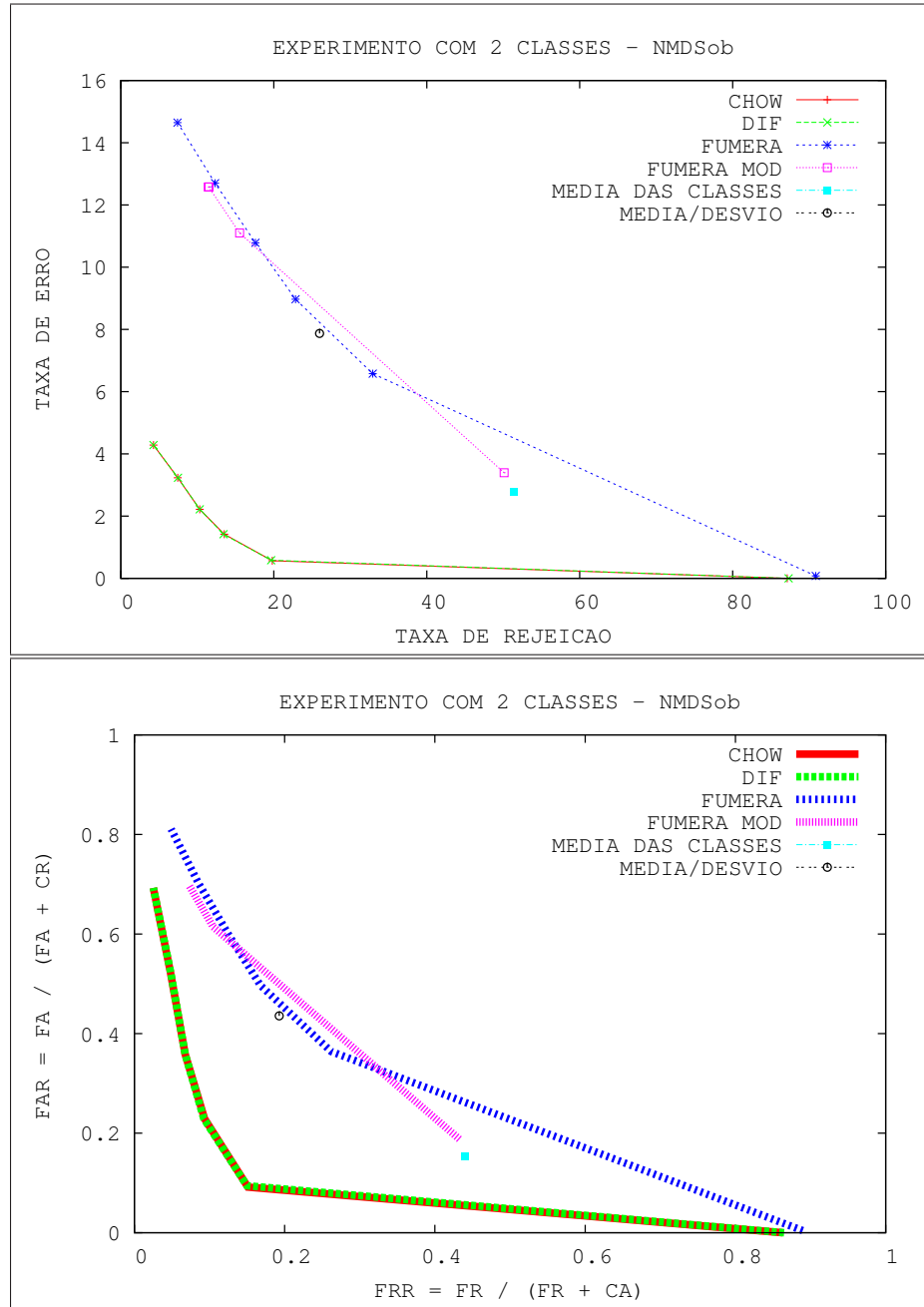


Figura 4.12: Conjunto de dados NMD Sob - Gráfico Erro/Rejeição e Gráfico ROC para as seis estratégias implementadas.

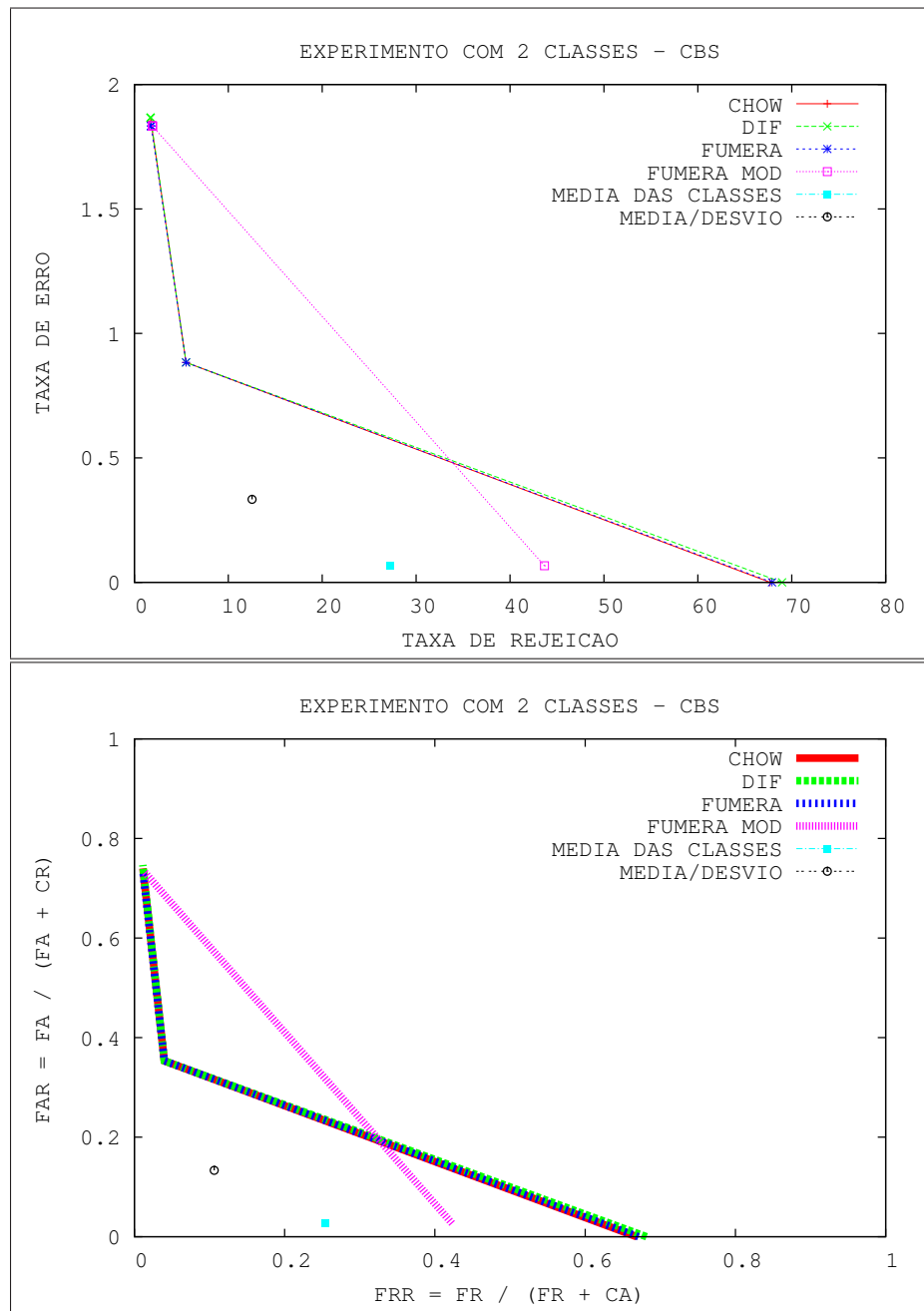


Figura 4.13: Conjunto de dados CBS - Gráfico Erro/Rejeição e Gráfico ROC para as seis estratégias implementadas.

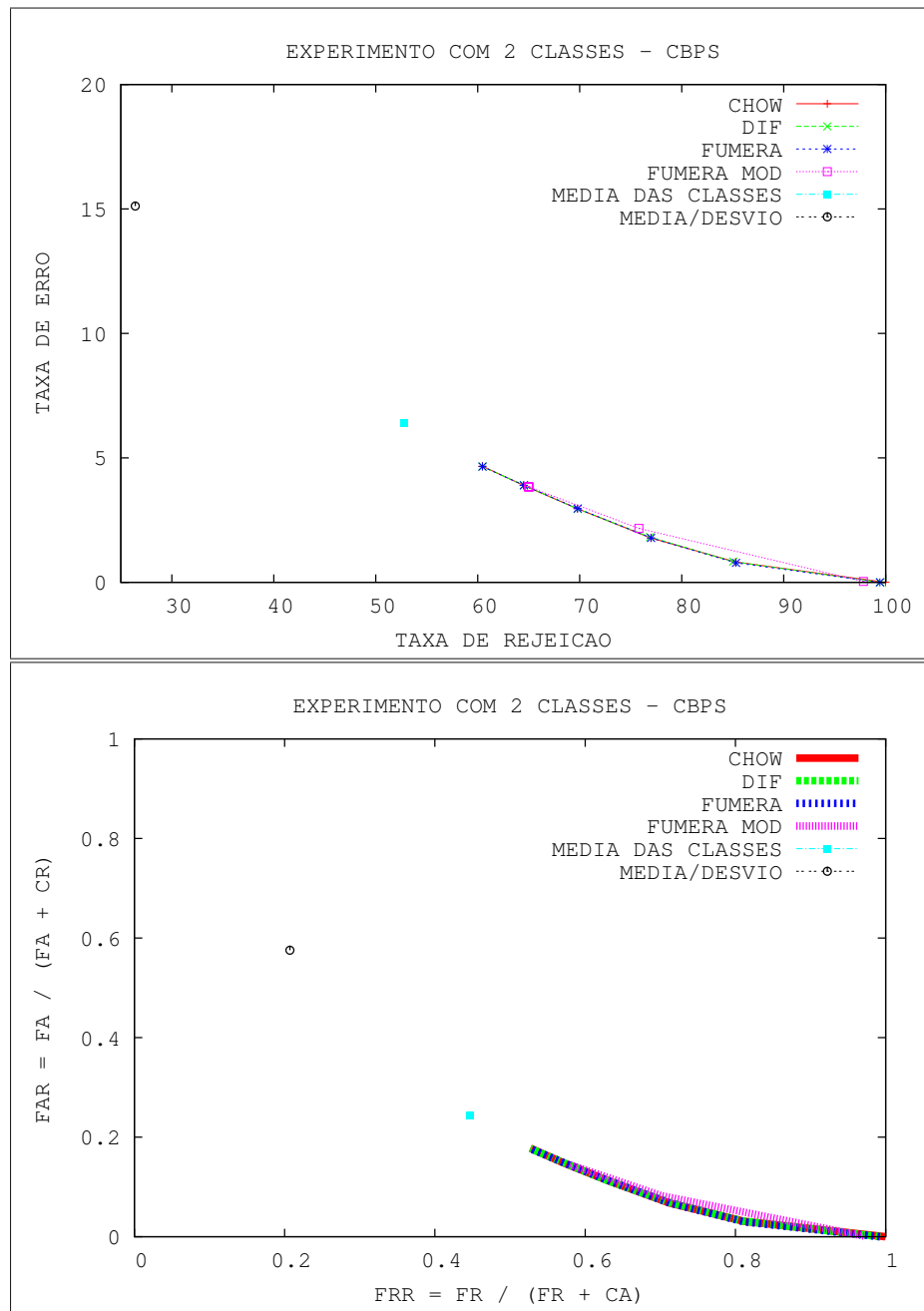


Figura 4.14: Conjunto de dados CBPS - Gráfico Erro/Rejeição e Gráfico ROC para as seis estratégias implementadas.

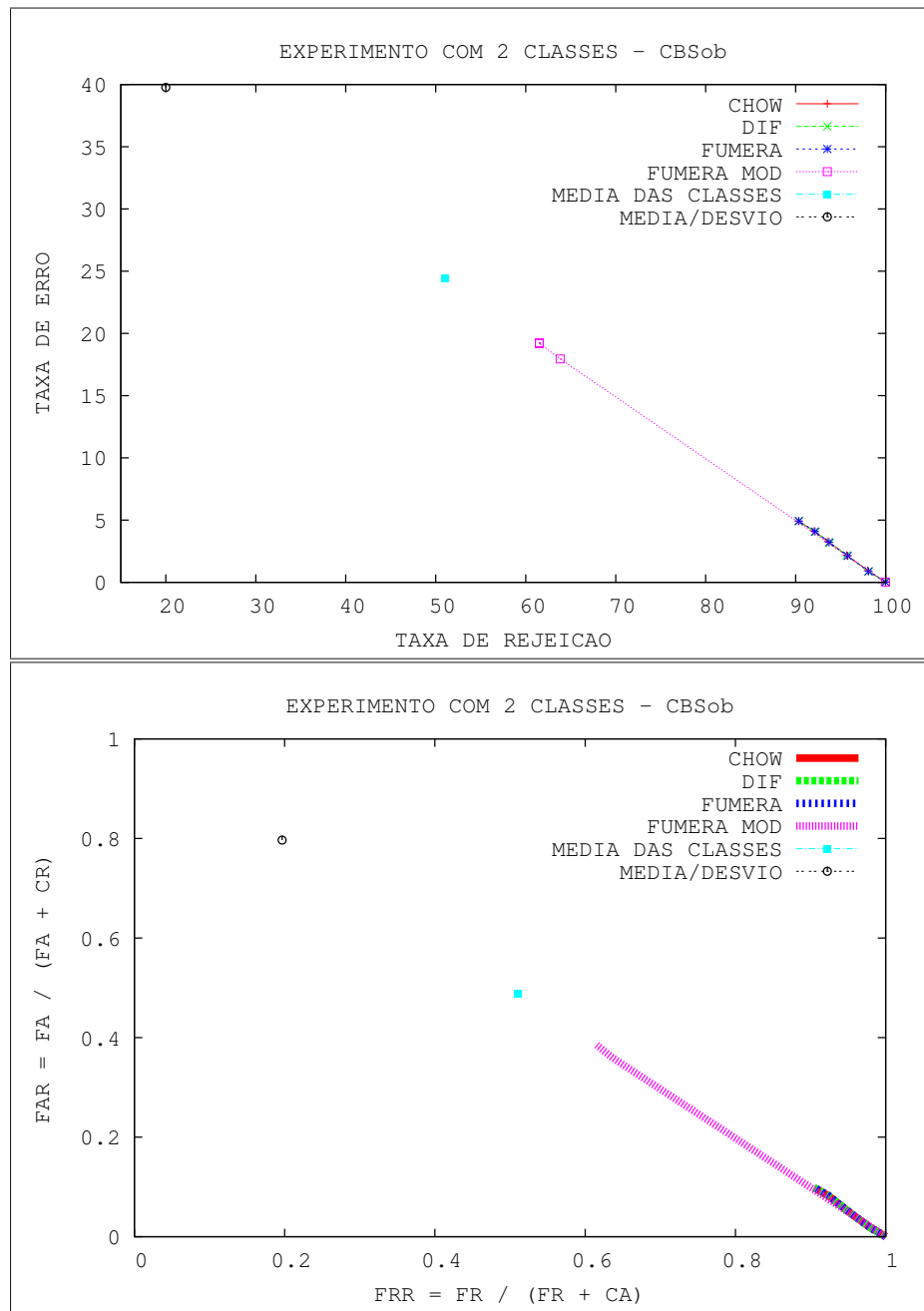


Figura 4.15: Conjunto de dados CBSob - Gráfico Erro/Rejeição e Gráfico ROC para as seis estratégias implementadas.

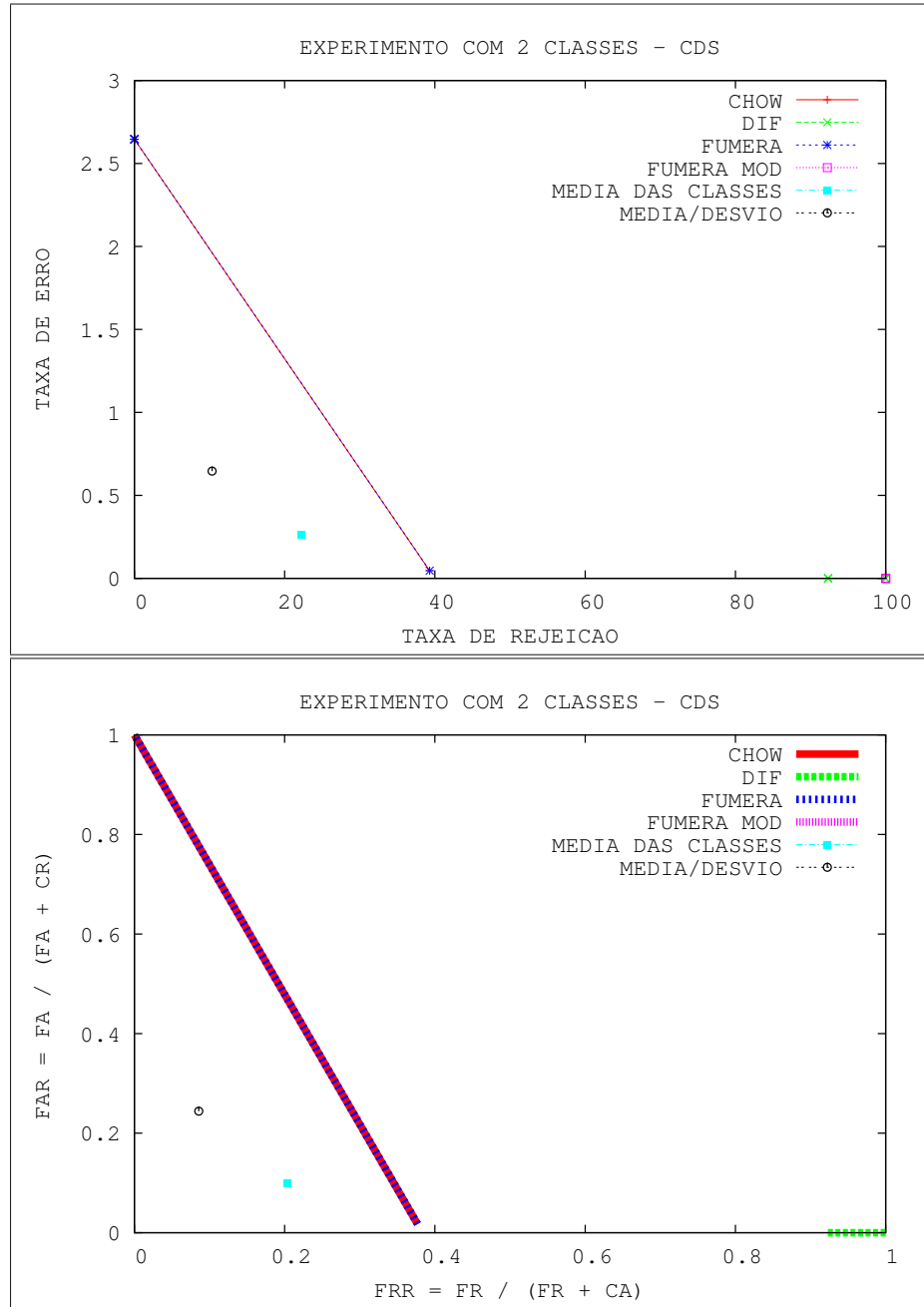


Figura 4.16: Conjunto de dados CDS - Gráfico Erro/Rejeição e Gráfico ROC para as seis estratégias implementadas.

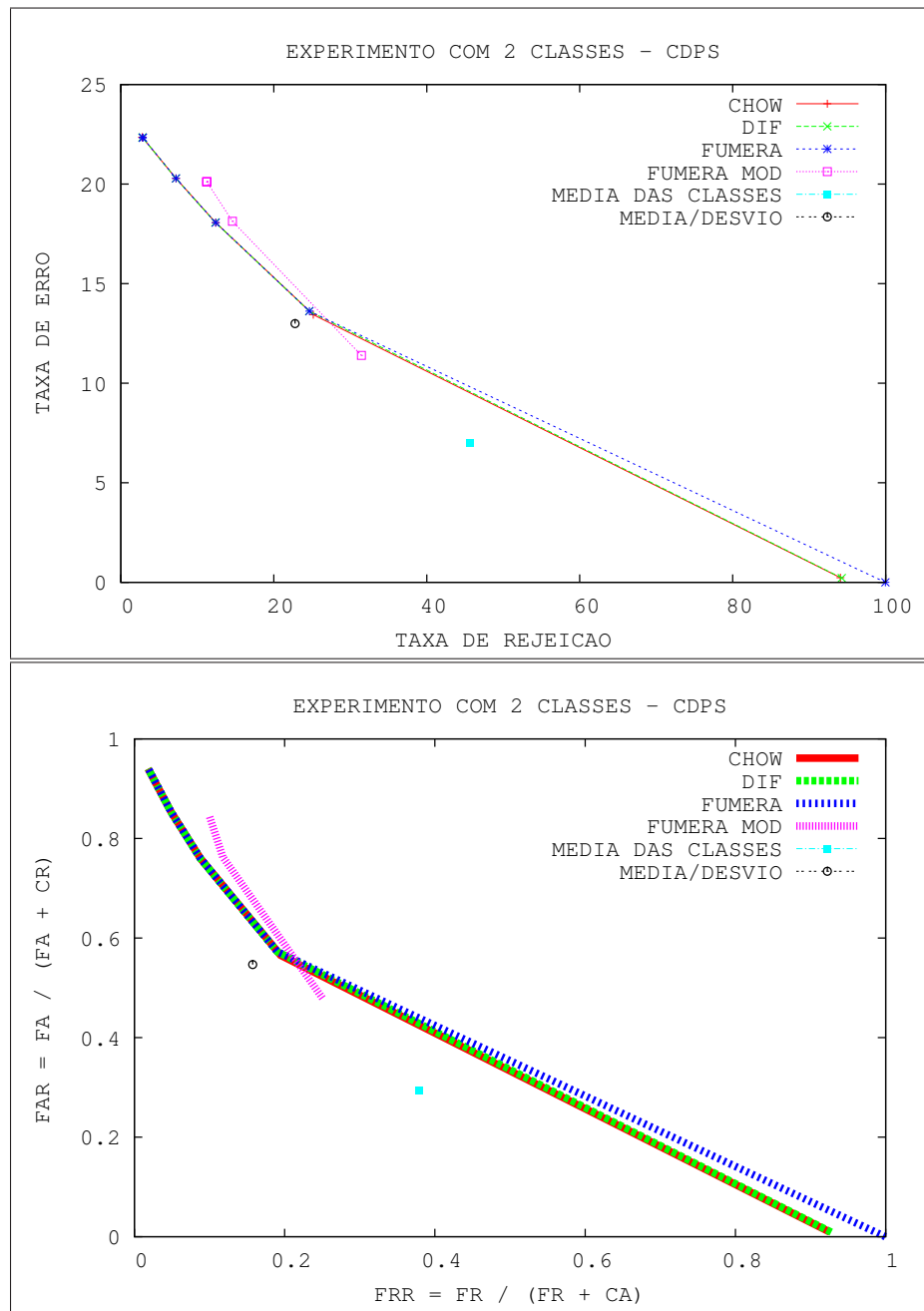


Figura 4.17: Conjunto de dados CDPS - Gráfico Erro/Rejeição e Gráfico ROC para as seis estratégias implementadas.

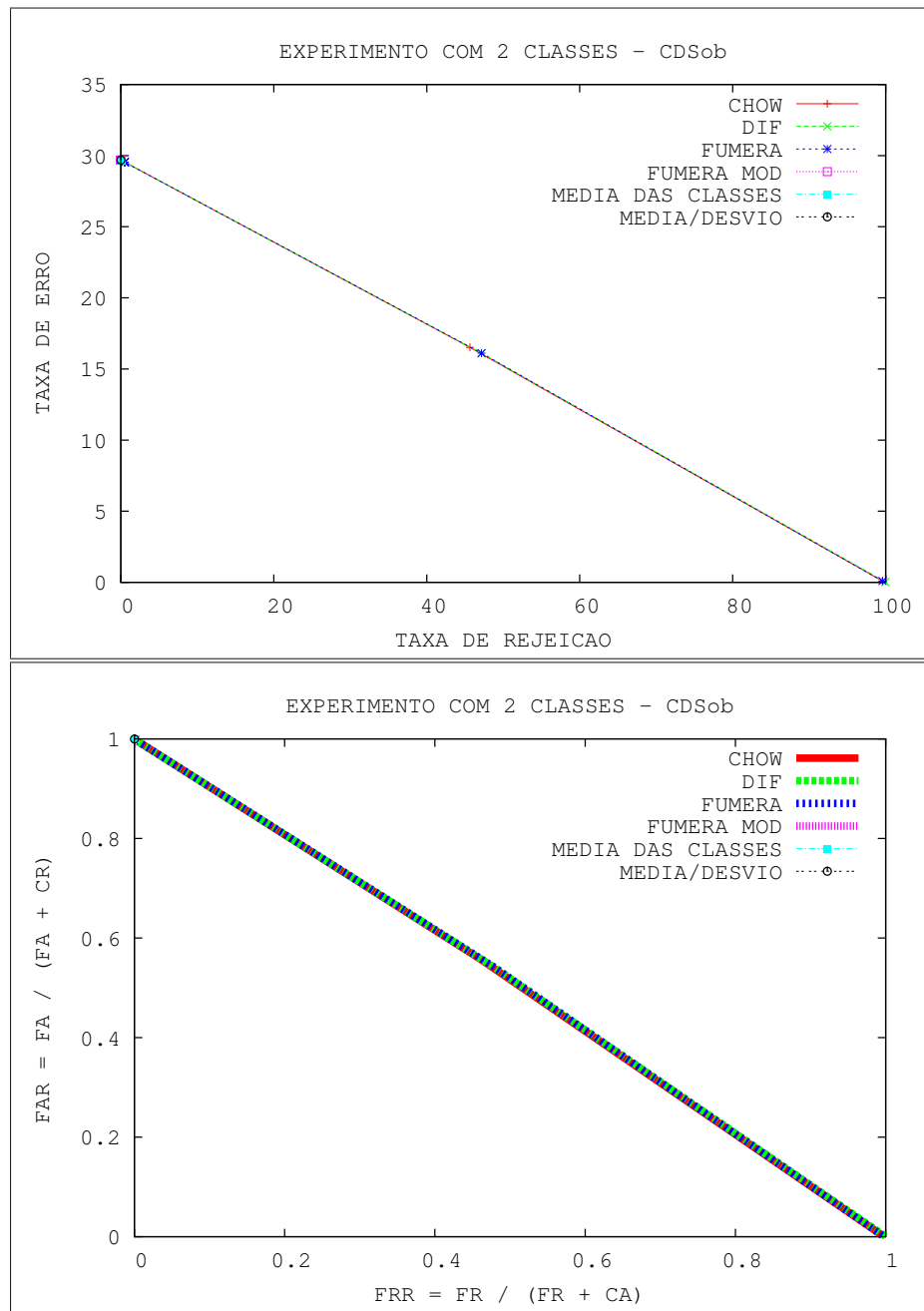


Figura 4.18: Conjunto de dados CDSob - Gráfico Erro/Rejeição e Gráfico ROC para as seis estratégias implementadas.

Tabela 4.1: Resumo do comportamento das estratégias de rejeição em relação a problemas com duas classes gerados através de uma Distribuição Normal.

Conjunto	Compromisso Erro/Rejeição	Análise ROC
NBS	Chow Fumera FumeraMod DIF Média/Desvio Média das Classes	DIF Média/Desvio Média das Classes Chow Fumera FumeraMod
NBPS	FumeraMod Chow DIF Fumera Média/Desvio Média das Classes	Chow DIF Fumera Média/Desvio FumeraMod Média das Classes
NBSoB	FumeraMod Média/Desvio Média das Classes DIF Chow Fumera	FumeraMod Média das Classes DIF Chow Fumera Média/Desvio
NDS	Chow DIF Fumera Média/Desvio Média das Classes FumeraMod	Chow Média/Desvio Média das Classes DIF Fumera FumeraMod
NDPS	Chow Fumera DIF Média/Desvio Média das Classes FumeraMod	Média/Desvio Média das Classes Chow Fumera DIF FumeraMod
NDSoB	Fumera Chow DIF Média/Desvio Média das Classes FumeraMod	Fumera FumeraMod DIF Chow Média das Classes Média/Desvio

Tabela 4.2: Resumo do comportamento das estratégias de rejeição em relação a problemas com duas classes gerados através de uma Distribuição Normal Multivariável.

Conjunto	Compromisso Erro/Rejeição	Análise ROC
NMBS	FumeraMod Chow DIF Fumera Média/Desvio Média das Classes	Chow DIF Fumera FumeraMod Média/Desvio Média das Classes
NMBPS	Chow Fumera FumeraMod DIF Média/Desvio Média das Classes	Média/Desvio Média das Classes FumeraMod Chow DIF Fumera
NMBSoB	Média/Desvio FumeraMod DIF Chow Fumera Média das Classes	FumeraMod Fumera DIF Chow Média das Classes Média/Desvio
NMDS	Média das Classes FumeraMod DIF Chow Fumera Média/Desvio	FumeraMod Chow DIF Fumera Média/Desvio Média das Classes
NMDPS	DIF Chow Fumera Média/Desvio FumeraMod Média das Classes	Média/Desvio FumeraMod Média das Classes Chow DIF Fumera
NMDSoB	Chow DIF Fumera FumeraMod Média/Desvio Média das Classes	Chow DIF Fumera Média das Classes FumeraMod Média/Desvio

Tabela 4.3: Resumo do comportamento das estratégias de rejeição em relação a problemas com duas classes gerados através de uma Distribuição *Chi-square*.

Conjunto	Compromisso Erro/Rejeição	Análise ROC
CBS	DIF Fumera Chow FumeraMod Média/Desvio Média das Classes	Média/Desvio Média das Classes Chow DIF Fumera FumeraMod
CBPS	Média/Desvio Média das Classes Chow DIF Fumera FumeraMod	Média das Classes DIF Fumera Chow FumeraMod Média/Desvio
CBSoB	Média/desvio Média das Classes FumeraMod Chow DIF Fumera	Média da Classes FumeraMod Média/Desvio Chow DIF Fumera
CDS	Chow Fumera Média/Desvio Média da Classes DIF FumeraMod	Média das Classes Media/Desvio Chow Fumera DIF FumeraMod
CDPS	DIF Chow Fumera FumeraMod Média/Desvio Média das Classes	Média das Classes FumeraMod Média/Desvio Chow DIF Fumera
CDSoB	Chow DIF Fumera Média das Classes Média/Desvio FumeraMod	DIF Chow Fumera Média das Classes FumeraMod Média/Desvio

4.2 Experimentos em problemas com múltiplas classes

Para realizar os experimentos com múltiplas classes utilizamos a Base NIST com parâmetros apresentados na seção 3.2. Antes de implementarmos as estratégias de rejeição esses conjuntos foram aplicados em uma rede neural cujas características e resultados foram apresentadas na seção 3.3.2.

4.2.1 Base NIST Caracteres Maiúsculos - (Upper)

O compromisso erro/rejeição para o conjunto de teste Upper é apresentado na Figura 4.19. Esse conjunto quando aplicado ao classificador de base apresenta uma taxa de erro 7,5%. Pretendemos identificar a estratégia de rejeição que consiga prover o melhor compromisso erro/rejeição, ou seja, a menor taxa de erro associada a menor taxa de rejeição. Observando o gráfico, verificamos que as estratégias DIF e Média/Desvio apresentam comportamento similares rejeitando 13% dos exemplos. A primeira apresenta taxa de erro de 1,8%, rejeitando 13,43%, a segunda precisa rejeitar 13,51% dos exemplos para obter uma taxa de erro de 1,7%. As demais estratégias, buscando um erro de até 2%, necessitam rejeitar um número maior de exemplos, como é o caso das estratégias Chow e Fumera que para obter uma taxa de erro de 1,6% precisam rejeitar 15,6% dos exemplos. A estratégia FumeraMod, caso, rejeite 15,79% dos exemplos vai proporcionar uma taxa de erro de 2,54% e a estratégia Média das Classes precisa rejeitar 22,76% dos exemplos para obter uma taxa de erro de 0,86%.

A comparação do desempenho das seis estratégias rejeição utilizadas para o conjunto Upper em termos de taxa de falsa rejeição (FRR) e taxa de falsa aceitação (FAR) pode ser vista na Figura 4.20. A curva ROC em cada estratégia refere-se aos exemplos das 26 classes pertencentes ao conjunto de teste, ou seja 11941 exemplos. A análise visual do gráfico ROC aponta a estratégia mais adequada para este conjunto foi a Média das Classes, provendo uma correta aceitação de 10116 dos exemplos e 686 são corretamente rejeitados. As estratégias Chow, Fumera, Média/Desvio apresentam resultados similares. A verificação das taxas de falsa aceitação e falsa rejeição apresenta para a estratégia Média das Classes valores 11,5% para FAR e 17,4% para FRR, e uma taxa de erro 0,86%. Utilizando a Média/Desvio como estratégia os valores são alterados para FAR 23,5% e FRR 8,4% para uma taxa de erro de 1,76%. As estratégias FumeraMod e DIF apresentam índices elevados de FAR em média 43,6% e índice baixos de FRR em média 2,1%, caracterizando os piores desempenho neste conjunto. Os valores referentes FAR para a estratégia Chow variam entre 0,3% e 44,8% e FRR entre 65,3% e 3,5% . Para Fumera,

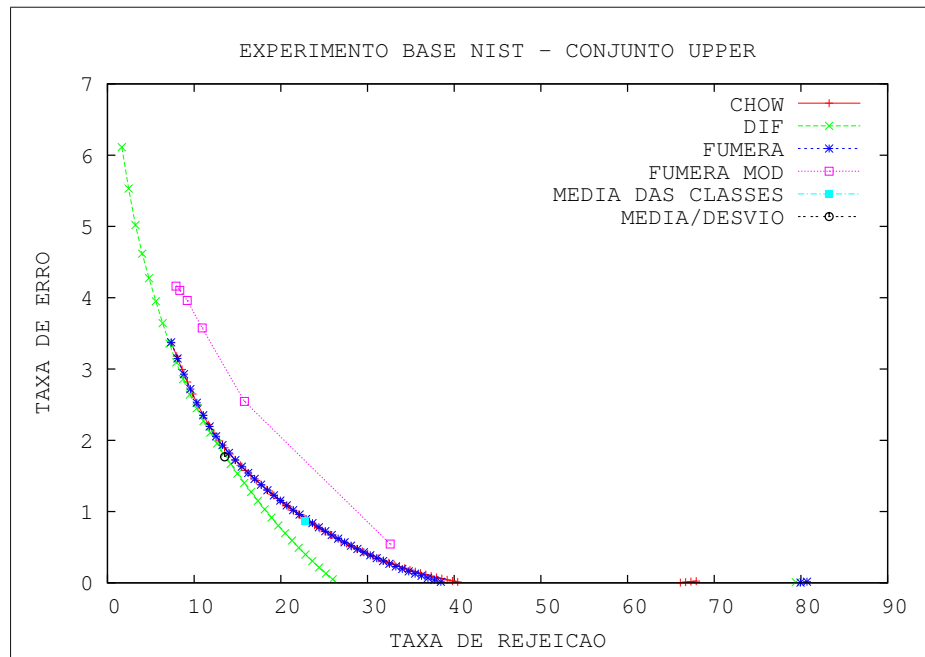


Figura 4.19: Avaliação das estratégias de rejeição utilizando o compromisso erro/rejeição para o conjunto Upper.

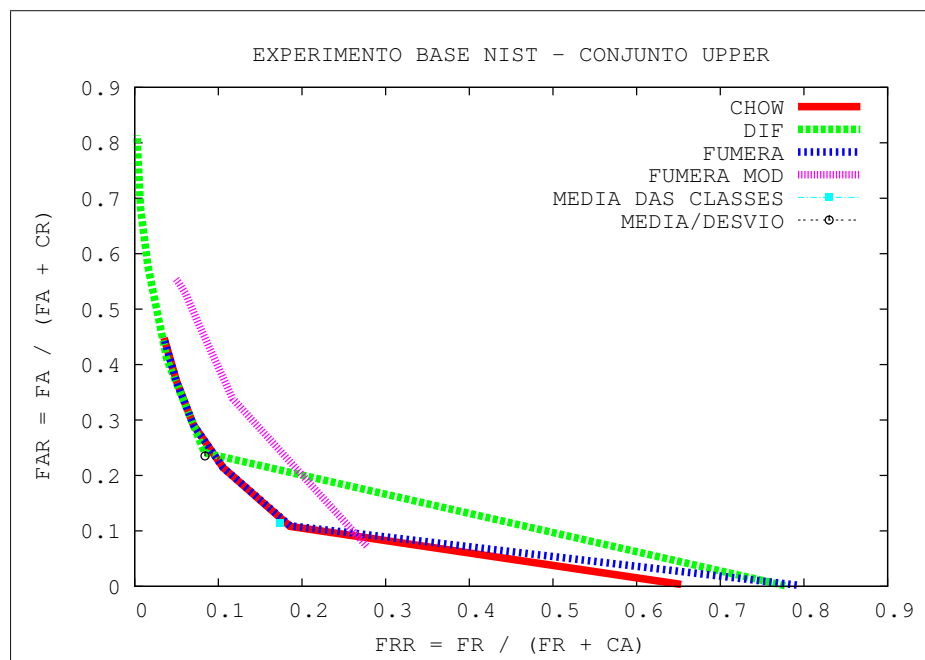


Figura 4.20: Avaliação das estratégias de rejeição utilizando Curva ROC para o conjunto Upper.

FAR varia de 0,2% a 44,9% e FRR entre 79% e 3,5% ambos relacionados a taxa erro entre 0% e 5%.

Resumindo, para este conjunto de dados a estratégia mais adequada é Média das Classes. Um fator importante com relação a essa escolha, além de apresentar o melhor

compromisso, ela também proporciona a obtenção do limiar de forma automática, justificando tal escolha já que outras estratégias como Chow e Fumera propiciam resultados similares. A implantação de qualquer uma das estratégias de rejeição se apresenta de forma adequada para aplicação neste problema, pois, a diferença de desempenho é pequena.

4.2.2 Base NIST Caracteres Minúsculos - (Lower)

A avaliação do compromisso erro/rejeição para o conjunto Lower é apresentada na Figura 4.21. Verificamos que as taxas de erro e rejeição aumentam significativamente em relação ao conjunto Upper. Para este conjunto o classificador de base apresenta taxa de erro de 13,27%. A estratégia escolhida como mais adequada para o conjunto anterior agora para obter um índice similar de rejeição de 16,71% apresenta taxa de erro 5,85%, comportando-se de forma completamente diferente. Neste caso, são obtidos resultados similares pelas estratégias DIF, Chow, Fumera, FumeraMod, dificultando a escolha da estratégia com melhor desempenho. A estratégia Média das Classes precisa rejeitar 28,02% para obter taxa de erro 3,16%. Com a utilização da estratégia Fumera a taxa de rejeição varia entre 16,92% e 69,49% e as taxas de erro de 0,25% até 5,61%. A estratégia FumeraMod apresenta uma variação na taxa de rejeição de 49,39% até 15,40%, porém a taxa de erro atinge 6,35%. A estratégia Chow precisa rejeitar 67,90% dos exemplos para obter uma taxa de erro de 0,25%. A estratégia DIF necessita rejeitar um número elevado de exemplos cerca de 93,88% para conseguir taxa de erro próxima de 0%, porém a análise dos demais valores sugerem esta estratégia como a mais adequada a este problema.

A análise visual do gráfico ROC aponta o desempenho das seis estratégias de rejeição na Figura 4.22, apresentando a estratégia DIF como a mais próxima do ponto (0,0) com FAR variando de 0% até 50,5%. Verificando taxas de falsa aceitação e falsa rejeição observamos que a estratégia Média/Desvio apresenta valores de 44,2% para FAR e 10,7% para FRR, e uma taxa de erro 5,85%. Utilizando a Média das Classes como estratégia os valores são alterados para FAR e 23,9% e FRR 20,7% para uma taxa de erro de 3,16%. Os valores referentes as estratégia Chow e Fumera novamente são similares, ambas apresentam FAR variando entre 1,9% e 42,3% e FRR entre 10,7% e 65,1%. A estratégia FumeraMod apresenta variação de 6,3% e 47,% para FAR e 42,6% e 9,8% para FRR.

A escolha da estratégia DIF como mais adequada é de fácil identificação visual no gráfico ROC, detalhes deste desempenho podem ser resgatados em relação ao índice de correta aceitação. Este índice aponta a correta classificação de 9636 dos 12000 exemplos

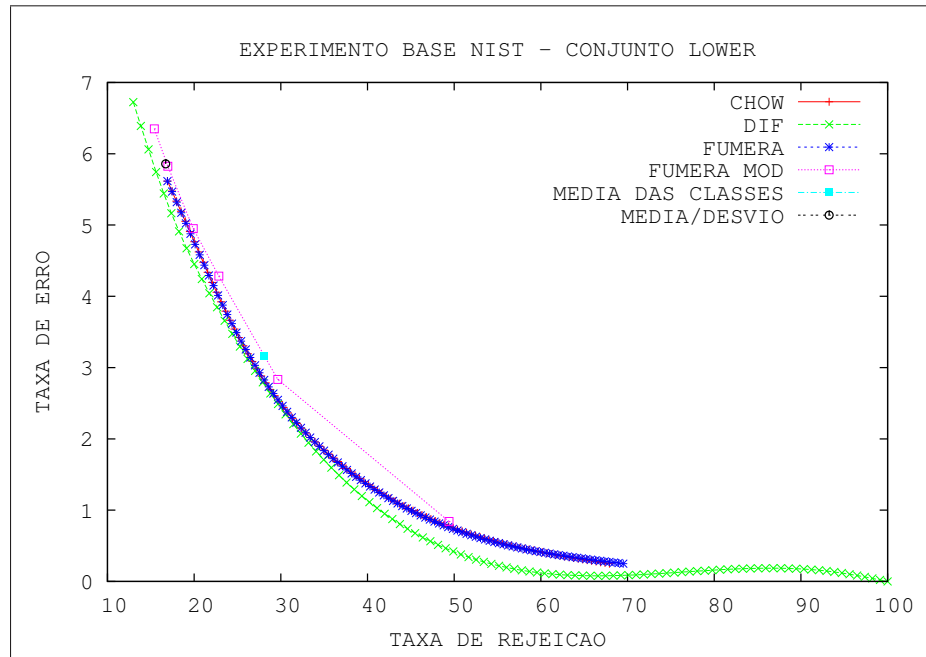


Figura 4.21: Avaliação das estratégias de rejeição utilizando o compromisso erro/rejeição para Base Lower.

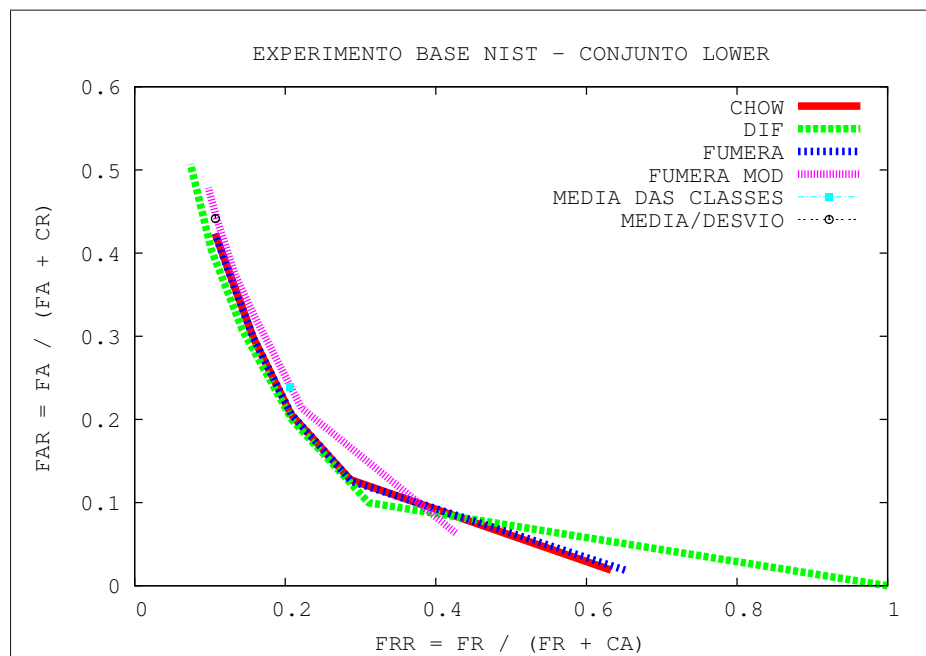


Figura 4.22: Avaliação das estratégias de rejeição utilizando Curva ROC para Base Lower.

pertencentes ao conjunto e a correta rejeição de 807 exemplos, proporcionando FAR de 50,7% e FRR de 7,4% e uma taxa de erro de 6,72% para 12,97 % de taxa de rejeição. Para este conjunto a avaliação em relação a erro/rejeição tornou-se bem mais complexa para as seis estratégias, sendo imprescindível o uso da curva ROC para verificar o desempenho.

4.2.3 Base NIST Caracteres Maiúsculos e Minúsculos - (UpperLower)

A Figura 4.23 apresenta o compromisso erro/rejeição para as estratégias implementadas para o conjunto UpperLower com 52 classes. Neste problema a complexidade da classificação é significativamente maior que nos anteriores, a taxa de erro apresentada pelo classificador de base é 30,08%. A busca por uma taxa de erro próxima de 5% neste conjunto implica em taxas de rejeições extremamente altas, em média 50% do conjunto precisa ser rejeitado. Em um problema como esse é necessário estabelecer os critérios desejados, por exemplo, caso o custo de uma classificação incorreta seja o critério mais importante seria viável trabalhar com taxas de rejeição elevadas. Caso o custo de uma classificação incorreta não proporcione maiores danos, podemos trabalhar com taxas de erro mais elevadas, porém, ainda assim menores que a apresentada pelo classificador de base. Entretanto, para nossa análise definimos para todos os conjuntos de dados utilizados valores fixos de taxa de erro, afim de verificar o comportamento das estratégias de rejeição.

A estratégia Média/Desvio para conseguir taxa de erro 16,60% precisa rejeitar 27,57% dos exemplos pertencentes ao conjunto apontando o melhor desempenho neste conjunto. A estratégia Média das Classes apresenta taxa de erro igual ao classificador de base. Com a utilização da estratégia Fumera a taxa de rejeição varia entre 89,18% e 49,04% e as taxas de erro de 0,2% até 6,88%. A estratégia FumeraMod apresenta uma variação na taxa de rejeição de 85,13% até 31,58%, e taxa de erro entre 0,5% e 14,05%. A estratégia DIF também necessita rejeitar um número elevado de exemplos, cerca de 50% para conseguir taxa de erro próxima de 5%, entretanto, parece ser a mais adequada na avaliação geral.

A comparação do desempenho das seis estratégias de rejeição em termos de taxa de falsa rejeição (FRR) e taxa de falsa aceitação (FAR) para o conjunto UpperLower pode ser vista na Figura 4.24. A análise visual do gráfico ROC aponta a estratégia DIF com melhor desempenho e a estratégia Média/Desvio com o pior desempenho, concordando com os resultados encontrados na avaliação do conjunto anterior. Depois da estratégias DIF os melhores desempenhos são apontados por Chow e Fumera seguidos de FumeraMod e Média das Classes apresentando resultados similares. A verificação das taxas de falsa

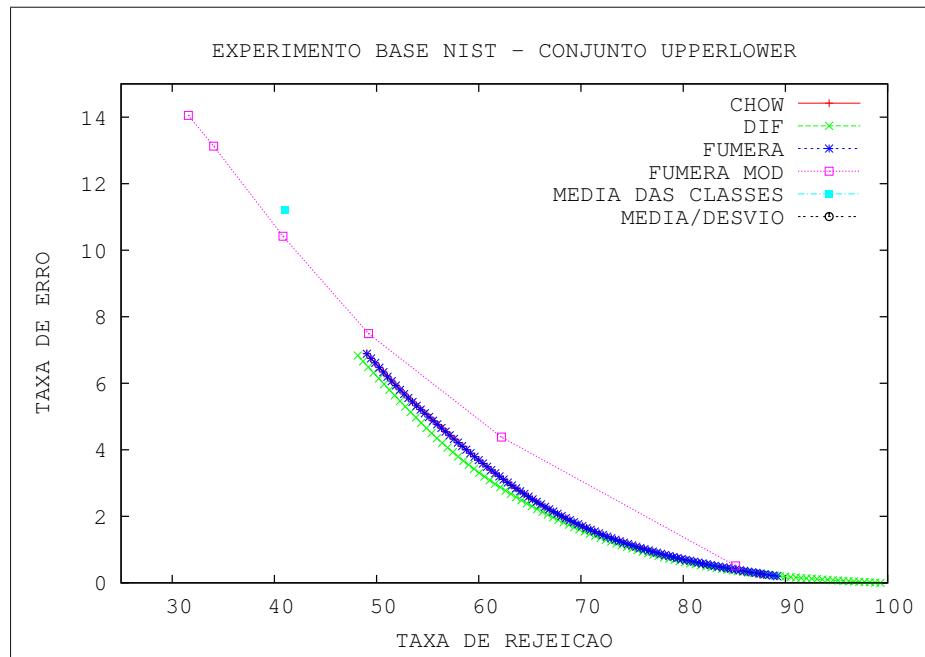


Figura 4.23: Avaliação das estratégias de rejeição utilizando o compromisso erro/rejeição para Base UpperLower.

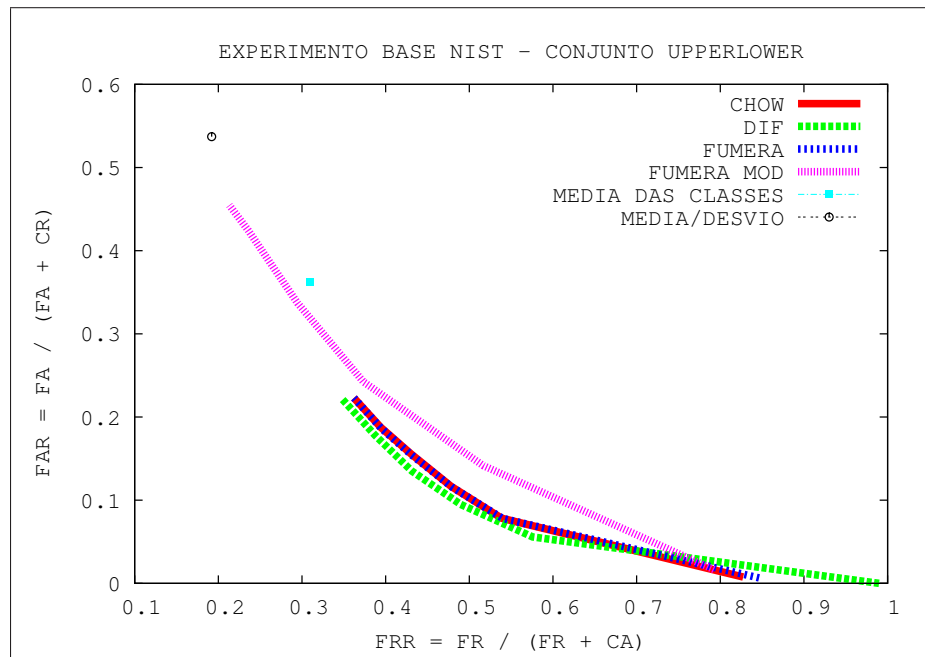


Figura 4.24: Avaliação das estratégias de rejeição utilizando Curva ROC para Base UpperLower.

aceitação e falsa rejeição apontam a estratégia Média/Desvio apresentando 55,82% para FAR e 27,57% para FRR, e uma taxa de erro 16,60%, ou seja o pior desempenho para o conjunto. Utilizando a Média das Classes como estratégia os valores são alterados para FAR e 36,2% e FRR 30,9% para uma taxa de erro de 30,91%. A estratégia FumeraMod

apresenta FAR entre 1,6% e 45,5% e FRR entre 79,2 % e 21,3%. Os valores referentes FAR para a estratégia Chow variam entre 0,8% e 22,3% e FRR entre 82,7% e 36,2%. Para Fumera, FAR varia de 0.6% a 22,3% e FRR entre 84,6% e 32,2% ambos variando a taxa erro entre 0% e 5%.

4.2.4 Análise dos resultados utilizando múltiplas classes

Os gráficos erro/rejeição evidenciam na maioria dos casos a similaridade entre as estratégias e a complexidade na escolha de qual estratégia apresentou melhor desempenho. Visualizar o pior e melhor compromisso erro/rejeição quando se avalia apenas um modelo é uma tarefa fácil, mas tende a ficar mais complexa à medida que outros modelos são adicionados. O uso da análise visual das curvas ROC, também apresentou-se complexa em vários casos de sobreposição. Por isso o processo de análise não foi fundamentado somente em análise visual, mas também nos demais valores que geram os gráficos. O único conjunto que apresentou-se neutro ao método de avaliação escolhido foi o Lower apresentando resultados equivalentes para análise ROC e a avaliação do compromisso erro/rejeição.

A regra heurística representada pela estratégia DIF obteve desempenho na maioria dos casos melhor ou igual a estratégias estatísticas ditas como "ótimas" na literatura, sendo o caso do Chow e Fumera. A idéia de encontrar os limiares ótimos locais em cada classe representada pela estratégia FumeraMod apresentou na maioria dos casos como uma alternativa inadequada, sugerindo que nem sempre um ótimo local é também um ótimo global, pois, outros métodos com menores custos de implementação apresentam resultados mais adequados. As demais regras heurísticas representadas pelas estratégias Média das Classes e Média/Desvio também apresentam resultados viáveis para aplicação em um sistema de reconhecimento, principalmente pelo fato do processo de obtenção dos limiares para cada classe apresenta-se de forma simples e automática. Entretanto, parecem ser as estratégias que mais são influenciadas pelo formação do conjunto de dados. A Tabela 4.4 apresenta as estratégias ordenadas de acordo com a distância em relação ao ponto (0,0) dos seus respectivos gráficos em conjuntos de dados com múltiplas classes. Analisando a tabela percebemos que a análise ROC aponta a mesma ordem de desempenho para as estratégias implementadas para os conjuntos Lower e UpperLower. Apontando a seguinte seqüência: DIF com o melhor desempenho, Chow e Fumera com resultados equivalentes, seguidos de FumeraMod, Média da Classes e Média/Desvio. Nestes conjuntos a dificuldade do classificador é evidente nas taxas sugeridas pela RNA-MLP.

Percebemos que a estratégia DIF destacou-se principalmente trabalhando com pro-

blemas de classificação mais complexos. Quando esta estratégia foi utilizada por Pitrelli e Perrone (2002) seu desempenho em relação as demais não se destacou. Neste caso, eles aproveitam a flexibilidade de uma tarefa de pós-processamento e realizaram teste com oito diferentes estratégias relacionadas as medidas de confiança oferecidas pelo classificador.

Acredita-se que o pior desempenho atribuído à estratégia Média/Desvio deve-se ao fato desta estratégia ser diretamente indexada pelo desvio padrão influenciado pelas características dos conjuntos Lower e UpperLower. O classificador considera o conjunto UpperLower o de mais difícil classificação e o conjunto Upper sua tarefa mais simples. O conjunto Lower é considerado um problema de complexidade intermediário entre os demais. De forma geral, todas as estratégias melhoraram significativamente o desempenho do sistema de reconhecimento, pois, as taxas de erros diminuíram e as taxas de acerto aumentaram.

Tabela 4.4: Resumo do comportamento das estratégias de rejeição em relação a problemas multiclases.

Conjunto	Compromisso Erro/Rejeição	Análise ROC
Upper	DIF Fumera Chow FumeraMod Média/Desvio Média das Classes	Média das Classes Chow Fumera Média/Desvio DIF FumeraMod
Lower	DIF Chow ou Fumera FumeraMod Média das Classes Média/Desvio	DIF Chow ou Fumera FumeraMod Média das Classes Média/Desvio
UpperLower	Média/Desvio FumeraMod Média das Classes DIF Chow ou Fumera	DIF Chow ou Fumera FumeraMod Média das Classes Média/Desvio

Capítulo 5

Conclusão

Este Capítulo apresenta conclusões obtidas verificando em paralelo a relação com os objetivos iniciais propostos. Para finalizar, a última Seção discute possíveis trabalhos futuros.

5.1 Conclusões

A partir dos resultados obtidos e apresentados nas seções anteriores foi possível avaliar o desempenho e as diferenças entre as estratégias de rejeição quando aplicadas a conjuntos de dados diferentes. Analisando os resultados podemos concluir que o uso das estratégias de rejeição conseguiram produzir sistemas classificadores mais confiáveis, além de melhorar significativamente os resultados medidos em relação a taxa de erro apresentada pelo classificador de base. Depois de testados experimentalmente as implementações dessas estratégias em duas bases de dados sintéticos e dados reais identificamos:

- Quanto à hipótese de que o balanceamento, ou não, dos dados pode influenciar o comportamento das estratégias de rejeição, percebemos que de fato existe essa influência, mais acentuada com relação a dados desbalanceados. Entretanto, essa característica não é muito expressiva com relação ao comportamento das estratégias de rejeição.
- Quanto à hipótese de que a separação dos dados pode influenciar o comportamento das estratégias de rejeição, percebemos claramente que essa característica é influente com relação ao desempenho da estratégia de rejeição. Podemos verificar essa situação em todos os conjuntos através dos resultados e os gráficos apresentados no Capítulo 4.
- Quanto à utilização de conjuntos de dados pré-definidos com características diferen-

tes: número de amostras por classe, distribuição de dados e quantidade de classes, observamos que os resultados em raros casos apresentam comportamento similar, evidenciando a existência da influência das características no desempenho das estratégias de rejeição.

- O comportamento das estratégias de rejeição para os conjuntos de dados com duas classes ou múltiplas classes é completamente influenciado pelas características dos conjuntos de dados utilizados.

O resultados mostram claramente para os experimentos utilizando múltiplas classes, ou duas classes que estratégias clássicas na literatura como Chow e Fumera obtiveram resultados similares, em muitos casos até inferiores aos métodos heurísticos e estatísticos implementados. Por exemplo, para os conjuntos com múltiplas classes a estratégia heurística DIF apresenta os melhores desempenhos para a maioria dos casos. Para os conjuntos com duas classes a estratégia de destaque e a FumeraMod.

Enfatiza-se, porém que as porcentagens de erro obtidas no conjunto de teste podem ainda ser ajustadas para as mesmas fixadas no conjunto de validação, estabelecendo novos critérios de obtenção dos limiares para cada estratégia. Entretanto, tais ajustes não foram realizados respeitando o objetivo principal deste trabalho o estudo de como as características do problema podem impactar nos resultados apresentados pelas diferentes estratégias de rejeição.

A apresentação dos experimentos realizados pode ser visualizada de duas formas: o compromisso entre as taxas de erro e rejeição, através dos gráfico erro/rejeição e o compromisso entre a falsa aceitação e a falsa rejeição através da análise ROC a utilização de dois métodos foi valida devido ao fato de ambos expressarem informações diferentes em relação ao mesmo conjunto, buscando encontrar indícios para resgatar exemplos que poderiam ser reconhecidos de forma incorreta pelo mecanismo de rejeição, mas que na verdade podem ser corretamente reconhecidos. Todas essas considerações são validas, pois, estratégias de rejeição bem elaboradas podem ajudar a resolver muitos problemas reais considerados complexos.

Para finalizar, considerando as estratégias estudadas e uma suposta necessidade de escolha entre alguma, para aplicação em um problema real. Em ambos os casos, duas classes ou múltiplas classes, a estratégia que apresentou-se como uma solução viável é a estratégia DIF, pois, obteve desempenho na maioria dos casos melhor ou igual a estratégias estatísticas ditas “ótimas” na literatura; trata-se uma estratégia pouco influenciada pelas condições do conjunto de dados; e ainda sua implementação não possui grau de dificuldade elevado.

5.2 Trabalhos Futuros

Durante o desenvolvimento desta dissertação, não se teve a oportunidade de testar outras estratégias de rejeição, devido ao espaço de tempo para a conclusão desta. Aqui delineamos em traços gerais outras verificações do comportamento das estratégias de rejeição que acredita-se ser merecedora de investigação, por exemplo, manipular os limiares utilizando informação contextual e heurística baseada na matriz de confusão (*Contextual Rejection Approach*). Poderíamos proceder da seguinte forma: Utilizando classificador de base previamente treinado e o conjunto de validação, gerar a matriz de confusão e em seguida analisar e definir as confusões mais críticas entre as classes e a partir desta análise estabelecer:

- Critérios de rejeição mais rígidos para classes que tendem a apresentar maior confusão;
- Critérios de rejeição mais folgados para classes que não tendem a apresentar ambigüidade ou confusão.
- Caso houverem probabilidades similares, ou seja, não haver uma probabilidade dominante para a classe, verificar quais as classes associadas as probabilidades possuem o primeiro e o segundo maior valor.
- Realizar uma pré-análise da matriz de confusão sobre a base de validação ou mesmo de treinamento para identificar classes ambíguas e definir novos critérios.

Referências Bibliográficas

CHOW, C. K. On optimum recognition error and reject tradeoff. *In IEEE Transactions in Information Thoery*, v. 16, n. 1, p. 41–46, 1970.

DUDA, R.; HART, P.; STORK, D. *Pattern Classification*. New York: John Wiley & Sons, 2000.

DUIN, R. et al. *PRTools4, A Matlab Toolbox for Pattern Recognition*. 2004. Delft University of Technology.

FAWCETT, T. An introduction to ROC analysis. *Pattern Recognition Letters*, v. 227, n. 8, p. 861–874, 2006.

FAWCETT, T.; PROVOST, F. Adaptive fraud detection. *Data Min. Knowl. Discov.*, v. 1, n. 3, p. 291–316, 1997.

FUMERA, G.; ROLI, F.; GIACINTO, G. Reject option with multiple thresholds. *Pattern Recognition Letters*, v. 33, n. 12, p. 2099–2101, 2000.

GORSKI, N. Optimizing error-reject trade off in recognition systems. In: *Proc. 4th International Conference Document Analysis and Recognition*. Ulm, Germany: IEEE Computer Society, 1997. p. 1092–1096.

HA, T. *An optimum class-selective rejection rule for pattern recognition*. 1996. P. 75-80.

HAND, D. J.; TILL, R. J. A simple generalisation of the area under the roc curve for multiple class classification problems. *Machine Learning*, v. 45, n. 2, p. 171–186, 2001.

JAIN, A.; DUIN, R. P. W.; MAO, J. Statistical pattern recognition: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 22, n. 1, p. 4–37, 2000.

KOERICH, A. L. *Unconstrained Handwritten Character Recognition Using Different Classification Strategies*. 2003. In Proc. of the IAPR TC3 International Workshop on Artificial Neural Networks in Pattern Recognition, Firenze, Italy.

KOERICH, A. L. *Rejection Strategies for Handwritten Word Recognition*. 2004. IWFHR, pp. 479-484, Ninth International Workshop on Frontiers in Handwriting Recognition (IWFHR'04).

MARUKATAT, S. et al. *Rejection measures for handwriting sentence recognition*. 2002. In Proc. 8th International Workshop on Frontiers in Handwriting Recognition, pages 24-29, Niagara-on-the-Lake, Canada.

MASSAD, E. et al. *Métodos quantitativos em medicina*. Barueri: Manole, 2004.

MITCHELL, T. *Machine Learning*. New York: McGraw-Hill, 1997.

MOUCHERE, H.; ANQUETIL, E. A unified strategy to deal with different natures of reject. *18th International Conference on Pattern Recognition (ICPR 2006), 20-24 August 2006, Hong Kong, China*, IEEE Computer Society, p. 792–795, 2006.

PITRELLI, J. F.; PERRONE, M. P. Confidence modeling for verification post-processing for handwriting recognition. In: *In Proc. 8th International Workshop on Frontiers in Handwriting Recognition*. Niagara-on-the-Lake, Canada, 2002.: IEEE Computer Society, 2002. p. 30–35.

PROVOST, F.; FAWCETT, T. Analysis and visualization of classifier performance: Comparison under imprecise class and cost distributions. In: *Knowledge Discovery and Data Mining*. Huntington Beach, CA: American Association for Artificial Intelligence, 1997. p. 43–48.

PUDIL, P. et al. Multistage pattern recognition with reject option. In: *Proc. 11th Int. Conf. Pattern Recognition, The Hague, The Netherlands, IEEE Computer Society Press*. The Hague, Netherlands: IEEE Computer Society Press, 1992. v. 2, p. 92–95.

RICHARD, M.; LIPPMANN, R. Neural network classifiers estimate bayesian a posteriori probabilities. *Neural Computation*, v. 3, n. 4, p. 461–483, 1991.

RUSSEL, S. J.; NORVIG, P. *Artificial Intelligence: A Modern Approach*. New Jersey: New Jersey, 2003.

SCHURMANN, J. *Pattern Classification: A Unified View of Statistical and Neural Approaches*. New York: JohnWiley and Sons, 1996.

ZHANG, G. P. Neural networks for classification: a survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, v. 30, n. 4, p. 451–462, 2000.

ZIMMERMANN, M.; BERTOLAMI, R.; BUNKE, H. Rejection strategies for offline handwritten sentence recognition. In: *ICPR '04: Proceedings of the Pattern Recognition, 17th International Conference on (ICPR'04)*. Washington, DC, USA: IEEE Computer Society, 2004. v. 2, p. 550–553.

Apêndice A

Experimentos com duas classes

Este Apêndice apresenta detalhes da realização dos experimentos em dados sintéticos. No início do trabalho pretendia-se utilizar conjuntos de dados de diferentes tamanhos, por exemplo, 20, 50, 100, 200, 500, 5.000, 10.000 em cada classe, variando o tamanho do conjunto, tanto em classes balanceadas ou desbalanceadas. Entretanto, no decorrer do trabalho percebeu-se que tal variação não influenciava nos resultados obtidos. Então, optou-se pela utilização de conjunto balanceados possuindo 20.000 amostras, sendo 10.000 para a classe w_1 e outras 10.000 para a w_2 . Os conjuntos de dados desbalanceados também possuem 20.000 amostras sendo 16.000 para a classe w_1 e outras 4.000 para w_2 . A Tabela A.1 apresenta os valores utilizados para gerar os conjuntos de dados.

A.1 Distribuição Normal

A distribuição Normal é uma das distribuições fundamentais da teoria estatística. Sua maior vantagem é facilidade de definição com apenas dois parâmetros: (μ, σ) . Para esta distribuição apresentamos as seis combinações que geram seis diferentes conjuntos de dados. A Figura A.1 apresenta a separação dos dados para o experimento NBS. A Figura A.2 apresenta a facilidade do classificador em trabalhar com dados balanceados e separados. A Figura A.3 utiliza um conjunto do mesmo tamanho para o experimento NBPS, porém, os valores de média e variância são alterados de forma proporcionar uma sobreposição pequena entre os dados. A Figura A.4 apresenta superfície de separação traçada pelo classificador para o experimento NBPS. A separação dos dados sobreposta utilizada no experimento NBSob pode ser visualizada na Figura A.5 e a respectiva superfície de separação realizada pelo classificador na Figura A.6, percebe-se que a complexidade de classificação, neste caso, é significativamente alterada. A Figura A.7 apresenta um problema com dados desbalanceados separados, experimento NDS. A superfície de

Tabela A.1: Os conjuntos são construídos alterando os parâmetros de média e variância na Distribuição Normal e Normal Multivariável e graus de liberdade na distribuição *Chi-square* pré-definindo suas características de apresentação.

Experimento	Valores	Experimento	Valores
NBS	w_1 (3,1)	NMDS	w_1 (3,1,1)
	w_2 (8,1)		w_2 (8,1,1)
NBPS	w_1 (5,2)	NMDPS	w_1 (5,2,1)
	w_2 (9,2)		w_2 (9,2,1)
NBSOB	w_1 (7,3)	NMDSOB	w_1 (7, 3, 1)
	w_2 (9,3)		w_2 (9,3,1)
NDS	w_1 (3,1)	CBS	w_1 (8)
	w_2 (8,1)		w_2 (1)
NDPS	w_1 (5,2)	CBPS	w_1 (8)
	w_2 (9,2)		w_2 (5)
NDSOB	w_1 (7,3)	CBSOB	w_1 (8)
	w_2 (9,3)		w_2 (8)
NMBS	w_1 (3,1,1)	CDS	w_1 (8)
	w_2 (8,1,1)		w_2 (1)
NMBPS	w_1 (5,2,1)	CDPS	w_1 (8)
	w_2 (9,2,1)		w_2 (5)
NMBSOB	w_1 (7,3,1)	CDSOB	w_1 (8)
	w_2 (9,3,1)		w_2 (8)

separação proposta pelo classificador no conjunto de validação é apresentada na Figura A.8, deixando clara a desproporção de instâncias entre as classes. A Figura A.9 superfície de separação em relação ao conjunto de teste. A Figura A.10 apresenta o experimento NDPS. O comportamento do classificador para o conjunto de teste pode visto na Figura A.11. Para finalizar os experimentos com a Distribuição Normal de dados balanceados e desbalanceados apresentados. Na Figura A.12 visualizamos o conjunto NDSob e na Figura A.13 a superfície de separação obtida pelo classificador, neste caso, não sendo uma tarefa trivial.

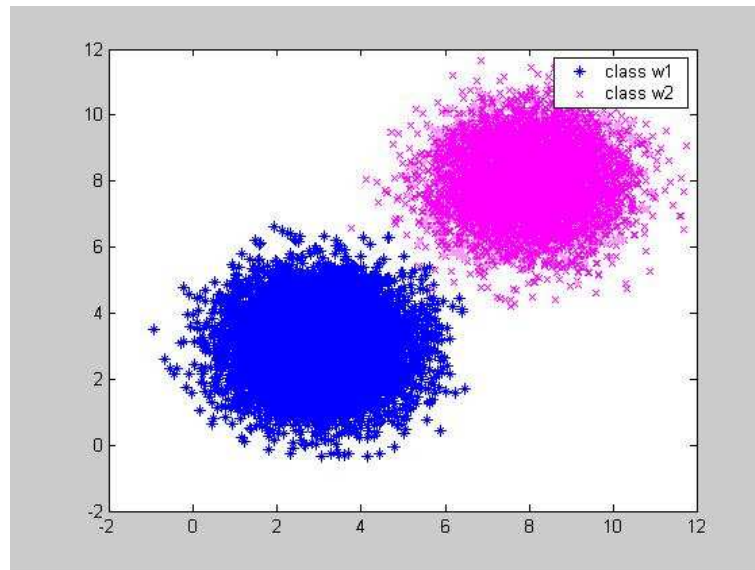


Figura A.1: Separação dos dados para uma Distribuição Normal - NBS.

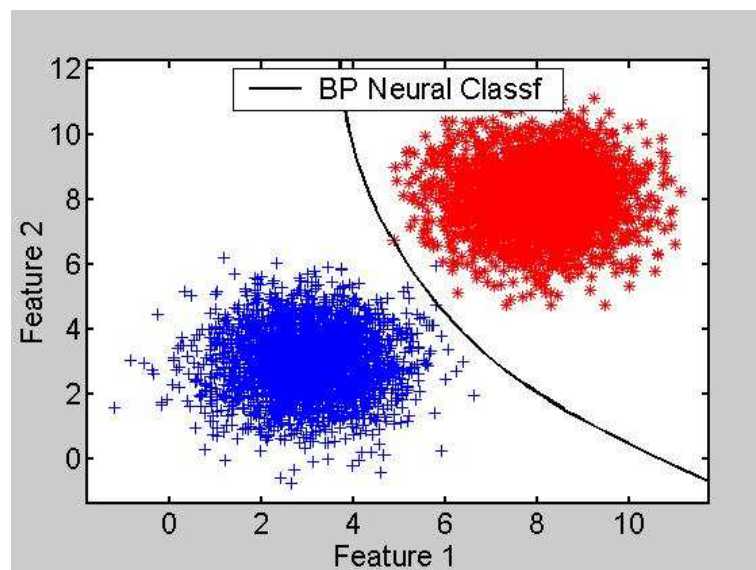


Figura A.2: Separação dos dados para o conjunto de teste e respectiva superfície de separação traçada pela RNA-MLP para dados NBS.

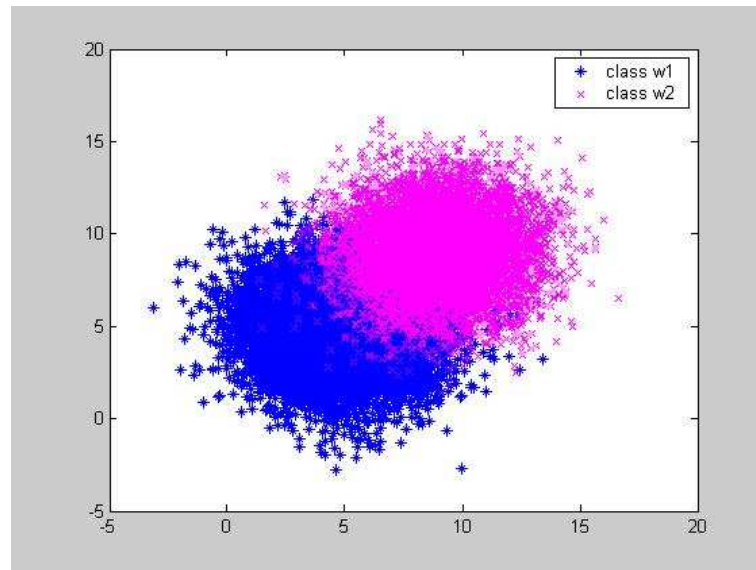


Figura A.3: Separação dos dados para uma Distribuição Normal - NBPS.

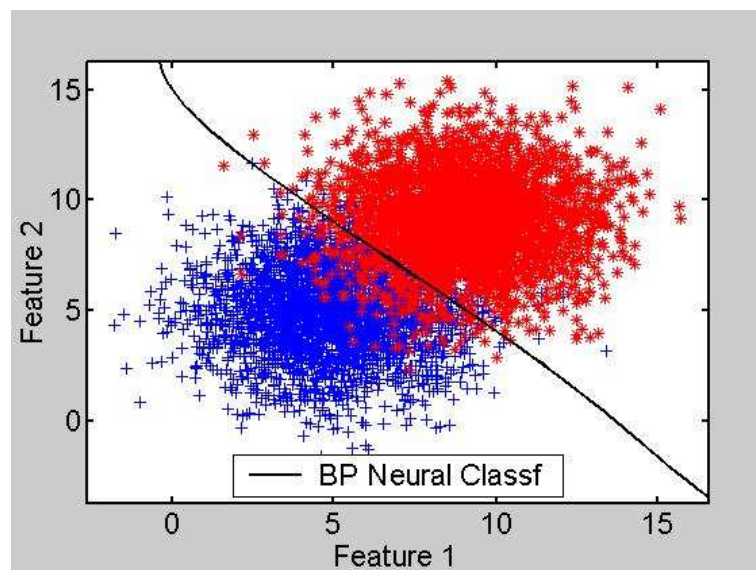


Figura A.4: Separação dos dados para o conjunto de teste e respectiva superfície de separação traçada pela RNA-MLP para dados NBPS.

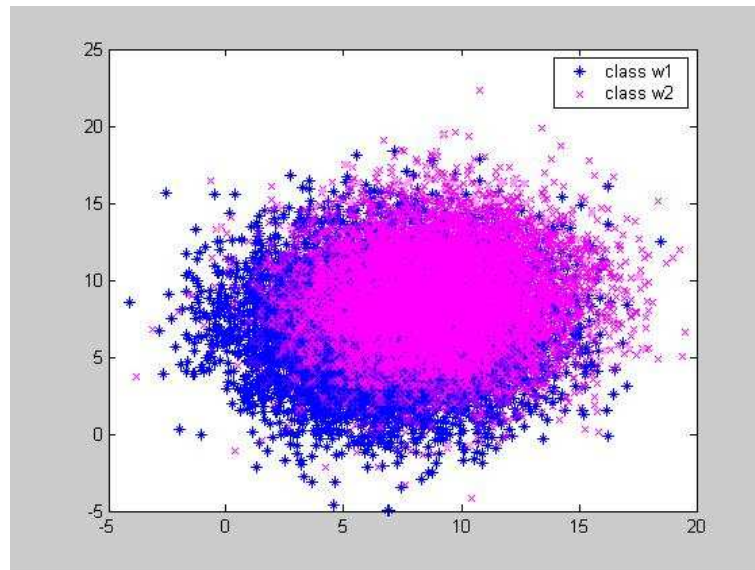


Figura A.5: Separação dos dados para uma Distribuição Normal - NBSob.

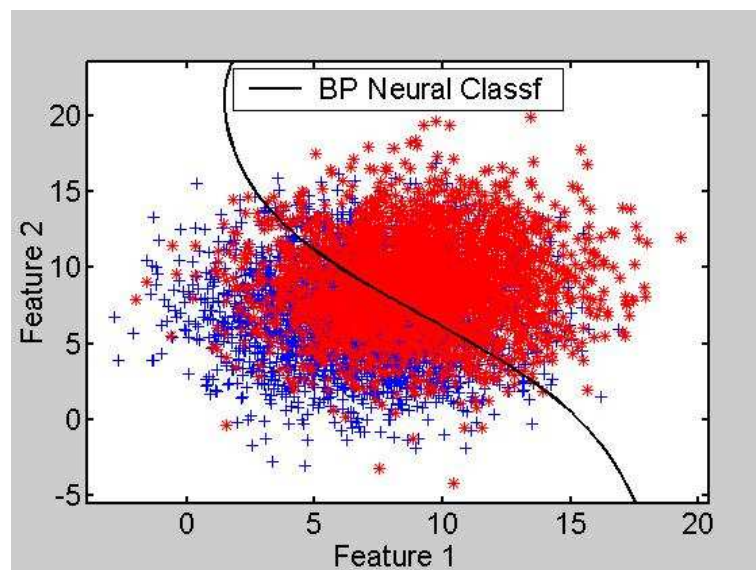


Figura A.6: Separação dos dados para o conjunto de teste e respectiva superfície de separação traçada pela RNA-MLP para dados NBSob.

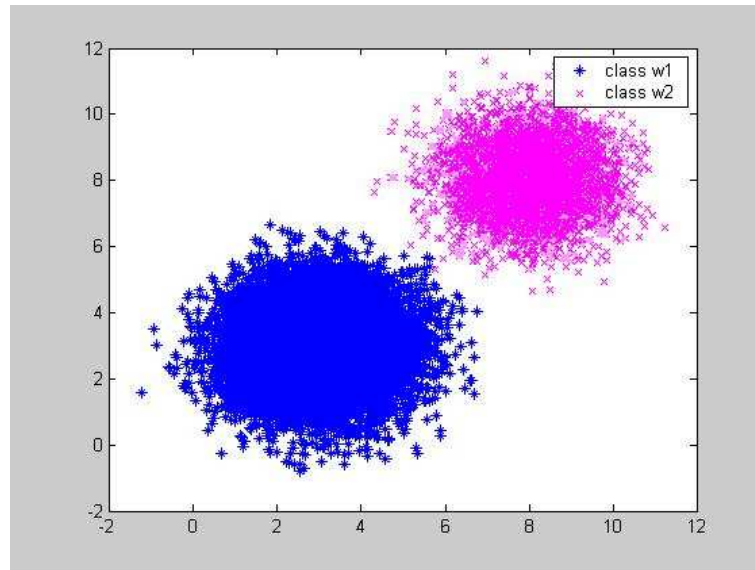


Figura A.7: Separação dos dados para uma Distribuição Normal - Desbalanceados - Separados - NDS.

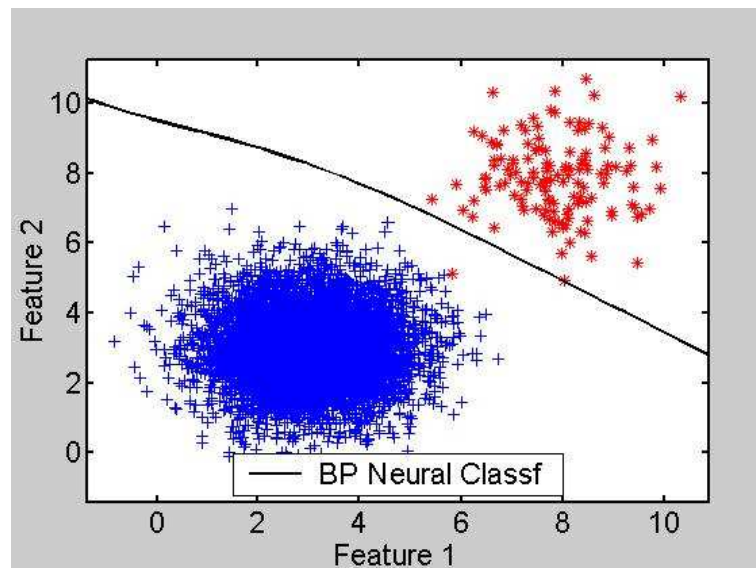


Figura A.8: Separação dos dados para o conjunto de validação e respectiva superfície de separação traçada pela RNA-MLP para o experimento NDS.

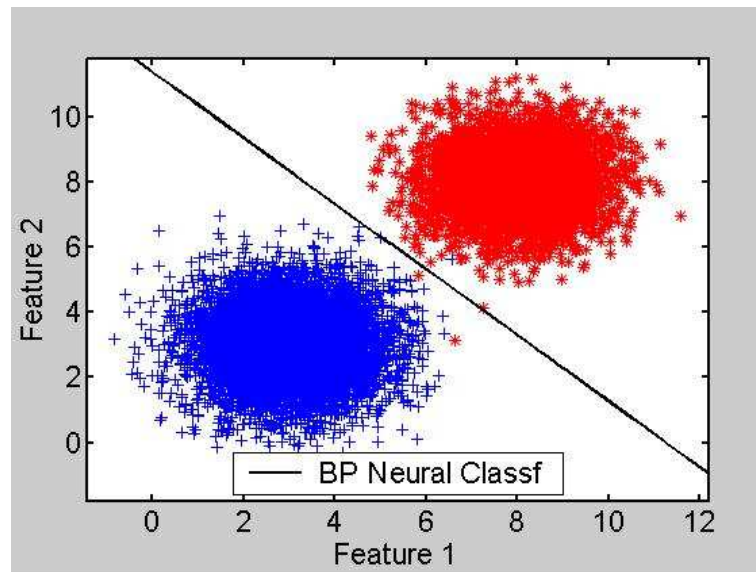


Figura A.9: Separação de dados para o conjunto de teste e respectiva superfície de separação traçada pela RNA-MLP para o experimento NDS.

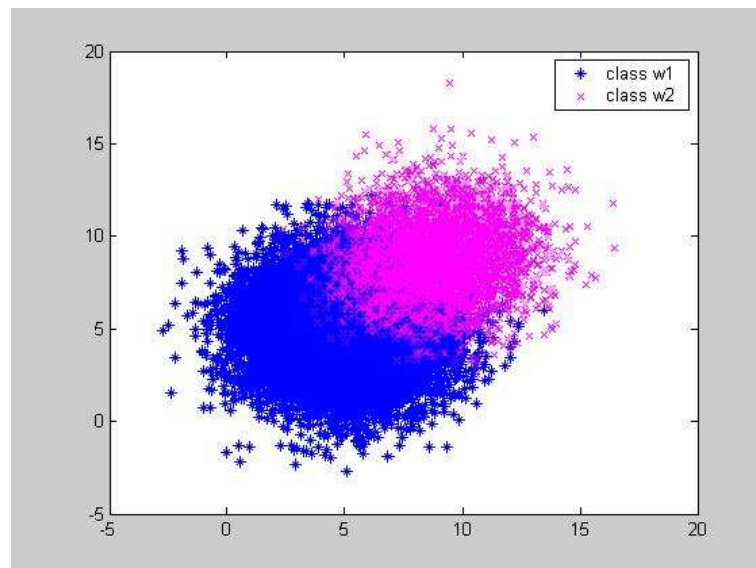


Figura A.10: Separação de dados para uma Distribuição Normal - Desbalanceados - Parcialmente Sobrepostos - NDPS.

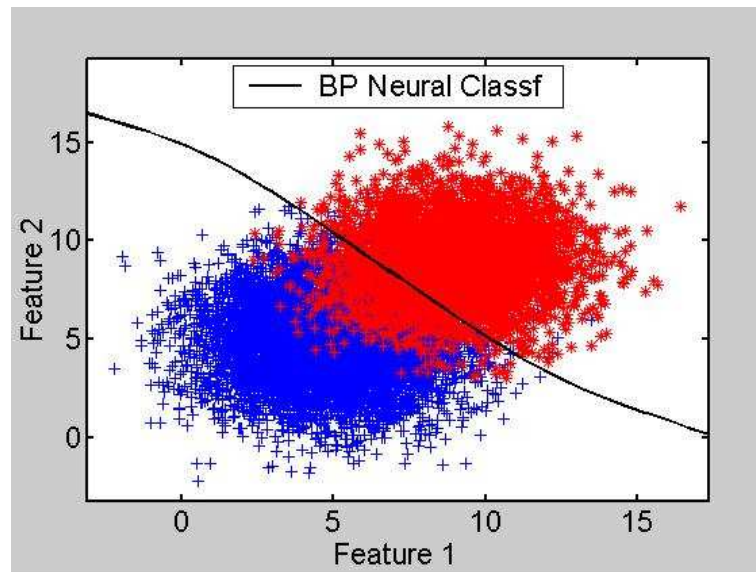


Figura A.11: Separação de dados para o conjunto de teste e respectiva superfície de separação traçada pela RNA-MLP para o experimento NDPS.

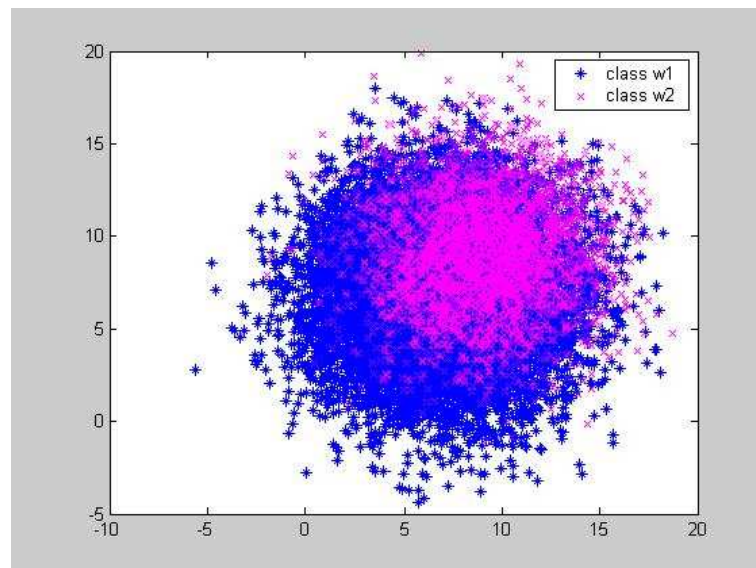


Figura A.12: Separação de dados para uma Distribuição Normal - Desbalanceados - Sobrepostos - NDSob.

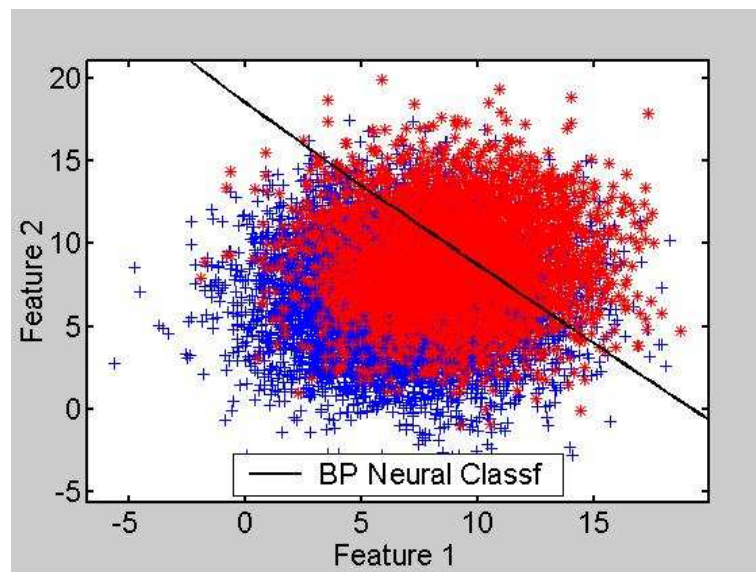


Figura A.13: Separação de dados para o conjunto de teste e respectiva superfície de separação traçada pela RNA-MLP para o experimento NDSob.

A.2 Distribuição Normal Multivariável

Dizemos que um conjunto de n variáveis aleatórias reais X_i ($i = 1, n$) apresenta uma distribuição Normal Multivariável se e só se todas as combinações lineares dessa variáveis tiver uma distribuição normal, quaisquer que sejam os coeficientes a_i dessa combinação: $\sum_{i=1}^n a_i X_i \sim N(\mu, \sigma) = 0$. As Figuras A.13, A.14 apresentam o conjunto com dados desbalanceados e sobrepostos.

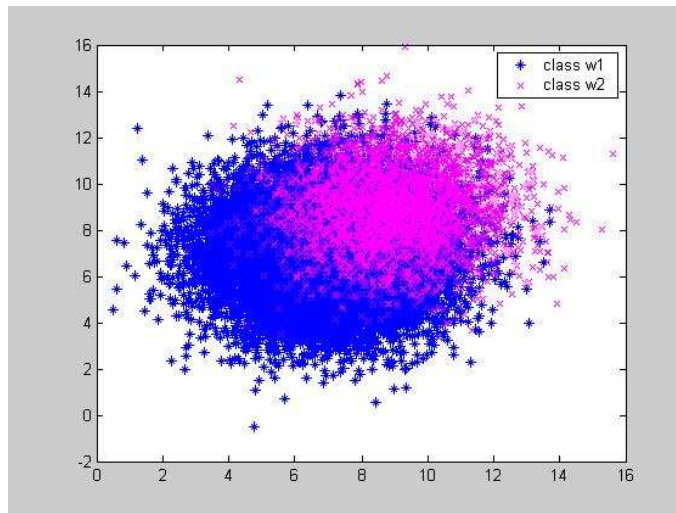


Figura A.14: Separação de dados em uma distribuição Normal Multivariável - Dados Desbalanceados - Sobrepostos - NMDSOB.

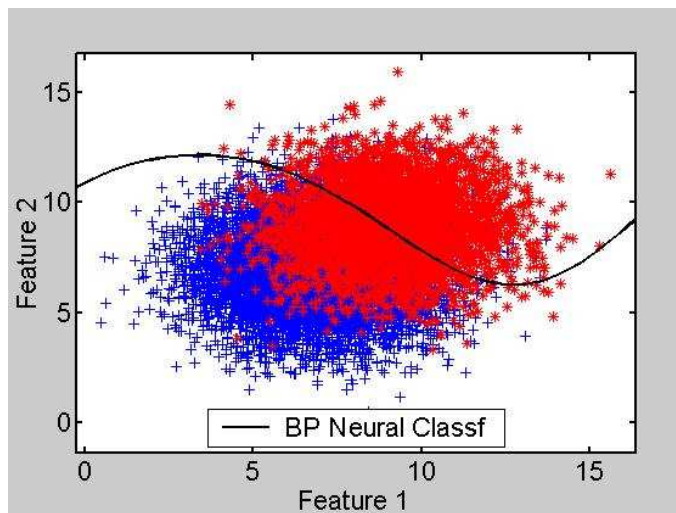


Figura A.15: Separação de dados para o conjunto de teste e respectiva superfície de separação traçada pela RNA-MLP para o experimento NMDSOB.

A.3 Distribuição Chi-square

A distribuição *Chi-square* é indexada por um parâmetro que representa o número de graus de liberdade, trata-se de uma distribuição assimétrica e positiva. A separação de dados para os conjuntos CDS pode ser observada na Figura A.16. Os demais experimentos são semelhantes as duas distribuições anteriores. A Figura A.17 apresenta superfície de divisão realizada pelo classificador para dados desbalanceado e separados.

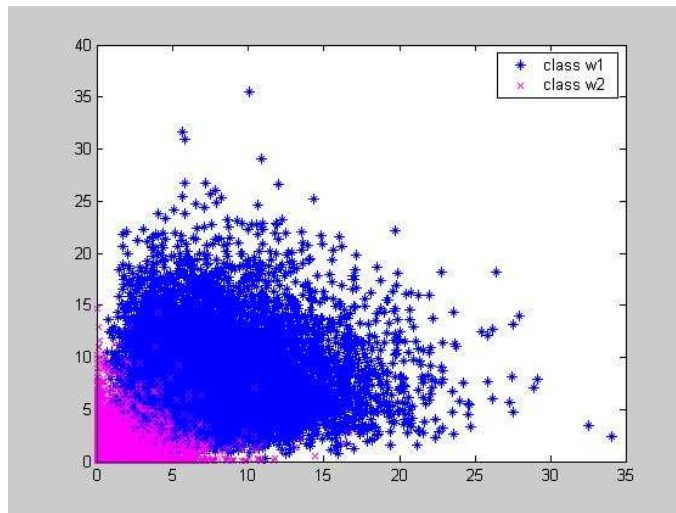


Figura A.16: Separação de dados em uma Distribuição *Chi-square* - Balanceados - Dados Separados - CDS.

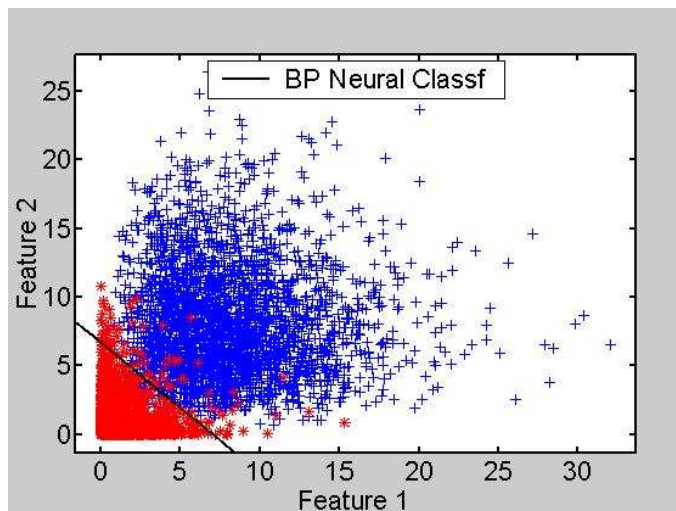


Figura A.17: Separação de dados para o conjunto de teste e respectiva superfície de separação traçada pela RNA-MLP para o experimento CDS.