

FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO



FEUP

Classificação Ordinal com Opção de Rejeição

Ezilda Duarte Almeida

Mestrado Integrado em Engenharia Informática e Computação

Orientador: Professor Doutor Jaime S. Cardoso

Responsável no INESC Porto: Mestre Ricardo Sousa

31 de Julho de 2010

Classificação Ordinal com Opção de Rejeição

Ezilda Duarte Almeida

Mestrado Integrado em Engenharia Informática e Computação

Aprovado em provas públicas pelo Júri:

Presidente: Doutor Rui Camacho

Vogal Externo: Doutor Manuel Ferreira

Orientador: Professor Doutor Jaime S. Cardoso

31 de Julho de 2010

Resumo

Nos últimos anos, várias técnicas têm vindo a ser desenvolvidas com o objectivo de encontrar melhores soluções para sistemas inteligentes. A busca é direccionada em atribuir à máquina capacidade de aprendizagem e generalização semelhantes às capacidades humanas, sendo a classificação uma das principais tarefas que fazem parte das tais técnicas.

Para melhorar o desempenho de sistemas de reconhecimento de padrões introduz-se o conceito de opção de rejeição. As opções descritas na literatura referem-se a problemas bem definidos em conjuntos de dados e objectivos específicos, em geral considerando apenas duas classes. Sendo a opção de rejeição uma ferramenta útil para melhorar o desempenho de um processo de classificação, então será uma grande valia considerar a opção de rejeição não só para problemas binários mas também para problemas multi-classe.

Esta dissertação apresenta um novo método que faz a classificação multi-classe com a opção de rejeição usando dados ordinais. O método proposto é uma extensão de um algoritmo disponível na literatura para a classificação multi-classe com dados ordinais.

Palavras-chave: Classificação multi-classe, opção rejeição, dados ordinais, Support Vector Machine.

Abstract

In recent years, several techniques have been developed and tested in order to find better results for intelligent systems. The goal is to construct an intelligent system with learning capabilities like a human being. Classification is one of the main tasks of such techniques. To improve the performance of learning system one introduces the concept of reject option. Methodologies described in the literature refer to well-defined problems in data sets and specific objectives, usually considering only two classes. Since the reject option is a useful tool to improve the performance of a classification process, it will then be of great value considering the reject option not only on binary problems but also on multi-class problems.

This thesis presents a new method that makes the multi-class classification with reject option on ordinal data. The proposed method is an extension of a method available in the literature for multi-class classification with ordinal data.

Key words: Multi-class Classification, reject option, ordinal data, Support Vector Machine.

Agradecimentos

Agradeço às pessoas e entidades que de alguma forma contribuíram para a realização desta dissertação, com especial agradecimento ao meu orientador Professor Doutor Jaime S. Cardoso e ao responsável no INESC Porto Mestre Ricardo Sousa, pela oportunidade que me deram em realizar este trabalho e por me ajudarem a elaborar e desenvolver um bom trabalho.

Aos meus pais, Maria Helena Duarte Almeida e Teófilo Eduardo Almeida, por todo o apoio e força que me deram ao longo destes anos.

Ao Instituto de Engenharia de Sistemas e Computadores do Porto (INESC Porto), pelo acolhimento e fornecimento de todos os meios necessários para a realização deste trabalho.

Aos professores, Francisco Restivo, Raul Vidal e Ricardo Sousa pela disponibilidade demonstrada que se traduziram em bons ensinamentos para este trabalho.

Ao meu colega, Lindomar Bandeira Rocha, pelos bons conselhos e pela disponibilidade demonstrada.

Aos meus restantes familiares e amigos pelo apoio incondicional que sempre me deram.

Para terminar, gostaria de agradecer aos meus colegas e companheiros desta longa caminhada, pelos momentos que compartilhamos juntos e pelo apoio que sempre demos uns aos outros.

Ezilda Duarte Almeida

Conteúdo

1. Introdução.....	1
1.1 Enquadramento da Dissertação	1
1.2 Descrição do Problema.....	2
1.3 Motivação e Objectivos.....	3
1.4 Estrutura da Dissertação.....	3
2. Conceitos Fundamentais.....	4
2.1 Problemas de Classificação	4
2.1.1 Classificação Binária.....	5
2.1.2 Classificação Multi-Classe	8
2.2 Classificação Binária com a Opção de Rejeição	9
2.3 Classificação Multi-Classe com a Opção de Rejeição	11
3. Revisão Bibliográfica	12
3.1 Algoritmo de classificação de Frank and Hall	12
3.2 Regra de Chow	13
3.3 Regra de Fumera	14
3.4 Rejeição utilizando a diferença entre os valores de confiança (DIF).....	14
3.5 Rejeição com múltiplos limiares (FumeraMod).....	14
3.6 Rejeição utilizando Média das Classes	14
3.7 Rejeição utilizando Média das Classes e Desvio Padrão	15
3.8 Revisão Tecnológica	15
3.9 Análise.....	16
4. Metodologias.....	17
4.1 SVM- Suport Vector Machine	17
4.1.1 SVM para Problemas Multi-Classes	20
4.1.2 SVM com Opção de Rejeição	20
4.2 Validação Cruzada	21
4.3 Classificação Multi-classe com Dados Ordinais.....	21
4.4 Opção de Rejeição.....	23
5. Resultados	27
5.1 Dataset's Utilizados.....	27
5.2 Resultados da Classificação Binária com a Opção de Rejeição	27
5.2.1 Dataset BCCT	28

5.2.2	Dataset Sintético.....	30
5.3	Resultados da Classificação Multi-classe com a Opção de Rejeição.....	33
5.3.1	Dataset BCCT	33
5.3.2	Dataset Sintético.....	36
5.4	Discussão dos Resultados Obtidos.....	39
6.	Conclusões e Futuros Trabalhos	40
6.1	Conclusões	40
6.2	Trabalhos Futuros.....	40
	Referências Bibliográficas	41

Lista de Figuras

Figura 1. 1: Classificação Multi-classe com a Opção de Rejeição.	2
Figura 2. 1: Rede Neuronal [13].	6
Figura 2. 2: Hiperplanos de separação [18].	8
Figura 2. 3: Classificação Multi-Classe.	9
Figura 2. 4: Classificador Binário com a Opção de Rejeição [3].	10
Figura 4. 1: Exemplo em duas dimensões da utilização de SVM como classificador [20].	17
Figura 4. 2: Exemplo em duas dimensões dos vectores de suporte [20].	18
Figura 4. 3: Classificadores Binários usando diferentes funções Kernel [32].	19
Figura 4. 4: Demonstração gráfica do procedimento de validação cruzada 10-fold [22]. ...	21
Figura 4. 5: Transformação de um problema multi-classe em k-1 problemas binárias [2].	22
Figura 4. 6: Rejeição de uma instância.	24
Figura 5. 1: Variação de Rejeição em relação ao Custo de Rejeição.	28
Figura 5. 2: Variação do Erro em relação ao Custo de Rejeição.	29
Figura 5. 3: Variação do Erro em relação à taxa de Rejeição.	29
Figura 5. 4: Variação do Desempenho em relação da taxa de Rejeição.	30
Figura 5. 5: Variação de Rejeição em relação ao Custo de Rejeição.	31
Figura 5. 6: Variação do Erro em relação ao Custo de Rejeição.	32
Figura 5. 7: Variação do Erro em relação da taxa de Rejeição.	32
Figura 5. 8: Variação do Desempenho em relação da taxa de Rejeição.	33
Figura 5. 9: Variação de Rejeição em relação ao Custo de Rejeição.	34
Figura 5. 10: Variação do Erro em relação ao Custo de Rejeição.	35
Figura 5. 11: Variação do Erro em relação da taxa de Rejeição.	35
Figura 5. 12: Variação do Desempenho em relação da taxa de Rejeição.	36
Figura 5. 13: Variação de Rejeição em relação ao Custo de Rejeição.	37

Figura 5. 14: Variação do Erro em relação ao Custo de Rejeição.	38
Figura 5. 15: Variação do Erro em relação dá taxa de Rejeição.	38
Figura 5. 16: Variação do Desempenho em relação da taxa de Rejeição.	39

Lista de Tabelas

Tabela 2. 1: Definição do Problema de Classificação.....	5
Tabela 4. 1: Exemplos de Kernels.	19
Tabela 4. 2: transformação de um problema multi-classe em k-1 problemas binárias.	23
Tabela 4. 3: Exemplo de Modelos.	24
Tabela 4. 4: Processo de atribuição dos custos às classes.....	24
Tabela 4. 5: Processo de Aprendizagem com $W_r=0.04$	25
Tabela 4. 6: Processo de Aprendizagem com $W_r=0.24$	25
Tabela 4. 7: Processo de Aprendizagem com $W_r=0.44$	25
Tabela 5. 1: Dataset BCCT.	28
Tabela 5. 2: Dataset Sintético.	31
Tabela 5. 3: Dataset BCCT.	34
Tabela 5. 4: Dataset Sintético.	37

Abreviaturas e símbolos

BCCT	Breast Cancer Conservative Treatment
CA	Class Average
DIF	Diferença entre os valores de confiança
IDE	Integrated Development Environment
GNU	General Public License
LIBSVM	Library for Suporte Vector Machine
MATLAB	MATrix LABoratory
RNA	Redes Neurais Artificiais
SRM	Minimização do Risco Estrutural
SVM	Suporte Vector Machine

Capítulo 1

1. Introdução

O processo da classificação de dados é uma necessidade real presente em diversas áreas. Ao longo da história, surgiram diversas técnicas de classificação alternativas aplicadas em diversos contextos de aprendizagem automática. No entanto, os classificadores não garantem um bom desempenho nas suas diversas aplicações, quer a nível de robustez, quer a nível de eficácia. Este problema tem vindo a ser em parte solucionado pelas técnicas de inteligência artificial, através de heurísticas específicas para cada caso. Nesta dissertação será apresentada mais uma nova abordagem para melhorar o desempenho dos classificadores.

1.1 Enquadramento da Dissertação

A classificação é uma tarefa inerente ao ser humano e de uso corrente no seu quotidiano. Efectivamente, a informação do mundo exterior é percebida pelos sentidos e é facilmente classificada em determinado contexto usando classes pré-estabelecidas pela experiência.

Este fenómeno pode ocorrer de forma imediata e sem esforço, no caso de a prática relacionada com a classificação já tenha sido adquirida posteriormente num processo de aprendizagem.

Não obstante, em casos em que a tarefa de classificação é efectuada sobre informação complexa e de grande dimensão, bem como em casos que apresentam atributos com pouco poder discriminante ou pouco claros, a aptidão humana mostra-se limitada. Deste modo, a automatização do processo de classificação passa a ser de grande interesse.

Na classificação que envolve um conjunto de dados cujas classes são conhecidas, são usados algoritmos de aprendizagem automática para prever a classe correcta, à qual um dado arbitrário do conjunto pertence.

Algumas técnicas de aprendizagem são originalmente concebidas para a solução de problemas com apenas duas classes, que se designa por classificação binária. Contudo, nem sempre os problemas podem ser linearmente separáveis.

Com objectivo de atingir uma maior eficácia na aprendizagem efectuada, o conceito da opção de rejeição foi aplicado, o que introduziu uma região dúbia para além das regiões de dados “positivos” e “negativos”. Os dados que sejam atribuídos a esta zona não são classificados.

A escolha por estratégias que incluam a opção de rejeição é revelada em muitos trabalhos existentes na comunidade científica. Uma consideração importante em relação a estes trabalhos é o facto de eles fazerem referência a opção de rejeição apenas para problemas de classificação binárias ou para problemas com dados onde não existe uma ordem entre as classes, isto é, dados nominais.

Esta dissertação aborda esta lacuna, propondo um método para problemas de classificação multi-classe com a opção de rejeição usando dados onde existe uma ordem entre as classes, isto é, designados por dados ordinais.

1.2 Descrição do Problema

Os métodos de apoio à decisão têm evoluído de forma a reproduzir as acções de tomada de decisão característica do ser humano. No entanto, estes métodos apresentam uma problemática associada a erros de classificação. Para torná-los mais eficientes tem-se vindo a desenvolver paradigma onde a opção de rejeição introduz uma alternativa viável quando a eficácia na decisão é baixa. Nestes casos o sistema não emite nenhuma solução e remete-a para um especialista, habitualmente um decisor humano.

Na comunidade científica, diferentes métodos têm vindo a ser apresentados, no entanto não existem métodos que usam a opção de rejeição para problemas multi-classe sobre dados ordinais. Deste modo, o desenvolvimento de métodos para estes tipos de problemas é de grande importância.

Nesta dissertação pretende-se desenvolver um método que faça a classificação multi-classe com a opção de rejeição sobre dados ordinais. A abordagem tomada consistiu em implementar um método que existe na literatura para realizar uma classificação multi-classe com dados ordinais, e posteriormente incorporar as regiões de rejeição.

A Figura 1.1 ilustra o conceito da classificação multi-classe com a opção de rejeição.

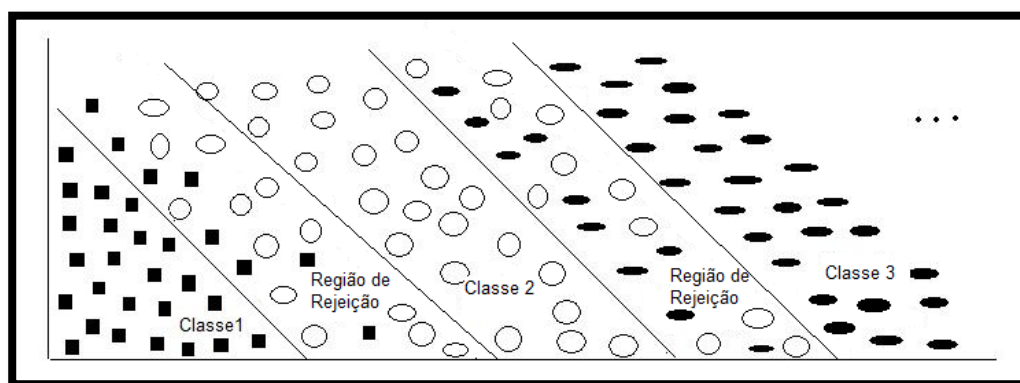


Figura 1. 1: Classificação Multi-classe com a Opção de Rejeição.

O seguinte exemplo apresenta uma aplicação do problema de classificação multi-classe com a opção de rejeição:

- O tratamento conservador do cancro da mama pode induzir diversas assimetrias entre a mama tratada da não tratada. Assim, um sistema de apoio à decisão pode tomar partido de diversas características como alterações visíveis na pele, no volume, ou ainda na forma da mama para determinar o resultado estético final deste tratamento para cada paciente: “Excelente”, “Bom”, “Suficiente” e “Mau”. Todavia, nem sempre é trivial discernir um caso Excelente dum Bom. Consequentemente, e para os casos dúbios as regiões de rejeição podem aparecer naturalmente entre as regiões Excelente e Bom, Bom e Suficiente, entre outras.

1.3 Motivação e Objectivos

Os sistemas de apoio à decisão têm um interesse crescente em diversas áreas, tais como medicina, finanças, ciências naturais e pesquisas. Estes sistemas estão a tentar reproduzir, tanto quanto possível, a decisão humana.

A classificação, que consiste no processo de prever a classe à qual um objecto pertence, é um processo destacado como uma das disciplinas de aprendizagem automática que confere esta propriedade aos sistemas de apoio à decisão.

Portanto, o desenvolvimento de algoritmos de classificação em problemas multi-classe e com a aplicação de opção rejeição é bastante motivador.

Esta dissertação tem como objectivo desenvolver um método que soluciona este tipo de problema, isto é, estender a classificação binária com a opção de rejeição para a classificação multi-classe sobre dados ordinais.

1.4 Estrutura da Dissertação

Esta dissertação encontra-se dividida em seis Capítulos. No presente Capítulo é contextualizado e apresentado o problema, no qual é descrito e enquadrado em diversas aplicações alvo. Em seguida procede-se à descrição da motivação e dos objectivos do estudo bem como a estrutura geral da dissertação. No segundo Capítulo é realizado um enquadramento com a apresentação dos principais elementos teóricos que enquadram e sustentam o estudo empírico, entre os quais foram destacados os seguintes tópicos: problema de classificação, classificação binária, classificação multi-classe, classificação binária com a opção de rejeição e classificação multi-classes com a opção de rejeição. No terceiro Capítulo é efectuado um breve enquadramento dos estudos existentes na literatura no âmbito desta dissertação, nomeadamente as estratégias de rejeição e o algoritmo de *Frank and Hall*. É também apresentada uma revisão tecnológica sobre as principais metodologias usadas para solucionar este tipo de problema. Ainda, é realizada uma análise que enfatiza as principais concretizações inovadoras desta dissertação. No quarto Capítulo apresentam-se alguns conceitos fundamentais relativos à aprendizagem automática e ao algoritmo proposto por *Frank and Hall*, no qual esta tese se baseou. É também apresentada a estratégia para fazer a integração da opção de rejeição. No quinto Capítulo estão apresentados e analisados os resultados obtidos ao longo de todo o estudo. Ainda neste capítulo referem-se os *datasets* utilizados nomeadamente o *dataset BCCT (Breast Cancer Conservative Treatment)* e o *dataset* sintético.

Finalmente, no sexto Capítulo, apresentam-se as conclusões e os futuros trabalhos, enumerando os objectivos da dissertação que foram atingidos.

Capítulo 2

2. Conceitos Fundamentais

Nesta secção são apresentados alguns conceitos básicos relativos a aprendizagem automática. Para cada uma das abordagens mencionadas, vão ser analisados alguns métodos do estado da arte para a sua resolução. Em seguida, será descrita aplicação da opção de rejeição em classificadores binários, multi-classe e os actuais métodos existentes.

2.1 Problemas de Classificação

Na vida quotidiana, frequentemente existe a necessidade de classificar objectos que ocorre de forma inconsciente. Por exemplo, numa frutaria, as pessoas não seleccionam arbitrariamente a fruta. No caso da compra de bananas, por exemplo, ocorre um processo de escolha segundo vários critérios. Normalmente analisam a aparência das bananas, rigidez, cor, preço e outras características, de forma a levarem para casa as bananas com melhor qualidade. Apesar da qualidade das bananas ser um critério subjectivo, os compradores classificam as bananas em duas classes: aquelas que estão e aquelas que não estão em bom estado.

Desta forma, problemas de classificação são definidas como aquelas em que, a partir de um conjunto de atributos de um dado objecto de estudo (no caso das bananas: aparência, cor, rigidez e preço), deve-se definir um grupo ao qual este objecto pertença (no caso das bananas: os grupos de bananas que estão ou que não estão em bom estado). A cada um desses grupos é atribuído o nome de classes e ao responsável pelo processo de classificação é atribuído o nome de classificador (que no caso das bananas corresponde ao comprador na frutaria). Os problemas de classificação estão presentes em diversas áreas, como por exemplo na determinação do diagnóstico médico de acordo com os resultados de exames de um paciente (se ele está ou não doente, ou qual o tipo de uma doença), seja na determinação de um tipo de tumor de acordo com as suas características (se o tumor é benigno ou maligno).

Em aprendizagem automática, existem inúmeras propostas para tratamento destes problemas de classificação, que se baseiam tanto no uso individual de classificadores como na combinação destes classificadores. A Tabela 2.1 apresenta a uma ideia geral do problema de classificação.

Tabela 2. 1: Definição do Problema de Classificação.

<p>Considere um conjunto de objectos definidos por :</p> <ul style="list-style-type: none"> • a - atributos • $C = \{c_1, \dots, c_i\}$ - um conjunto de classes <p>Suponha que existe uma função que associa cada objecto a uma classe $f: a \rightarrow c_i$</p> <p>O problema de classificação consiste em :</p> <ul style="list-style-type: none"> • Dado um conjunto de treino definido por pares atributos-classes $\langle a_i c_i \rangle$, • Encontrar uma função $f: a \rightarrow c_i$
--

Como é referido anteriormente, para obter-se a classe com base nos atributos a que pertence o exemplo, é utilizado um classificador. Este classificador é treinado através de um processo de aprendizagem realizado com base em informação histórica (exemplos para os quais são conhecidos os atributos e a respectiva classe). Este histórico de exemplos define um conjunto de treino que é aplicado a um algoritmo de aprendizagem que retorna a informação que permite a classificação. Esta é a fase de treino.

Novos exemplos não classificados são recebidos pelo (s) classificador (es) que se denominam como conjunto de teste. Este conjunto terá o mesmo tipo de atributos do conjunto de treino, no entanto estes não foram usados na fase de treino com o objectivo de não introduzir tendência na classificação. Esta classificação é usada para obter informação de desempenho usando métricas de sucesso.

2.1.1 Classificação Binária

Na classificação binária, o classificador desenvolvido tem como objectivo dividir o espaço das características em duas regiões que correspondem a cada uma das classes. Considerando o exemplo das “bananas” da secção anterior, a classificação binária se concretizaria nas duas classes: uma classe das bananas que compensam adquirir e outra classe das bananas que não compensam ser adquiridas.

Os rótulos para cada uma das classes podem ser definidos da seguinte forma: $y_i \in \{-1, 1\}$, outros valores como $y_i \in \{0, 1\}$ poderiam ser usados, mas nesta dissertação convencionou-se a primeira forma. A interpretação do significado de cada uma das duas classes pode variar consideravelmente.

Na literatura existem várias implementações de classificadores binários, entre os quais, redes neuronais [34], classificador de *Bayes*, Máquinas de Vectores de Suporte [11], entre outras técnicas.

2.1.1.1 Redes Neurais Artificiais (RNA)

As redes neuronais artificiais têm tido uma evolução histórica consideravelmente particular. Apesar das redes neuronais terem causado grande impacto positivo nos meios científicos [32] nos anos 60, foram votadas ao abandono até aos anos 80, em que se assistiu ao seu ressurgimento com um devido ao método de aprendizagem *backpropagation*.

No seu exemplo mais típico, uma rede neuronal artificial é formada por três tipos de elementos a que se dá o nome de neurónios: os neurónios da camada de entrada que introduzem os dados na rede, os neurónios das camadas internas e os neurónios da camada de saída que produzem o resultado. Cada ligação entre neurónios possui um peso, tendo cada neurónio, em geral, um nível de patamar que influencia directamente a sua saída [13]. A Figura 2.1 mostra a estrutura típica de uma rede neuronal.

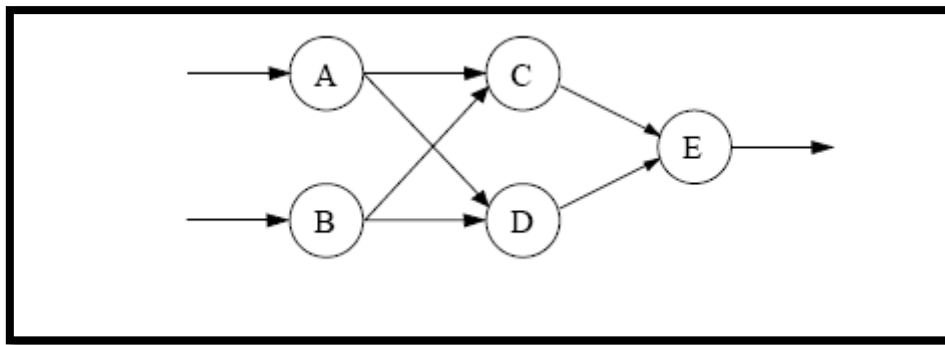


Figura 2. 1: Rede Neuronal [13].

Uma rede neuronal pode possuir uma ou múltiplas camadas. No exemplo da Figura 2.1 mostra uma rede neuronal com três camadas e seis neurónios. Cada neurónio possui ligações para outros neurónios, nas quais recebem e enviam sinais.

Na primeira camada definida pelos neurónios A e B tem como função receber os padrões, esta camada é designada por camada de entrada.

Na camada intermédia definida por C e D realiza o processamento e extracção de características.

Por fim, a camada de saída que conclui e apresenta o resultado final.

Quanto maior o número de camadas, melhor a capacidade de aprendizagem. A camada de entrada deve possuir uma unidade especial conhecida como *bias*, usada para aumentar os graus de liberdade, permitindo uma melhor adaptação, por parte da rede neuronal, ao conhecimento a ela fornecido.

2.1.1.2 Máquinas de Vectores de Suporte (SVM)

A máquina de vector suporte é uma técnica de aprendizagem automática, baseada no princípio da Minimização do Risco Estrutural (SRM), e pode ser usada para resolver problemas de classificação.

Num problema de classificação binária, o funcionamento de uma SVM é descrito da seguinte forma:

1. Atribuir duas classes e um conjunto de pontos que pertencem a essas classes.

2. Determinar o hiperplano óptimo que separa os pontos de forma a colocar o maior número de pontos da mesma classe do mesmo lado, enquanto maximiza a distância de cada classe a esse hiperplano.

A distância de uma classe a um hiperplano é a menor distância entre ele e os pontos dessa classe. Classificações binárias usando máquinas de vectores de suporte realizam-se pelo uso da função $g(x)$, com o seguinte modo: Os dados pertencem à classe positiva, se $g(x) \geq 0$, e caso contrário, pertencem à classe negativa. A superfície de decisão será representada por um hiperplano na forma [16]:

$$g(x)=(w^t x)+b=0, \tag{2.1}$$

Onde $w \in \mathbb{R}^m$ é o vector de pesos, e $b \in \mathbb{R}$ é o *bias*. Assim aplica-se o seguinte modo de decisão:

$$g(x)=(w^t x) + b \geq 0, \quad \text{para } y = +1 \tag{2.2}$$

$$g(x)=(w^t x) + b < 0, \quad \text{para } y = -1 \tag{2.3}$$

A seguinte forma canónica utiliza-se para descrever o lugar geométrico dos hiperplanos separados onde o vector de peso w e o *bias* b são novamente escalados de forma a atender as desigualdades [16]:

$$g(x)=(w^t x) + b \geq 0, \quad \text{para } y = +1 \tag{2.4}$$

$$g(x)=(w^t x) + b \leq 0, \quad \text{para } y = -1 \tag{2.5}$$

A Figura 2.2 ilustra os hiperplanos de separação para o caso onde o espaço $m=2$. A separação entre $g(x)=(w^t x)+b=0$ e o dado de entrada mais perto, para um dado vector de pesos w e *bias* b , é chamado de margem de separação, denotada por ρ . Sempre que ρ for positivo, existirão infinitos hiperplanos, dentre os quais se estima o hiperplano particular em que a margem de separação ρ é maximizada [16]. De acordo com esta condição, a superfície de decisão é considerada como o hiperplano óptimo, [18] [15].

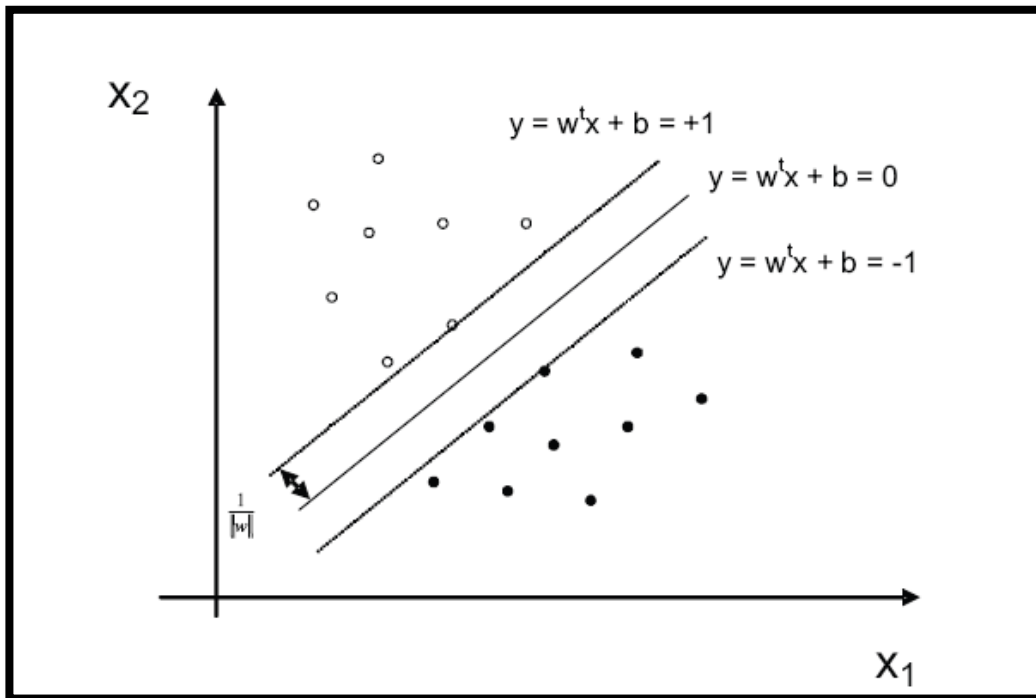


Figura 2. 2: Hiperplanos de separação [18].

2.1.1.3 Classificador de Bayes

Este tipo de classificador, é baseado no teorema de *Bayes*:

$$P(A|x) = \frac{P(A)P(x|A)}{P(x)} \quad (2.6)$$

Onde $P(x)$ é a probabilidade da classe que se define como a frequência relativa da classe, definida pelo problema. $P(x|A)$ é a probabilidade de uma amostra pertencer a determinada classe, que corresponde a uma medida de verosimilhança da amostra x como pertencente à classe A . $P(x)$ é a probabilidade da amostra, isto é, a frequência relativa dessa amostra em relação a todas as amostras possíveis.

O classificador de *Bayes* é utilizado de seguinte forma [18]:

$$C = P(A|x) - P(B|x) \quad (2.7)$$

Se C for positivo, então x pertence à classe A se não, x pertence à classe B .

2.1.2 Classificação Multi-Classe

Muitos dos problemas de classificação envolvem mais do que duas classes, a qual se designa como classificação multi-classe. Podemos facilmente pensar no caso do estado de saúde de um paciente onde se poderia definir, por exemplo, as classes “Mau”, “Razoável”, “Bom”, “Muito Bom”. Neste caso, pretende-se classificar o estado de saúde do paciente em uma das classes.

O que distingue a classificação binária da classificação multi-classe, é o número de classes existentes para a classificação. Na classificação binária são definidas exactamente duas classes e na classificação multi-classe, são definidas n classes ($n > 2$). A Figura 2.3 ilustra um exemplo de classificação multi-classe.

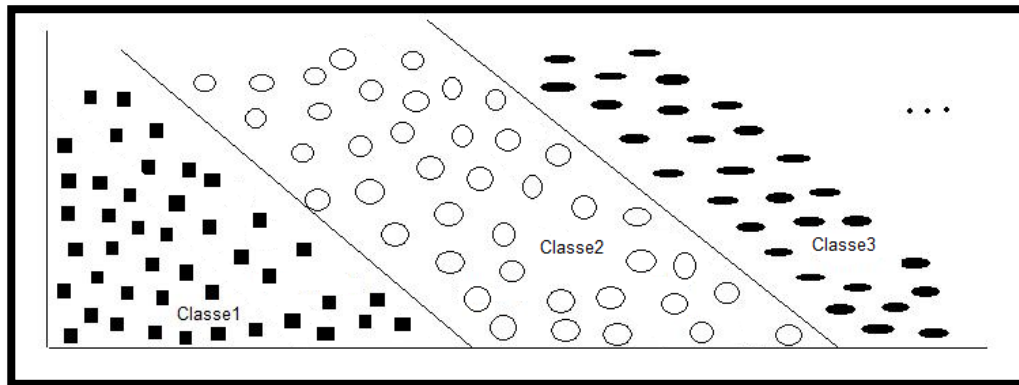


Figura 2. 3: Classificação Multi-Classe.

Para solucionar problemas de classificação multi-classe decompõe-se o problema em vários problemas de classificação binária. Na literatura existem diversos estudos sobre o modo de decompor os problemas multi-classe em problemas binários. Uma solução consiste em estender a SVM binária a problemas de classificação multi-classe. Existem duas abordagens básicas para esta extensão. A primeira é a redução do problema multi-classe a um conjunto de problemas de classificação binária. Dois métodos usam essa abordagem: decomposição “**um-contra-todos**” e “**todos-contra-todos**”. A decomposição “**um-contra-todos**” consiste em comparar cada uma das k classes do problema original contra todas as outras. Um problema multi-classe com k classes é decomposto em $k-1$ problemas binários: uma classe principal contra as restantes classes. Todas as classes k do problema original são seleccionadas uma a uma [18].

A decomposição “**todos-contra-todos**” consiste em criar conjuntos de dados com todas as combinações possíveis das classes do problema original, juntando-as duas a duas. Para cada uma das combinações é gerada informação para a classificação. Um problema multi-classe com k classes é transformado em $k-1$ problemas binários, cada um com o seu classificador b , independente. Cada problema binário é definido por um dado de treino constituído pelos exemplos de apenas duas classes. Por exemplo: para um problema original com quatro classes (A,B,C,D), seriam geradas seis combinações e consequentemente seis problemas binários: A-B, A-C, A-D, B-C, B-D, C-D [18].

A segunda abordagem é a generalização de SVMs binários para mais de duas classes. O método que utiliza essa abordagem é o método de *Crammer e Singer* [1]. Na secção 4.1.1, será dada mais ênfase a esta abordagem.

2.2 Classificação Binária com a Opção de Rejeição

Como mencionado anteriormente a introdução do conceito da opção de rejeição, permite não só separar os dados “positivos” dos “negativos”, como também a inclusão de uma zona onde os pontos que aí caíam não sejam classificados.

Na classificação binária com a opção de rejeição, o classificador desenvolvido tem como objectivo dividir o espaço das características em três regiões, das quais duas regiões

correspondem a cada uma das classes e a outra região corresponde à região de rejeição. A Figura 2.4 ilustra um classificador binário com a opção de rejeição.

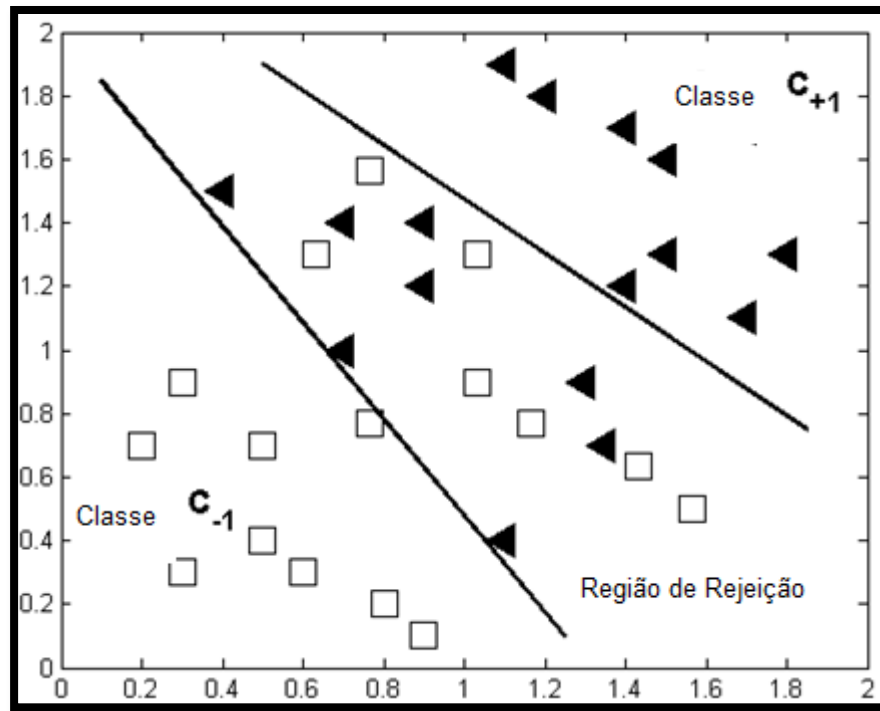


Figura 2. 4: Classificador Binário com a Opção de Rejeição [3].

Os elementos quadrados da Figura 2.4 pertencem à classe C_{-1} , os elementos triângulos pertencem à classe C_{+1} e na zona central temos uma região que inclui os elementos de ambas as classes e que se denomina por região de rejeição.

Existem três abordagens diferentes para analisar um classificador binário com a opção de rejeição:

- **Dois classificadores independentes.** O primeiro classificador é treinado para saída C_{-1} , somente quando a probabilidade de C_{-1} , é elevada, um segundo classificador é treinado para saída C_{+1} somente quando a probabilidade de C_{+1} é elevada. A simplicidade desta estratégia tem a desvantagem do cruzamento das fronteiras, levando à ocorrência de regiões de uma decisão não lógica [3].
- **Um único classificador binário padrão.** Esta abordagem já não fornece cruzamento das fronteiras. Um dado é rejeitado se o máximo das duas probabilidades for menor que um determinado limiar. Se o classificador não fornecer saídas probabilísticas utiliza-se um limiar de rejeição. A região de rejeição é determinada após o treino do classificador, definindo valores limite adequados sobre a saída do classificador [3].
- **Um único classificador com opção de rejeição incorporado.** Esta abordagem consiste no desenvolvimento de algoritmos especificamente adaptados para este tipo de problemas [4], [12]. Um exemplo desta abordagem é a SVM com opção de rejeição. Uma SVM com a opção de rejeição é implementada usando duas abordagens. Na primeira

abordagem, a regra de rejeição consiste na rejeição de dados para os quais a distancia $d(x)$ seja inferior a um limiar pré-definido. Na segunda abordagem a opção de rejeição em SVMs consiste no mapeamento de suas saídas para probabilidades posterior, de modo que a regra de *Chow* pode ser aplicada.

Nos problemas de classificação com a opção de rejeição, o objectivo é de obter um erro mínimo, minimizando o número de casos mal classificados. No entanto, quando o número dos casos rejeitados diminui, o erro de classificação aumenta. Por outro lado, se o erro de classificação diminui, a região de rejeição aumenta. O equilíbrio entre a rejeição e o erro depende da relação dos custos associados.

2.3 Classificação Multi-Classe com a Opção de Rejeição

O problema da classificação com a opção de rejeição estende-se também ao problema multi-classe.

Para formalizar a rejeição usa a notação de confiança associada a uma classe. Considera-se um classificador C que fornece na saída uma medida de confiança c_i para cada uma das “ c ” classes pertencentes ao problema. 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 [6].

Posteriormente, conhecidas as medidas de confiança ($c_1, c_2... c_c$) fornecidas pelo classificador, um modo de rejeição pode ser aplicado. 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 aceite, enquanto exemplos com $\hat{c} < T$ são rejeitados. O objectivo de aplicar uma estratégia de rejeição em um dado problema é encontrar um valor óptimo para T . Quando este valor é encontrado uma estratégia de rejeição consegue excluir todos os exemplos incorrectamente classificados pelo classificador e também aceitar todos os exemplos correctamente classificados [6].

Algumas das estratégias de rejeição descritas na literatura são apresentadas:

- Rejeição com um limiar (*Chow*)
- Rejeição com múltiplos limiares (*Fumera*)
- Rejeição utilizando a diferença entre os valores de confiança (DIF)
- Rejeição com múltiplos limiares (*FumeraMod*)
- Rejeição utilizando Média das Classes (CA)
- Rejeição utilizando Média das Classes e Desvio Padrão.

Estas estratégias de rejeição, serão apresentadas nas secções mais á frente.

Capítulo 3

3. Revisão Bibliográfica

O trabalho proposto foi iniciado por um levantamento do estado da arte com o objectivo de identificar os problemas, revendo os métodos básicos relacionados com a classificação multi-classe e opção de rejeição que mais se aproximam do objectivo desta dissertação.

Deste modo, este capítulo apresenta e descreve pormenorizadamente o algoritmo de Frank and Hall e as estratégias de opção de rejeição. As ferramentas disponíveis que implementam estas técnicas foram também discutidas numa análise que visou a selecção do software mais adequado para constituir um ambiente de desenvolvimento.

3.1 Algoritmo de classificação de Frank and Hall

Na literatura existe um algoritmo que implementa a classificação multi-classe ordinal designado por *Frank and Hall*. O algoritmo de *Frank and Hall* [2] considera que os dados estão expressos numa escala ordinal. Os dados ordinais caracterizam-se por apresentarem uma ordem específica das classes tendo um significado implícito. Como exemplo de escalas ordinais temos as notas qualitativas obtidas numa disciplina, que assumir os valores de “Mau”, “Suficiente”, “Bom” e “Muito Bom”. Nesta escala está implícito o grau de sucesso do aluno.

Em oposição, existem os dados nominais que se caracterizam, ao contrário dos dados ordinais, pelo facto de não apresentarem uma ordem nas classes. Como exemplos de escalas nominais temos a religião, a raça, a localização geográfica, o sexo, a profissão, tipo de residência, preferências, ocupações, etc.

Para exemplificar o princípio de funcionamento deste algoritmo, é usado um exemplo de tipos de dados ordinais cujas classes são “Quente”, “Ameno” e “Frio”. No início, o problema ordinal K-Classe é decomposto em K-1 problemas de classificação binária. Durante a fase do treino, os conjuntos de novos dados são derivados a partir do conjunto de dados originais. Existem dois conjuntos de dados derivados: o primeiro tem um atributo de classe que representa $Target > Frio$ e o segundo tem um atributo de classe que representa $Target > Ameno$. O conjunto original e o conjunto resultante da decomposição, contém o mesmo número de atributos. Nas próximas etapas, o algoritmo de classificação é aplicado para gerar um modelo para cada um dos novos conjuntos de dados. Para prever o valor de classe de uma instância é preciso estimar as probabilidades de K classes ordinais originais usando o modelo K-1. A estimação da probabilidade para o primeiro e último valor da classe ordinal depende de um único classificador. A probabilidade do primeiro valor ordinal (*Frio*) é dada por $1 - Pr(Target > Frio)$. Da mesma forma, o último valor ordinal (*Quente*) é calculado a partir da $Pr(Target > Ameno)$.

Para valores de classe no meio da escala há apenas um (*Ameno*) cuja probabilidade depende de um par de classificadores. Neste exemplo é dado pelo $\Pr(\text{Target} > \text{Frio}) - \Pr(\text{Target} > \text{Ameno})$.

Uma instância da classe desconhecida é processada por cada um dos $K-1$ classificadores e a probabilidade de cada um dos valores da classe K ordinal é calculada utilizando o método acima. A classe com probabilidade máxima é atribuída à instância [2]. Na secção 4.3 será dado mais ênfase a este método.

3.2 Regra de Chow

Chow é um autor que tem dado várias contribuições relacionadas com este tema de trabalho, na literatura. Entre os trabalhos desenvolvidos por *Chow*, a regra de *Chow* para a rejeição de uma dada instância é a que mais se enquadra ao tema desta dissertação.

Geralmente, sistemas de reconhecimento aplicam uma decisão global que decide entre aceitar ou recusar o resultado do reconhecimento. Em classificação, um dado é considerado ambíguo se este não pode ser associado a uma classe com determinada certeza, enquanto um dado associado a uma baixa confiança para todas as classes também pode ser tratado como um “dado falso”.

O objectivo do mecanismo de rejeição é minimizar o número de erros de reconhecimento para um dado número de rejeições. Uma forma simples de rejeição é excluir um dado que tem uma probabilidade global menor do que um determinado limiar, isto é, rejeitar as observações menos prováveis, como denotado pela regra de *Chow* [21].

De acordo com a regra de *Chow*, um dado x é rejeitado se:

$$\max_{k=1,\dots,N} P(w_k | x) = P(w_i | x) < T \quad (3.1)$$

Onde $T \in [0,1]$. Por outro lado, o padrão x é aceite e associado a classe w_i , se:

$$\max_{k=1,\dots,N} P(w_k | x) = P(w_i | x) \geq T \quad (3.2)$$

Segundo *Chow*, o espaço de características é subdividido em $N+1$ regiões. A região de rejeição D_n é definida de acordo com a Equação (3.1), enquanto a região de decisão D_0, \dots, D_{n-1} é definida de acordo com a Equação (3.2). A probabilidade de um dado ser rejeitado pode ser calculado deste modo:

$$P(\text{Rejeicao}) = \int p(x) dx \quad (3.3)$$

Em que $p(x)$ representa uma função densidade de probabilidade. A precisão do classificador é definida como a probabilidade condicional de um dado classificado correctamente, dado que este foi aceite, e é representada pela equação (3.4):

$$\text{Precisao} = P(\text{correcto} | \text{aceite}) = \frac{P(\text{correcto})}{P(\text{correcto}) + P(\text{erro})} \quad (3.4)$$

3.3 Regra de Fumera

Fumera fez uma análise cuidadosa da regra de *Chow*, de onde conclui que a regra de rejeição de *Chow* prevê um erro de rejeição óptimo se as probabilidades a posteriori são exactamente conhecidas. Não obstante, tais probabilidades são afectadas por erros de estimativas em aplicações reais. Tendo em conta, este problema da regra de *Chow*, *Fumera* propôs o uso de múltiplos limiares de rejeição, em que é utilizada um limiar para cada classe do problema. Para esta estratégia, uma instância x é rejeitada caso obedeça a equação 3.5. Enquanto uma instância x é aceite e associada á classe w_i , caso obedeça a Equação 3.6 [33].

$$\max_{k=1,\dots,N} \hat{P}(w_k | x) = \hat{P}(w_i | x) < T_i \quad (3.5)$$

Onde $T \in [0,1]$

$$\max_{k=1,\dots,N} \hat{P}(w_k | x) = \hat{P}(w_i | x) \geq T_i \quad (3.6)$$

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

Sua implementação é obtida calculando a diferença entre os dois maiores níveis de confiança obtidos na saída da RNA. A regra heurística representada pela estratégia DIF obteve desempenho melhor ou igual a estratégias estatísticas dita como “óptimas” na literatura, como por exemplo as regras de *Chow* e *Fumera*. Esta estratégia é pouco influenciada pelas condições do conjunto de dados e ainda sua implementação não possui grau de dificuldade elevado.

3.5 Rejeição com múltiplos limiares (FumeraMod)

Esta estratégia é baseada na ideia da estratégia “gulosa” (em inglês, *greedy*), ou seja, encontra uma solução óptima para cada classe pretendendo chegar a uma solução óptima global para o problema da rejeição. Neste caso, busca os limiares óptimos locais considerando uma classe de cada vez e obtém os limiares necessários para dividir o conjunto de validação em vários subconjuntos, afectando um para cada classe. A aplicação do mecanismo de rejeição é realizada da mesma forma que as estratégias anteriores [6].

3.6 Rejeição utilizando Média das Classes

Esta estratégia é um método heurístico para calcular um suposto limiar óptimo a ser adoptado como critério de rejeição. O limiar é obtido de forma automática em relação ao

conjunto de dados para encontrar a média de um conjunto de valores. Neste caso, extraí a média de um subconjunto (classe seleccionada) considerando somente os exemplos correctamente classificados de uma determinada classe [6].

3.7 Rejeição utilizando Média das Classes e Desvio Padrão

Esta estratégia é derivada da anterior. Neste caso, além de calcular a média, calcula-se também o desvio padrão para os dados correctamente classificados. O limiar é obtido através da diferença desses dois valores [6].

3.8 Revisão Tecnológica

Nesta secção apresenta-se os resultados de uma revisão da literatura referente às principais ferramentas adequadas para o âmbito da dissertação com os respectivos argumentos de utilização.

No âmbito do projecto, várias tecnologias podem ser usadas, nomeadamente Weka, Matlab, PRTools, GNU/Octave e a biblioteca LibSVM.

Em termos das tecnologias descritas na literatura pretende-se usar o MATLAB e a biblioteca LibSVM para o desenvolvimento da dissertação.

MATLAB (*Mathworks Inc.*) é uma IDE (*Integrated Development Environment*) com ambiente interactivo que possibilita o desenvolvimento de algoritmos, visualização de análise de dados, computação numérica ou simbólica, possuindo adicionalmente excelentes recursos gráficos 2D e 3D. Pode ser usado em variadas aplicações, que inclui *toolboxes* específicas para processamento de sinal, comunicação, controle, simulação, redes neuronais, modelagem financeira e muitos outros. O código de cada uma dessas aplicações apresenta elevadas eficiência e rapidez de execução.

Os códigos fontes dos mesmos são robustos e internacionalmente reconhecida como *standard* em suas áreas de aplicação. O código gerado em MATLAB pode ser integrado noutras linguagens de programação e aplicações desenvolvidas em C, C++, Delphi, Java, Excel etc., e os seus algoritmos podem, assim, ser livremente distribuídos [8].

O **LIBSVM** é uma biblioteca de implementações de SVM desenvolvida por *Chin-Chung Chang* [17] com várias finalidades: classificação, regressão e estimativa de distribuição. Um dos algoritmos de classificação implementado na biblioteca é identificado como C-SVC.

Os motivos que levaram à escolha do MATLAB para o desenvolvimento da dissertação foram o facto de possuir o código das aplicações eficientes e robustos, ser internacionalmente reconhecido como *standard* e permitir a integração noutras linguagens de programação.

A utilização da biblioteca libSVM é justificada pelo facto de permitir criar aplicações com SVM, que é o método de aprendizagem máquina é referido como um dos mais eficientes na literatura [17].

3.9 Análise

Considerando o estudo efectuado nas áreas referidas nas secções anteriores, compreende-se que nos últimos anos, várias técnicas têm sido desenvolvidas e testadas visando encontrar melhores resultados para problemas de classificação com opção de rejeição. Apesar de muitos estudos sobre problemas de classificação binários e multi-classe nominais, é possível afirmar que relativamente aos problemas de multi-classe ordinais não existem muitos estudos.

O interesse por estratégias de rejeição é encontrado em muitos trabalhos reportados na literatura. Uma consideração importante em relação a esses trabalhos, é o facto da maioria deles utilizarem formas estatísticas, isto é, probabilidades e heurísticas.

Capítulo 4

4. Metodologias

Neste capítulo é explicado o processo de adaptação da opção de rejeição ao algoritmo de *Frank and Hall* que constitui a inovação do trabalho bem como o uso dos SVM para as fases de treino e de teste.

Assim, são apresentados os conceitos fundamentais relativos à aprendizagem automática baseada no conjunto de métodos dos SVM. Para além disso, é descrito o método validação cruzada que é utilizado na fase de teste para dar mais consistência aos resultados de desempenho.

4.1 SVM- Support Vector Machine

Aprendizagem automática é um campo de pesquisa da Inteligência Artificial, que estuda o desenvolvimento de métodos capazes de extrair conhecimento, a partir de amostras de dados.

Em geral, os diversos algoritmos de aprendizagem automática são utilizados de forma a gerar classificadores para um conjunto de exemplos. Como algoritmos de aprendizagem automática dispomos dos SVMs, redes neuronais e árvores de decisão. Na realização da dissertação foi utilizada a SVM.

As SVM têm obtido sucesso em um grande número de aplicações, que variam desde identificação de partículas, identificação de face, etc. O modo como a aproximação é realizado como base, a teoria estatística [19]. As SVM funcionam construindo um hiperplano de N dimensões que separa os dados em duas categorias de forma otimizada [20]. A Figura 4.1 demonstra um exemplo de como são separadas duas classes, usando as SVM.

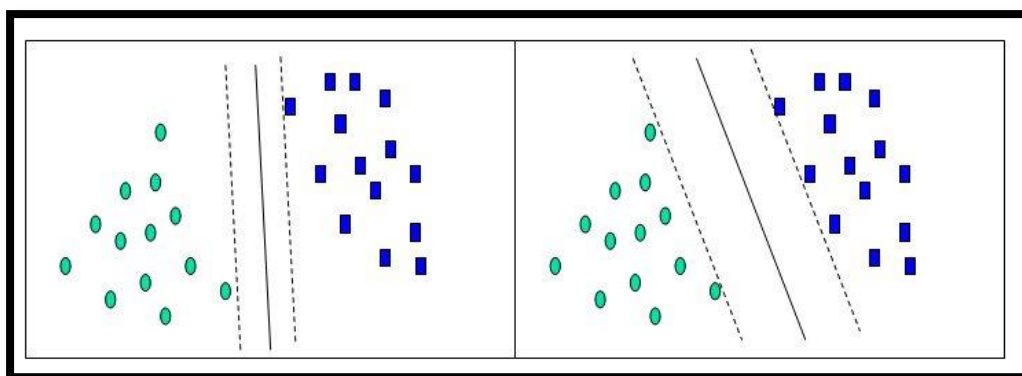


Figura 4. 1: Exemplo em duas dimensões da utilização de SVM como classificador [20].

Na literatura, a tarefa de escolher a representação mais satisfatória é conhecida como selecção de característica. O conjunto de características que descrevem uma determinada situação é denominado de vector. Assim, o objectivo de SVM é modelar um hiperplano óptimo. Os vectores perto do hiperplano denominam-se como vectores de suporte. A Figura 4.2 ilustra bem esses conceitos.

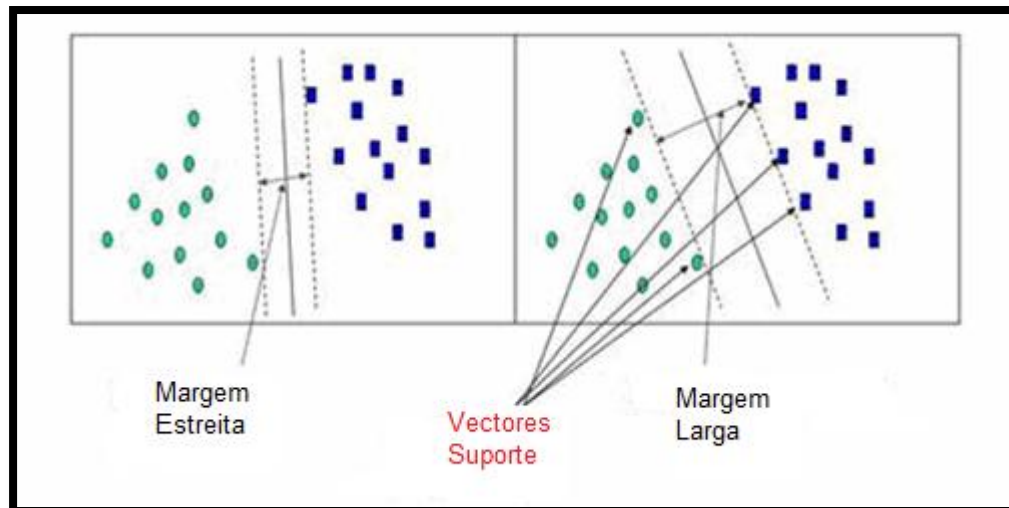


Figura 4. 2: Exemplo em duas dimensões dos vectores de suporte [20].

Algumas das principais características das SVMs são:

- **Boa capacidade de generalização** – os classificadores gerados por uma SVM em geral alcança bons resultados em termo de generalização. Essa capacidade é medida por sua eficiência na classificação de dados que não pertençam ao conjunto utilizado no seu treino.
- **Robustez em grandes dimensões** – as SVMs são robustas mediante uma grande quantidade de dados.
- **Teoria bem definida** – as SVMs possuem uma base teórica bem estabelecida dentro da Matemática e Estatística.
- **Flexibilidade** – é possível adaptar o problema de classificação binária, que foi a abordagem base na formulação da SVM, para resolver muitos outros tipos de problemas, através do uso dos conceitos básicos de maximização de margem, dualidade e produto interno.

O exemplo descrito nas Figuras 4.1 e 4.2 é simplificado, pois só possui duas dimensões. Nesse exemplo, foi assumido que existem duas classes, sendo uma representada por rectângulos e a outra por círculos. Nesse caso, uma classe localiza-se no canto inferior esquerdo e a outra localiza-se no canto superior direito. As SVM tentam determinar um hiperplano com uma dimensão que separe as duas classes. Existem uma infinidade de hiperplanos que podem separar essas classes [20].

O hiperplano do lado esquerdo da Figura 4.1 tem uma margem de separação reduzida, enquanto o hiperplano do lado direito da mesma figura tem uma margem de separação superior. As linhas pontilhadas paralelas à linha divisória marcam a distância entre essa e os vectores mais próximos da linha. A distância entre as linhas pontilhadas é designada por margem [20]. Os vectores mais próximos às linhas pontilhadas são designados por vectores de suporte, como descrito no parágrafo anterior e ilustrados na Figura 4.2.

Uma SVM tenta achar um hiperplano de forma que o tamanho da margem seja máximo para que haja um maior limite de decisão para dados não lineares dentro do espaço de entrada.

As SVM utilizam funções denominadas kernel. Essas funções são capazes de mapear o conjunto de dados em diferentes espaços, fazendo com que um hiperplano possa ser usado para fazer a separação. Os principais tipos de funções kernel estão representados na tabela seguinte.

Tabela 4. 1: Exemplos de Kernels.

Kernel	Expressão	Parâmetros
Linear	$K(x_i, x_j) = x_i^T x_j$	τ
Polinomial	$K(x_i, x_j) = (\gamma x_i^T x_j + r)^d, \gamma > 0$	τ, d, γ
Sigmóide	$K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r)$	τ, r, γ
RBF	$K(x_i, x_j) = \exp(\gamma \ x_i - x_j\ ^2), \gamma > 0$	γ

A Figura 4.3 ilustra o exemplo de classificadores binários usando diferentes funções de kernel (Na ordem: Linear, RBF e Polinomial).

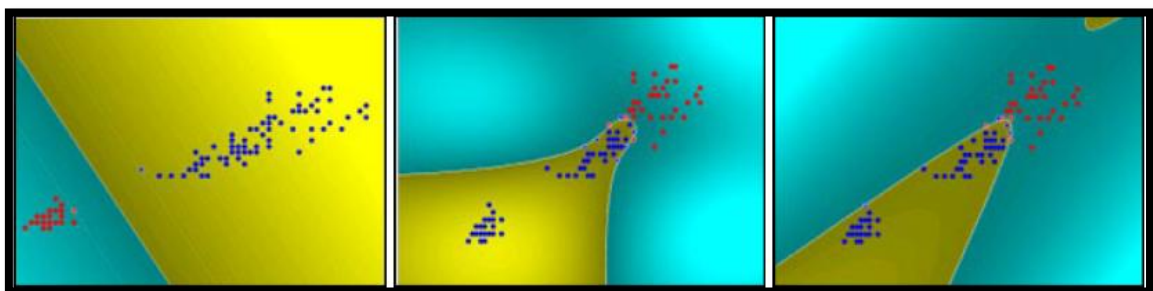


Figura 4. 3: Classificadores Binários usando diferentes funções Kernel [32].

A SVM tem vindo a resolver vários tipos de problemas. Alguns problemas mais conhecidos são abaixo referidos:

- **Regressão** – que foi o segundo problema a ser abordado por SVM, apenas com a modificação na formulação original da função objectivo para uma em que o erro medido pela distância do valor estimado em relação ao valor real é igual a zero para valores pequenos desta distância, e de valor crescente para quando a distância ao valor real é maior do que um determinado limiar, [27] [28];
- **Classificação Multi-Classe**, [19] [26];
- **Previsão de Séries Temporais**, [29];
- **Detecção de Novidade** (*Novelty Detection*), [30];
- **Estimação de Densidades**, [19] [31].

4.1.1 SVM para Problemas Multi-Classes

As SVMs são originalmente utilizadas para classificação dos dados em duas classes distintas. Estas podem ser denominadas “positivas” e “negativas”. Contudo, muitas aplicações envolvem o agrupamento em mais de duas classes. Diversas técnicas são propostas para estende-las em problemas multi-classe.

As duas abordagens usuais para a realização dessa tarefa, são denominadas, respectivamente, decomposição “**um-contra-todos**” e “**todos-contra-todos**”:

- **Decomposição “um-contra-todos”** - Uma abordagem usual para a solução de um problema multi-classe a partir de SVMs, consiste na geração de k SVMs, onde k é o número de classes [20]. Na criação de cada uma dessas máquinas, uma classe é fixada como positiva e as restantes como negativas. Este método é independente do algoritmo de aprendizagem utilizado no treino dos classificadores. O método 1-c-t tem a desvantagem de não ser possível prever limites no erro de generalização através de seu uso. Além disso, o seu tempo de treino é usualmente longo [20].
- **Decomposição “todos-contra-todos”** - Outra abordagem para solução de problemas multi-classe a partir de classificadores binários envolve a construção de $k(k-1)/2$ SVMs, separando cada classe da outra. Para unir estes classificadores, propõe-se o uso de um esquema de votação por maioria, em que cada um dos classificadores fornece uma classe como resultado. A solução final é dada pela classe que recebeu mais votos. Esta metodologia também não prevê limites no erro de generalização. Além disso, o tamanho dos classificadores gerados é, em geral, elevada e a avaliação de seu resultado pode ser lenta [20].

4.1.2 SVM com Opção de Rejeição

O facto de a opção de rejeição se revelar útil para proteger aplicações contra excesso de erros de classificação, torna-se importante ter uma SVM com a opção de rejeição incorporada.

Segundo o estudo feito por *Fumera* e *Roli* [4], uma SVM com a opção de rejeição é implementada usando duas abordagens. A primeira abordagem utiliza como medida de confiança de classificação, a distância $d(x)$ entre o dado de entrada x e o hiperplano de separação óptima que se situa no espaço características induzido pelo kernel escolhido.

A regra de rejeição consiste na exclusão de dados para os quais $d(x)$ seja inferior a um limiar pré-definido.

Yves Grandvalet e seus colaboradores [14] referiram que a abordagem proposta por *Fumera* e *Roli* não contemplavam algumas das características mais vantajosas das SVMs como convexidade e a dispersão. Estes autores introduziram um critério de formação linear por partes, dedicada ao problema de classificação com a opção de rejeição. A proposta é inspirada na interpretação probabilística de montagem da SVM e reflecte os dois limiares na regra de *Chow*.

4.2 Validação Cruzada

Validação cruzada [24] é uma técnica para estimar a capacidade de generalização de um classificador. Essa técnica consiste em dividir o conjunto de treino em K-partes aproximadamente iguais. Uma dessas partes constitui o subconjunto a ser utilizado para a fase de teste. Em cada execução, esse conjunto se modifica.

Um exemplo de validação cruzada [22, 23] consiste em dividir o conjunto total de dados em k grupos de tamanhos aproximadamente iguais (*k-fold cross validation*). Por isso, o treino é realizado k vezes deixando um dos grupos para teste em cada iteração. Ou seja, se $K = 5$, o treino ocorre cinco vezes e na primeira vez, o primeiro grupo será usado para teste e os outros quatro serão usados para o treino. Na segunda vez, o segundo grupo é usado para teste e os outros restante quatro são usados para o treino, e assim sucessivamente. Uma demonstração gráfica é apresentada na Figura 4.4.

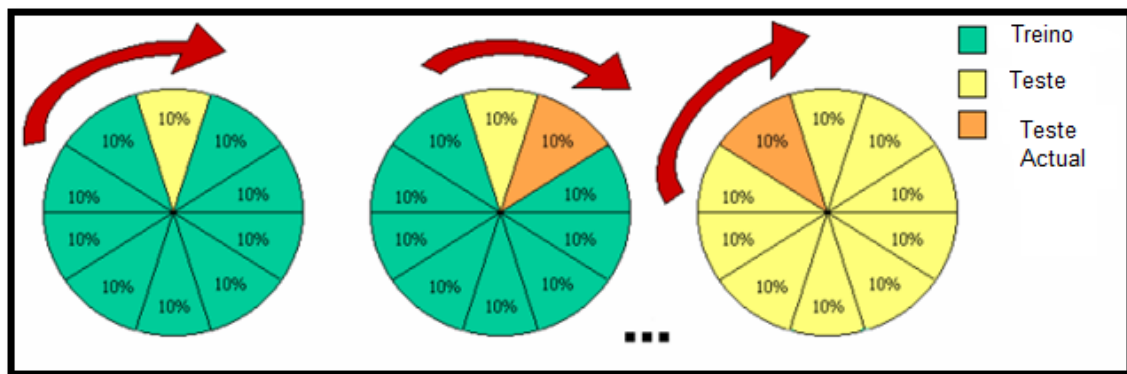


Figura 4. 4: Demonstração gráfica do procedimento de validação cruzada 10-fold [22].

A cada treino é calculado um erro de classificação, e no final é calculada a média desses erros para se obter um resultado final. Neste trabalho, os dados serão divididos em grupos de cinco, o que na literatura é denominada por validação cruzada 5-fold [22].

4.3 Classificação Multi-classe com Dados Ordinais

Como referido nas secções anteriores, um método existente na literatura que resolve o problema de classificação multi-classe com dados ordinais, é o método de *Frank and Hall* [2].

Este método foi implementado em ambiente Matlab e os respectivos testes conduziram aos mesmos resultados de desempenho obtidos pelo algoritmo *Frank and Hall* implementado na Weka.

Segundo *Frank and Hall*, o problema ordinal k-Classe é decomposto em k-1 problemas de classificação binária. A Figura 4.5 mostra o processo da decomposição de um problema multi-classe em k-1 problemas binários. Para treinar o classificador, é convertido o conjunto de dados ordinais com classes c_1, \dots, c_k num conjunto de dados binários, discriminando c_1, \dots, c_i contra c_{i+1}, \dots, c_k . Dada uma nova instância para prever o valor da classe, combina-se as k-1 saídas binárias para obter uma única estimativa.

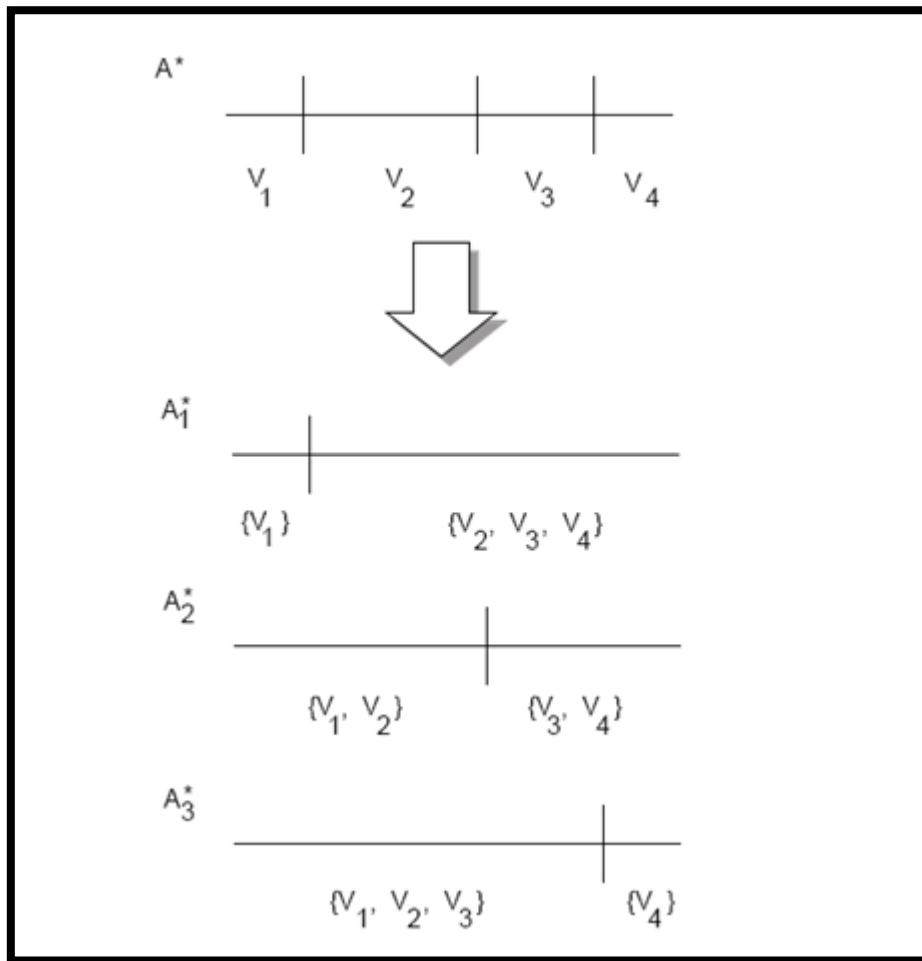


Figura 4. 5: Transformação de um problema multi-classe em $k-1$ problemas binárias [2].

Para melhor compreensão do método de *Frank and Hall*, considerou-se um exemplo em que o problema multi-classe contém quatro classes: as classes 1, 2, 3 e 4. Portanto, o problema de K classes (4 classes), é decomposto em $k-1$ problemas binários e cada uma destas decomposições denomina-se modelo.

Convencionou-se os rótulos -1 e 1 para representar o problema binário, isto é, para discriminar uma classe c_i das restantes classes, a classe c_i vai passar a ser representada por -1 e as restantes por 1.

A Tabela 4.2 apresenta o método de *Frank and Hall*, considerando um exemplo de quatro classes.

Tabela 4. 2: Exemplo de transformação de um problema multi-classe em k-1 problemas binárias.

Classes	{C1} {C2,C3,C4}	{C1,C2} {C3,C4}	{C1,C2,C3} {C4}
1	-1	-1	-1
1	-1	-1	-1
1	-1	-1	-1
1	-1	-1	-1
1	-1	-1	-1
2	1	-1	-1
2	1	-1	-1
2	1	-1	-1
3	1	1	-1
3	1	1	-1
3	1	1	-1
3	1	1	-1
4	1	1	1
4	1	1	1
4	1	1	1
4	1	1	1
4	1	1	1

A Tabela 4.2 apresenta um problema de quatro classes que foi decomposto em três problemas binários, isto é, três modelos. O primeiro modelo, $\{C1\} \{C2,C3,C4\}$, discrimina a classe C1 das restantes Classes: a classe C1 fica representada por -1 e as restantes por 1. O segundo modelo, $\{C1,C2\} \{C3,C4\}$, discrimina as classes C1 e C2 das restantes: as classes C1 e C2 ficam representadas por -1 e as outras por 1. Por último, o terceiro modelo, $\{C1,C2,C3\} \{C4\}$, discrimina as classes C1, C2 e C3 da classe C4: as classes C1, C2 e C3 ficam representados por -1 e a classe C4 por 1.

Neste caso, para prever o valor da classe a que uma nova instância pertence, realiza-se a combinação das três saídas binárias para obter uma estimativa.

4.4 Opção de Rejeição

Para integrar a opção de rejeição num problema de classificação multi-classe, transformou-se o problema multi-classe em k-1 problemas binárias, usando o algoritmo de *Frank and Hall*, o qual foi abordado na secção anterior. O equilíbrio entre a rejeição e o erro de uma classificação errada depende dos custos de rejeição associados.

A estratégia usada para integrar a opção de rejeição no problema multi-classe explicado anteriormente consistiu em associar os respectivos custos de rejeição W_r , para cada um dos modelos obtidos com o algoritmo de *Frank and Hall*. Se o processo de treino não considerar os custos de rejeição, pode-se obter o mesmo limite de classificação para todos os dados. Portanto, para evitar este facto, o custo de rejeição é alterado. Para uma melhor compreensão do processo da atribuição dos custos, é considerado o exemplo anterior. A Tabela 4.3 representa os modelos obtidos.

Tabela 4. 3: Exemplo de Modelos.

Modelos	Pontos da classe C1	Pontos da classe C2
Modelo 1	{C1}	{C2,C3,C4}
Modelo 2	{C1,C2}	{C3,C4};
Modelo 3	{C1,C2,C3}	{C4}

O problema de classificação multi-classe tem K modelos passa a ter 2K modelos quando integra-se a opção de rejeição, na medida em que uma vez atribuído um custo elevado (C_h) para as instâncias da classe C1 e um custo baixo (C_l) para as instâncias da classe C2, é necessário inverter os custos para as instâncias das classes deste mesmo modelo.

A Tabela 4.4 mostra o processo de atribuição do custo, no qual se verifica que o número de modelos se duplicou.

Tabela 4. 4: Processo de atribuição dos custos às classes.

Modelos	Pontos da classe C1	Pontos da classe C2
Modelo 1	{C1}; C_l	{C2,C3,C4}; C_h
Modelo 2	{C1}; C_h	{C2,C3,C4}; C_l
Modelo 3	{C1,C2}; C_l	{C3,C4}; C_h
Modelo 4	{C1,C2}; C_h	{C3,C4}; C_l
Modelo 5	{C1,C2,C3}; C_l	{C4}; C_h
Modelo 6	{C1,C2,C3}; C_h	{C4}; C_l

A Figura 4.6 ilustra o processo de rejeição de um dado.

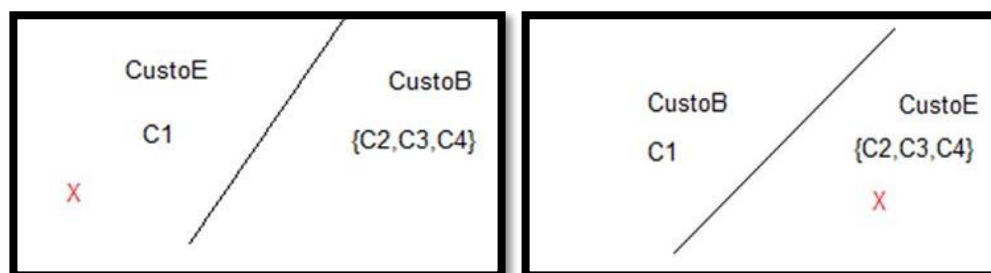


Figura 4. 6: Rejeição de uma instância.

Como mostra a Figura 4.6 dado uma instância X, se para o modelo 1 a instância pertence á classe C1 e para o modelo 2 a instância pertence á classe {C2,C3,C3}, ela vai ser rejeitada. Por este motivo é que foi necessário duplicar os modelos.

Depois de atribuir os custos aos modelos, para cada iteração da validação cruzada obtém-se um erro (E) e um *ratio* de rejeição (R), que é utilizada no cálculo do erro (P) para classificadores com a opção de rejeição. A equação 4.1 apresenta a fórmula do cálculo do erro (P).

$$P = W_r * R + E \quad (4.1)$$

No final do processo da validação cruzada determina-se a média dos erros (P) para comparar o melhor erro obtido até ao momento, e escolher os melhores parâmetros para a sua utilização

na fase de teste. O classificador é testado considerando estes parâmetros. Todos estes procedimentos vão ser repetidos para cada valor do W_r .

Como referido anteriormente, o processo de aprendizagem varia de acordo com os custos de rejeição associados.

Nas tabelas 4.5, 4.6 e 4.7 estão apresentados os processos de aprendizagem, com diferentes custos $W_r=0.04$, $W_r=0.24$ e $W_r=0.44$, respectivamente.

Tabela 4. 5: Processo de Aprendizagem com $W_r=0.04$.

Partição 1	Partição 2	Partição 3	Partição 4	Partição 5
5	3	4	4	3
5	5	5	2	3
2	3	2	1	7
5	3	3	3	3
5	3	6	3	2
3	7	2	3	3
3	5	3	1	5
5	3	5	3	3
3	4	3	3	
5	5	2	3	

Tabela 4. 6: Processo de Aprendizagem com $W_r=0.24$.

Partição 1	Partição 2	Partição 3	Partição 4	Partição 5
5	3	3	5	3
5	5	5	1	3
2	3	1	1	7
5	3	3	3	3
5	3	7	3	2
3	7	3	3	3
3	5	3	1	5
5	3	5	3	3
3	5	3	1	
5	6	2	3	

Tabela 4. 7: Processo de Aprendizagem com $W_r=0.44$

Partição 1	Partição 2	Partição 3	Partição 4	Partição 5
5	3	3	5	3
5	5	5	1	3
1	3	1	1	7
5	3	3	3	3
5	3	7	3	2
3	7	3	3	3
3	5	3	1	5
5	2	5	3	3
3	5	3	1	
5	7	1	3	

Foi considerado um *dataset* com 960 observações, numa experiência com 5% dos dados de treino e validação cruzada 5-fold, com uma gama de custo de rejeição num intervalo [0.04,0.24,0.44]. Cada coluna representa as classes de saída do classificador para uma determinada partição.

Analisando as três tabelas anteriores, verifica-se que o processo de aprendizagem está a variar de acordo com os custos de rejeição associados.

Capítulo 5

5. Resultados

Neste Capítulo são apresentados os resultados globais obtidos na execução do algoritmo implementado. Numa primeira secção são apresentados os *datasets* utilizados, nomeadamente o *dataset* BCCT e o *dataset* sintético. Ainda, são demonstrados os resultados da classificação binária, com opção de rejeição, existentes na literatura. Por último, será feita uma discussão dos resultados obtidos.

5.1 Dataset's Utilizados

O desempenho do algoritmo da classificação multi-classe com a opção de rejeição foi avaliado usando dois *datasets*: o *dataset* BCCT e o *dataset* sintético.

O *dataset* BCCT engloba 960 observações [25]. Este *dataset* apresenta a avaliação estética do tratamento conservador do cancro da mama. Para cada paciente submetido a BCCT, foram consideradas 30 medidas, relativamente a alterações visíveis na pele ou alterações no volume da mama ou ainda na forma da mama. O resultado estético do tratamento para todos os pacientes foi classificado numa das 4 classes: “Excelente”, “Bom”, “Suficiente” e “Mau” [3].

O segundo *dataset* foi sinteticamente gerado. Para o conjunto de dados sintéticos, foram gerados 400 pontos exemplo $x=[x_1 \ x_2]^t$ em $[0, 1] \times [0, 1] \subset \mathbb{R}^2$ de acordo com uma distribuição uniforme. Esta distribuição cria planaltos uniformemente distribuídos e uma zona de transição de probabilidade que diminui de forma linear e é delimitado por fronteiras hiperbólicas [3]. Foram considerados 3 classes.

5.2 Resultados da Classificação Binária com a Opção de Rejeição

Conforme citado anteriormente, a classificação multi-classe com a opção de rejeição usando dados ordinais é um problema pouco abordado na literatura, apesar de existirem vários estudos. O método de classificação binária com a opção de rejeição usando dados ordinais é o método que mais se enquadra neste tema de trabalho. Os resultados obtidos na execução deste método não são suficientes para fazer uma comparação rigorosa com os resultados obtidos no método desenvolvido nesta dissertação, visto que são problemas diferentes.

Na literatura, verifica-se que independentemente de ser um problema multi-classe ou um problema binário, os parâmetros Erro, Rejeição, Custo de Rejeição e o Desempenho, apresentam o mesmo comportamento para ambos os métodos.

Para este método também usaram-se o *dataset* BCCT e o *dataset* sintético. Para o *dataset* BCCT o problema multi-classe foi transformado em binário, sendo as classes “Excelente” e “Bom” transformadas em apenas uma, e as classes “Suficiente” e “Mau” transformadas noutra.

Os dados foram divididos em conjuntos de treino e de teste, com 5% e 95% dos dados respectivamente. Esta divisão dos dados em conjuntos de treino e de teste foi repetida cem vezes a fim de obter resultados mais estáveis.

5.2.1 Dataset BCCT

Os resultados obtidos usando o dataset BCCT numa experiência com 5% dos dados de treino e 95% dos dados de teste, validação cruzada 5-fold, com uma gama de custo de rejeição num intervalo [0.04, 0.24, 0.44]. A Tabela 5.1 apresenta, de uma forma geral, os resultados obtidos usando o *dataset* BCCT.

Tabela 5. 1: Dataset BCCT.

Wr	Rejeição	Erro	Desempenho
0.04	0.99	0.007	0.934
0.24	0.78	0.082	0.918
0.44	0.08	0.335	0.665

As Figuras 5.1, 5.2, 5.3 e 5.4 ilustram as relações existentes entre o custo da rejeição W_r , a rejeição, o erro e o desempenho.

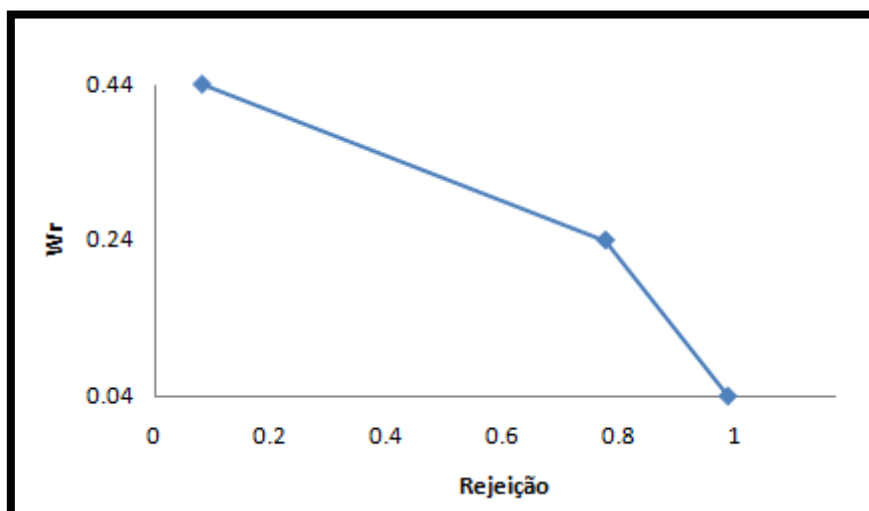


Figura 5. 1: Variação de Rejeição em relação ao Custo de Rejeição

A Figura 5.1 ilustra a relação existente entre o custo de rejeição W_r e a rejeição. Como mostra a figura, à medida que o custo de rejeição aumenta, o número de casos rejeitados diminui.

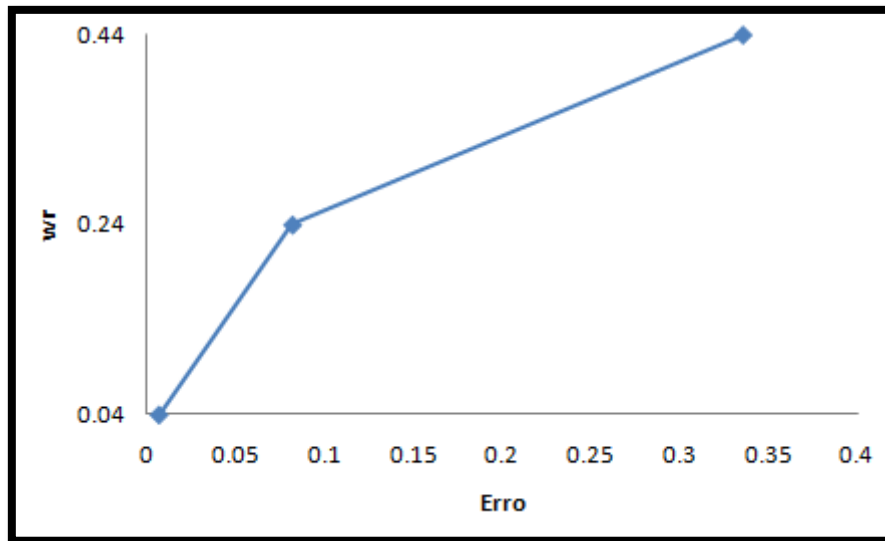


Figura 5. 2: Variação do Erro em relação ao Custo de Rejeição.

A Figura 5.2 ilustra a relação existente entre o erro e o custo de rejeição W_r . Quando o custo de rejeição aumenta o erro diminui, na medida em que quanto maior é o custo de rejeição W_r , menor é o número de casos rejeitados e logo maior é a percentagem de dados mal classificados.

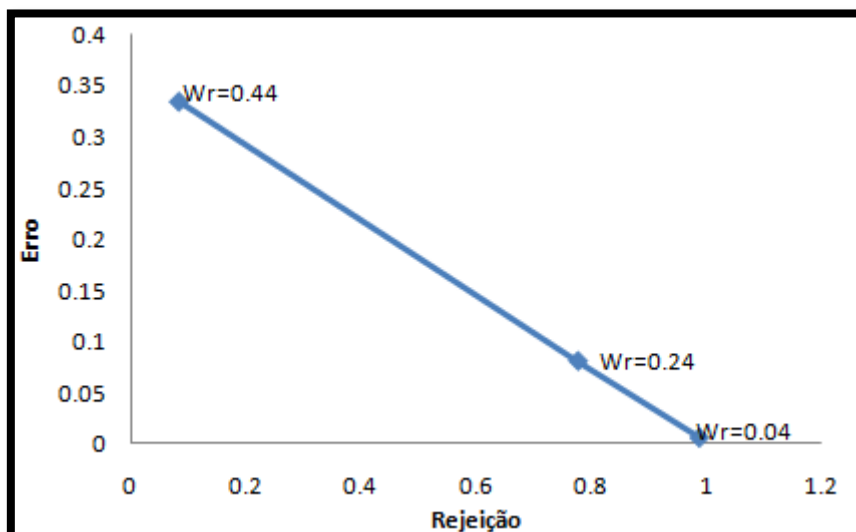


Figura 5. 3: Variação do Erro em relação à taxa de Rejeição.

A Figura 5.3 ilustra a variação do erro em relação à taxa de rejeição. A relação existente entre o erro e a rejeição depende do custo de rejeição. Como mostra a figura, à medida que a

rejeição aumenta o número de erros diminui, na medida em que quanto maior é o número de casos rejeitados menor será a probabilidade de verificar instâncias mal classificadas. Como explicado nas secções anteriores, a grande vantagem da opção de rejeição é precisamente reduzir o erro de classificação.

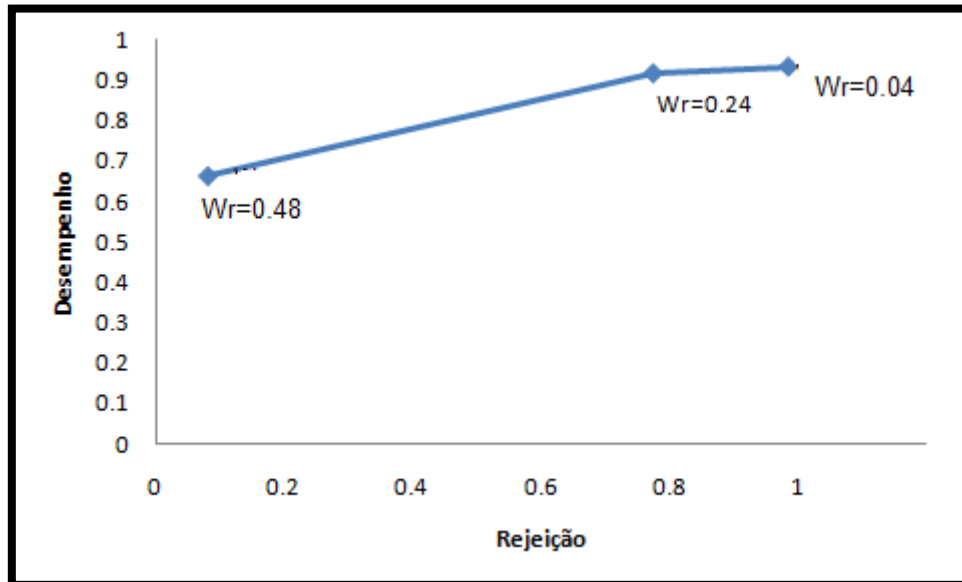


Figura 5. 4: Variação do Desempenho em relação da taxa de Rejeição.

A Figura 5.4 ilustra a relação existente entre o desempenho e a taxa de rejeição. Esta relação depende do custo de rejeição. Quando maior é a taxa de rejeição maior é o desempenho, dado que quanto maior é o número de casos rejeitados menor é o erro de classificação e portanto, maior será o desempenho do classificador.

5.2.2 Dataset Sintético

Usando o *dataset* sintético numa experiência com 5% dos dados de treino e 95% dos dados de testes, validação cruzada 5-fold, com uma gama de custo de rejeição num intervalo de [0.04;0.48] com passos de 0.04. A tabela 5.2 apresenta os resultados obtidos.

Tabela 5. 2: Dataset Sintético.

Wr	Rejeição	Erro	Desempenho
0.04	0.87	0.0304	0.970
0.08	0.76	0.048	0.952
0.12	0.64	0.073	0.927
0.16	0.44	0.109	0.891
0.2	0.35	0.132	0.868
0.24	0.23	0.154	0.846
0.28	0.14	0.183	0.817
0.32	0.09	0.201	0.799
0.36	0.06	0.213	0.787
0.4	0.05	0.214	0.786
0.44	0.03	0.221	0.779
0.48	0.01	0.232	0.768

As Figuras 5.5, 5.6, 5.7 e 5.8 ilustram as relações existentes entre o custo da rejeição W_r , a rejeição, o erro e o desempenho.

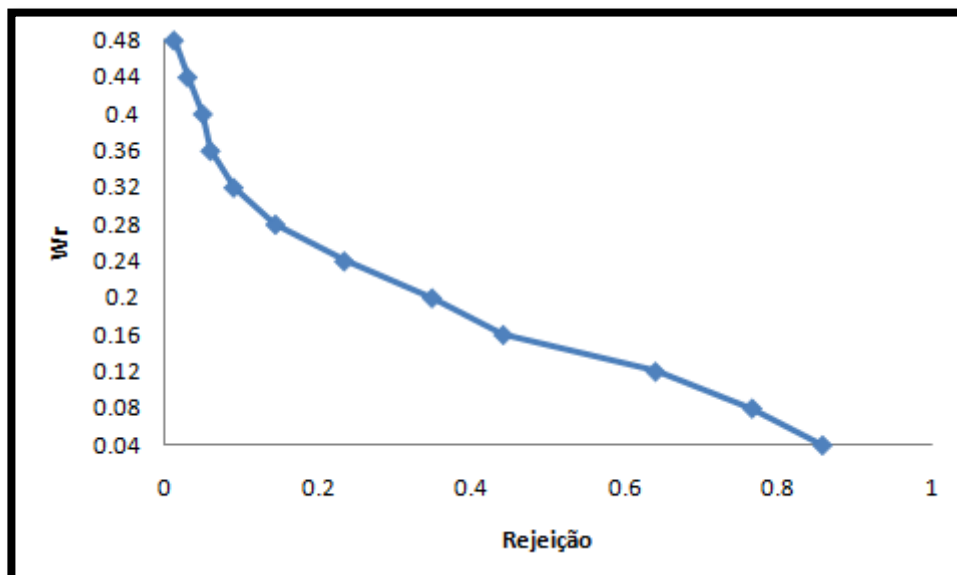


Figura 5. 5: Variação de Rejeição em relação ao Custo de Rejeição.

Na Figura 5.5 ilustra a variação de rejeição em relação ao custo de rejeição. Como mostra a curva do gráfico da figura, a rejeição está a diminuir com o aumento do custo de rejeição.

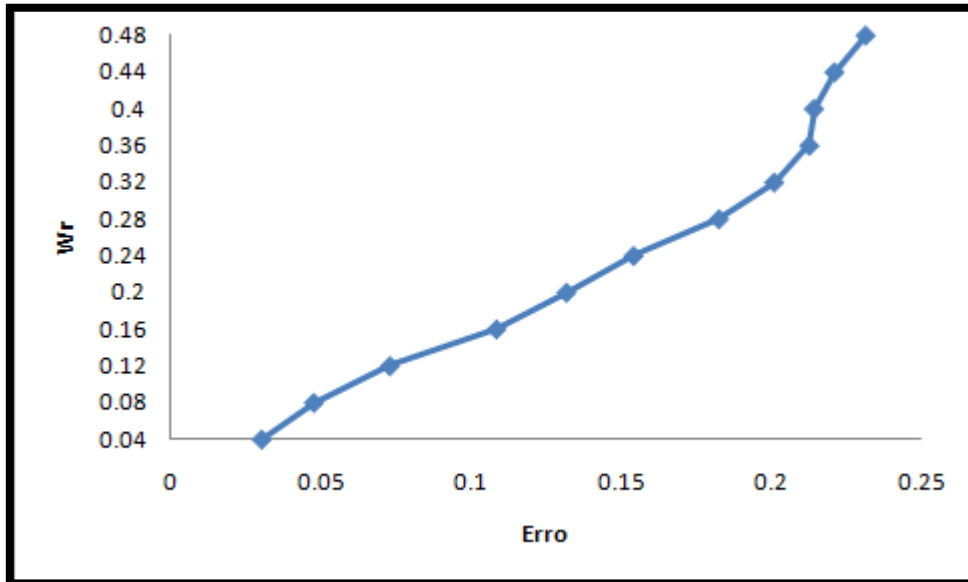


Figura 5. 6: Variação do Erro em relação ao Custo de Rejeição.

A Figura 5.6 ilustra a relação existente entre o erro e o custo de rejeição. Quando o custo de rejeição aumenta o erro diminui, devido às razões descritas anteriormente na Figura 5.2.

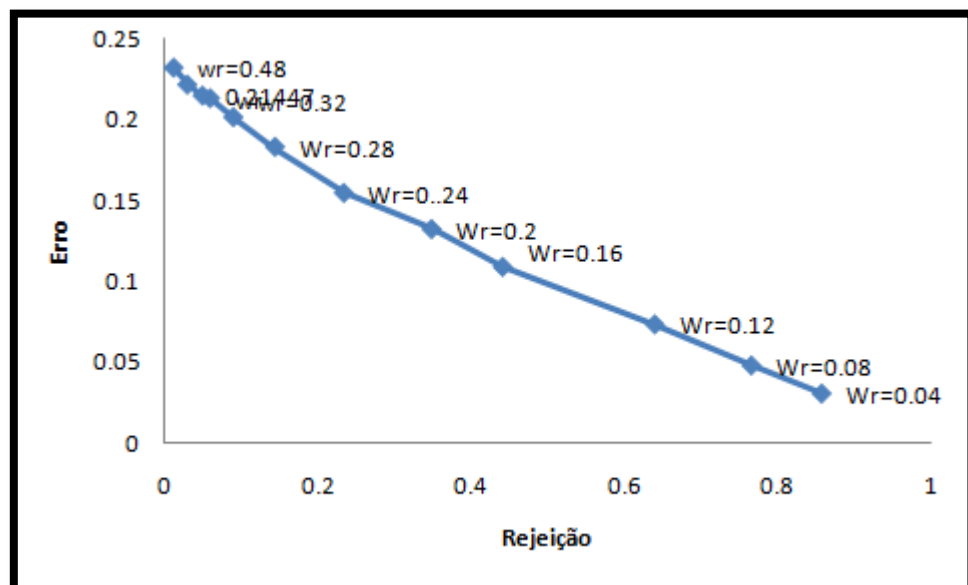


Figura 5. 7: Variação do Erro em relação da taxa de Rejeição.

A Figura 5.7 ilustra a variação do erro em função da taxa de Rejeição. O erro diminui à medida que a taxa de rejeição aumenta, devido às razões descritas anteriormente na Figura 5.3.

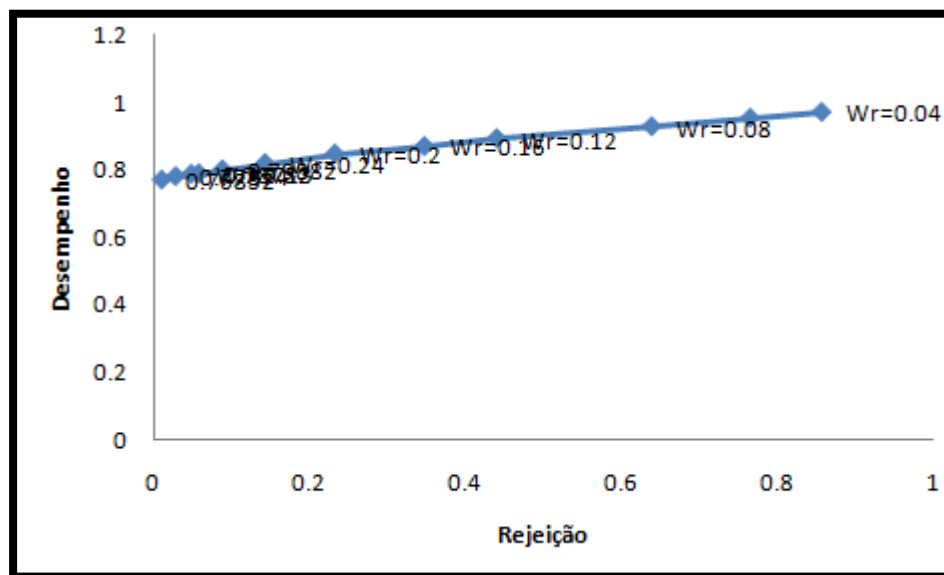


Figura 5. 8: Variação do Desempenho em relação da taxa de Rejeição.

A Figura 5.8 representa a relação existente entre o desempenho e a taxa de rejeição. Esta relação depende do custo de rejeição. Quando maior é a taxa de rejeição maior é o desempenho, pelas razões descritas na Figuras 5.4.

5.3 Resultados da Classificação Multi-classe com a Opção de Rejeição

Como referido anteriormente, o desempenho do algoritmo da classificação multi-classe com a opção de rejeição foi avaliado usando dois *datasets*: o *dataset* BCCT e o *dataset* Sintético.

Os gráficos das figuras apresentadas a seguir tem os mesmos comportamentos que os gráficos da classificação binária com a opção de rejeição.

5.3.1 Dataset BCCT

Os resultados obtidos usando o dataset BCCT numa experiência com 5% dos dados de treino e 95% dos dados de testes, validação cruzada 5-fold, com uma gama de custo de rejeição num intervalo [0.04, 0.24, 0.44], são apresentados na Tabela 5.3.

Tabela 5. 3: Dataset BCCT.

Wr	Rejeição	Erro	Desempenho
0.04	0.43	0.102	0.898
0.24	0.16	0.184	0.819
0.44	0.06	0.236	0.793

As Figuras 5.9, 5.10, 5.11 e 5.12 ilustram, de uma forma geral, a variação do custo da rejeição W_r , rejeição, erro e do desempenho.

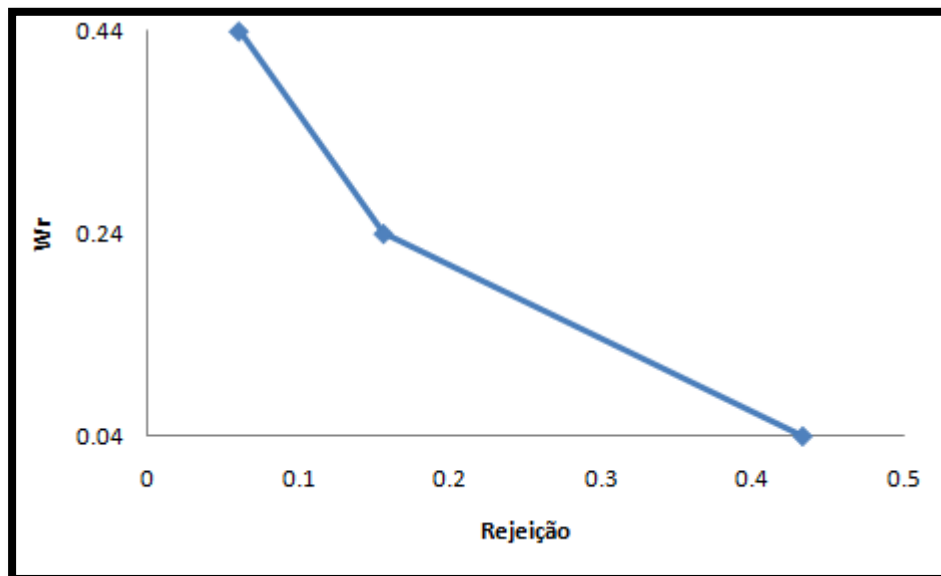


Figura 5. 9: Variação de Rejeição em relação ao Custo de Rejeição.

Na Figura 5.9 ilustra a variação de rejeição em relação ao custo de rejeição W_r . Como mostra a curva, a rejeição diminui em função custo de rejeição.

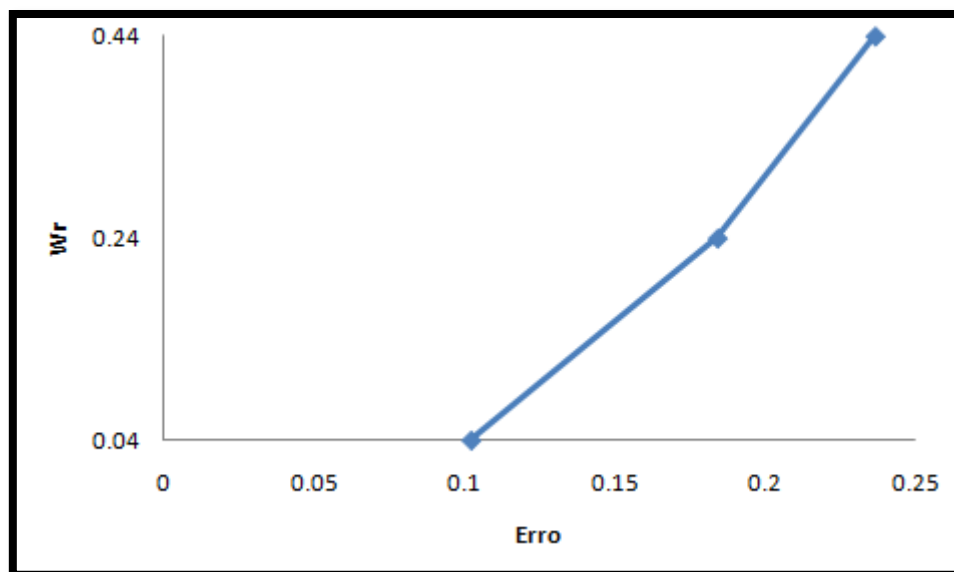


Figura 5. 10: Variação do Erro em relação ao Custo de Rejeição.

A Figura 5.10 ilustra a relação existente entre o erro e o custo de rejeição. Quando o custo de rejeição aumenta, o erro diminui, na medida em que quanto maior é o custo de rejeição W_r , menor é o número de casos rejeitados e logo maior é a percentagem de dados mal classificados.

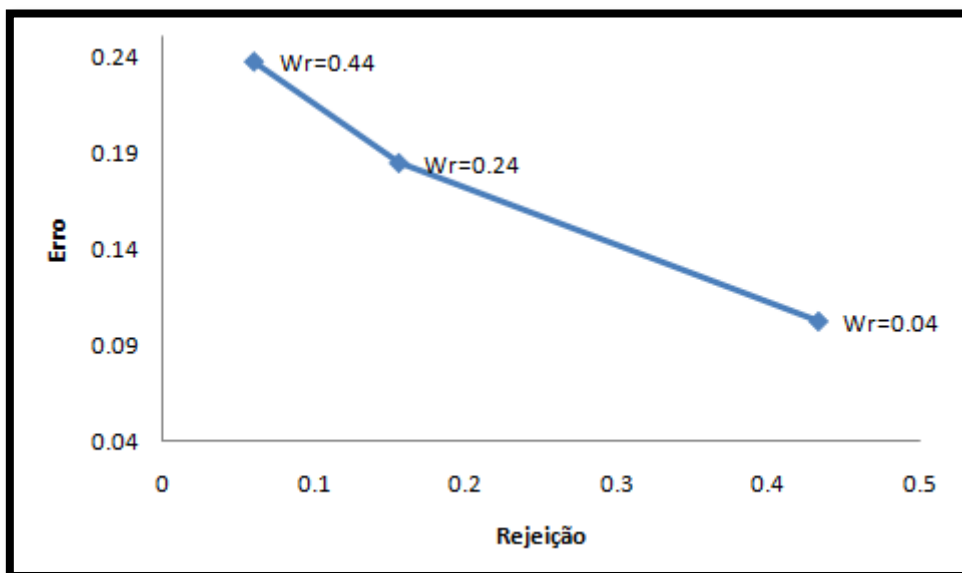


Figura 5. 11: Variação do Erro em relação da taxa de Rejeição.

A Figura 5.11 ilustra a variação do erro em relação à taxa de rejeição. À medida que a rejeição aumenta, o erro diminui. Como mostra a figura, à medida que a rejeição aumenta o número de erros diminui, na medida em que quanto maior é o número de casos rejeitados menor será a probabilidade de verificar instâncias mal classificados.

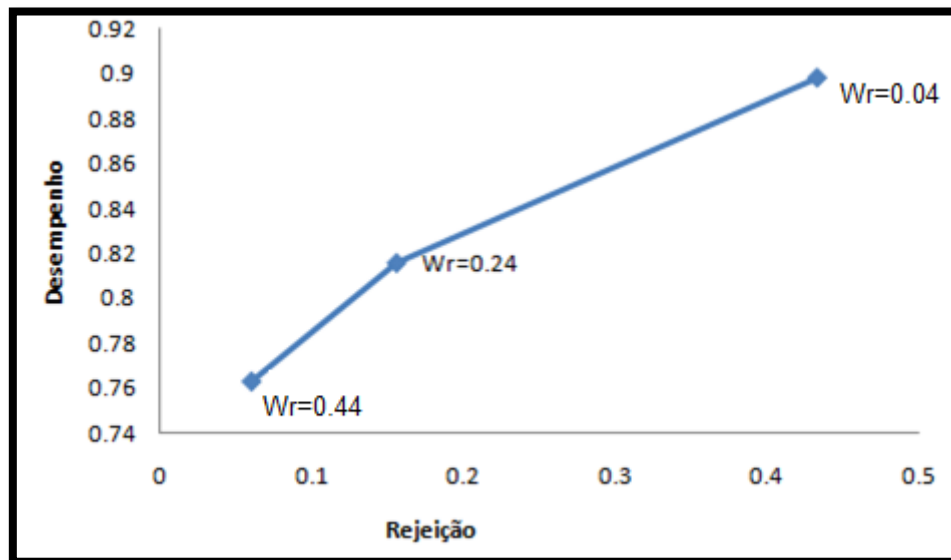


Figura 5. 12: Variação do Desempenho em relação da taxa de Rejeição.

A Figura 5.12 ilustra a relação existente entre o desempenho e a taxa de rejeição. Esta relação depende do custo de rejeição. Quanto maior é a taxa de rejeição, maior é o seu desempenho, dado que quanto maior é o número de casos rejeitados menor é o erro de classificação e portanto, maior será o desempenho do classificador.

5.3.2 Dataset Sintético

Os Resultados obtidos usando o dataset sintético, numa experiência com 5% dos dados de treino e 95% dos dados de testes, validação cruzada 5-fold, com uma gama de custo de rejeição num intervalo [0.04;0.48] com passo de 0.04. são apresentados na Tabela 5.4.

Tabela 5. 4: Dataset Sintético.

Wr	Rejeição	Erro	Desempenho
0.04	0.55	0.099	0.900
0.08	0.52	0.106	0.894
0.12	0.47	0.102	0.898
0.16	0.36	0.123	0.877
0.2	0.26	0.141	0.859
0.24	0.20	0.155	0.845
0.28	0.13	0.169	0.830
0.32	0.10	0.179	0.821
0.36	0.07	0.194	0.806
0.4	0.04	0.205	0.794
0.44	0.03	0.216	0.784
0.48	0.01	0.224	0.776

As Figuras 5.13, 5.14, 5.15 e 5.16 ilustram, de uma forma geral, a variação do custo de rejeição, a rejeição, o erro e do desempenho.

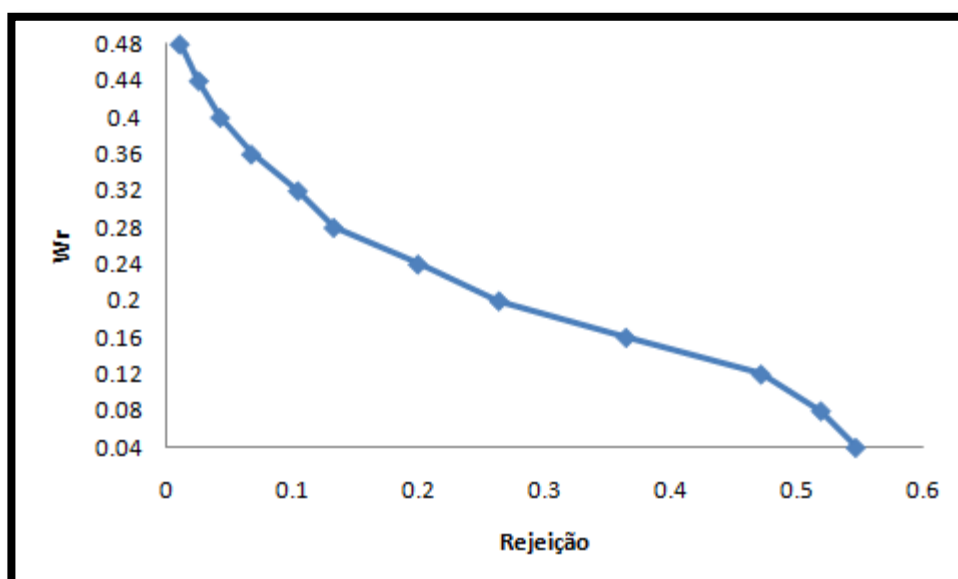


Figura 5. 13: Variação de Rejeição em relação ao Custo de Rejeição.

Na Figura 5.13 ilustra a variação de rejeição em relação ao custo de rejeição W_r . Como demonstra a curva descrita no gráfico, a rejeição diminui com o aumento do custo de rejeição.

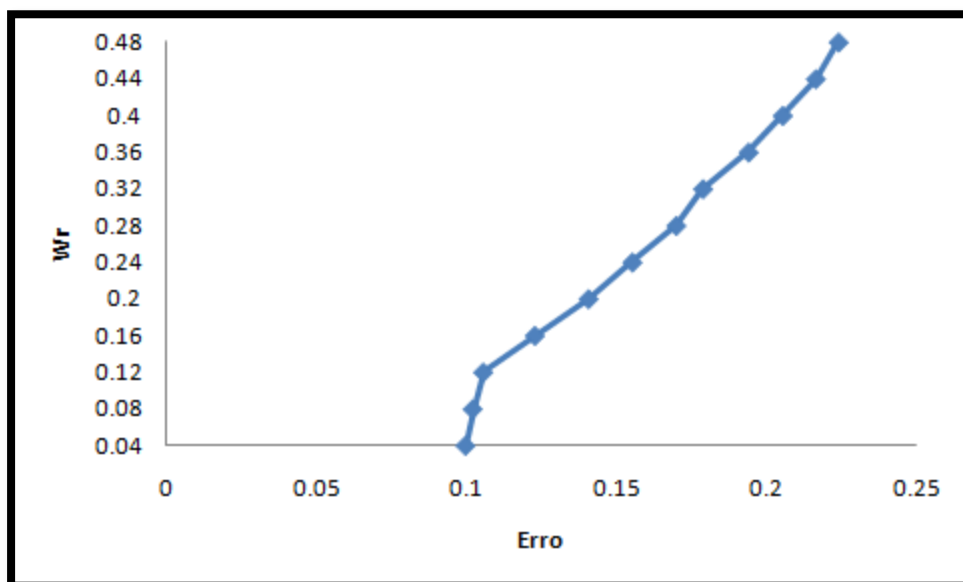


Figura 5. 14: Variação do Erro em relação ao Custo de Rejeição.

A Figura 5.14 ilustra a relação existente entre o erro e o custo de rejeição W_r . Quando o custo de rejeição aumenta, o erro diminui, devido às razões descritas anteriormente na Figura 5.10.

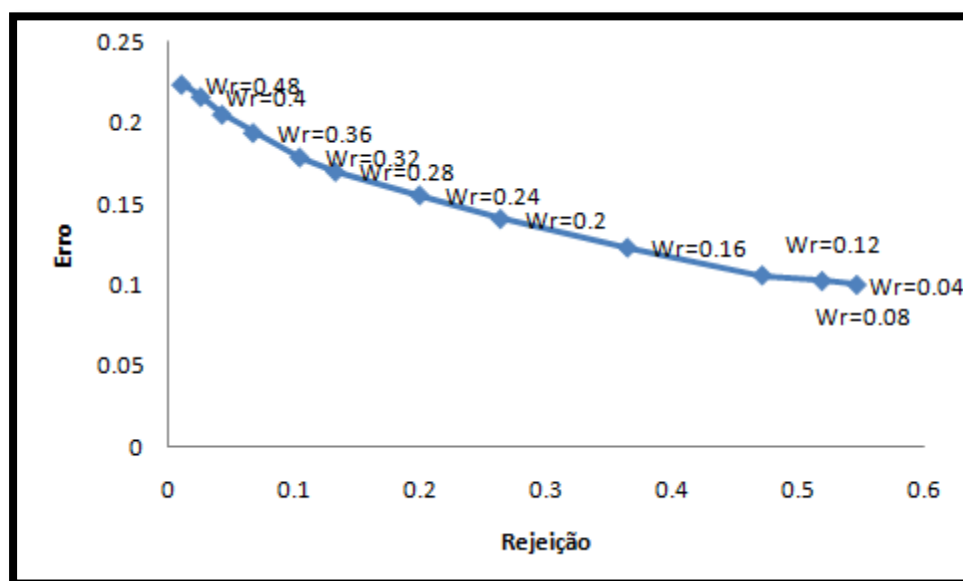


Figura 5. 15:Variação do Erro em relação dá taxa de Rejeição.

A Figura 5.15 ilustra a variação do erro em função da taxa de rejeição. A relação existente entre o erro e a rejeição depende do custo de rejeição. À medida que a rejeição aumenta, o número de erro diminui, devido às razões descritas anteriormente na Figura 5.11.

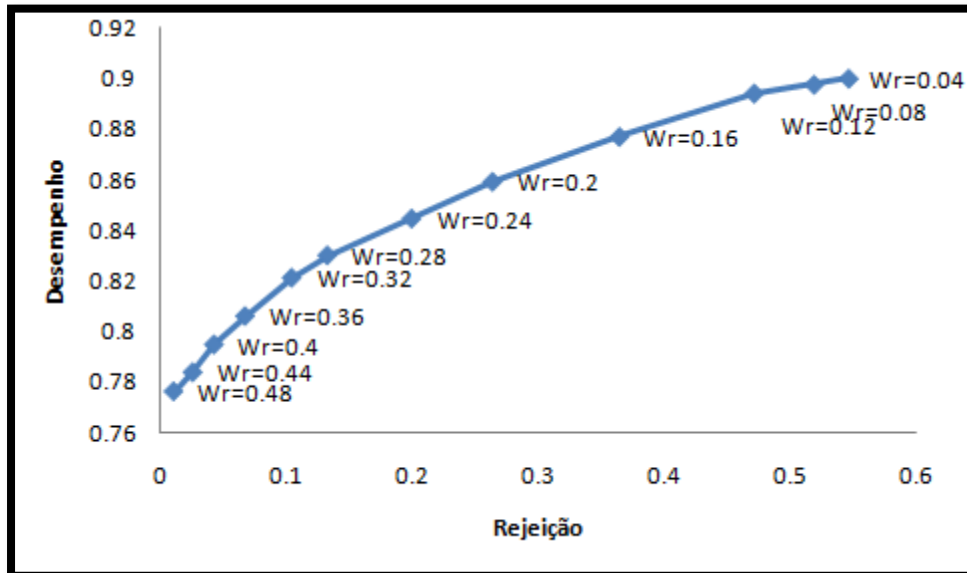


Figura 5. 16: Variação do Desempenho em relação da taxa de Rejeição.

A Figura 5.16 ilustra a relação existente entre o desempenho e a taxa de rejeição. Esta relação depende do custo de rejeição. Quanto maior é a taxa de rejeição maior é o desempenho, pelas razões descritas na Figuras 5.12.

5.4 Discussão dos Resultados Obtidos

Na literatura, o problema da classificação multi-classe com a opção de rejeição usando dados ordinais é pouco referido não sendo possível realizar uma comparação rigorosa dos resultados obtidos com a execução do método desenvolvido nesta dissertação.

O método de classificação binária com a opção de rejeição usando dados ordinais contribui significativamente para a validação dos resultados obtidos do método desenvolvido nesta dissertação.

Os resultados não foram comparados em termos quantitativos, dado que um método trata de problemas binários e o outro está relacionado com os problemas multi-classe.

Contudo, foi possível analisar as relações existentes entre os valores do custo de rejeição, a rejeição, o erro e o desempenho. Para ambos os métodos, a relação existente entre estes parâmetros tiveram o mesmo comportamento.

Os resultados obtidos confirmaram as teorias sobre a importância da opção de rejeição nos problemas de classificação, quer o problema seja binário ou multi-classe. É possível afirmar que a opção de rejeição efectivamente diminui o erro de rejeição, aumentando assim o desempenho.

Capítulo 6

6. Conclusões e Futuros Trabalhos

6.1 Conclusões

No decurso desta dissertação foi implementado um algoritmo de classificação multi-classe com a opção de rejeição usando dados ordinais. A realização desta tese, traz contribuições para os problemas de classificação multi-classe. Estas contribuições são: Capacidade de detectar regiões de rejeição nos problemas de classificação multi-classe onde as classes de saída estão ordenadas; não sendo necessário definir limiares e atribuir probabilidades para definir regiões de confiança e proporcionando a eliminação do problema em que um dado pode pertencer a várias classes.

Efectivamente, a opção de rejeição é um método muito útil para melhorar o desempenho no processo de classificação.

Com maior ou menor grau de dificuldade, os objectivos foram cumpridos deixando sempre soluções plausíveis de optimização para futuros trabalhos.

6.2 Trabalhos Futuros

Como referido anteriormente, devido ao facto de ainda ser pouco referido na literatura os métodos que implementem o problema da classificação multi-classe com a opção de rejeição usando dados ordinais, não foi possível fazer uma análise comparativa rigorosa dos resultados obtidos com a execução do método desenvolvido nesta dissertação.

Existe ainda espaço de investigação para elaborar novas abordagens da classificação multi-classe com a opção de rejeição usando dados ordinais, podendo num estudo futuro realizar-se uma comparação rigorosa do método desenvolvido nesta dissertação com outros futuros métodos a serem desenvolvidos.

Referências Bibliográficas

- [1] Adriana Chaves, “Extracção de Regras de Fuzzy para Máquinas de Vector Suporte (SVM) para Classificação em Múltiplas Classes”, pp. 13-16, 2004.
- [2] Eibe Frank and Mark Hall, “A Simple Approach to Ordinal Classification”, Working Paper Series ISSN 1170-487X, pp. 2-5, 2001.
- [3] Ricardo Sousa, Beatriz Mora and Jaime S. Cardoso, “An Ordinal Data Method for the Classification with Reject Option”, pp.1-5.
- [4] G. Fumera and F. Roli, “Support Vector Machines with Embedded Reject Option,” in SVM '02: Proceedings of the First International Workshop on Pattern Recognition with Support Vector Machines. London, UK: Springer-Verlag, pp. 68–82, 2002.
- [5] A. Bounsiar, P. Beuseroy, and E. Grall-Maes, “General solution and learning method for binary classification with performance constraints”, Pattern Recognition Letters, vol. 29, no. 10, pp. 1455–1465, 2008.
- [6] Andréia Marini and Alessandro L. Koerich, “Estratégias de Rejeição para Problemas Multiclasses”, XXXIV Conferencia Latinoamericana de Informática (CLEI 2008), pp. 162-171, 2008.
- [7] C. K. Chow, “On Optimum Recognition Error and Reject Tradeoff”, In IEEE Transactions in Information Thoery, 16, pp.41-46, 1970.
- [8] Daniel Gomes, “Detecção de intrusão em Redes de Computadores utilizando classificadores One-Class ”, Tese, pp.31-39, 2006.
- [9] André Gonçalves, “Aplicações de Técnicas de Aprendizagem de Máquinas”, Tese, pp.30-31, 2005.
- [10] Ilueny Santos and Robinson Alves, “Estudo para Identificação de Promotores Procarióticos Através de Classificação Bayesiana”, Tese, pp.1-3, 2008.
- [11] V. V. Corina Cortes. “Support-vector networks. Machine Learning “, pp.7-8, 1995.
- [12] A. Bounsiar, P. Beuseroy, and E. Grall-Maes, “General solution and learning method for binary classification with performance constraints,” Pattern Recognition Letters, vol. 29, no. 10, pp. 1455–1465, 2008.
- [13] José Fonseca, “Indução de Arvores de Decisão”, Tese, pp.22. 1994.
- [14] Yves Grandvalet, Alain Rakotomamonjy, Joseph Keshet and Stéphane Canu; “Support Vector Machines with a Reject Option,” in Computer Science, vol. 2388/2002, pp.2-6, 2002.
- [15] Vapnik, VN e Chervonenkis, “A convergência uniforme das frequências relativas dos eventos de suas probabilidades. Probabilidade Teórica e suas Aplicações ”, vol. 17, pp. 264-280, 1971.

- [16] Marcelo Corrêa, Cairo Lúcio and Thiago Marques, “Técnicas de Redes Neurais para Reconhecimento de Caracteres: Um estudo comparativo”, Relatório Anual, pp.1-4, 2006.
- [17] C.-C. Chang and C.-J.Lin,” LIBSVM: a library for support vector machines”, 2001. Disponível em: <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.
- [18] Breno Flesch Franco,” Classificação Hierárquica de Caracteres Matemáticos”, Tese, 2009, pp.7.
- [19] Vapnik, V., “Statistical Learning Theory, Wiley”, New York, pp.156-160, 1998.
- [20] “SVM- Suport Vector Machines”. Disponível em: <http://www.dtreg.com/svm.htm> .
- [21]C. K. Chow. “On optimum error and reject tradeoff. IEEE Trans.Inform. Theory”, pp.41–46, 1970.
- [22] Carolina Baldisserotto, “Técnicas de aprendizagem máquina para previsão de sucesso em implantes dentários”, Tese, pp. 29-30, 2005.
- [23] Duda, R. O., Hart, P. E. e Stork, D. G. “Pattern Classification”, Wiley Interscience, vol.27, pp.305-307, 2007.
- [24] Peña, J. M.; BJORKEGREN J.; TEGNER, J.” Learning dynamic Bayesian network models via cross-validation”, Pattern Recognition Letters, vol. 26, pp. 2295-2308, 2005.
- [25] J. S. Cardoso and M. J. Cardoso, “Towards an intelligent medical systemfor the aesthetic evaluation of breast cancer conservative treatment”, Artificial Intelligence in Medicine, vol. 40, pp. 115–126, 2007.
- [26] Weston, J. and Watkins, C. “Multi Class Support Vector Machines, in Proceedings of ESANN99, ed. M. Verleysen, D. Facto Press, Brussels”, pp. 219-224, 1999.
- [27] Vapnik, V.N., “The Nature of Statistical Learning Theory. Springer” ,1995. Disponível em: <http://cscs.umich.edu/~crshalizi/reviews/vapnik-nature/>
- [28] Cristianini, N. and Shawe-Taylor, J., “An Introduction to Support Vector Machines. Cambridge U.P”, pp.5-6, 2000.
- [29] Muller, K., Smola, A., Ratsh, G., Scholkopf, B., Kohlmorgen, J. and Vapnik, V.N., “Predicting time series with support vector machines. Advances in Kernel Methods - Support Vector Learning”, pp. 243-253, 1999.
- [30] Bennett, K. and Campbell, C., “Support Vector Machines: Hype or Hallelujah. SIGKDD Explorations”, vol. 2, pp.1-13, 2000.
- [31] Vapnik, V.N. and Mukherjee, S., “Support vector method for multivariant density estimation. In Neural Information Processing Sytems”, pp.1-4, 1999.
- [32] Nilsson, N. J. “Learning Machines. New York: McGraw Hill”,pp.5-6, 1965.
- [33] G. Fumera and F. Roli and G. Giacinto, “Reject option with multiple thresholds”, Pattern Recognition Letters, vol.33, pp.2099-2101, 2000.
- [34] C. M. Bishop. “Neural Networks for Pattern Recognition”. Oxford : Clarendon, pp.5-6, 1996.