

DISSERTAÇÃO DE MESTRADO Nº 989
DEFINIÇÃO DE PARÂMETROS DE RBF UTILIZANDO GRAFO DE GABRIEL

Marcelo de Oliveira Queiroz

DATA DA DEFESA: 06/07/2017

Universidade Federal de Minas Gerais

Escola de Engenharia

Programa de Pós-Graduação em Engenharia Elétrica

**DEFINIÇÃO DE PARÂMETROS DE RBF UTILIZANDO GRAFO DE
GABRIEL**

Marcelo de Oliveira Queiroz

Dissertação de Mestrado submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em Engenharia Elétrica da Escola de Engenharia da Universidade Federal de Minas Gerais, como requisito para obtenção do Título de Mestre em Engenharia Elétrica.

Orientador: Prof. Frederico Gualberto Ferreira Coelho

Coorientador: Prof. Luiz Carlos Bamberira Torres

Coorientador: Prof. Antônio de Pádua Braga

Belo Horizonte - MG

Julho de 2017

Q3d

Queiroz, Marcelo de Oliveira.

Definição de parâmetros de RBF utilizando grafo de Gabriel
[manuscrito] / Marcelo de Oliveira Queiroz. – 2017.

xx, 89 f., enc.: il.

Orientador: Frederico Gualberto Ferreira Coelho.

Coorientadores: Luiz Carlos Bamber Torres, Antônio de Pádua
Braga.

Dissertação (mestrado) - Universidade Federal de Minas Gerais,
Escola de Engenharia.

Bibliografia: f. 85-89.

1. Engenharia elétrica - Teses. 2. Classificação - Teses. 3. Teoria dos grafos - Teses. 4. Teoria dos conjuntos - Teses. I. Coelho, Frederico Gualberto Ferreira. II. Torres, Luiz Carlos Bamber. III. Braga, Antônio de Pádua. IV. Universidade Federal de Minas Gerais. Escola de Engenharia. V. Título.

CDU: 621.3(043)

"Definição de Parâmetros de RBF Utilizando Grafo de Gabriel"

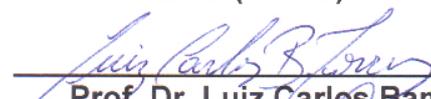
Marcelo de Oliveira Queiroz

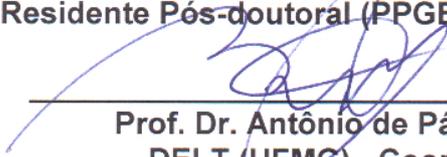
Dissertação de Mestrado submetida à Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação em Engenharia Elétrica da Escola de Engenharia da Universidade Federal de Minas Gerais, como requisito para obtenção do grau de Mestre em Engenharia Elétrica.

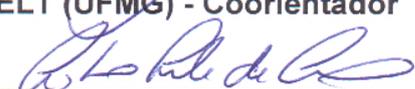
Aprovada em 06 de julho de 2017.

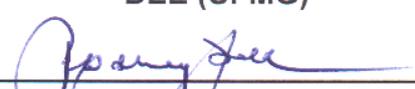
Por:


Prof. Dr. Frederico Gualberto Ferreira Coelho
DEE (UFMG) - Orientador


Prof. Dr. Luiz Carlos Bambirra Torres
DELT (Residente Pós-doutoral (PPGEE/UFMG)) - Coorientador


Prof. Dr. Antônio de Pádua Braga
DELT (UFMG) - Coorientador


Prof. Dr. Cristiano Leite de Castro
DEE (UFMG)


Prof. Dr. Rodney Rezende Saldanha
DEE (UFMG)

Resumo

O uso de RBF (*Radial Basis Function*) em problemas de classificação de amostras tem sido uma abordagem muito discutida na literatura. Suas diversas aplicações permitem tratar problemas de complexidades diferentes. Dentre as funções radiais mais aplicadas a gaussiana é considerada uma das mais eficientes pela sua simplicidade de configuração. No entanto um dos desafios no uso deste tipo de função é definir os parâmetros c e σ , centro e raio respectivamente, adequados para evitar sub ou superdimensionamento na solução do problema. Este trabalho propõe uma metodologia baseada no grafo de Gabriel e na teoria dos conjuntos dominantes para encontrar estes parâmetros para um dado conjunto de amostras, sem a necessidade de arbitragem. Em uma segunda etapa, aplicam-se as gaussianas encontradas em uma arquitetura de redes neurais artificiais com o objetivo de classificar estas amostras. Em uma terceira etapa, comparam-se os resultados encontrados com os de classificadores clássicos conhecidos na literatura. Uma vez confrontados estes resultados, analisam-se as particularidades de cada problema estudado e suas influências sobre as métricas da metodologia proposta o que pode criar uma necessidade de adaptá-las para o tratamento de algumas destas particularidades. Dentre as particularidades estudadas estão a sobreposição de amostras classificadas, definida como ruídos, e o desbalanceamento de amostras, muito comum nos problemas reais. Para tratar a sobreposição foi proposto um processo de filtragem que objetiva melhorar a acurácia na classificação. Para o desbalanceamento foi usada na metodologia a técnica de *undersampling* que procura melhorar as acurácias conforme descrito na literatura. Como objetivo de dar uma nova a

bordagem em um problema de classificação de despacho de usinas, foi proposto utilizar o método em um sistema de seis barras com duas usinas Termoelétricas, uma Hidroelétrica fio d'água e duas Eólicas classificadas como geração distribuída, para validar sua aplicação em um problema que é bem atual, ou seja, qual tipo de usina deve ser despachada a depender das condições climáticas. Por fim, o método proposto objetiva tratar os pontos descritos de forma a obter os melhores resultados possíveis para cada problema, sem a necessidade de ajustes de parâmetros das gaussianas a priori, o que permite sua aplicação, de forma geral, ser simples de configurar, não necessitando de conhecimentos técnicos profundos para sua utilização.

Abstract

The use of RBF (*Radial Basis Function*) in sample classification problems has been a much discussed approach in the literature. Its diverse applications allow you to deal with problems of different complexities. Among the most applied radial functions Gaussian is considered one of the most efficient for its simplicity of configuration. However one of the challenges in using this type of function is to define the parameters c and σ , center and radius respectively, suitable to avoid sub or oversize in the solution of the problem. This work proposes a methodology based on Gabriel's graph and the theory of the dominant sets to find these parameters for a given set of samples, without the necessity of arbitration. In a second step, the gaussians found in an artificial neural network architecture are applied in order to classify these samples. In a third step, we compare the results found with those of classical classifiers known in the literature. Once these results are analyzed, the particularities of each problem studied and their influences on the metrics of the proposed methodology are analyzed, which may create a need to adapt them to the treatment of some of these particularities. Among the particularities studied are the overlap of classified samples, defined as noise, and the unbalance of samples, very common in real problems. In order to deal with the overlap a filtering process was proposed that aims to improve the accuracy in the classification. For the unbalance the technique was used in the technique of undersampling which seeks to improve accuracy as described in the literature. In order to re-embroider a problem of classification of dispatch of plants, it was proposed to use the method in a six-bar system with two thermoelectric plants, one hydroelectric

plant and two wind farms classified as distributed generation, to validate Its application in a problem that is very current, ie, which type of plant should be dispatched depending on the climatic conditions. Finally, the proposed method aims to treat the described points in order to obtain the best possible results for each problem, without the need to adjust Gaussian parameters a priori, which allows its application, in general, to be simple to configure, not needing deep technical knowledge for its use.

Agradecimentos

Primeiramente à lei universal do retorno onde Deus nos presenteia com boas conquistas, resultado do esforço e dedicação empreendido em nossa caminhada profissional. À minha família por todo o apoio nos momentos em que esta dedicação nos deixou ausente ou privados da companhia das pessoas amadas. Aos meus colegas do curso de pós-graduação, especialmente aos do LITC pela paciência e boa vontade em ajudar nos momentos de dúvidas e discussões produtivas. Aos professores que além de contribuir com o ensinamento também o fizeram com a troca de experiências. Agradeço ainda aos funcionários da UFMG pelos serviços de qualidade prestados. Por fim, agradeço à FAPEMIG - Fundação de Amparo à Pesquisa de Minas Gerais pelo apoio financeiro.

Conteúdo

Lista de Símbolos	xiv
Lista de Abreviaturas	xvi
Lista de Figuras	xviii
Lista de Tabelas	xx
1 Introdução	1
2 Referencial Teórico	5
2.1 Teoria dos Grafos	5
2.1.1 Graus de relevância dos vértices do Grafo	6
2.1.2 Conexão dos vértices do Grafo	7
2.1.3 Grafo Isomorfos	7
2.1.4 Sub-Grafos	8
2.1.5 Vértices e arestas independentes	8
2.2 Diagrama de Voronoi	9
2.3 Triangulação de Delaunay	11
2.4 Grafo de Gabriel	13
2.5 Conjunto Dominante	14
2.6 Conceito de Kernel e Verossimilhança	20
2.7 Estimadores de densidade de Kernels	21
2.8 Funções RBF	21
2.8.1 Aplicação de RBF em Redes Neurais Artificiais	23
2.9 Conclusão do capítulo	27

3	Abordagem Proposta	28
3.1	Formação do Grafo de Gabriel e Dominância	28
3.2	Seleção dos centros e raios da RBF	30
3.3	Formação da RBF proposta	31
3.4	Demonstração em problemas com bases bidimensionais	33
3.5	Tratando a Sobreposição	36
3.5.1	Problemas com o tratamento da Sobreposição	37
3.5.2	Bases com classes balanceadas	38
3.5.3	Bases com classes desbalanceadas	39
3.6	Conclusão do capítulo	41
4	Experimentos	42
4.1	Explicando os cenários estudados	42
4.2	Bases testadas	43
4.3	Preparação para os experimentos	45
4.4	Apresentação dos resultados	45
4.5	Cenário I - Modelos Gaussianos de classificação	46
4.5.1	Experimento I	47
4.5.2	Experimento II	52
4.5.3	Análise de Significância dos experimentos	55
4.6	Cenário II - Modelos de Classificação clássicos	57
4.6.1	Experimento I	58
4.6.2	Experimento II	62
4.6.3	Análise de Significância dos experimentos	65
4.7	Conclusão do capítulo	66
5	Aplicação	68
5.1	Explicando o problema	68
5.2	Modelagem do Artigo estudado	70
5.3	Aplicação do método proposto	72
5.4	Bases testadas	73
5.5	Experimentos	74
5.5.1	Cenário I - Período com chuva	74
5.5.2	Cenário II - Período com estiagem	77

CONTEÚDO

5.6	Conclusão do capítulo	81
6	Conclusões	82
6.1	Proposta de continuidade	84
	Referências	89

Lista de Símbolos

\mathbb{R}	Conjunto dos números reais
\mathcal{E}_S	Esfera
$\phi_{\mathcal{E}}$	Diâmetro da esfera
\mathcal{H}	Espaço Hiper-dimensional
\mathcal{S}	Superfície plana
N	Conjunto de amostras de dados
N_d	Número de dimensões do conjunto de dados
$N_p(v)$	Números de vértices de um intervalo aberto do grafo
$N_p[v]$	Números de vértices de um intervalo fechado do grafo
N_{v+}	Número de vértices da classe 1
N_{v-}	Número de vértices da classe 2
σ_N	Desvio padrão dos testes
Φ	Matriz de interpolação dos espaços
\hat{x}	Vetor de entrada da RNA
\hat{w}	Vetor de pesos
\hat{y}_{teste}	Vetor de teste encontrado no método
\hat{y}_D	Vetor de amostras rotuladas, dado no problema
K	Matriz de adjacências de kernel
A	Matriz de adjacências do grafo
T	Tabela booleana de vértices do grafo
$\ \cdot\ $	Norma Euclidiana
$\text{sgn}(\cdot)$	Função Sinal
c	Centro da gaussiana
σ	Raio da gaussiana
ζ	<i>Shape parameter</i>
$\mathcal{B}(\cdot)$	Bissetriz
$\mathcal{DV}(\cdot)$	Diagrama de Voronoi de um conjunto de dados
$\mathcal{R}(\cdot)$	Região de Voronoi de um conjunto de dados
$\mathcal{Del}(\cdot)$	Diagrama de Delaunay de um conjunto de dados
$\mathcal{TD}(\cdot)$	Triangulação de Delaunay

$\delta(\cdot)$	Fornece a distância Euclidiana entre dois pontos
G	Grafo
$\tilde{G}_{\delta_{\mathcal{X}}}$	Grafo Planar
\dot{G}	Grafo de Gabriel
\min_{Do}	Mínimo conjunto dominante
\mathcal{V}_{Do}	Vértice Dominante do grafo
\mathcal{V}_{Adj}	Vértice Adjacente do grafo
γ_{Do}	Grau mínimo de dominância
Γ_{Do}	Grau máximo de dominância
$v_{i,j}$	Vértices do grafo
$\alpha_{i,j}$	Arestas do grafo
$T_{i,j}$	Elementos da tabela de adjacências do grafo
$V_{i,j}$	Conjunto de vértices
\hat{v}_{v_r}	Vetor de vértices de ruídos ou sobrepostos
S_N	Conjunto de Amostras desequilibradas
S_{min}	Conjunto minoritário de amostras desbalanceadas
S_{maj}	Conjunto majoritário de amostras desbalanceadas
E_q	Conjunto resultante do equilíbrio de amostras desbalanceadas
$\varphi(\cdot)$	Função gaussiana
p_w	p-valor de Wilcoxon
z_w	Valor do z paramétrico de Wilcoxon
σ_w	Desvio padrão do teste de Wilcoxon
\bar{w}_w	Média do valor de rank de Wilcoxon

Lista de Abreviaturas

RNA	Rede Neural Artificial
MLP	Multilayer Perceptron
ELM	Extream Learning Machine
LS-SVM	Least square Support Vector Machine
SVM	Support Vector Machine
RBF	Radial Basis Function
<i>Bayes</i>	Modelo com regra de Bayes
SOM	Self-Organizing Maps
WTA	Winner-takes-all
<i>K-Silhouette</i>	Método Silhoueta com k-médias
<i>K-Affinity propagation</i>	Método Afinidade com k-médias
<i>K-Aleatório</i>	Método Aleatório com k-médias
<i>F-Silhouette</i>	Método Silhoueta com c-médias
<i>F-Affinity propagation</i>	Método Afinidade com c-médias
<i>F-Aleatório</i>	Método Aleatório com c-médias
Alias	Nome da base abreviada
<i>Dom</i>	Modelo proposto com sem filtragem de ruído
<i>Dom-SR</i>	Modelo proposto com filtragem de ruído
AUC	Area under the curve
EPE	Empresa de Pesquisa Energética
GD	Geração distribuída
UHT	Usina Hidroelétrica Tradicional
PCH	Pequena Central Hidroelétrica
UTE	Usina Termoelétrica
UEOL	Usina Eólica
UFV	Usina Foto-Voltaica

Lista de Figuras

1.1	Montagem do Grafo de Gabriel e sub-grafo dominante	2
2.1	Modelos de Grafos	6
2.2	Modelos de Grafos com n-cliques	6
2.3	Graus do vértice de um Grafo	7
2.4	Grafos Isoformos	8
2.5	Representação do Teorema de Menger	9
2.6	Definição da Região de Voronoi	10
2.7	Definição do Diagrama de Voronoi	11
2.8	Definição da Triangulação de Delaunay	12
2.9	Formação do Grafo de Gabriel	14
2.10	Visualização da dominância em um grafo	15
2.11	Busca do conjunto dominante	16
2.12	Tabela booleana de do grafo da Figura 2.11	17
2.13	Processo booleano de seleção do conjunto dominante mínimo . . .	18
2.14	Processo booleano de seleção do conjunto dominante mínimo . . .	19
2.15	Conjunto dominante mínimo da Figura 2.11	20
2.16	Rede neural com RBF	24
3.1	Grafo de Gabriel para classificador bidimensional	29
3.2	Formação da Dominação para classificador de duas classes	30
3.3	Seleção dos centros e raios das gaussianas	31
3.4	Legenda das figuras	33
3.5	Método proposto aplicado ao problema Corners	34
3.6	Método proposto aplicado ao problema Clusters	34

LISTA DE FIGURAS

3.7	Método proposto aplicado ao problema Two moons	35
3.8	Método proposto aplicado ao problema Full moons	35
3.9	Método proposto aplicado ao problema Half kernels	36
3.10	Método proposto aplicado ao problema Spiral	36
3.11	Método proposto aplicado às amostras sobrepostas	38
3.12	Método proposto aplicado a base desbalanceada com <i>undersampling</i>	40
5.1	Sistema genérico de seis barras	71
5.2	Curva de carga do sistema	72

Lista de Tabelas

2.1	Comparação de uma rede <i>RBF</i> com uma <i>MLP</i> convencional . . .	24
4.1	Bases Reais - Repositório UCI	44
4.2	Métodos de cálculo de centros e estimação de Kernels	46
4.3	<i>Benchmarking</i> com classificadores Gaussianos	47
4.4	<i>Benchmarking</i> com classificadores Gaussianos	48
4.5	<i>Benchmarking</i> com classificadores Gaussianos	49
4.6	<i>Benchmarking</i> com classificadores Gaussianos	50
4.7	<i>Benchmarking</i> com classificadores Gaussianos	51
4.8	<i>Benchmarking</i> com classificadores Gaussianos	52
4.9	<i>Benchmarking</i> com classificadores Gaussianos	53
4.10	<i>Benchmarking</i> com classificadores Gaussianos	54
4.11	<i>Benchmarking</i> com classificadores Gaussianos	55
4.12	<i>Benchmarking</i> de Acurácia de teste de classificadores Gaussianos .	56
4.13	<i>Benchmarking</i> com classificadores clássicos	58
4.14	<i>Benchmarking</i> com classificadores clássicos	59
4.15	<i>Benchmarking</i> com classificadores clássicos	60
4.16	<i>Benchmarking</i> com classificadores clássicos	60
4.17	<i>Benchmarking</i> com classificadores clássicos	61
4.18	<i>Benchmarking</i> com classificadores clássicos	62
4.19	<i>Benchmarking</i> com classificadores clássicos	63
4.20	<i>Benchmarking</i> com classificadores clássicos	63
4.21	<i>Benchmarking</i> com classificadores clássicos	64
4.22	<i>Benchmarking</i> de Acurácia de teste dos classificadores clássicos . .	65

LISTA DE TABELAS

5.1	Preços de Leilões da EPE entre 2004 e 2016	69
5.2	Preços de geração em US\$	71
5.3	Curva de Carga diária	72
5.4	Bases para os cenários de despacho	73
5.5	Despacho por Carga considerando dia útil	75
5.6	Despacho por Usina	75
5.7	Despacho por Carga - finais de semana e feriado	75
5.8	Despacho por Usina	76
5.9	Despacho por Custo	76
5.10	Despacho por Usina	77
5.11	Despacho por Carga considerando dia útil	78
5.12	Despacho por Usina	78
5.13	Despacho por Carga - finais de semana e feriado	79
5.14	Despacho por Usina	79
5.15	Despacho por Custo	80
5.16	Despacho por Usina	80

Capítulo 1

Introdução

Diversas técnicas em RNA (*Rede Neural Artificial*) são aplicadas para problemas de classificação e reconhecimento de padrões de amostras. Uma abordagem largamente discutida na literatura é o uso de funções radiais para separação de conjuntos que não são linearmente separáveis conforme descrito no Teorema de Cover (Haykin, 2005) abrindo um campo de aplicação que permite tratar problemas de diferentes dimensões e níveis de complexidade. Entre as várias funções aplicáveis para este tipo de abordagem, o uso de RBF com Gaussianas é considerada uma das mais eficazes conforme descrito no Teorema de Micchelli e corroborado em (Braga et al., 2007; Haykin, 2009). Todavia, o ajuste das Gaussianas é sempre um desafio devido à possibilidade de *overfitting* na solução do problema, explicado em (Braga et al., 2007). Para evitar este tipo de ocorrência, estudos recentes na literatura (Torres, 2012, 2016; Wanderley, 2013) têm apresentado diferentes métodos para se encontrar o número adequado de Gaussianas a serem utilizadas nas RBFs. Os ajustes dos parâmetros neste caso se limitam ao posicionamento, relacionado ao centro c e ao raio de abrangência definido como σ . O modelo proposto neste trabalho, vem corroborar com estes estudos para a literatura, visando apresentar uma nova abordagem de ajustes destes parâmetros de forma automática, ou em outras palavras sem arbitrar-los. A hipótese apresentada objetiva utilizar a teoria do conjunto dominante aplicada ao grafo de Gabriel para encontrar os ajustes adequados da função radial a ser utilizada. De forma resumida a proposta do método parte dos seguintes pressupostos: em um espaço hiper-dimensional \mathcal{H}

onde há duas classes de amostras de tamanhos N_{v^+} e $N_{v^-} \in \mathcal{S}_{\mathcal{H}}$, onde \mathcal{S} é o espaço "separável", é possível considerar que cada conjunto separável de amostras possa ser dividido em sub-conjuntos agrupados por suas verossimilhanças. Estes clusters são agrupados através de ligações fortes denominadas links ou arestas entre suas amostras que são consideradas vértices de um grafo. Neste cenário é possível traçar o grafo de Gabriel $\tilde{G}_{\mathcal{S}}$, considerando dois grafos planares tal que cada sub-conjunto seja um sub-grafo que permita aplicar sobre este a teoria do conjunto dominante. Cada sub-conjunto elegerá amostras dominantes definidas como vértices dominantes e amostras adjacentes definidas como vértices dominados. A partir deste ponto parte-se do pressuposto que os vértices dominantes serão os centros (c) das Gaussianas e a média das distâncias entre os vértices dominantes e dominados os raios (σ). A Figura 1.1, demonstra graficamente o processo.

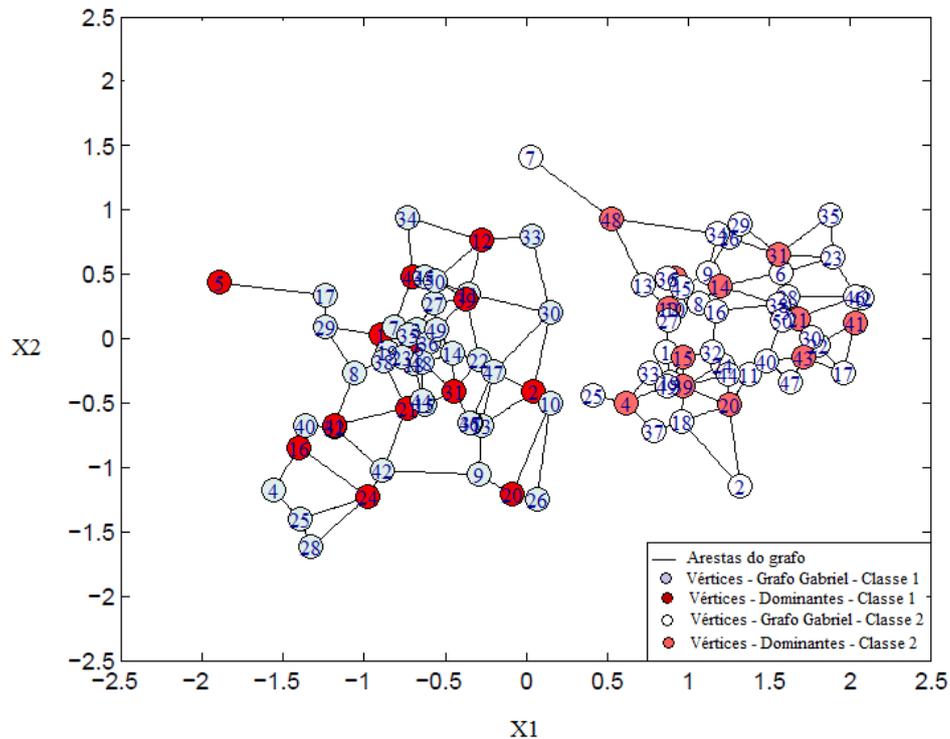


Figura 1.1: Montagem do Grafo de Gabriel e sub-grafo dominante

Uma vez ajustados os parâmetros, aplicam-se as Gaussianas encontradas em RBFs de uma rede neural com uma modelagem similar a uma MLP (*Multi Layer Perceptron*) (Haykin, 2005; Ruck et al., 1990) com o objetivo de classificar as amostras. Em uma RNA este tipo de comparação é possível devido à proximidade entre uma modelagem MLP tradicional e uma RBF conforme descrito em (Braga et al., 2007). Para demonstrar esta teoria foram preparados dois cenários comparativos: o primeiro com métodos de classificação utilizando funções Gaussianas onde os parâmetros são calculados também sem a intervenção do usuário; o segundo com classificadores clássicos da literatura onde há a necessidade dessa arbitragem. Em ambos os cenários, os modelos escolhidos, todos na linguagem R, possuem processos internos de otimização e alguns deles com abordagens recursivas e/ou iterativas para retornarem os melhores valores possíveis das métricas desejadas para cada problema. É importante destacar que se tratando de problemas de classificação, a depender do nível de complexidade, outros desafios que vão além da questão de linearização se apresentarão. Dentre os que podem afetar o resultado das métricas foram estudados neste trabalho dois fortemente intrincados: o problema da sobreposição das classes amostradas, conhecido muitas vezes por ruídos. Como possível solução, técnicas são aplicadas para a eliminação desta sobreposição, no entanto, a depender do problema esta filtragem pode acabar por eliminar muitas amostras e reduzir demais seu tamanho, comprometendo o resultado do classificador. O problema do desbalanceamento das classes do conjunto amostrado. Este problema é explicado em (Castro & Braga, 2013) que descreve seu surgimento como uma consequência direta de assumir custos (perdas) iguais aos erros de classificação, assim, sob a hipótese de custo igual, o único objetivo se torna a minimização de um critério geral que corresponde à probabilidade de erro geral (ou taxa de erro de classificação esperada). Em outras palavras, um classificador tende a prever a saída de acordo com o padrão de entrada que por ser muito desbalanceado, afeta a acurácia aumentando o erro de classificação. Como solução, ainda citada em (Castro & Braga, 2013), há diferentes técnicas que são aplicadas para este tipo de problema, uma delas a de *undersampling*, descrita em (Haibo & Garcia, 2009) que visa estudar uma redistribuição estatística das classes separando-as em duas classes de igual dimensão definida pelo conjunto de menor representatividade de amostras. Fica claro que os desafios vão

além da separabilidade, de fato, esta só será possível se forem tratadas as particularidades de cada problema, conforme descritas mais a frente nesta dissertação. Neste sentido este estudo visa demonstrar que o método proposto não deve tratar a mesma abordagem generalizando os problemas de classificação, e para tal, 18 bases retiradas do repositório UCI (Asuncion & Newman, 2007) com características diferentes de balanceamento, dimensão e complexidade e que vêm sendo exaustivamente utilizadas em problemas de classificação foram escolhidas para um *benchmarking* nos dois cenários. Ainda no sentido de validação do método, foi proposta uma aplicação deste em um problema de classificação de despacho de usinas. O problema foi apresentado no artigo (Queiroz & Junior, 2016), onde um sistema elétrico com seis barras possui duas usinas Termoelétricas e duas Eólicas, simuladas como GD (Geração distribuída), onde o objetivo é analisar o despacho mais econômico no contexto das quatro usinas. Neste sistema simulado, o método foi aplicado em conjunto com a inserção de um PCH (Pequena Usina Hidroelétrica), ou fio d'água, com o objetivo de classificar o despacho por usina, ou em outras palavras, quais usinas são despachadas dados diferentes situações de entrada. Ao final deste trabalho, pretende-se provar que o método pode ser eficaz como classificador quando aplicado em problemas distintos nos dois cenários estudados, e na mesma direção, sendo validado através do teste paramétrico de Wilcoxon sugerido em (Demsar, 2006), demonstrando que sua abordagem simples, de fácil configuração e sem processos recursivos e de otimização pode ser assertivo quando comparado o resultado de suas métricas com as de métodos robustos estudados na literatura.

Capítulo 2

Referencial Teórico

Neste capítulo, apresentaremos toda a fundamentação teórica do modelo sugerido. Esta fundamentação passa pelo pressuposto que amostras podem ser separadas por padrões definidos de comportamento através de sub-conjuntos (kernels) separáveis (Haykin, 2005) por suas verossimilhanças, ou seja, por amostras que estejam agrupadas mesmo que em uma distribuição estatística aleatória. Considerando cada amostra um componente deste kernel, pode-se interpretar este sub-conjunto como um grafo onde as amostras mais próximas podem ser ligadas trigonometricamente. A partir deste pressuposto, as RBFs são utilizadas como metodologia aplicada a uma RNA para a aprendizagem e separação dos padrões classificando-os.

2.1 Teoria dos Grafos

A teoria dos grafos é um ramo da matemática que estuda as relações entre os objetos pertencentes a um determinado conjunto. Para tal este conjunto é separado em duas entidades nomeadas vértices v e arestas α . Os vértices são os pontos a serem interligados enquanto as arestas representam um subconjunto de pares dos vértices não ordenados, ou seja, as linhas de interligação, tornando-se em um problema geométrico. O objetivo é formar poliedros convexos conforme a Figura 2.2 (Sheehan, 1986). Com este conceito, pode-se formar grafos em quaisquer conjuntos de dados que se queira selecionar. Algumas estruturas simples são designadas como *Grafo-Nulo*, que não possui vértices ou arestas; *Grafo-vértice*,

que não possui arestas sendo apenas o próprio vértice Figura 2.1(i); *Grafo-looping* demonstrado na Figura 2.1(ii), cuja ligação é o mesmo ponto e *Grafo-link* demonstrado na Figura 2.1(iii) que possui uma interligação entre dois vértices.

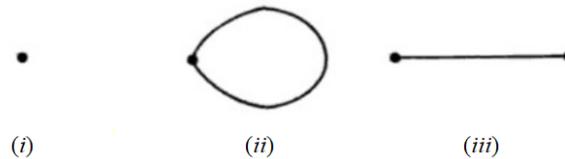


Figura 2.1: Modelos de Grafos

Outras estruturas pouco mais complexas podem ser definidas a partir de n cliques, arcos ou circuitos, tal que n seja positivo e inteiro. O *clique* ou circuito n é definido como um *Grafo-looping* de n vértices e $\frac{1}{2}n(n-1)$ arestas, sendo que cada par de vértices se une formando um *Grafo-link*, assim, n -*cliques* se tornam um Grafo convexo (Sheehan, 1986). Neste caso, um *Grafo-nulo* possui 0-*clique*, um Grafo-vértice possui 1-*clique* e um Grafo-link possui 2-*cliques*. A Figura 2.2 mostra outros Grafos com 2, 4 e 5-*cliques*. Note que na Figura 2.2(iii), há um cruzamento de arestas (Sheehan, 1986).

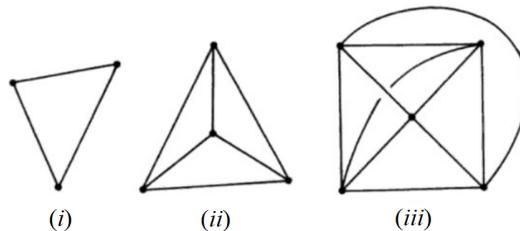


Figura 2.2: Modelos de Grafos com n-cliques

2.1.1 Graus de relevância dos vértices do Grafo

O grau de um vértice é o número de arestas que compartilham este vértice, sejam as arestas que incidem ou as que saem deste vértice. Este grau é definido em máximo ou mínimo a depender de quantas arestas são compartilhadas. A Figura 2.3 mostra os conceitos descritos dando uma visão mais clara do grau. Na Figura 2.3(i) possível observar que há 7 vértices, sendo que o vértice 7 é isolado, não

tendo arestas. O grafo neste caso é não direcionado e o vértice 3 possui grau 4 ($\alpha_a, \alpha_b, \alpha_c$ e α_d).

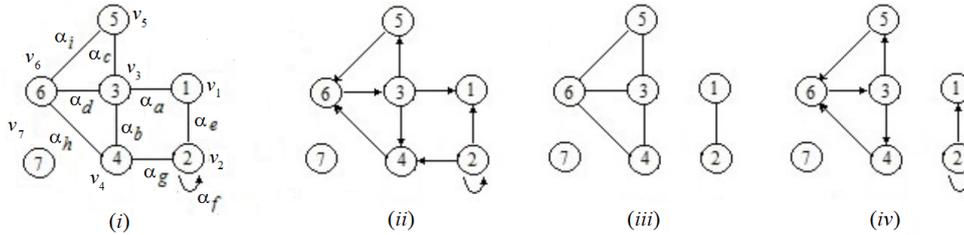


Figura 2.3: Graus do vértice de um Grafo

Quando se considera a incidência ou direção, o grau pode aumentar ou diminuir. Na Figura 2.3(ii) pode ser observado que o vértice 3 possui grau 4 com uma aresta incidindo sobre ele (α_d) e três saindo dele (α_a, α_b e α_c). Nesta figura ainda observa-se o grau máximo do Grafo que é 4 e o grau mínimo que é 2, representado pelos vértices 1 (α_e e α_a) e 5 (α_c e α_i) (Sheehan, 1986).

2.1.2 Conexão dos vértices do Grafo

Um Grafo não direcionado é conectado quando cada par de vértice estiver ligado por um caminho ou ainda se possui um componente conectado. Considerando a Figura 2.3(iii), os componentes são: $\{3, 4, 5, 6\}$, $\{1, 2\}$ e $\{7\}$. Já um grafo direcionado possui conexões camadas fortes. O conceito de conexões fortes vem da conexão dos vértices tal que se cada par de vértice pode ser alcançado por outro par, ele é considerado fortemente conectado podendo se dizer que são conjuntos do vértices mutuamente alcançáveis, como pode ser observado na Figura 2.3(iv) cujos os componentes $\{3, 4, 5, 6\}$ são fortemente conectados e os componentes $\{1, 2\}$ e $\{7\}$, não pois não podem ser alcançados pelo outro vértice.

2.1.3 Grafo Isomorfos

O conceito de isomorfismo vem da correlação de vértices entre Grafos. Dado dois Grafos, A e B , sendo $f \in \alpha(A)$ e $g \in \alpha(B)$ e cada link $(f, g) = \theta$, conforme (Tutte, 1998) se diz que A e B são isomórficos se somente se: $|v(A)| =$

$|v(B)|$ e $|\alpha(A)| = |\alpha(B)|$. Pode-se afirmar ainda que $\theta(A) = B$ e por consequência que $\theta_v = f_v$ e que $\theta_\alpha = g_\alpha$ para cada vértice v e aresta α de A e de B . A Figura 2.4 mostra o conceito graficamente.

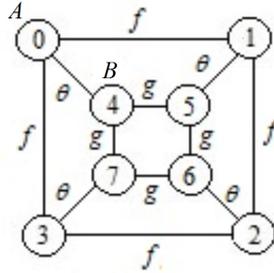


Figura 2.4: Grafos Isoformos

2.1.4 Sub-Grafos

Na teoria dos Grafos pode-se considerar que alguns Grafos contêm outros Grafos. Na figura 2.4, pode-se dizer que o Grafo B está contido no Grafo A , sendo B um sub-grafo de A . Em outras palavras que $v(A) \subseteq v(B)$ e que $\alpha(A) \subseteq \alpha(B)$, ou ainda que cada aresta de A possui o mesmo final em A como em B e sob esta condição B está contido em A (Sheehan, 1986).

2.1.5 Vértices e arestas independentes

Definindo um dos o polígonos ou poliedros (Tutte, 1998) formados na Figura 2.2 como um Grafo G com vértices v e arestas α_a e α_b podemos nos basear no teorema de Menger (Bohme et al., 2001) que descreve:

Teorema 1 *Um grafo G com dois vértices v_1 e v_2 , tal que estes vértices sejam interligados por uma aresta α_a , esta aresta será independente se nenhuma outra aresta compartilha os mesmos vértices v_1 e v_2 .*

Desta forma, é garantida que as arestas não compartilharão mais que um par de vértices e este conjunto dado pelos vértices v_1 e v_2 e aresta α_a passa a ser uma conexão integrante do grafo, sendo removidas outras arestas de mesmo caminho.

Na Figura 2.5 o resultado do teorema. Pode ser visto na Figura 2.5(i) o grafo completo. O par de vértices $(v_1 - v_2)$ é interligado pelas arestas $(\alpha_a$ e $\alpha_b)$ e o par de vértices $(v_2 - v_3)$ é interligado pelas arestas $(\alpha_c$ e $\alpha_d)$. Após aplicar o Teorema 1, o novo grafo G é apresentado na Figura 2.5(ii).

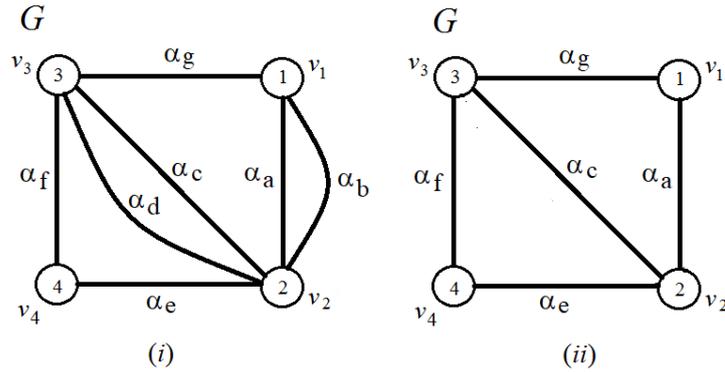


Figura 2.5: Representação do Teorema de Menger

2.2 Diagrama de Voronoi

Voronoi estudou a separação de um espaço n -dimensional em subespaços decompostos através da distância entre seus pontos. O diagrama de Voronoi \mathcal{DV} foi introduzido na computação por Shanos e Hoey (Klein, 1989). Dado um espaço $\mathcal{H} \in \mathbb{R}^{m \times n}$ e três pontos p, q e z aleatórios neste espaço tal que $p, q, z \in \mathcal{S}_{\mathcal{H}}$ onde \mathcal{S} representa um plano com um conjunto de pontos e α sejam possíveis caminhos entre os três pontos no plano \mathcal{S} tal que $\overline{pq} \in \mathcal{S}_{\mathcal{H}}$ e $z \notin \{p, q\}$, então pode ser descrito que a distância euclidiana δ entre p e q é dada por:

$$\delta(p, q) = \delta(p, z) + \delta(z, q) \quad (2.1)$$

Através do Teorema de Menger 1, sabe-se que dois pontos em plano podem ser unidos pelos caminhos α , assim, conforme equação 2.1 os pontos p, q e z de α serão os segmentos de distância δ mais curtos entre eles, formando assim um dos limites da região de Voronoi $\mathfrak{R}_{\mathcal{S}}(p, q)$ dada pela bissetriz \mathcal{B} de:

$$\mathcal{B}(p, q) = \{z; \delta(p, z) = \delta(q, z)\} \quad (2.2)$$

2.2 Diagrama de Voronoi

A bissetriz $\mathcal{B}(p, q)$ transforma o segmento em dois "meio planos", ou dois novos segmentos de $\mathfrak{R}_S(p, q)$ através da distância euclidiana entre dois pontos vetoriais do segmento \overline{pq} (Torres, 2016). O Diagrama de Voronoi $\mathcal{DV}(\mathcal{S})$ pode então ser considerado como a divisão do hiperplano \mathcal{H} em n polígonos convexos compostos por \mathcal{L} caminhos (Berg et al., 2008; Torres, 2016). Um polígono $\mathcal{L}(x_i)$ é chamado polígono de Voronoi relativo ao ponto x_i e é formado pela interseção do conjunto de bissetrizes $\mathcal{B}(x_i, \mathcal{L}(x_i))$ e um ponto $p \in \mathcal{S}_{\mathcal{H}} \subset \mathcal{L}(x_i)$ se somente se a seguinte desigualdade for satisfeita:

$$\delta(p, x_i) \leq \delta(p, x_j), \quad x_i, x_j \in \mathcal{S}_{\mathcal{H}}, \quad \forall j \neq i \quad (2.3)$$

Seja $\mathcal{S} = \{p_1, \dots, p_n\}$ um plano com um conjunto de $n = \{x_1, \dots, x_n\}$ diferentes pontos do hiperplano \mathcal{H} . Para a distância euclidiana de ordem δ^n a $\mathfrak{R}_S(p, q)$ com $p \neq q$ pode ser escrito como:

$$\mathcal{B}(p, q) = \{x_i \in \mathbb{R}^2; |p - x_i| = |q - x_i|\} \quad (2.4)$$

$$\mathfrak{R}(p, q) = \{x_i \in \mathbb{R}^2; |p - x_i| < |q - x_i|\} \quad (2.5)$$

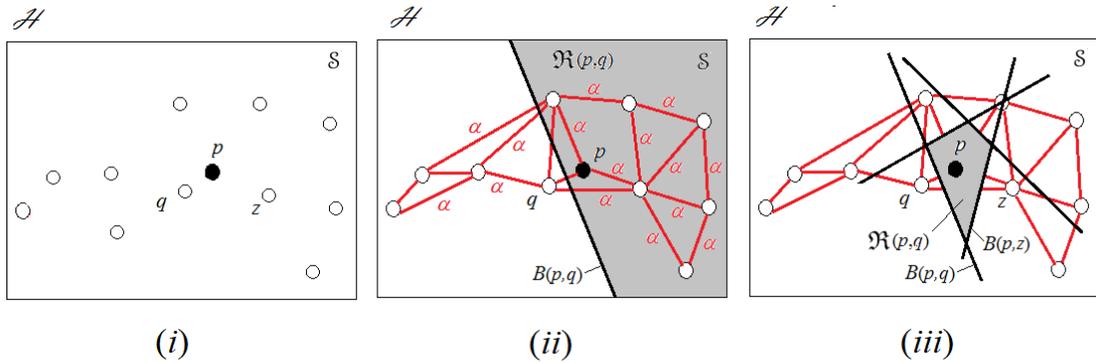


Figura 2.6: Definição da Região de Voronoi

A Figura 2.6 demonstra a construção de um diagrama de Voronoi. Na Figura 2.6(ii) a bissetriz $\mathcal{B}(p, q)$ divide o plano \mathcal{S} em dois semiplanos onde $\mathfrak{R}(p, q)$ é o que contém p ou seja:

$$\mathfrak{R}(p, \mathcal{S}) = \bigcap_{\substack{q \in \mathcal{S} \\ p \neq q}} \mathfrak{R}(p, q) \quad (2.6)$$

O conjunto de todos os pontos x_i que estão mais próximos de p que os demais do plano \mathcal{S} formam um polígono convexo possivelmente limitado é chamado de região de Voronoi $\mathfrak{R}(p, \mathcal{S})$. A união dos pontos do contorno da região é chamada de Diagrama de Voroni (Klein, 1989), mostrada na Figura 2.7 definida como:

$$\mathcal{DV}(\mathcal{S}) = \bigcup_{p \in \mathcal{S}} \partial \mathfrak{R}(p, \mathcal{S}) \quad (2.7)$$

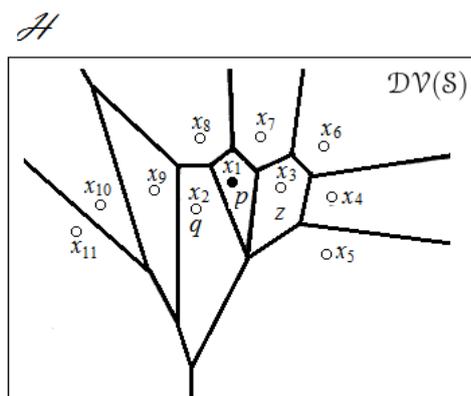


Figura 2.7: Definição do Diagrama de Voronoi

2.3 Triangulação de Delaunay

Conforme visto na Figura 2.7, o diagrama de Voroni é dado pela união dos polígonos convexos formados na região de Voronoi. Seja $\mathcal{S} = \{p_1, \dots, p_n\}$ o plano com um conjunto de n pontos tal que $n = \{x_1, \dots, x_n\} \forall n \rightarrow x_{(i)} = p_{(i)}$ do diagrama de Voronoi considerando cada ponto p como um vértice de um grafo, demonstrado também na Figura 2.7, de uma $\mathcal{DV}(p, \mathcal{S})$. Assume-se então que os vértices $p_i \in v_1$ compartilham limites de região com outros vértices $p_j \in v_j$ do grafo. O resultado ao se interligar as arestas entre os vértices, é a formação de um novo conjunto convexo ou um Grafo não orientado tal que $v = \{p_1, \dots, p_n\}$ definido

2.3 Triangulação de Delaunay

como Diagrama de Delaunay $\mathcal{D}el(\mathcal{S})$. A Figura 2.8(i) demonstra graficamente a região descrita como um polígono convexo definido pela equação 2.8 (Gallier, 2008; Torres, 2016):

$$\mathcal{D}el(\mathcal{S}) \subset v = \{p_1, \dots, p_{v+1}\} \iff v_1 \cap \dots \cap v_{v+1} \neq \emptyset \text{ sujeito a } 0 \leq v \leq m \quad (2.8)$$

Onde m é a dimensão de \mathcal{H} . A triangulação de Delaunay $\mathcal{T}\mathcal{D}(\mathcal{S})$ pode ser representada por um grafo planar $\tilde{G}(v, \alpha)$ em uma triangulação dos vértices v (Torres, 2016) do grafo conforme Lema 1 (Aurenhammer & Klein, 1989; Gallier, 2008).

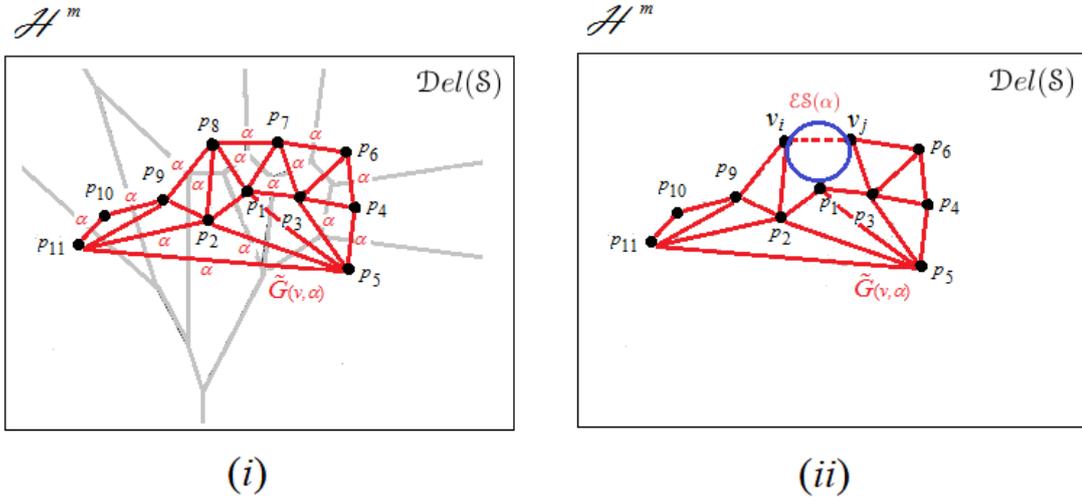


Figura 2.8: Definição da Triangulação de Delaunay

Lema 1: Seja $\tilde{G} = \{p_1, \dots, p_n\}$ um conjunto de n vértices em \mathcal{H}^m dispostos genericamente, então a triangulação de Delaunay de um polígono convexo de \tilde{G} é de fato uma triangulação associada com \tilde{G} se somente se satisfaz as seguintes propriedades:

1. O Limite de $\mathcal{D}el(\tilde{G})$ é o polígono convexo de \tilde{G} ; Figura 2.8(i)

2. A triangulação $\mathcal{TD}(\mathcal{S})$ associada com \tilde{G} é a Triangulação de Delaunay $\mathcal{TD}(\tilde{G}_{\mathcal{S}})$ se uma esfera $\mathcal{E}_{\mathcal{S}}(\alpha)$ possuir o par (v_i, v_j) circunscrita sob um m -polígono de arestas α e todos os outros vértices p de \tilde{G} estiverem fora de $\mathcal{E}_{\mathcal{S}}(\alpha)$, o que pode ser observado na Figura 2.8(ii).

2.4 Grafo de Gabriel

O Grafo de Gabriel \ddot{G} é interpretado como um conjunto de pontos de \mathcal{H} em um espaço euclidiano representado pelo plano \mathcal{S} expresso pela medida da distância média entre estes pontos. Este conceito foi introduzido por (K. R. Gabriel e R. R. Sokal) em 1969. Considerando um grafo com $p = \{x_1, \dots, x_n\}$ pontos (vértices) e dois destes x_i e x_j conectados por uma aresta α , o grafo \ddot{G} existirá apenas se ao se traçar uma esfera $\mathcal{E}_{\mathcal{S}}$ de diâmetro (x_i, x_j) não existir nenhum outro ponto $p = \{x_z\}$ interno à esfera tal que (Zhang & King, 2002):

$$\phi_{\mathcal{E}}(x_i, x_j) = \max_{\substack{x \in p \\ r \neq x_i, x_j}} (\delta(x_i, r), \delta(x_j, r)) \leftrightarrow (x_z \supset \mathcal{E}_{\mathcal{S}}) \notin \phi_{\mathcal{E}_{\mathcal{S}}}(x_i, x_j) \quad (2.9)$$

Onde $\phi_{\mathcal{E}_{\mathcal{S}}}$ é o diâmetro dado pela distância euclidiana δ e r o raio da esfera. O Grafo de Gabriel \ddot{G} pode ser descrito como um sub-conjunto de pontos do Diagrama de Voroni $\mathcal{DV}(\mathcal{S})$ e também um sub-grafo, definido na seção 2.1.4 da triangulação de Delaunay $\mathcal{TD}(\mathcal{S})$ (Souza et al., 2014; Zhang & King, 2002). Conforme (Berg et al., 2008) O Grafo de Gabriel de um conjunto de pontos de um plano $\tilde{G}(v, \alpha) \in \mathcal{S}$ deve obedecer à seguinte definição (Torres, 2012, 2016):

$$(v_i, v_j) \in \alpha \leftrightarrow \delta^2(v_i, v_j) \leq [\delta^2(v_i, x) + \delta^2(v_j, x)] \quad \forall x \in v_{(\mathcal{S})}, v_i, v_j \neq x \quad (2.10)$$

onde $v_{(\mathcal{S})}$ é o conjunto de vértices de $\ddot{G}_{(\mathcal{S})}(v, \alpha)$. A Figura 2.9 demonstra o processo de formação do $\ddot{G}_{(\mathcal{S})}$. Na Figura 2.9(i) há uma distribuição genérica de pontos $p = \{x_1, \dots, x_n\} \forall (x = v)$ no plano $\tilde{G}(\mathcal{S})$ em \mathcal{H} . Conforme Voronoi, calcula-se a distância euclidiana entre os pontos. Por Delaunay, considera-se cada um dos pontos como vértice de um grafo não orientado e trassa-se $\mathcal{E}_{\mathcal{S}}$ (circulo pontilhado) de diâmetro $\phi_{\mathcal{E}_{\mathcal{S}}}$ dado pela distância euclidiana entre cada par dos vértices demonstrado na Figura 2.9(ii), se não existir nenhum outro ponto $x \in p$ entre

os vértices dentro de \mathcal{E}_S é traçada uma aresta α entre estes vértices formando $\ddot{G}_{(S)}(v, \alpha)$ demonstrado na Figura 2.9(v). Na formação do grafo, alguns pontos podem estar dentro de \mathcal{E}_S ou externos, muito próximos a esta, como demonstrado 2.9(ii) e não satisfazem a equação 2.10. Neste caso por Delaunay não haverá uma aresta entre os vértices, sendo refeito o processo traçando novas $\phi_{\mathcal{E}_S}$ como demonstrado na Figura 2.9(iii).

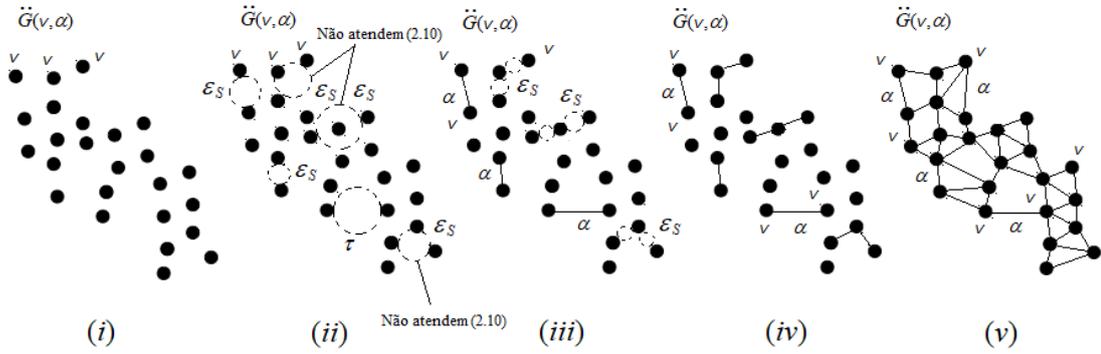


Figura 2.9: Formação do Grafo de Gabriel

2.5 Conjunto Dominante

Dado um conjunto de vértices V de um grafo, um conjunto dominante deste é um sub-grafo $G_{(S)}(v, \alpha)$ formado por um conjunto $\mathcal{S} \subseteq V(G_{(S)})$, tal que todo vértice do grafo, ou pertence a \mathcal{S} , ou é adjacente a pelo menos um elemento de \mathcal{S} (Sousa, 2015). Segundo (Haynes & Slater, 1998), dado $N_p = \{x_1, x_2, \dots, x_n\}$ pontos para um conjunto $\mathcal{S} \subseteq V$, um vértice $v \in \mathcal{S}$ é chamado de fronteiro se estiver em um intervalo fechado de \mathcal{S} , ou seja, se $N_p[v] \subseteq \mathcal{S}$, e chamado de isolado se estiver em um intervalo aberto de \mathcal{S} tal que $N_p(v) \subseteq V - \mathcal{S}$. Há ainda outras maneiras de se definir um conjunto dominante em um grafo, cada uma com um aspecto diferente de dominação. Em (Haynes & Slater, 1998), um conjunto $\mathcal{S} \subseteq V$ com vértices em um $G_{(S)}(v, \alpha)$ é dominante se somente se:

- i. Para todo vértice $v \in V - \mathcal{S}$, existe um vértice $u \in \mathcal{S}$, tal que v é adjacente a u ;

- ii. Para todo vértice $v \in V - \mathcal{S}$, $d(v, \mathcal{S}) \leq 1$;
- iii. $N_p[\mathcal{S}] = V$;
- iv. Para todo vértice $v \in V - \mathcal{S}$, $|N_p(v) \cap \mathcal{S}| \geq 1$, ou seja, todo vértice $v \in V - \mathcal{S}$ é adjacente a pelo menos um vértice em \mathcal{S} ;
- v. Para todo vértice $v \in V$, $|N_p(v) \cap \mathcal{S}| \geq 1$ e;
- vi. $V - \mathcal{S}$ é uma região sem fronteira.

O problema do conjunto dominante consiste em determinar a cardinalidade de um conjunto dominante mínimo em um grafo $G_{(\mathcal{S})}(v, \alpha)$ (Sousa, 2015), em outras palavras, o número de dominação de $G_{(\mathcal{S})}(v, \alpha)$ é o número de vértices em um conjunto menor de $G(\mathcal{S})$. Isto leva ao conceito de mínimo conjunto dominante cardinal ou $min_{D_o}(G_{(\mathcal{S})})$ mostrado na Figura 2.10, com $\gamma_{D_o} = \{1, 3, 5\}$ Figura 2.10(ii), com $\gamma_{D_o} = \{3, 6, 7, 8\}$ Figura 2.10(iii) e com $\gamma_{D_o} = \{2, 4, 6, 7, 8\}$ Figura 2.10(iv).

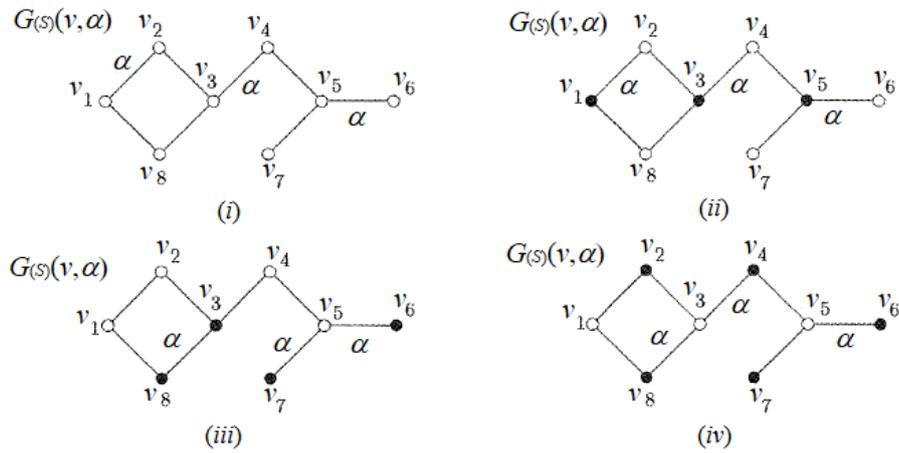


Figura 2.10: Visualização da dominância em um grafo

Onde γ_{D_o} é o número de vértices dominantes, ou de dominação do grafo. Conforme (Ore, 1962), o número de $\gamma_{D_o}(G_{(\mathcal{S})})$ do grafo é igual ao mínimo cardinal de conjunto em $min_{D_o}(G_{(\mathcal{S})})$, enquanto $\Gamma_{D_o}(G_{(\mathcal{S})})$ é considerado o número máximo de dominação. Na Figura 2.10, $\gamma_{D_o}(G_{(\mathcal{S})}) = 3$ e $\Gamma_{D_o}(G_{(\mathcal{S})}) = 5$. Conjuntos

2.5 Conjunto Dominante

dominantes estão intimamente relacionados aos conjuntos independentes. Um conjunto independente é aquele que possui dois vértices de $G_{(S)}(v, \alpha)$ em S tal que cada vértice v que não está em S é adjacente a pelo menos um membro de $G_{(S)}$. O número de dominação n segundo (Haynes & Slater, 1998) é dado pela expressão 2.11.

$$n = \gamma(G) + E_v(G_{(S)}) \quad (2.11)$$

Onde $E_v = \{\alpha_{v_1} + \alpha_{v_2} + \dots + \alpha_{v_n}\}$ é o número máximo de arestas a partir do vértice v de $G_{(S)} = (v, \alpha)$. Um conjunto independente é também um conjunto dominante se, e somente se, é um conjunto independente máximo, de modo que qualquer conjunto independente máximo em um grafo é necessariamente também um conjunto dominante mínimo. Figura 2.11, demonstra um grafo de Gabriel com $\min_{Do}(G_{(S)})$ selecionados, sendo para este caso, $\gamma_{Do}(G_{(S)}) = 4$.

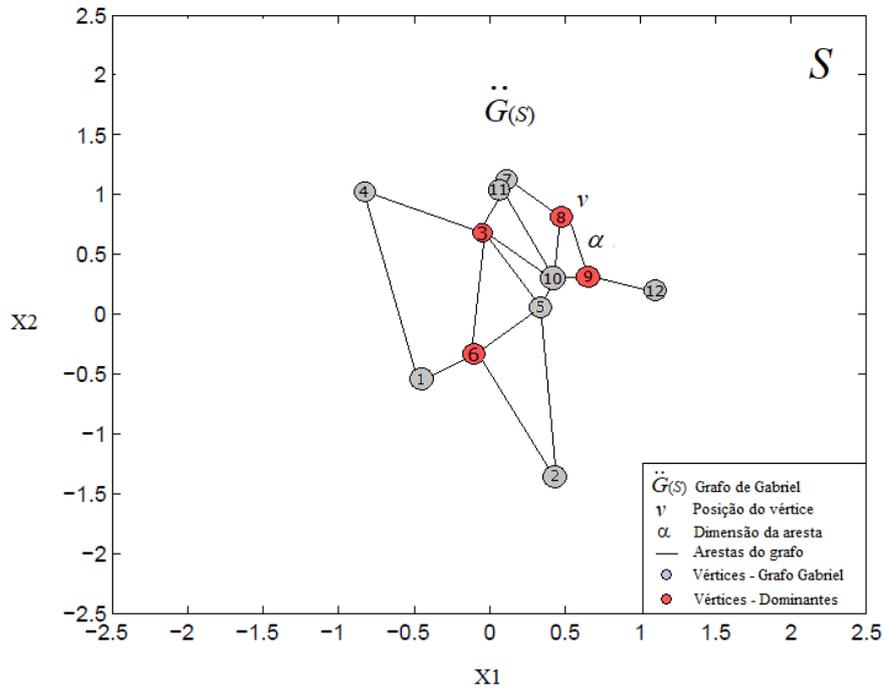


Figura 2.11: Busca do conjunto dominante

Uma metodologia eficaz para se encontrar vértices dominantes de um grafo é a

2.5 Conjunto Dominante

busca através da estrutura de árvore, proposta em (Even, 1973) demonstrada nas Figuras (2.12, 2.13, 2.14 e 2.15) que encontra o conjunto dominante usando a estratégia da determinação da cobertura máxima de conjuntos descrita na equação 2.11, onde um conjunto de vértices $v_{G(S)}$ é dominante se todo vértice $v \in G(S)$ ou é adjacente a um de seus elementos. Nesta situação, torna-se natural a inclusão da matriz de adjacências A , conforme descreve (Barroso, 2001), como elemento fundamental do processo. Posto isto, monta-se a tabela booleana de A com a matriz identidade de ordem n_v , que corresponde ao número de vértices do grafo. Cria-se a partir deste pressuposto uma tabela T onde a i -ésima linha associa-se ao i -ésimo vértice v_i , enquanto que cada coluna associa-se aos vértices adjacentes à v_j correspondente. A solução do problema da cobertura de conjuntos então passa a ser feito através da manipulação da tabela T diminuindo-a até que se encontre o mínimo conjunto dominante dentro de T . O processo pode ser mais claramente compreendido a partir do exemplo a seguir: Dada uma distribuição aleatória de pontos representada pelo grafo de Gabriel $\ddot{G}_{(S)}(v, \alpha)$ da Figura 2.11, procura-se determinar conjunto dominante mínimo, ou $min_{D_o}(\ddot{G}_{(S)})$, sendo o primeiro passo transformar as amostras aleatórias em vértices do grafo preenchendo a tabela T inicial descrita na Figura 2.12 que neste caso terá ordem $n_v = 12$ por existirem 12 vértices.

T		A_j											
		$v1$	$v2$	$v3$	$v4$	$v5$	$v6$	$v7$	$v8$	$v9$	$v10$	$v11$	$v12$
A_i	$v1$	1	0	0	1	0	1	0	0	0	0	0	0
	$v2$	0	1	0	0	1	1	0	0	0	0	0	0
	$v3$	0	0	1	1	1	1	0	0	0	1	1	0
	$v4$	1	0	1	1	0	0	0	0	0	0	0	0
	$v5$	0	1	1	0	1	1	0	0	0	1	0	0
	$v6$	1	1	1	0	1	1	0	0	0	0	0	0
	$v7$	0	0	0	0	0	0	1	1	0	0	1	0
	$v8$	0	0	0	0	0	0	1	1	1	1	0	0
	$v9$	0	0	0	0	0	0	0	1	1	1	0	1
	$v10$	0	0	1	0	1	0	0	1	1	1	1	0
	$v11$	0	0	1	0	0	0	1	0	0	1	1	0
	$v12$	0	0	0	0	0	0	0	0	1	0	0	1

Figura 2.12: Tabela booleana de do grafo da Figura 2.11

O próximo passo é a manipulação da próxima tabela T através da eliminação de linhas e colunas na matriz de adjacências $A(v_i, v_j)$. Este processo de eliminação

2.5 Conjunto Dominante

de linhas e colunas segue as seguintes premissas (Barroso, 2001), explicitadas nas Figuras 2.13 e 2.14:

1. Existindo uma linha i da tabela T com um único valor unitário, garante-se a seleção do vértice j associado à coluna que contém esse elemento para pertencer ao conjunto dominante, já que tal escolha é forçada pela condição do problema, deste modo, assumi-se que $v = v \cup \{j\}$ e pode-se excluir a coluna j de T , bem como todas as linhas i que possuem elementos unitários pertencentes à esta coluna que já foram contempladas;
2. Para cada coluna j , seja $A(i) = \{i \text{ o elemento } T_{ij} = 1\}$; dadas duas colunas (j,k) , se $A(k) \subseteq A(j)$, então a coluna k é definida como redundante e pode ser excluída de T já que qualquer solução mínima que contenha j cobre todos os elementos que a coluna k cobriria.

T		A_j									
		$v1$	$v3$	$v4$	$v5$	$v6$	$v7$	$v8$	$v9$	$v10$	$v11$
A_i	$v1$	1	0	1	0	1	0	0	0	0	0
	$v2$	0	0	0	1	1	0	0	0	0	0
	$v3$	0	1	1	1	1	0	0	0	1	1
	$v4$	1	1	1	0	0	0	0	0	0	0
	$v5$	0	1	0	1	1	0	0	0	1	0
	$v6$	1	1	0	1	1	0	0	0	0	0
	$v7$	0	0	0	0	0	1	1	0	0	1
	$v8$	0	0	0	0	0	1	1	1	1	0
	$v9$	0	0	0	0	0	0	1	1	1	0
	$v10$	0	1	0	1	0	0	1	1	1	1
	$v11$	0	1	0	0	0	1	0	0	1	1
	$v12$	0	0	0	0	0	0	0	1	0	0

(i)

T		A_j									
		$v1$	$v3$	$v4$	$v5$	$v6$	$v7$	$v8$	$v10$	$v11$	
A_i	$v1$	1	0	1	0	1	0	0	0	0	
	$v2$	0	0	0	1	1	0	0	0	0	
	$v3$	0	1	1	1	1	0	0	1	1	
	$v4$	1	1	1	0	0	0	0	0	0	
	$v5$	0	1	0	1	1	0	0	1	0	
	$v6$	1	1	0	1	1	0	0	0	0	
	$v7$	0	0	0	0	0	1	1	0	1	
	$v11$	0	1	0	0	0	0	0	1	1	

(ii)

Figura 2.13: Processo booleano de seleção do conjunto dominante mínimo

Seguindo a premissa número 2 na tabela inicial verifica-se que: $A_j(v_2) \subseteq A_j(v_5)$ e $A_j(v_{12}) \subseteq A_j(v_9)$, assim, monta-se a nova tabela eliminando as colunas $A_j(v_2)$ e $A_j(v_{12})$ o que pode ser observado na Figura 2.13(i), onde se verifica ainda que a linha $A_i(v_{12})$ possui apenas um elemento unitário (marcado de vermelho) na coluna $A_j(v_9)$, esta coluna então deve fazer parte do conjunto dominante conforme a premissa número 1. Deste modo, $A(j) = \{\emptyset\}$, ou seja, que não possuía

2.5 Conjunto Dominante

T		A_j					
		$v1$	$v3$	$v4$	$v6$	$v8$	$v11$
A_i	$v1$	1	0	1	1	0	0
	$v2$	0	0	0	1	0	0
	$v3$	0	1	1	1	0	1
	$v4$	1	1	1	0	0	0
	$v5$	0	1	0	1	0	0
	$v6$	1	1	0	1	0	0
	$v7$	0	0	0	0	1	1
	$v11$	0	1	0	0	0	1

(i)
(ii)

T		A_j			
		$v3$	$v8$	$v11$	
A_i	$v3$	1	0	0	
	$v7$	0	1	1	
	$v11$	1	0	1	

(iii)
(iv)

T		A_j	
		$v7$	$v11$
A_i	$v7$	1	1
	$v11$	0	1

Figura 2.14: Processo booleano de seleção do conjunto dominante mínimo

nenhum dominante, passa a ter o primeiro dominante $A_1(j) = \{v_9\}$. Seguindo o procedimento o próximo passo é a eliminação da coluna $j(v_9)$ e das linhas $i(v_8)$, $i(v_9)$, $i(v_{10})$ e $i(v_{12})$ que possuem elementos unitários na coluna criando a nova tabela da Figura 2.13(ii). Nesta nova tabela, a coluna $A_j(v_5) \subseteq A_j(v_6)$, $A_j(v_7) \subseteq A_j(v_8)$ e $A_j(v_{10}) \subseteq A_j(v_3)$, desta forma a nova tabela é montada eliminando as colunas, $A_j(v_5)$, $A_j(v_7)$ e $A_j(v_{10})$ conforme demonstrado na Figura 2.14(i), onde se percebe que a linha $A_i(v_2)$ possui um elemento unitário (marcado de vermelho) na coluna $A_j(v_6)$, encontra-se então o segundo dominante $A_2(j) = \{v_6\}$. Repete-se o processo para a montagem da próxima tabela demonstrada na Figura 2.14(ii), a coluna $A_j(v_6)$ e as linhas: $A_i(v_1)$, $A_i(v_2)$, $A_i(v_3)$, $A_i(v_5)$ e $A_i(v_6)$ são excluídas. Observa-se nesta tabela que a coluna $A_j(v_1) \subseteq A_j(v_3)$ e $A_j(v_4) \subseteq A_j(v_3)$, desta forma estas colunas são excluídas montando a tabela mostrada na Figura 2.14(iii) onde se observa que a linha $A_i(v_3)$ possui um elemento unitário (marcado de vermelho) na coluna $A_j(v_3)$. Esta coluna é então o terceiro dominante, $A_3(j) = \{v_3\}$. Dando continuidade, cria-se a tabela descrita na Figura 2.14(iv) onde pode ser visto um elemento unitário (marcado de vermelho) na linha $A_i(v_7)$ com coluna $A_j(v_8)$, sendo esta o quarto e último dominante deste conjunto, $A_4(j) =$

2.6 Conceito de Kernel e Verossimilhança

$\{v_8\}$. Por fim, para o grafo demonstrado na Figura 2.11 o mínimo conjunto dominante encontrado pelo processo booleano é: $min_{Do}(A_{(T)}) = \{v_3, v_6, v_8, v_9\}$, o que pode ser observado pelos vértices na cor vermelha da mesma figura. A tabela T dominante final é demonstrada na Figura 2.15 onde se percebe que cada linha $A(i)$ possui pelo menos um valor unitário permitindo a vinculação dos elementos T_{ij} como exigido.

T		A_j			
		v_3	v_6	v_8	v_9
A_i	v_1	0	1	0	0
	v_2	0	1	0	0
	v_3	1	1	0	0
	v_4	1	0	0	0
	v_5	1	1	0	0
	v_6	1	1	0	0
	v_7	0	0	1	0
	v_8	0	0	1	1
	v_9	0	0	1	1
	v_{10}	1	0	1	1
	v_{11}	1	0	0	0
	v_{12}	0	0	0	1

Figura 2.15: Conjunto dominante mínimo da Figura 2.11

2.6 Conceito de Kernel e Verossimilhança

Para os fins deste trabalho estes conceitos se aproximam. O Kernel K é mais relacionado a um núcleo, ou concentração de eventos. Neste caso concentração de conjuntos do Grafo $G_{(S)} = (v, \alpha)$, ou sub-grafos deste. Verossimilhança no tratamento estatístico está mais direcionado ao conjunto de amostras que possuem características semelhantes em classificação de padrões. Assim, para os fins deste trabalho, as verossimilhanças serão utilizadas para se identificar $K(\ddot{G}_N)$, através do grafo de Gabriel e do Conjunto Dominante dada a base com N amostras. Estes conceitos serão usados para encontrar os parâmetros c e σ das Gaussianas a serem utilizadas na funções RBFs das RNAs nos métodos de classificação estudados.

2.7 Estimadores de densidade de Kernels

Na literatura há diferentes formas de se encontrar os centros das gaussianas. Dentre algumas mais eficientes estão o *k-means* (Sing et al., 2003) ou k-médias como é conhecido e o *c-means* (Chiu, 1994) que aplica técnicas baseadas em rede Fuzzy. Estas técnicas podem ser utilizadas para estimar o número e a posição dos centros *c* sendo excelentes estimadores de densidades dos Kernels Gaussianos e podem ser usadas em conjunto com abordagens do tipo: Aleatória, Silhoueta e por Afinidade. Estas três abordagens, melhor detalhadas abaixo, foram utilizadas para uma análise comparativa mais aproximada da hipótese proposta no método a ser apresentado no Capítulo 3 que também é Gaussiano.

- *Aleatório* - Nesta técnica a escolha da localização das Gaussianas ocorre de forma aleatória dos clusters, ou em outras palavras, através de seleção aleatória das densidades, encontram-se os centros e acha-se os raios nas amostras aleatórias selecionadas.
- *Affinity propagation* - Esta abordagem é recursiva e tem como objetivo encontrar os centros mais próximos ou que têm mais afinidade com as amostras mais próximas, ou com densidades mais próximas. Esta técnica visa minimizar o erro na escolha dos centros, desta forma, os centros escolhidos são os que possuem o menor erro de aproximação dos centros "ótimos" ou otimizados durante o processo de escolha. Esta técnica está mais claramente explicada em (Frey & Dueck, 2007).
- *Silhouette* - Também conhecido como método da silhueta, é mais utilizado quando os kernels são formados por amostras com densidades muito próximas ou compactas e estão mais claramente separados de outras amostras densamente compactadas. Este método está bem esclarecido em (Rousseeuw, 1987).

2.8 Funções RBF

Em RNA (*Redes Neurais Artificiais*) as RBFs são conhecidas pela simplicidade de sua abordagem matemática e sua larga aplicabilidade em problemas multi-

dimensionais e diferentes complexidade de solução. Uma RBF pode ser usada em diferentes camadas de uma rede neural nos conjuntos de treinamento, comumente chamados de aprendizagem e nos conjuntos de teste às vezes chamados de conjunto de validação. Há diferentes tipos de funções radiais, mas com uma característica em comum: todas necessitam de uma distância a partir de sua origem (Haykin, 2009) $\varphi(x) = \varphi(\|x\|)$ ou da distância euclideana de seu centro até uma origem, $\varphi(x) = \varphi(\|x - c\|)$. Esta característica habilita o seu uso em problemas não linearmente separáveis através do uso de Kernels 2.6. Dado certo tipo de problema não linearmente separável, os Kernels K , são sub-conjuntos separáveis do conjunto inicial projetados em uma outra dimensão onde este sub-conjunto pode ser separado linearmente ou seja, $\mathcal{S} : \mathbb{R}^{m_o} \longrightarrow \mathbb{R}^1$. Esta definição segue o teorema de (Cover,1985), citado abaixo (Haykin, 2005):

Teorema 2 *Um problema complexo de classificação de padrões, lançado em um espaço de alta dimensão não-linear, é mais provável que seja linearmente separável do que em um espaço de baixa dimensão, desde que o espaço não seja densamente povoado.*

Michelli, estudou tipos de funções radiais para aplica-las em conjuntos de planos \mathcal{S} em \mathcal{H} . O teorema 3 descreve suas deduções:

Teorema 3 *Seja $\{x_i\}_{i=1}^N$ um conjunto de pontos distintos em $\mathcal{S} : \mathbb{R}^{m_o} \longrightarrow \mathbb{R}^1$. Então a interpolação $N - N$ da matriz Φ , da qual o enésimo elemento $i - j$ é dado por: $\varphi_{ij} = \varphi(\|x_i - x_j\|)$, é não singular.*

Onde Φ é a matriz de interpolação dos espaços dada como: $\Phi = \{\varphi_{ij} | = \varphi(\|x_i - x_j\|)\}$, também conhecida como matriz K ou matriz de afinidade (Wanderley, 2013). Das diversas funções RBFs cobertas pelo teorema de Michelli, as mais aplicadas em estudos de redes neurais que usam RBF são as do tipo: Multiquadrática equação 2.12, multiquadrática inversa equação 2.13 e Gaussiana equação 2.14 (Haykin, 2009).

$$\varphi(r) = (r^2 + \zeta^2)^{\frac{1}{2}} \text{ para algum } \zeta > 0 \text{ e } r \in \mathbb{R} \quad (2.12)$$

$$\varphi(r) = \frac{1}{(r^2 + \zeta^2)^{\frac{1}{2}}} \text{ para algum } \zeta > 0 \text{ e } r \in \mathbb{R} \quad (2.13)$$

$$\varphi(r) = \exp\left(-\frac{r^2}{2\sigma^2}\right) \text{ para algum } \sigma > 0 \text{ e } r \in \mathbb{R} \quad (2.14)$$

As equações demonstram que uma RBF requer o ajuste dos parâmetros descritos nas equações (2.12, 2.13, e 2.14): r que equivale ao posicionamento do seu centro, σ que é o raio de sua abertura ou cobertura e do *shape parameter* ζ que é uma constante que estabelece o formato da função, para esta seja ajustada adequadamente evitando um sub ou super-dimensionamento em sua aplicação. Conforme descrito no Teorema 3 e corroborado em (Haykin, 2009), a equação 2.14 que representa uma distribuição do tipo Gaussiana é uma solução muito eficiente para aplicação em problemas não linearmente separáveis e será a função cujos parâmetros serão calculados conforme definido em:

$$r_i = (x_i - c_i) \therefore \varphi(r) = \exp\left(-\frac{(x_i - c_i)^2}{2\sigma_i^2}\right) \quad (2.15)$$

Onde o termo numerador equivale ao quadrado da distância euclideana. A partir de 2.15, os parâmetros a serem encontrados passam a ser: c_i , os centros das Gaussianas e σ_i os raios, já que x_i são os pontos ou amostras, portanto, dados no problema.

2.8.1 Aplicação de RBF em Redes Neurais Artificiais

Uma boa abordagem aplicada em RNA é a utilização de RBF para encontrar kernels em problemas de classificação de diferentes dimensões, por este motivo é comum associar uma RBF com uma MLP (*Multi layer Perceptron*) largamente usada em classificadores clássicos, devido à algumas semelhanças conforme descrito na Tabela 2.1 (Braga et al., 2007):

Tabela 2.1: Comparação de uma rede *RBF* com uma *MLP* convencional

<i>MLP</i>	<i>RBF</i>
Pode ter uma ou mais camadas intermediárias.	Geralmente possui apenas uma camada intermediária
Neurônios das camadas intermediárias e de saídas têm funções semelhantes.	Neurônios das camadas intermediárias têm funções diferentes dos neurônios da camada de saída
Entrada da função de ativação é o produto interno dos vetores de entrada e de pesos.	Entrada da função de ativação é a distância euclidiana entre os vetores de entrada e de pesos.
Separa os padrões de entrada com hiper-planos.	Separa os padrões de entrada com hiper-elipsoides
Constrói aproximadores globais para o mapeamento entrada-saída.	Constrói aproximadores locais para o mapeamento entrada-saída.

A Figura 2.16 demonstra a arquitetura de uma RNA de uma camada escondida com funções radiais.

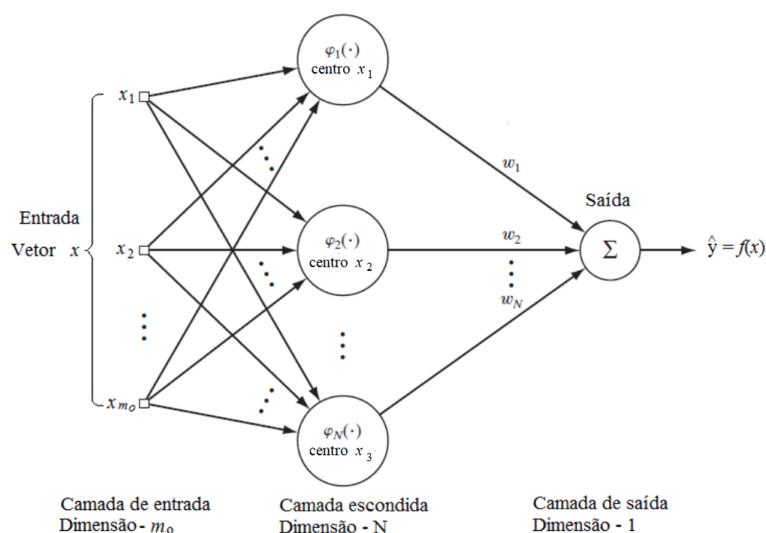


Figura 2.16: Rede neural com RBF

Conforme descrito em (Braga et al., 2007), ao se definir o número de neurônios como o número de padrões de entrada e o número de funções radiais como o número de padrões de treinamento, cada centro pode ser situado sobre um vetor de entrada, com isso, a RBF mapeia com exatidão o vetor de entrada para a saída correta. O problema desta decisão é causar *overfitting* na solução devido à interpolação exata dos padrões. Para evitar este problema uma boa abordagem é dada em (Haykin, 2009), proposta por Broomhead e Lowe em 1988 que é a

interpolação na fase de treinamento e de validação em duas etapas:

- A fase de treinamento otimiza o processo para a separação dos Kernels com base em pontos de dados de entrada e saída conhecidos.
- A fase de validação interpola os pontos de dados dos Kernels com o ajuste dos pesos w permitindo uma aproximação ótima à superfície verdadeira dos Kernels.

Posto isto, a RBF da Figura 2.16 pode ser representada pela função $F(x)$ tal que:

$$F(x) = \sum_{i=1}^N \hat{w}_i \varphi(\|x_i - x_j\|) \quad (2.16)$$

Aplicando os teoremas: 2 da separabilidade proposta por Cover e 3 da interpolação dos espaços proposto por Michelli, na rede neural RBF da Figura 2.16, pode-se deduzir que a saída \hat{y} da rede será uma soma linear das interpolações das entradas com as RBFs ou kernels e os pesos \hat{w}_i levando a equação(2.16) ser reescrita como:

$$F(x) = \sum_{i=1}^K \hat{w}_i \varphi(\|x_i - x_j\|) \quad (2.17)$$

Onde o K representa a matriz de adjacência dos kernels dado por $\varphi(K)$ dos agrupamentos ou sub-conjuntos separáveis (Johnson, 1967; Wanderley, 2013) descrito como:

$$\varphi(K) = \exp\left(-\frac{(x_i - c_i)^2}{2\sigma_i^2}\right) \quad (2.18)$$

$$K = \begin{bmatrix} K_{11} & K_{12} & \dots & K_{1n} \\ K_{21} & K_{22} & \dots & K_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ K_{n1} & K_{n2} & \dots & K_{nn} \end{bmatrix} \quad (2.19)$$

e n é o número de sub-conjuntos verossimilhantes das amostras N . Cada membro K_{ij} da matriz é uma sub-matriz formada pela afinidade entre as amostras x_i e

x_j . Assim, conforme (Wanderley, 2013), a representação das afinidades através da matriz de kernel permite não só construir modelos de redes neurais para classificação e regressão a partir dos dados (Vapnik, 1995), como também contém informação da densidade do conjunto N de amostras. De acordo com estes pressupostos, a fase de treinamento da rede pode ser elaborada a partir de algumas abordagens como: K-médias (Sing et al., 2003), Fuzzy-C-médias (Chiu, 1994), rede SOM (Kohonen, 1990) e WTA (Duda et al., 2012). Nesta fase o estimador de densidade do kernel dado por: $k(x_i, x_j)$ possui apenas um parâmetro a ser arbitrado, ou seja, o σ descrito em (2.18), uma vez que os centros c são estimados pela distância euclidiana entre as amostras. As aproximações descritas abaixo são algumas das abordagens para se arbitrar o σ mais utilizadas pelas RNAs com Kernel radial. A primeira é a dada pela média da distância euclidiana dos centros de cada x_i e os centros de cada x_j (Braga et al., 2007):

$$\sigma_i = \frac{1}{m} \sum_{i=1}^m \|\mu_j - \mu_i\| \quad (2.20)$$

Onde μ_i é o centro mais próximo de μ_i e $\|\mu_j - \mu_i\|$, a distância euclidiana entre os centros. Outra abordagem é atribuir a cada σ_i uma largura diferente dada por:

$$\sigma_i = \alpha \|\mu_j - \mu_i\| \quad (2.21)$$

Onde α é uma constante entre 1, $0 \leq \alpha \leq 1,5$. Uma terceira opção é associar o σ_j à média da distância entre os μ_j aos N vetores de entrada mais próximos de μ_j :

$$\sigma_i = C \left(\frac{1}{N} \sum_{i=1}^N \|\mu_j - x_i\| \right) \quad (2.22)$$

Onde x_i, x_j, \dots, x_n são os N vetores de entrada mais próximos de μ_j , N e C são constantes escolhidas de forma experimental tal que σ_j cubra a área dos vetores de entrada. Uma outra possibilidade a ser destacada é o σ de Silverman(1986) dado por (Wanderley, 2013):

$$\sigma_i = 0,79 Q n^{-\frac{1}{5}} \quad (2.23)$$

Onde Q é o mínimo do intervalo (inter-quartis) da Gaussiana e n o número de dimensões do problema. A proposta de arbitragem utilizada nos Experimentos 4 para a comparação da metodologia proposta 3 neste trabalho é dada por (Bishop, 1995) e apresentada em (Torres, 2016):

$$\sigma_i = \phi_{\varepsilon_s} \quad (2.24)$$

onde σ_i é igual à duas vezes o espaçamento médio entre os centros que equivale ao diâmetro das hiper-esferas traçadas entre cada vértice do grafo de Gabriel, conforme equação 2.9. Após encontrarmos os parâmetros c_i e σ_i das Gaussianas, o problema passa a ser encontrar um \hat{w} que ajuste a equação 2.17 através de:

$$\hat{w} = K^\dagger \hat{y}_D \quad (2.25)$$

Onde K^\dagger é a pseudo inversa de Moore-Penrose (Albert, 1972) da matriz K e \hat{y}_D o vetor de rótulos de cada classe.

2.9 Conclusão do capítulo

Neste capítulo descrevemos uma curta revisão bibliográfica para embasar nossa ideia de como o método proposto poderá usar os postulados discriminados para encontrar os parâmetros c e σ de uma Gaussiana, sem necessidade de ajustes, para depois aplica-la nas RBFs de classificadores de uma RNA. Na seção 2.4 demonstramos como o grafo de Gabriel é formado e na seção 2.5 como o grafo pode ser dividido em sub-grafos dominantes que são kernels agrupados por amostras verossimilhantes. Por fim, vimos que estes kernels podem ser utilizados em uma RNA na seção 2.8.1 e desta forma ser aplicada em um processo de classificação.

Capítulo 3

Abordagem Proposta

Neste capítulo colocamos em discussão o método proposto para o problema de classificação de padrões. O método sugerido utilizará dos conceitos e postulados matemáticos apresentados no capítulo 2 para encontrar os parâmetros c_i e σ_i das gaussianas sem a necessidade de arbitragem ou de ajustes no modelo. Será discutido ainda uma proposta de filtragem de ruídos, ou de amostras sobrepostas na aplicação do método em bases que possuem desbalanceamento e onde esta característica pode influenciar no resultado do classificador.

3.1 Formação do Grafo de Gabriel e Dominância

O algoritmo proposto seleciona as amostras de treinamento de cada classe da entrada x_i de N , após o tratamento da base estudada. Neste caso, o tratamento mencionado refere-se à retirada dos dados não necessários para o vetor \hat{x} de entrada como: rótulos e nomes. O treinamento passa a ser estruturado iniciando pela elaboração do grafo de Gabriel no conjunto amostrado. Conforme descrito na seção 2.4, através de Voronoi seção 2.2 e Delaunay seção 2.3, para cada classe de amostras selecionadas será elaborado \ddot{G} a partir do planar $\tilde{G}_{(S)}$. O grafo é traçado para os padrões de entrada tal que $\ddot{G} \subseteq v\{\hat{x}_i | i = 1, ..2, \dots, N_v\}$ e o conjunto de arestas α satisfaçam a equação (2.10). A Figura 3.1 mostra a formação de $\ddot{G}(v, \alpha)$ para um problema bidimensional com duas classes (Torres, 2012). O segundo passo após a elaboração do grafo é montagem da dominância. Feito através do conceito de conjunto dominante, descrito na seção 2.5. A partir do

3.1 Formação do Grafo de Gabriel e Dominância

$\ddot{G}_{(S)} = (v, \alpha)$ este será um subconjunto de tal modo que o número de vértices v equivalha ao número de dominação que respeita à equação (2.11).

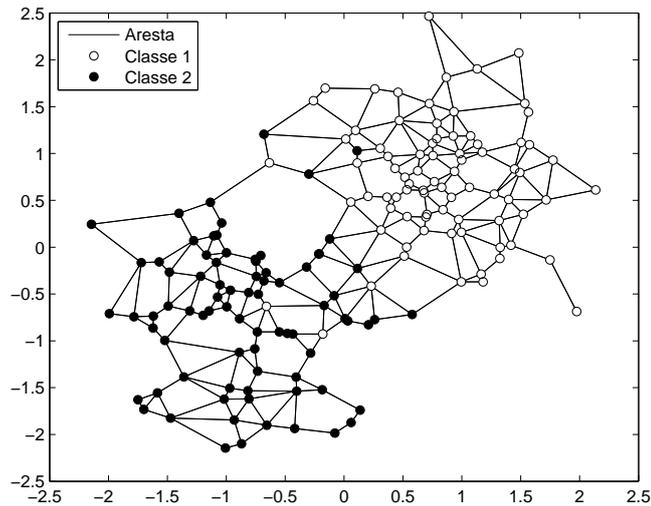


Figura 3.1: Grafo de Gabriel para classificador bidimensional

A Figura 3.2 mostra a abordagem sugerida pelo método proposto. Na Figura 3.2(i) é demonstrada a formação do grafo de Gabriel, para cada classe, diferenciando da Figura 3.1 na qual o grafo foi feito para todo conjunto. Cada par de vértices do grafo é interligado pela aresta que os conecta, conforme os critérios de Voroni e Delaunay, descritos no capítulo 2. Os vértices do grafo são os elementos da entrada x_i para o módulo do algoritmo que calcula os dominantes conforme o critério descrito por (Barroso, 2001) e demonstrado nas Figuras (2.12 e 2.13) encontrando o conjunto mínimo de dominância máxima conforme descrito na seção 2.5 e demonstrado na Figura (2.15), colorindo os vértices dominantes em vermelho, (mais escuro para a classe 1 e mais claro para a classe 2), o que pode ser observado na Figura 2.5(ii). Após serem encontrados os vértices dominantes continua-se o processo.

3.2 Seleção dos centros e raios da RBF

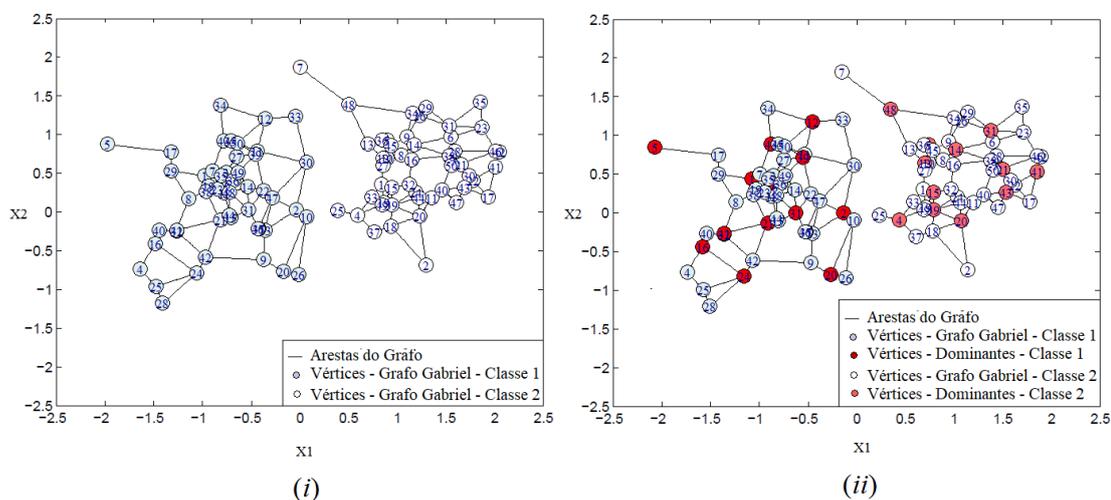


Figura 3.2: Formação da Dominação para classificador de duas classes

3.2 Seleção dos centros e raios da RBF

Como descrito na seção 2.8.1, estudos recentes apresentam diferentes maneiras de se calcular os parâmetros c e σ das gaussianas. A abordagem proposta parte dos pressupostos descritos anteriormente e que podem ser aproveitados para a hipótese colocada. Como descrito em (Braga et al., 2007) o número de neurônios de uma RBF pode ser associado ao número de padrões de entrada, isto significa que o número de gaussianas está diretamente ligado ao parâmetro x e à matriz K demonstrada em 2.19. Pelo conceito de verossimilhanças e formação de kernels, estas gaussianas podem de fato representar o centro das concentrações destes kernels sendo sua aproximação perfeitamente plausível pelo cálculo da distância euclidiana destas amostras $K(\hat{\mathbf{x}}_i, \hat{\mathbf{x}}_j)$. Posto isto, o método propõe usar os vértices dominantes do grafo de Gabriel como os centros c das gaussianas. Esta hipótese vem do princípio descrito na seção 2.5, onde vértices de um grafo são dominantes em relação a outros vértices adjacentes que para a abordagem são as amostras verossimilhantes. Os raios σ das gaussianas serão encontrados pela distância euclidiana média destas adjacências que são os vértices dominados de cada conjunto dominante, utilizando a equação 2.20.

3.3 Formação da RBF proposta

A RBF utiliza a equação gaussiana descrita em 2.14, onde r é o termo relacionado aos centros c . O método proposto encontra os parâmetros c e σ , conforme já descrito, dado um problema de classificação. Os centros correspondem ao número de vértices dominantes $\mathcal{V}_{Do(i)}$ do grafo de cada classe descrito na equação 3.1 e os raios σ encontrados pela equação 2.20 (Torres, 2016) correspondem à média distância euclidiana substituindo o termo μ_j pelos vértices de cada classe e o termo μ_i pelas adjacências $\mathcal{V}_{Adj(i)}$ de cada vértice conforme equação 3.2:

$$c_i(x_1) = \mathcal{V}_{Do(i)}(x_1) \text{ e } c_i(x_2) = \mathcal{V}_{Do(i)}(x_2) \quad \forall i = \{1, 2, \dots, n\} \text{ de } x_1 \text{ e } x_2 \quad (3.1)$$

$$\sigma_i = \frac{1}{m} \sum_{i=1}^m \|\mathcal{V}_{Do(i)}(x) - \mathcal{V}_{Adj(i)}(x)\| \quad (3.2)$$

Figura 3.3 mostra graficamente o processo, onde pode ser observado 6 vértices ou centros para a classe 1 com as gaussianas traçadas e 7 para a classe 2.

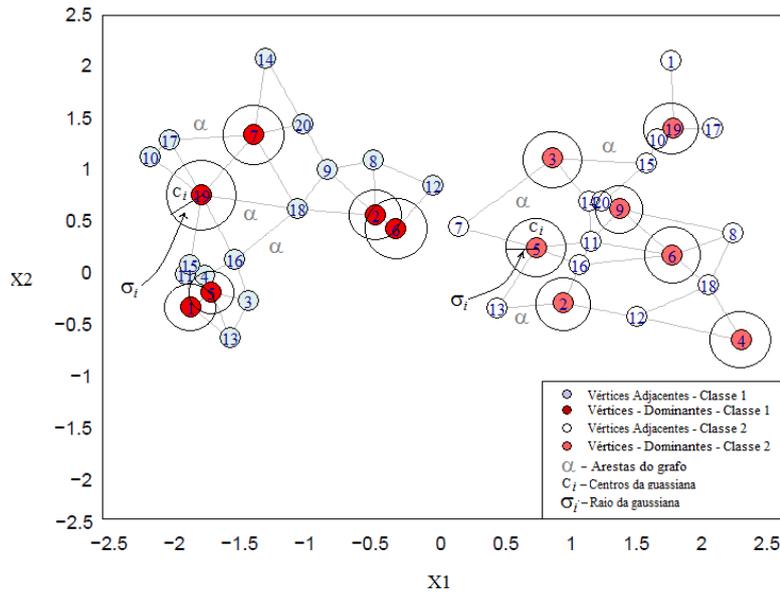


Figura 3.3: Seleção dos centros e raios das gaussianas

3.3 Formação da RBF proposta

A hipótese é que o método, ao ser utilizado em problemas de classificação, evite *overfitting* uma vez que os vértices dominantes que serão os centros ou neurônios ou ainda RBFs do conjunto de treinamento, Figura 3.3 e este número será sempre inferior ao número de amostras do vetor de entrada x ou vértices (adjacentes e/ou outros dominantes) evitando uma sobreposição de amostras entre a entrada e saída. O algoritmo do método proposto é apresentado abaixo:

Algoritmo 1: Metodologia proposta

```
1 Carrega a base de dados
2 para Repete até o critério de parada faça
3   gg( $\hat{x}_1, \hat{x}_2$ ) - "Montando o Grafo de Gabriel"
4   para  $\hat{x}_1 = 1$  a  $N$  faça
5     para  $\hat{x}_2 = 1$  a  $N$  faça
6       evolui  $\ddot{G}(\hat{x}_1, \hat{x}_2)$  usando a equação 2.10;
7     fim
8   fim
9    $V_{Do}(gg1, gg2)$  - "Vértices dominante do Grafo de Gabriel"
10  para gg1 = 1 a  $N$  faça
11    para gg2 = 1 a  $N$  faça
12      evolui  $\gamma(gg1, gg2)$  usando o método de busca por estrutura de
13        árvore 2.5;
14    fim
15  fim
16 Separa os Dominantes  $\gamma(gg1, gg2)$  e encontra as adjacências  $\alpha$ ;
17 Calcula a distância euclideana  $dist(V_{Do}, V_{Adj})$ ;
18 Encontra as RBF com os centros  $c = \gamma$  e  $\sigma = mean(dist(V_{Do}, V_{Adj}))$ ;
```

O uso de kernels, como já descrito, permite aplicar o método em problemas com bases de diferentes dimensões. O treinamento da rede é feito com a matriz K descrita em 2.19 utilizando as RBFs encontradas pelas equações 3.1 e 3.2 aplicando-se a equação 2.24, solução muito parecida com uma MLP de uma camada escondida conforme descrito no quadro 2.1. Uma vez a rede treinada, a validação é feita utilizando:

$$\hat{y}_{teste} = \hat{w}K \quad (3.3)$$

3.4 Demonstração em problemas com bases bidimensionais

Onde \hat{y}_{teste} , corresponde à saída da rede testada, \hat{w} os pesos encontrados no treinamento da rede e K a matriz de kernel com as RBFs encontradas pelo método. O erro de teste tende a ser mínimo indicando que o classificador alcançou seu objetivo de classificação dada a métrica escolhida:

$$Erro_{teste} = sign(\hat{y}_{teste} - \hat{y}_D) \approx 0 \quad (3.4)$$

Onde \hat{y}_D corresponde ao vetor de rótulos dado no problema. As equações 3.3 e 3.4 são utilizadas pelas métricas: *AUC* (*area under the curve*) de teste e *Acurácia* de teste. Estas métricas foram as escolhidas para comparar com os resultados dos métodos Gaussianos e de classificadores clássicos estudados para validar hipótese levantada e utilizada na abordagem proposta.

3.4 Demonstração em problemas com bases bidimensionais

Foram escolhidos seis problemas clássicos de classificação da literatura para demonstração dos passos descritos nas seções: 3.1, 3.2 e 3.3. Para uma melhor visualização do processo a legenda foi retirada das figuras. Isso se faz necessário devido à sobreposição e proximidade das amostras de cada classe. A Figura 3.4, mostra a legenda das figuras 3.5 à 3.10 que demonstram o processo em três etapas:

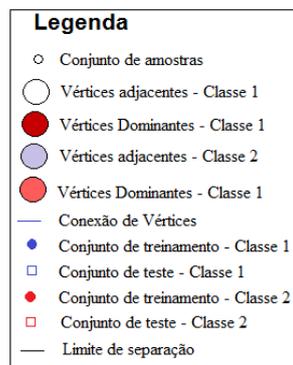


Figura 3.4: Legenda das figuras

3.4 Demonstração em problemas com bases bidimensionais

A primeira identificada como (i), mostra a disposição das amostras de entrada no problema. A segunda etapa identificadas nas figuras como (ii), refere-se à formação do grafo de Gabriel e dos mínimos conjuntos dominantes de cada classe. A terceira etapa identificadas nas figuras como (iii), demonstra a solução do problema, feito através das gaussianas encontradas, aplicadas ao conjunto de treinamento e teste. Nesta terceira etapa, a rede treinada e testada é separada demonstrando a capacidade do modelo atuar como classificador.

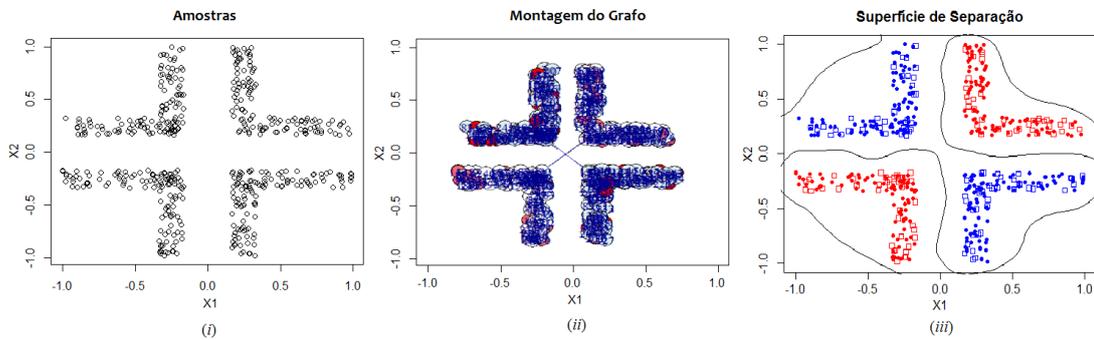


Figura 3.5: Método proposto aplicado ao problema Corners

A Figura 3.5 refere-se ao problema Corners. Neste problema pode ser observado a grande densidade de amostras em cada classe limitando muito a visualização do grafo e da dominância. Na terceira etapa pode ser visto a finalização do processo com as amostras de treinamento representadas por pontos (*vermelhos e azuis*) de teste, representadas por quadrados (*vermelhos e azuis*) e a superfície de separação traçada pela equação 3.3.

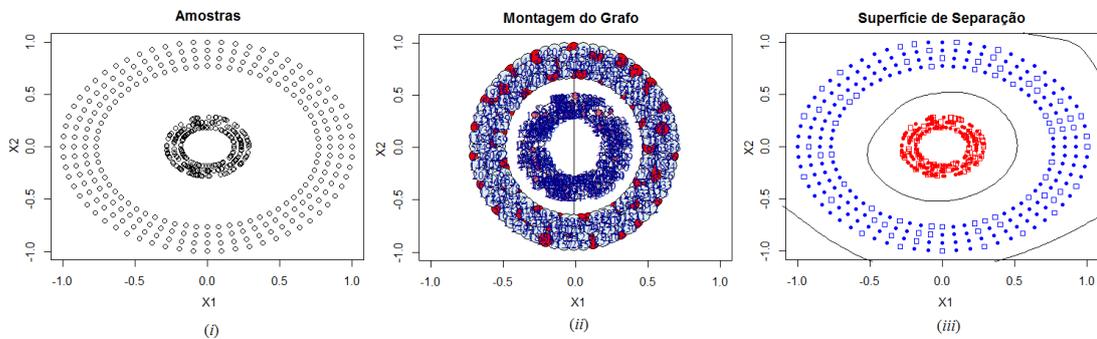


Figura 3.6: Método proposto aplicado ao problema Clusters

3.4 Demonstração em problemas com bases bidimensionais

A Figura 3.6, demonstra a aplicação do método para o problema de classificação Cluster e a Figura 3.7 para o problema Two moons. Para ambos os problemas, é possível visualizar os dominantes embora o grafo de Gabriel ainda seja de difícil de ser identificado.

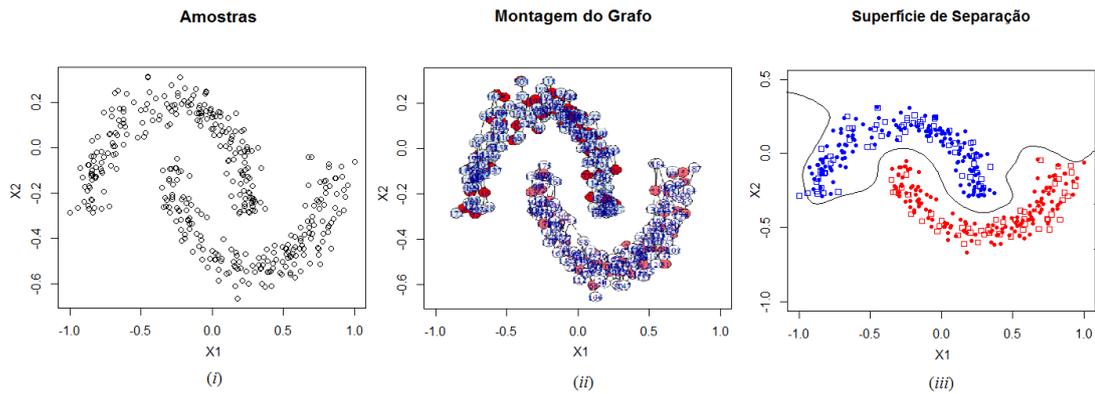


Figura 3.7: Método proposto aplicado ao problema Two moons

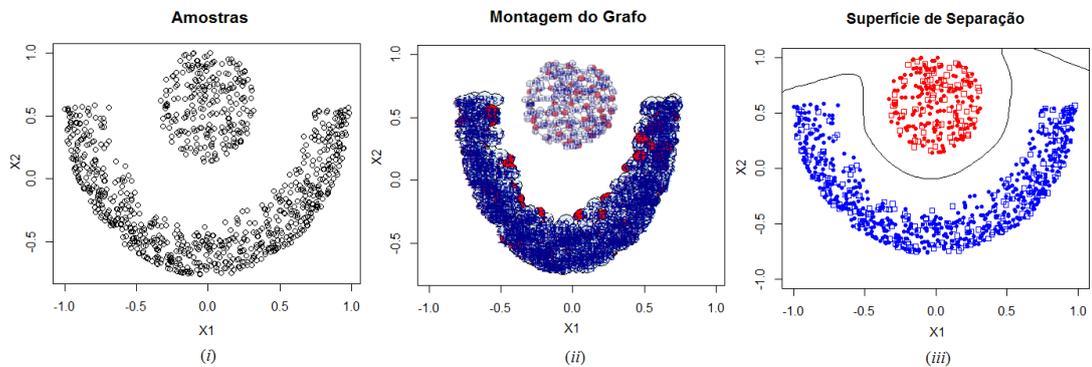


Figura 3.8: Método proposto aplicado ao problema Full moons

A Figura 3.8 apresenta o problema Full moons, a Figura 3.9 o problema Half Kernel e a Figura 3.10 o problema Spiral. Todas as figuras demonstram que o algoritmo proposto foi capaz de separar as classes de amostras, traçar o grafo de Gabriel em cada classe, eleger os vértices dominantes e separa a matriz K . Após esta etapa ele treina a rede e valida apresentando a superfície de separação em uma solução parecida com a MLP de uma camada a ser discutida na próxima seção.

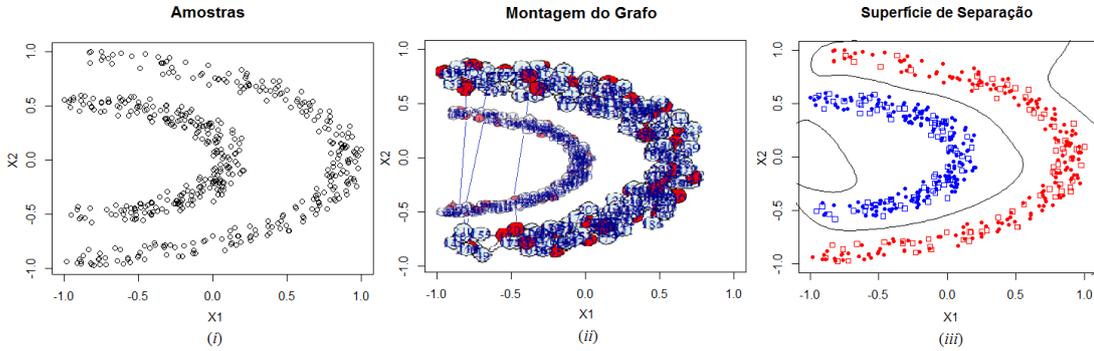


Figura 3.9: Método proposto aplicado ao problema Half kernels

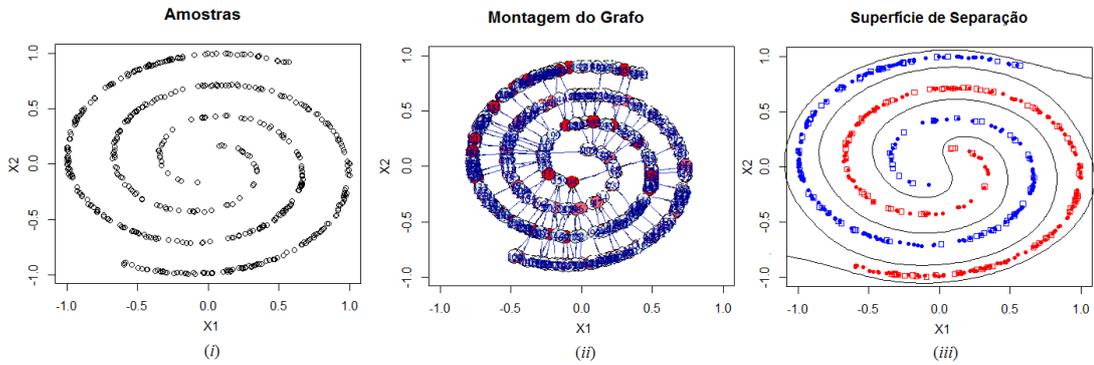


Figura 3.10: Método proposto aplicado ao problema Spiral

3.5 Tratando a Sobreposição

Nesta seção será apresentada uma sugestão de filtragem dos ruídos, ou em outras palavras a sobreposição de amostras de classes diferentes. Em problemas reais, a superposição de amostras, ou seja, amostras rotuladas com erro, podem representar uma diminuição da acurácia do classificador. Isto tende a ser mais impactante quando o "*universo*" amostrado é menor, pois à medida que este aumenta, um bom método apresenta ferramentas de filtragem destes erros apresentando uma melhor eficácia. Esta filtragem pode ser feita através de diferentes técnicas como as apresentadas em (Carvalho et al., 2007; Torres, 2016; Wanderley, 2013). A abordagem proposta apresentada a seguinte hipótese: Dado um Grafo $\dot{G}_{(S)} = (v_i, \alpha_i)$ traçado a partir de um problema de classificação, onde $S = \{(v_i, y_i) | i = 1, \dots, N\}$, $y_i \in \{+1, -1\}$ são os rótulos de cada classe e $v_i \in \mathbb{R}$ os

vértices; Sejam v_j os vértices adjacentes a v_i ligados por suas arestas α_i , v_{j+} e v_{j-} as adjacências de cada classe, parte-se do pressuposto que estes vértices estão na região próxima à separação das classes e que as adjacências podem representar uma sobreposição. Assim para o conjunto formado pelo vetor de vértices \hat{v}_i e suas adjacências tal que: $\hat{v}_{(v_i)} \subseteq \{v_{j1}, v_{j2}, \dots, v_{jn}\} \forall v_i$ e $v_j \in \mathbb{R}$, temos:

$$\hat{v}_{v_r}(i) = \sum_{j=1}^{N_{v_{j+}}} v_{j+} > \frac{1}{2} \left(\sum_{j=1}^{N_{v_{j-}}} v_{j-} \right) + 1 \quad \forall y_i = -1 \quad (3.5)$$

$$\hat{v}_{v_r}(i) = \sum_{j=1}^{N_{v_{j-}}} v_{j-} > \frac{1}{2} \left(\sum_{j=1}^{N_{v_{j+}}} v_{j+} \right) + 1 \quad \forall y_i = 1 \quad (3.6)$$

Onde \hat{v}_{v_r} é o vetor com os vértices relacionados aos ruídos e que serão excluídos do conjunto de treinamento, v_{j+} e v_{j-} são as adjacências da classe 1 e da classe 2 respectivamente. A equação 3.5 é aplicada para a sobreposição da classe positiva e a equação 3.6 é aplicada para a sobreposição da classe negativa. A Figura 3.11 demonstra a o problema da sobreposição em uma base sintética com duas distribuições aleatórias normalizadas de 350 amostras cada. Na Figura 3.11(i) observa-se que o método ao encontra no treinamento e no teste amostras sobrepostas(ruídos). Na Figura 3.11(ii) observa-se que após a filtragem há a eliminação quase que total dos ruídos.

3.5.1 Problemas com o tratamento da Sobreposição

Alguns problemas podem ocorrer ao se tratar a sobreposição. As bases possuem características diferentes o que impede o mesmo tratamento de métodos para problemas com características distintas. Os vetores $\hat{v}_{v_r}(i)$ apresentados em 3.5 e 3.6 podem retornar resultados indesejados ou até mesmo inconsistentes com a análise e isso pode ocorrer principalmente por dois motivos: o problema apresentar uma área de separação onde há muita sobreposição e o número de amostras ser bem desequilibrado, ou seja, haverá mais N_{v+} ou mais N_{v-} . Na primeira situação a montagem dos grafos de cada classe poderá apresentar algum erro na classificação da classe que está mais sobreposta. Na segunda situação a montagem do grafo das classes poderá ser representada como uma sobre-montagem do grafo da menor

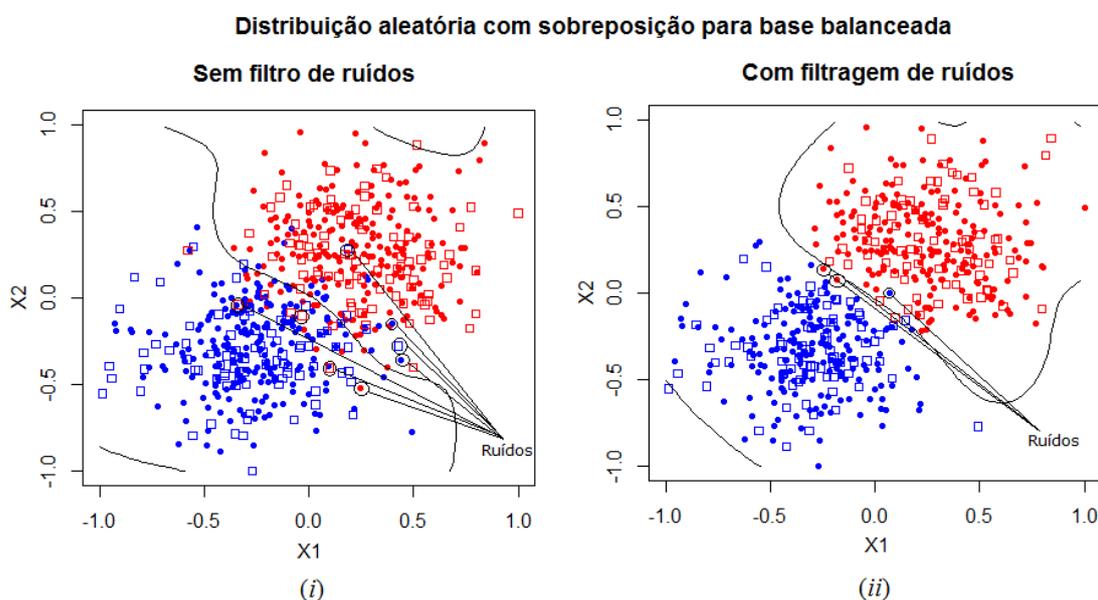


Figura 3.11: Método proposto aplicado às amostras sobrepostas

classe sobre o da maior e uma interpretação de que o grafo menor é representado por ruído. Como consequência, este pode ser totalmente eliminado. Este erro de interpretação poderá interferir na acurácia dos resultados e por conseguinte na diminuição da eficácia. Como solução para este problema foram sugeridas duas abordagens estatísticas:

- Abordagem para bases balanceadas;
- Abordagem para bases desbalanceadas.

3.5.2 Bases com classes balanceadas

Para este tipo de base, os grafos são montados de forma bem distintas e os vetores de ruído $\hat{v}_{v_r}(i)$ seguem às premissas descritas nas equações 3.5 e 3.6, no entanto, próximo ao limite de separação das classes ocorrerá situações em que $\|\hat{v}_{v_r}(i)\| = \frac{1}{2}$. As equações 3.7 e 3.8 descrevem este cenário onde ao se verificar que o vértice (amostra) está na classe oposta e o vetor de adjacências ao qual ele pertence tem seu módulo igual à 0,5, o que significa que metade de seus vértices está na mesma classe e metade na outra, este vértice é considerado um ruído. A Figura 3.11(ii)

mostra que as equações 3.5, 3.6, 3.7 e 3.8, foram eficazes na eliminação quase que total da sobreposição de amostras.

$$\|\hat{v}_{v_r}(i)\| = \sum_{j=1}^{N_{v_{j+}}} v_{j+} = \frac{1}{2} \wedge v_{j+} \supset N_{v_{j-}} \wedge N_{v_{j-}} \gtrsim N_{v_{j+}} \forall y_i = -1 \therefore \hat{v}_{v_r}(i) \notin N_{v_{j-}} \quad (3.7)$$

$$\|\hat{v}_{v_r}(i)\| = \sum_{j=1}^{N_{v_{j-}}} v_{j-} = \frac{1}{2} \wedge v_{j-} \supset N_{v_{j+}} \wedge N_{v_{j+}} \gtrsim N_{v_{j-}} \forall y_i = 1 \therefore \hat{v}_{v_r}(i) \notin N_{v_{j+}} \quad (3.8)$$

3.5.3 Bases com classes desbalanceadas

O problema de bases desbalanceadas é bem conhecido e vem sendo tratado em diferentes abordagens conforme (Castro & Braga, 2013) que descreve que tal problema surge como uma consequência direta de assumir custos (perdas) iguais aos erros de classificação, assim, sob a hipótese de custo igual, o único objetivo se torna a minimização de um critério geral que corresponde à probabilidade de erro geral (ou taxa de erro de classificação esperada). Neste sentido, modelos treinados de predição tendem a produzir regras de decisão tendenciosas para a classe maior o que causa um erro de acurácia nos resultados dos testes. Dentre as diferentes possibilidades de tratar o problema conforme descrito em (Castro & Braga, 2013), o método proposto utilizou a *undersampling* na qual se separa aleatoriamente do conjunto total S_N de amostras, uma quantidade igual da classe majoritária S_{maj} em relação à classe minoritária S_{min} de maneira a equilibrá-las, tal que $|S_N| = |S_{min}| + |S_{maj}| - |E_q|$. Conforme descrito em (Haibo & Garcia, 2009), estudos têm mostrado que para vários classificadores, um conjunto de dados equilibrados melhorou o desempenho geral da classificação em comparação com um conjunto de dados desequilibrados, o que justifica a utilização de métodos de amostragem para aprendizagem desequilibrada e para a maioria dos conjuntos desequilibrados estas técnicas de amostragem melhoram as acurácias do classificador. A escolha da técnica de *undersampling* se deu pelo fato desta não necessitar

de ajustes de parâmetros para a sua aplicação, deixando sua configuração mais simples na abordagem proposta que também não necessita de ajustes. Outras técnicas (Castro & Braga, 2013) usam ajustes para se encontrar a melhor forma de se equilibrar as bases. O critério utilizado no método foi aplicado às bases em que a diferença entre N_{v+} e N_{v-} é maior que 65% (valor definido empiricamente). A Figura 3.12 demonstra a eficácia do método para uma base sintética bidimensional de duas Gaussianas com distribuição aleatória desbalanceada.

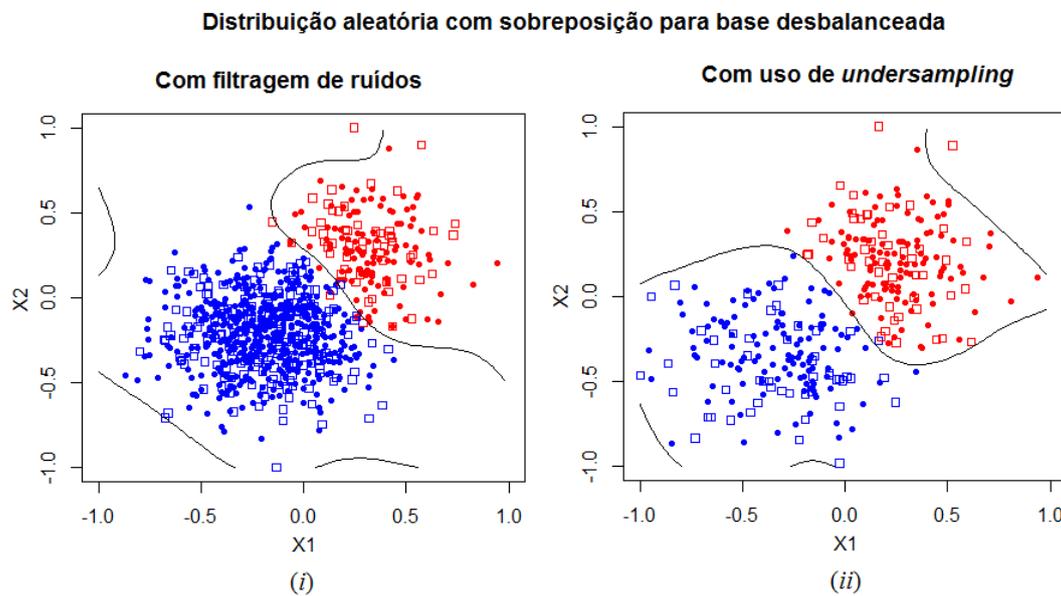


Figura 3.12: Método proposto aplicado a base desbalanceada com *undersampling*

Para uma distribuição aleatória de duas classes, sendo $N_{v+} = 700$ e $N_{v-} = 200$, na Figura 3.12(i), pode ser observado o resultado da classificação já com a retirada de ruídos, ou seja, amostras sobrepostas. Na Figura 3.12(ii) após aplicar a técnica de *undersampling* descrita em (Haibo & Garcia, 2009), há uma melhora da acurácia como será mostrado nos resultados em algumas bases como será demonstrado na próxima seção.

3.6 Conclusão do capítulo

Neste capítulo demonstramos nossa hipótese na qual o conjunto dominante mínimo do grafo de Gabriel pode eleger seus vértices como centros e a distância média entre os vértices e suas adjacências como raios das gaussianas das RBFs a serem utilizadas nas RNAs. Demonstramos em seis problemas bidimensionais a aplicação do método proposto. Analisamos na seção 3.5 problemas de sobreposição e apresentamos algumas maneiras de filtragem. Também descrevemos que a filtragem pode ser um problema quando a base tem um desbalanceamento superior a 65% e optamos pela técnica de *undersampling*, que não necessita de ajuste de parâmetros, para tratar o desbalanceamento.

Capítulo 4

Experimentos

Neste capítulo apresentamos os experimentos executados e discutimos a performance da abordagem proposta a fim de validar nossa hipótese, ou seja, que tratar amostras como um grafo de Gabriel, vértices Dominantes deste como centros de RBFs e a média de vértices dominados como os raios na parametrização da RNA em problemas de Classificação pode ser uma abordagem muito plausível. Explicamos como o algoritmo prepara a base para os experimentos. Aplicamos o método em 18 bases reais do repositório UCI (Asuncion & Newman, 2007). Confrontamos os resultados em dois cenários diferentes sendo o primeiro em uma comparativa com métodos em que não há arbitragem de parâmetros e o segundo em que a parametrização é necessária. Também analisamos a influência do processo de *undersampling* nas bases desbalanceadas e por fim, fazemos uma validação de significância dos resultados das acurácias de teste. Para avaliar a significância dos resultados existem testes como os descritos em (Demsar, 2006), dos quais optamos pelo teste não paramétrico de Wilcoxon devido à natureza dos experimentos não assumirem distribuições normais.

4.1 Explicando os cenários estudados

Com o objetivo de deixar os experimentos mais consistentes e coerentes com a abordagem proposta, dois cenários foram propostos para uma análise comparativa dos resultados. Conforme já descrito o método proposto encontra os

parâmetros das Gaussianas, sem arbitragem, para depois aplica-las nos classificadores. Posto isto, o Cenário I na seção (4.5) apresenta um *benchmarking* entre o modelo proposto com outros modelos Gaussianos nos quais não há arbitragem de parâmetros. Tais modelos usam as técnicas: K-médias (Sing et al., 2003), Fuzzy-C-médias (Chiu, 1994), para estimar os centros, a distância euclidiana das amostras até os centros encontrados como raios das Gaussianas e os métodos: *Affinity propagation* (Frey & Dueck, 2007), *Silhouette* (Rousseeuw, 1987) e aleatório para estimar as densidades de Kernel, para utiliza-los no classificador. O foco neste caso é a análise comparativa dos resultados de Acurácia de teste e AUC em métodos auto-ajustáveis dos parâmetros da gaussiana descritos na equação 2.14, e neste sentido, o objetivo do método proposto é aproximar destes resultados, demonstrando a simplicidade da abordagem e facilidade de configuração. O Cenário II na seção (4.6), apresenta um *benchmarking* do modelo proposto com os modelos clássicos de classificação: modelo MLP (Haykin, 2005; Ruck et al., 1990), devido ao tipo de abordagem comparativa demonstrada na Tabela 2.1. Modelos: RBF-clássico (Berthold & Diamond, 1995) e *Baeyiano* (Christian et al., 2014), devido ao tipo de proximidade das funções RBF a serem utilizadas no classificador. Modelo ELM (Huang et al., 2011, 2006), devido à simplicidade de configuração em problemas de classificação de dimensões diferentes e modelos: SVM (Vapnik, 1995) e LS-SVM (Suykens et al., 2002) devido à robustez de ambas as abordagens em problemas de classificação. O foco neste cenário é tentar demonstrar que o método proposto além de simples em sua abordagem e de fácil configuração, pode apresentar resultados satisfatórios de Acurácia de teste e AUC, quando comparados com modelos robustos e largamente utilizados na literatura servindo como parametrizador para futuros estudos que utilizam técnicas diversas, mas cujo objetivo é alcançar o menor erro de classificação possível, independente da dimensão do problema.

4.2 Bases testadas

Para validação do método proposto foram escolhidas 18 bases reais do repositório UCI (Bache & Lichman, 2013) descritas na Tabela 4.1. As bases escolhidas possuem diferentes dimensões sendo todas aplicáveis em problemas de classificação.

4.2 Bases testadas

Além das bases reais descritas, foi criada uma base sintética bidimensional com duas Gaussianas variando o número de amostras entre $100 \leq N \leq 2000$ com o objetivo de demonstrar graficamente o comportamento do método na formação do grafo de Gabriel (Fig.3.1), vértices dominantes (Fig.3.2), na retirada de ruídos (Fig.3.11) e uso da técnica de *undersampling* (Haibo & Garcia, 2009) (Fig.3.12).

Tabela 4.1: Bases Reais - Repositório UCI

Base de dados	Alias	N_d	N	N_{v+}	N_{v-}
<i>Appendicitis</i>	<i>Append</i>	7	106	21	85
<i>Australian</i>	<i>Austral</i>	14	690	307	383
<i>Banknote Authentication</i>	<i>Bank</i>	4	1372	610	762
<i>Blood</i>	<i>Blood</i>	4	784	178	570
<i>Breast Cancer</i>	<i>Br-Cancer</i>	9	683	444	239
<i>Breast Cancer Hess Probes</i>	<i>Br-Hess</i>	30	133	99	34
<i>Bupa</i>	<i>Bupa</i>	6	345	145	200
<i>Climate Model Simulation Craches</i>	<i>Climate</i>	18	540	494	46
<i>Pima Indian Diabetes</i>	<i>Diabetes</i>	8	768	500	268
<i>Fertility</i>	<i>Fertility</i>	9	100	12	88
<i>Stalog German Credit</i>	<i>German</i>	24	1000	700	300
<i>Glass Identification</i> ¹	<i>Glass</i>	9	214	29	185
<i>Haberman's Survival</i>	<i>Haberman</i>	3	306	225	81
<i>Stalog Heart</i>	<i>Heart</i>	13	270	150	120
<i>Indian Liver Patient</i>	<i>ILP</i>	10	579	414	165
<i>Ionosphere</i>	<i>Ionos</i>	33	351	225	126
<i>Parkinsons</i>	<i>Parkins</i>	22	195	147	48
<i>The Wisconsin Breast Cancer</i>	<i>WPBC</i>	33	194	46	148

Onde N como já descrito é o número de amostras, N_d o número de dimensão da base e N_{v+} e N_{v-} o número de amostras (vértices do grafo de Gabriel) de cada classe.

¹A base Glass icialmente possui sete classes. A utilizada neste trabalho foi tradada no artigo (Castro & Braga, 2013) que a transformou em um problema de duas classes

4.3 Preparação para os experimentos

Para testar a capacidade de generalização do classificador proposto foi utilizado o método de validação cruzada com 10 partições, ou seja, o algoritmo separa a base em 10 partições de maneira aleatória. A cada iteração, uma das partições é utilizada como teste e as demais utilizadas como treinamento. Ao final os valores médios das AUC (*treinamento* e *teste*) e da Acurácia do teste, com seus respectivos desvios padrões (σ_N) são obtidos. O algoritmo proposto foi elaborado na linguagem R e utilizado na plataforma R.Studio versão 1.0.136, para 64 Bits. Para os testes foram apresentadas duas versões: Para as bases que possuem menos desbalanceamento e onde a sobreposição não é filtrada é utilizada a versão *Dom*. Já a versão *Dom-SR* é utilizada para as bases com desequilíbrio acima de 65% (valor obtido empiricamente) onde se aplica os processos de filtragem e técnica de *undersampling* explicados em 3.5. Para os demais modelos estudados descritos em 4.1 foram utilizados diferentes pacotes do R. No primeiro cenário foram usados os pacotes: *Klar*, *apcluster*, *e1071* e, *cluster*. No segundo cenário, foi utilizado o pacote *Caret*. Um ponto a se destacar nestes cenários é que os pacotes usam técnicas diferentes de otimização, validação cruzada e em alguns casos técnicas recursivas para obter o melhor resultado para cada problema estudado o que nos permite destacar que o fato do método proposto não utilizar estes processos, seja de otimização seja de iteração, possa ser um diferencial na análise complexidade do problemas dada a simplicidade da abordagem proposta.

4.4 Apresentação dos resultados

Em cada um dos cenários estudados foram feitos dois experimentos. O primeiro analisando as bases balanceadas ou onde o desbalanceamento não possui efeitos significativos na sobreposição de amostras. O segundo em bases cujo desbalanceamento é maior que 65% entre as classes N_{v+} e N_{v-} . O procedimento dos testes para cada cenário seguiu a mesma parametrização de análise até mesmo para que se fosse feita uma comparação, ainda que com características diferentes de implementação pois, como já descrito, os modelos Gaussianos e de classificação estudados, possuem processos internos de otimização proprietários dos pacotes do R. Os resultados dos experimentos de cada cenário são apresentados nas ta-

4.5 Cenário I - Modelos Gaussianos de classificação

belas 4.3 à 4.22. As bases foram colocadas em ordem alfabética, não havendo correlação com níveis de complexidade do problema. As análises e discussões dos resultados focaram nas respostas do método proposto, com ou sem filtragem de ruídos, comparando-o em ambos os cenários usando três métricas: AUC do conjunto de treinamento, AUC do conjunto de teste e Acurácia dos testes. Para o teste não paramétrico de Wilcoxon foi utilizado o site da *Social Science Statistics* (Students & others working in the social sciences, 2017) e são apresentados os valores de: Média do valor do rank de Wilcoxon (\bar{w}_w), o valor paramétrico de Wilcoxon (z_w), o desvio padrão de Wilcoxon (σ_w) e o p-valor de Wilcoxon (p_w).

4.5 Cenário I - Modelos Gaussianos de classificação

Para os modelos baseados em Kernel Gaussianos foram utilizados os métodos: *Silhouette* (Rousseeuw, 1987), *Affinity propagation* (Frey & Dueck, 2007) e Aleatório para selecionar o número de centros c . A Tabela 4.2, retirada de (Torres, 2016), demonstra resumidamente as técnicas utilizadas e descritas na seção 2.7.

Tabela 4.2: Métodos de cálculo de centros e estimação de Kernels

Nome	Seleção de Centros	Seleção do número de Centros
K- <i>Silhouette</i>	k-médias	<i>Silhouette</i>
K- <i>Affinity propagation</i>	k-médias	<i>Affinity</i>
K-Aleatório	k-médias	Aleatório
F- <i>Silhouette</i>	c-médias	<i>Silhouette</i>
F- <i>Affinity propagation</i>	c-médias	<i>Affinity</i>
F-Aleatório	c-médias	Aleatório

Para encontrar os raios foi utilizada a metodologia encontrada em (Bishop, 1995) e proposta em (Torres, 2016), onde σ é igual à duas vezes o espaçamento médio entre os centros que equivale ao diâmetro das hiper-esferas traçadas entre cada par de vértice do grafo de Gabriel, dado pela equação 2.24. As técnicas *k-médias* (Sing et al., 2003) e *c-médias* (Chiu, 1994) baseado em Fuzzy foram utilizadas para a localização dos centros. As abordagens acima permitem fazer uma compa-

4.5 Cenário I - Modelos Gaussianos de classificação

ração mais próxima do modelo proposto pela maneira de encontrar os parâmetros c e σ . Neste sentido o foco aqui é perceber se a abordagem proposta que permite uma seleção automática evitando a arbitragem pode ser uma alternativa quando comparada com outras metodologias onde a arbitragem é necessária. Posto isto as métricas escolhidas serão comparadas para provar a hipótese levantada demonstrando a eficácia do método.

4.5.1 Experimento I

Neste experimento foram testadas as bases menos desbalanceadas ou em que o desbalanceamento não afeta diretamente o problema da sobreposição demonstrado na Figura 3.11. Foram feitas seis análises separadas para facilitar a interpretação dos resultados como demonstrado a seguir:

1^a Análise

Na tabela 4.3 pode ser observado que para a base *Appendicitis*, que apresenta um pouco de desbalanceamento, o método proposto apresentou uma média de AUC de teste acima dos demais métodos Gaussianos. Para os demais métodos Gaussianos, o modelo *F-Affinity propagation*, foi o que demonstrou melhor AUC de teste. Para a base *Australian*, os resultados seguiram na mesma direção que o anterior. O método proposto sem retirada de ruído teve um resultado superior ao que retira ruído bem como dos demais métodos Gaussianos dos quais o modelo *F-Affinity propagation* apresentou melhor resultado de AUC de teste.

Tabela 4.3: *Benchmarking* com classificadores Gaussianos

Modelo	<i>Append</i> - 106/7 N_{v+}/N_{v-} 21/85				<i>Austral</i> - 690/14 N_{v+}/N_{v-} 307/383			
	AUC				AUC			
	Treina	σ_N	Teste	σ_N	Treina	σ_N	Teste	σ_N
<i>Dom</i>	0,8043	0,0262	0,7672	0,1770	0,8765	0,0060	0,8722	0,0460
<i>Dom-SR</i>	0,9889	0,0351	0,7290	0,2141	0,9711	0,0054	0,8621	0,0462
<i>K-Silhouette</i>	0,7302	0,0164	0,7169	0,1759	0,8082	0,0815	0,8047	0,1026
<i>K-Affinity</i>	0,8307	0,0311	0,7356	0,1903	0,8981	0,0042	0,8544	0,0377
K-Aleatório	0,8628	0,0263	0,6986	0,2109	0,9430	0,0309	0,8143	0,0456
<i>F-Silhouette</i>	0,7269	0,0206	0,7169	0,1759	0,5000	0,0000	0,5000	0,0000
<i>F-Affinity</i>	0,8215	0,0164	0,7613	0,1519	0,8799	0,0064	0,8714	0,0398
F-Aleatório	0,8615	0,0408	0,7090	0,2098	0,8116	0,2160	0,7349	0,1663

Os resultados para ambas as bases demonstram que o método foi muito eficaz ao classificar a base de teste demonstrando sua robustez para problemas com

4.5 Cenário I - Modelos Gaussianos de classificação

dimensões diferentes. No caso da base *Appendicitis*, o método com filtragem de ruídos, proposto, se mostrou pouco eficaz provavelmente pelo desbalanceamento da base. Para esta situação, a utilização da técnica de *undersampling* descrita em 3.5.3, não melhorou o resultado provavelmente por ter menos sobreposição. Já no caso da base *Australian* o método que filtra os ruídos teve uma performance próxima do proposto e superior aos demais Gaussianos, perdendo apenas para o F-Affinity. Isto se deve, provavelmente, pela característica do problema ter pouca sobreposição e neste sentido a filtragem descrita em 3.5 e em 3.5.2, melhorou sua eficácia perante os métodos Gaussianos.

2ª Análise

Na tabela 4.4 pode ser observado que para a base *Banknote* ambos os métodos propostos tiveram resultados similares aos Gaussianos estudados com a metodologia *Affinity-propagation* e melhores em média que as metodologias *Silhouette* e Aleatório. O resultado de 100% de assertividade da AUC demonstrou a eficácia da metodologia proposta na classificação da base de teste. Para a base *Blood*, os resultados se assemelharam aos da base *Appendicitis* estudado anteriormente, ou seja, o método proposto que sem a filtragem de ruído teve uma resposta melhor quando comparado com os demais métodos Gaussianos, sendo que deste o K-Affinity-propagation foi o que teve a melhor resposta de AUC.

Tabela 4.4: *Benchmarking* com classificadores Gaussianos

Modelo	<i>Bankl</i> - 1372/4 N_{v+}/N_{v-} 610/762				<i>Blood</i> -748/4 N_{v+}/N_{v-} 178/570			
	AUC				AUC			
	Treina	σ_N	Teste	σ_N	Treina	σ_N	Teste	σ_N
<i>Dom</i>	1,0000	0,0000	1,0000	0,0000	0,6776	0,0091	0,7217	0,1171
<i>Dom-SR</i>	1,0000	0,0000	1,0000	0,0000	0,9697	0,0165	0,6763	0,1327
K- <i>Silhouette</i>	0,9348	0,1297	0,9394	0,1212	0,6038	0,0146	0,6479	0,1372
K- <i>Affinity</i>	1,0000	0,0000	1,0000	0,0000	0,6555	0,0103	0,6683	0,1382
K-Aleatório	0,9986	0,0043	0,9993	0,0022	0,6553	0,0203	0,6600	0,1383
F- <i>Silhouette</i>	0,9200	0,1191	0,9211	0,0996	0,6041	0,0186	0,6288	0,0961
F- <i>Affinity</i>	1,0000	0,0000	1,0000	0,0000	0,6542	0,0130	0,6558	0,1060
F-Aleatório	0,9350	0,1123	0,9432	0,0996	0,6595	0,0080	0,6500	0,1539

Os resultados da análise demonstram que o método proposto alcançou seu objetivo de classificação apresentando resultados melhores que as demais abordagens conhecidas. No caso da base *Banknote* que é mais balanceada, a filtragem de

4.5 Cenário I - Modelos Gaussianos de classificação

ruídos apresentou resultados tão bons, em média, quanto o método sem filtragem e que a metodologia *Affinity propagation*. Já na base *Blood*, a filtragem, embora tenha obtido melhores resultados que os métodos Gaussianos, foi menos eficaz que o sem filtragem. Este resultado valida o que foi descrito anteriormente em que bases com um maior desbalanceamento podem ter em sua sobreposição, retiradas amostras da classe minoritária comprometendo a eficácia do resultado. Neste caso a técnicas que tratam desbalanceamento podem ser usadas para tentar melhorar os resultados. conforme será mostrado no experimento 4.5.2.

3ª Análise

Na Tabela 4.5 estão os resultados para as bases *Breast-Cancer* e *Breas-tHess*, É possível observar que para a primeira base, maior e menos desbalanceada o método com filtragem foi mais eficaz que sem, contudo, a melhor eficácia ocorreu com o método *F-Affinity propagation*. As demais abordagens que utilizam *k-médias* e *c-médias*, baseada em Fuzzy, tiveram uma menor AUC de teste. Já para a base *Breast-Hess*, com característica de ter mais desbalanceamento, o método sem filtragem de ruídos apresentou melhor acurácia que o com filtragem, contudo, em média obteve respostas de AUC de teste menores que outros métodos Gaussianos como: *Affinity*, e *Silhouette*. De fato, para esta base, o método *F-Silhouette* foi o que apresentou melhor resultado.

Tabela 4.5: *Benchmarking* com classificadores Gaussianos

Modelo	<i>Br-Cancer</i> - 683/9 N_{v+}/N_{v-} 444/239				<i>Br-Hess</i> - 133/30 N_{v+}/N_{v-} 99/34			
	AUC				AUC			
	Treina	σ_N	Teste	σ_N	Treina	σ_N	Teste	σ_N
<i>Dom</i>	0,9712	0,0029	0,9687	0,0206	0,7934	0,0242	0,7197	0,1747
<i>Dom-SR</i>	0,9933	0,0025	0,9703	0,0187	0,9284	0,0407	0,6793	0,0989
<i>K-Silhouette</i>	0,9669	0,0030	0,9638	0,0224	0,7215	0,0365	0,7201	0,1814
<i>K-Affinity</i>	0,9747	0,0039	0,9656	0,0162	0,7658	0,0393	0,7281	0,1689
K-Aleatório	0,9891	0,0143	0,9336	0,0312	0,8862	0,0777	0,6576	0,1741
<i>F-Silhouette</i>	0,9669	0,0030	0,9654	0,0224	0,7855	0,0408	0,7456	0,2172
<i>F-Affinity</i>	0,9729	0,0034	0,9739	0,0204	0,7910	0,0312	0,7472	0,1783
F-Aleatório	0,9880	0,0052	0,9322	0,0300	0,8783	0,0828	0,6790	0,1806

Os resultados demonstraram que para ambas as bases o método proposto apresentou resultados muito próximos dos demais modelos Gaussianos, mas com resultados médios inferiores. Para a base *BreastCancer* o método proposto que

4.5 Cenário I - Modelos Gaussianos de classificação

filtra ruídos obteve um resultado pouco melhor em relação ao que não filtra, provavelmente se dá pela natureza do problema ter pouca sobreposição. O que também permitiu a melhor performance do método *Affinity propagation*. Já para a base *Breast-Hess* pouco mais desbalanceada o método que filtra ruídos teve sua performance comprometida pelos mesmos fatores descritos anteriormente. Ainda para esta base, o desempenho melhor do método *Silhouette*, demonstra que os kernels estão mais claramente separados.

4^a Análise

Na Tabela 4.6 pode ser observado que para ambas as bases, o método proposto sem filtragem obteve AUC de teste melhores que os demais métodos Gaussianos perdendo apenas para o F-*Affinity propagation*. Os resultados ainda mostram que para a base *Bupa* a metodologia *Silhouette* apresentou os piores resultados, quando comparado com as demais.

Tabela 4.6: *Benchmarking* com classificadores Gaussianos

Modelo	<i>Bupa</i> - 345/6 N_{v+}/N_{v-} 145/200				<i>Diabetes</i> - 768/8 N_{v+}/N_{v-} 500/268			
	AUC				AUC			
	Treina	σ_N	Teste	σ_N	Treina	σ_N	Teste	σ_N
<i>Dom</i>	0,7541	0,0137	0,6725	0,0722	0,7585	0,0096	0,7187	0,0566
<i>Dom-SR</i>	0,8959	0,0282	0,6430	0,0711	0,9496	0,0078	0,7100	0,0549
<i>K-Silhouette</i>	5,0000	0,0000	5,0000	0,0000	0,6522	0,0289	0,6469	0,0677
<i>K-Affinity</i>	0,7487	0,0166	0,6849	0,0870	0,7755	0,0062	0,7069	0,0586
<i>K-Aleatório</i>	0,8466	0,0210	0,6615	0,0673	0,8255	0,0417	0,6898	0,0585
<i>F-Silhouette</i>	5,0000	0,0000	5,0000	0,0000	0,5313	0,0350	0,5179	0,0302
<i>F-Affinity</i>	0,7527	0,0135	0,7030	0,0588	0,7459	0,0082	0,7132	0,0564
<i>F-Aleatório</i>	0,7773	0,0970	0,6581	0,1069	0,8268	0,0389	0,6948	0,0538

Aqui, mais uma vez podemos considerar que para ambas as bases, mesmo com características diferentes de balanceamento, a natureza do problema, por ter pouca sobreposição, a filtragem acaba por diminuir a performance, ou em outras palavras como já descrito anteriormente, a retirada de amostras sobrepostas, influencia no resultado o que significa que, para ambos os casos, este procedimento não deve ser executado.

4.5 Cenário I - Modelos Gaussianos de classificação

5^a Análise

Na Tabela 4.7 pode ser observado o resultado para outras duas bases. A *German*, embora com características de desbalanceamento, o método proposto com filtragem teve uma melhor performance que o proposto sem filtragem. Para esta base o método proposto teve melhores resultados de AUC de teste que os demais testados. O método *F-Silhouette* foi o que apresentou pior resultado e o *F-Affinity propagation* o que teve melhor entre os Gaussianos estudados. Para a base *Heart* o método proposto, com filtragem, também apresentou um resultado melhor que o sem filtragem e sobre os demais Gaussianos estudados. Para esta base o método *F-Silhouette* apresentou o pior resultado, embora o melhor entre os Gaussianos tenha sido o *K-Silhouette*.

Tabela 4.7: *Benchmarking* com classificadores Gaussianos

Modelo	<i>German</i> - 1000/24 N_{v+}/N_{v-} 700/300				<i>Heart</i> - 270/13 N_{v+}/N_{v-} 150/120			
	AUC				AUC			
	Treina	σ_N	Teste	σ_N	Treina	σ_N	Teste	σ_N
<i>Dom</i>	0,6526	0,0236	0,6241	0,0485	0,8390	0,0148	0,8145	0,0880
<i>Dom-SR</i>	0,7093	0,0267	0,6803	0,0442	0,9572	0,0219	0,8187	0,0814
<i>K-Silhouette</i>	0,5586	0,0456	0,5563	0,0570	0,8238	0,0172	0,8180	0,0683
<i>K-Affinity</i>	0,7545	0,0114	0,6710	0,0518	0,8587	0,0083	0,8014	0,0762
K-Aleatório	0,8919	0,0840	0,6304	0,0604	0,9591	0,0470	0,7405	0,1097
<i>F-Silhouette</i>	0,5000	0,0000	0,5000	0,0000	0,5000	0,0000	0,5000	0,0000
<i>F-Affinity</i>	0,7071	0,0108	0,6715	0,0505	0,8021	0,0591	0,7994	0,0762
F-Aleatório	0,7746	0,2066	0,6179	0,0869	0,9579	0,0553	0,7273	0,0836

Os resultados dos testes demonstraram que o método apresentou respostas médias muito próximas dos métodos já conhecidos, com AUC de teste maior que os demais. Para o caso da base *German*, com certo desbalanceamento, a filtragem melhorou a performance do método o que identifica que a retirada de amostras sobrepostas, permite uma melhor resposta no teste da base. Isto ocorreu também para a base *Heart* mais balanceada, onde a filtragem permitiu uma melhora, mesmo que pequena quando comparado com o que não filtra. Nesta base também foi possível perceber que a metodologia *Silhouette* mostrou ambiguidade de resultados, apresentando o melhor resultado entre os Gaussianos estudos quando se utilizou a técnica *k-médias* e o pior quando se utilizou a técnica *c-médias*.

4.5 Cenário I - Modelos Gaussianos de classificação

6^a Análise

Na Tabela 4.8 as duas últimas bases, com menor desbalanceamento, estudadas foram a *Indian Liver Patients* e a *Ionosphere*. Para a *Indian*, o método com filtragem e sem filtragem apresentou valores muito próximos ou praticamente iguais, embora tenham tido uma performance próxima, mais pouco pior que o *F-Affinity propagation* e *F-Aleatório*. Ainda para esta base o pior resultado foi para a metodologia *K-Silhouette*. Para a base *Ionosphere* a filtragem não representou melhora de performance apresentando resultados muito próximos da modelo sem filtragem, mesmo assim, apresentou melhor AUC de teste que os demais modelos Gaussianos. O melhor resultado gaussiano foi o do modelo *K-Affinity propagation*. Para esta base o melhor resultado foi obtido pelo método proposto sem filtragem de ruído.

Tabela 4.8: *Benchmarking* com classificadores Gaussianos

Modelo	<i>ILP</i> - 579/10 N_{v+}/N_{v-} 414/165				<i>Ionos</i> - 351/33 N_{v+}/N_{v-} 225/126			
	AUC				AUC			
	Treina	σ_N	Teste	σ_N	Treina	σ_N	Teste	σ_N
<i>Dom</i>	0,6442	0,0148	0,5731	0,0641	0,9596	0,0060	0,9430	0,0518
<i>Dom-SR</i>	0,6497	0,0739	0,5732	0,0341	0,9884	0,0058	0,9369	0,0391
<i>K-Silhouette</i>	0,5060	0,0112	0,4964	0,0058	0,8079	0,0750	0,7564	0,1130
<i>K-Affinity</i>	0,6329	0,0160	0,5707	0,0608	0,9411	0,0110	0,8834	0,0616
<i>K-Aleatório</i>	0,7073	0,0325	0,5702	0,0538	0,9666	0,0456	0,7877	0,1005
<i>F-Silhouette</i>	0,5071	0,0163	0,5209	0,0413	0,8721	0,0194	0,8685	0,0681
<i>F-Affinity</i>	0,6246	0,0181	0,5968	0,0691	0,9250	0,0082	0,8829	0,0560
<i>F-Aleatório</i>	0,7075	0,0402	0,5893	0,0681	0,9544	0,0255	0,8664	0,0617

Os resultados demonstraram que para a base *IndianLiverPatients* o método proposto teve resultados muito próximos aos métodos estudados o que possibilita dizer que foi eficiente na resolução do problema. Para a base *Ionosphere* o método proposto, sem filtragem, apresentou resultados médios superiores aos demais métodos, comprovando também sua robustez. Para esta base a filtragem não alterou significativamente o resultado, o que significa que a retirada de ruídos foi pouco relevante.

4.5.2 Experimento II

Neste experimento foram testadas as bases muito desbalanceadas ou em que o desbalanceamento entre as classes é superior à 65% conforme descrito na seção

4.5 Cenário I - Modelos Gaussianos de classificação

3.5.3. Foi observado o efeito da filtragem de ruído da aplicação da técnica de *undersampling* descrita em (Haibo & Garcia, 2009). Foram feitas três análises demonstradas a seguir:

1^a Análise

Na Tabela 4.9 pode ser observado que para ambas as bases o método de filtragem apresentou uma melhora na AUC de teste em relação ao método sem filtragem. Para a base *Climate* o método com filtragem obteve em média, resultados melhores que os demais modelos Gaussianos, excetuando-se os modelos Aleatórios. O melhor resultado foi alcançado pelo K-Aleatório. Para a base *Fertility* o método com filtragem foi mais eficiente apresentando uma AUC de teste maior que os demais modelos Gaussianos. O melhor modelo gaussiano foi o K-Aleatório.

Tabela 4.9: *Benchmarking* com classificadores Gaussianos

Modelo	<i>Climate</i> - 540/18 N_{v+}/N_{v-} 494/46				<i>Fertility</i> - 100/9 N_{v+}/N_{v-} 12/88			
	AUC				AUC			
	Treina	σ_N	Teste	σ_N	Treina	σ_N	Teste	σ_N
<i>Dom</i>	0,5012	0,0037	0,5000	0,0000	0,5045	0,0144	0,5000	0,0000
<i>Dom-SR</i>	0,6676	0,0754	0,6618	0,1231	0,6550	0,0919	0,6681	0,2727
<i>K-Silhouette</i>	0,5000	0,0000	0,5000	0,0000	0,5000	0,0000	0,5000	0,0000
<i>K-Affinity</i>	0,7142	0,0492	0,6022	0,0999	0,5719	0,0744	0,4882	0,0249
K-Aleatório	0,8856	0,1949	0,7611	0,1948	0,7386	0,1774	0,6014	0,2233
<i>F-Silhouette</i>	0,5000	0,0000	0,5000	0,0000	0,5000	0,0000	0,5000	0,0000
<i>F-Affinity</i>	0,5000	0,0000	0,5000	0,0000	0,5141	0,0227	0,5000	0,0000
F-Aleatório	0,9341	0,1552	0,7107	0,1980	0,8924	0,1432	0,5396	0,2381

Os resultados demonstraram que para bases com um desbalanceamento acima de 65% a técnica de *undersampling* pode representar uma melhoria de acurácia dos resultados conforme descrito em (Haibo & Garcia, 2009). Para as bases estudadas, isto de fato ocorreu, quando comparou-se os resultados entre o método sem filtragem com o método que aplica a filtragem, contudo, a acurácia de teste para este modelo foi pior, conforme pode ser observado na Tabela 4.12. Este resultado se deve, provavelmente, pelo fato do processo de aprendizagem ser feito com uma quantidade amostrada muito pequena quando se comparada ao ambiente de teste influenciando no resultado da acurácia. Por outro lado, a técnica de *undersampling*, embora objetive a melhoria dos resultados, nem sempre é a adequada para determinado tipo de base e neste sentido, outras técnicas, como

4.5 Cenário I - Modelos Gaussianos de classificação

a apresentada em (Castro & Braga, 2013) possam ser utilizadas para a melhoria de performance; Outra observação importante foi a performance do método Aleatório ter apresentado resultados em média, maiores dos demais Gaussianos. Isto ocorre provavelmente pela natureza amostral, uma vez que neste método as escolhas das RBFs pelas técnicas *k-médias* e *c-médias* se dá aleatoriamente dependendo menos das correlações entre as amostras de suas classes.

2^a Análise

Na Tabela 4.10 mostra os resultados para as bases: *Glass* e *Haberman*. Observa-se que embora bastante desbalanceadas, o método proposto obteve respostas diferentes. Para a base *Glass*, a filtragem não melhorou sua eficiência enquanto que para a base *Haberman* a filtragem melhorou o resultado. Ainda para a primeira base, o resultado do método proposto foi mais eficaz que os demais métodos Gaussianos, sendo que destes o *K-Affinity-propagation* obteve o melhor resultado. Para a segunda base, o método K-Aleatório foi o mais eficiente.

Tabela 4.10: *Benchmarking* com classificadores Gaussianos

Modelo	<i>Glass</i> - 214/9 N_{v+}/N_{v-} 29/185				<i>Haberman</i> - 306/3 N_{v+}/N_{v-} 225/81			
	AUC				AUC			
	Treina	σ_N	Teste	σ_N	Treina	σ_N	Teste	σ_N
<i>Dom</i>	0,9557	0,0172	0,9066	0,1027	0,6895	0,0162	0,5605	0,0695
<i>Dom-SR</i>	0,9828	0,0120	0,8841	0,1079	0,9315	0,0301	0,5679	0,0994
<i>K-Silhouette</i>	0,7431	0,1901	0,7254	0,2168	0,5795	0,0264	0,5610	0,0852
<i>K-Affinity</i>	0,9788	0,0111	0,9069	0,0995	0,6625	0,0234	0,5726	0,0621
K-Aleatório	0,9900	0,0200	0,9035	0,1050	0,6561	0,0576	0,5959	0,0971
<i>F-Silhouette</i>	0,7091	0,1829	0,7142	0,2162	0,5731	0,0285	0,5472	0,0632
<i>F-Affinity</i>	0,9666	0,0215	0,9038	0,1002	0,6619	0,0188	0,5893	0,0794
F-Aleatório	0,9945	0,0124	0,9010	0,1237	0,6739	0,0198	0,5950	0,1091

Os resultados demonstram que a metodologia ao usar a técnica de *undersampling* embora melhore os resultados de acurácia de teste demonstrados na Tabela 4.12 e AUC, quando comparado aos demais métodos Gaussianos, o que ocorreu para a base *Haberman*, isto não ocorreu para a base *Glass*, provavelmente pela característica desta base ter menos sobreposição. Em outras palavras, a depender do tipo de problema, outro tipo de técnica que atua sobre o desbalanceamento deva ser usada, como por exemplo as descritas em (Castro & Braga, 2013).

4.5 Cenário I - Modelos Gaussianos de classificação

3ª Análise

Na Tabela 4.11 pode ser observado os resultados para as duas últimas bases estudadas. Os testes seguiram a tendência da análise anterior. Para a base *Parkinsa* técnica de *undersampling* não melhorou os resultados e o método proposto sem filtragem obteve o melhor resultado dentre os demais. Dentre os Gaussianos estudados o *K-Affinity propagation* foi o mais eficiente. Para a base *WPBC*, a técnica obteve a melhoria esperada, embora tenha o método *K-Affinity* tenha sido o mais eficiente.

Tabela 4.11: *Benchmarking* com classificadores Gaussianos

Modelo	<i>Parkins</i> - 195/22 N_{v+}/N_{v-} 147/48				<i>WPBC</i> - 194/33 N_{v+}/N_{v-} 46/148			
	AUC				AUC			
	Treina	σ_N	Teste	σ_N	Treina	σ_N	Teste	σ_N
<i>Dom</i>	0,8278	0,0349	0,7925	0,1514	0,6231	0,0207	0,5484	0,0944
<i>Dom-SR</i>	0,9269	0,0337	0,6888	0,0953	0,6884	0,0459	0,5941	0,1139
<i>K-Silhouette</i>	0,6964	0,0179	0,7042	0,1474	0,5000	0,0000	0,5000	0,0000
<i>K-Affinity</i>	0,8520	0,0228	0,7898	0,1640	0,6986	0,0325	0,6451	0,1595
K-Aleatório	0,9620	0,0698	0,8230	0,1233	0,8859	0,1112	0,6106	0,0931
<i>F-Silhouette</i>	0,6944	0,0154	0,6979	0,1517	0,5000	0,0000	0,5000	0,0000
<i>F-Affinity</i>	0,7885	0,0136	0,7694	0,1530	0,5498	0,0322	0,5167	0,0602
F-Aleatório	0,8882	0,1187	0,7790	0,1050	0,7757	0,1793	0,5636	0,1210

Os resultados demonstram que haverá situações em que outras técnicas de tratamento de desbalanceamento devam ser utilizadas para determinado problema. A base *Parkinsosns*, por ter menos sobreposição, a técnica não obteve resultado esperado e ainda piorou a performance em relação aos demais. Já para a base *WPBC* a técnica melhorou a performance obtendo melhores resultados que as Gaussianas com exceção do método *K-Affinity propagation*.

4.5.3 Análise de Significância dos experimentos

Nesta seção discutiremos sobre os testes estatísticos de significância para a análise comparativa dos métodos. A Tabela 4.12, demonstra o *benchmarking* dos resultados médios para outra métrica escolhida, acurácia do teste, entre as bases testadas. Foram colocados em evidência (negrito) os melhores resultados para o método proposto com e sem filtragem comparando-o com o gaussiano (negrito). Ainda foi colocado em destaque (vermelho) os resultados para as bases: *Climate* e *Fertility* em que as acurácias foram piores em média, em relação às demais.

4.5 Cenário I - Modelos Gaussianos de classificação

Tabela 4.12: *Benchmarking* de Acurácia de teste de classificadores Gaussianos

Base	<i>Dom</i>	<i>Dom-SR</i>	<i>K-Silhou</i>	<i>K-Affin</i>	K-Aleat	<i>F-Silhou</i>	<i>F-Affin</i>	F-Aleat
<i>Append</i>	0,8673	0,8673	0,8513	0,8491	0,8028	0,8501	0,8781	0,8018
<i>Austral</i>	0,8710	0,8623	0,8159	0,8551	0,8159	0,5551	0,8667	0,7391
<i>Bank</i>	1,0000	1,0000	0,9395	1,0000	0,9275	0,9770	1,0000	0,9870
<i>Blood</i>	0,6400	0,6200	0,6200	0,6200	0,6000	0,6000	0,6000	0,6100
<i>Br-Cancer</i>	0,9668	0,9679	0,9692	0,9678	0,9575	0,9659	0,9707	0,9414
<i>Br-Hess</i>	0,8187	0,8115	0,8187	0,8352	0,7214	0,8187	0,8489	0,7747
<i>Bupa</i>	0,6869	0,6827	0,5797	0,6985	0,6691	0,5797	0,7157	0,6696
<i>Climate</i>	0,9148	0,6426	0,9148	0,9296	0,9407	0,9148	0,9148	0,9389
<i>Diabetes</i>	0,7682	0,7643	0,7147	0,7578	0,7369	0,6497	0,7670	0,7331
<i>Fertility</i>	0,8800	0,6400	0,8800	0,8600	0,8800	0,8800	0,8800	0,7600
<i>German</i>	0,7320	0,7000	0,7210	0,7550	0,7050	0,7000	0,7560	0,7170
<i>Glass</i>	0,9628	0,9530	0,9063	0,9628	0,9528	0,8779	0,9580	0,9532
<i>Haberman</i>	0,7056	0,7381	0,7378	0,7396	0,7480	0,7382	0,7383	0,7316
<i>Heart</i>	0,8148	0,8222	0,8185	0,8074	0,7444	0,6035	0,8074	0,7148
<i>ILP</i>	0,6994	0,7099	0,7099	0,7029	0,6614	0,7185	0,7270	0,6685
<i>Ionos</i>	0,9515	0,9431	0,7925	0,9117	0,8320	0,8740	0,9088	0,8918
<i>Parkin</i>	0,8613	0,8511	0,8197	0,8766	0,8674	0,8147	0,8561	0,8618
<i>WPBC</i>	0,7718	0,7895	0,7626	0,7876	0,7163	0,7626	0,7571	0,7105

Discussão dos Resultados

Os resultados demonstram que para a nova métrica, o método proposto obteve resultados muito próximos dos modelos Gaussianos da literatura. Foi possível observar que o método sem filtragem algumas vezes obteve resultados melhores que os demais para algumas bases. Já para as bases onde a sobreposição não é muito relevante a filtragem melhora um pouco a performance do método sem desbalancear a base. Outro ponto observado foi que dentre os métodos Gaussianos a abordagem *Affinity propagation* foi a que obteve as melhores acurácias, o que confirma que o uso de processos recursivos e técnica de otimização para encontrar os centros ótimos, demonstra a sua robustez para análise de problemas de diversas dimensões. Por fim, para as bases *Climate* e *Fertility* a técnica de *undersampling* embora tenha melhorado a AUC de teste conforme demonstrado na Tabela 4.9, a acurácia de teste foi bastante comprometida (destaque em vermelho). Este resultado, muito provavelmente foi afetado pelo resultado da predição do treinamento quando comparado com a parte da base testada, ou em outras palavras, o número de amostras treinadas e testadas ficaram muito próximos aumentando estatisticamente a possibilidade de erro no teste o que permite supor que outras

4.6 Cenário II - Modelos de Classificação clássicos

técnicas como a apresentada em (Castro & Braga, 2013) devam ser utilizadas para verificar se há melhoria de performance. Um ponto bem relevante aqui é perceber que ter mais de uma métrica ajuda a dar mais robustez à metodologia permitindo sua confiabilidade.

Teste de Wilcoxon

O teste de significância foi feito entre o método proposto *Dom*, que não usa filtragem e o método *F-Affinity propagation* que obtiveram em média, os melhores resultados. A hipótese colocada é que o método proposto seja significativo, com um nível igual à 0,05, em relação ao *Affinity propagation* que utiliza processo recursivo de otimização. O resultado do teste demonstrou um valor de $z_w = -0,3976$ com $\bar{w}_w = 53$ e $\sigma_w = 17,61$. O p-valor de Wilcoxon encontrado $p_w = 0,68916$ não é significativo para um $p_w \leq 0,05$. Isto equivale a dizer que o método proposto em comparação com o método *F-Affinity propagation* obteve a mesma significância ou em outras palavras, neste caso, se equivalem estatisticamente. Este resultado demonstra que a abordagem proposta, que não utiliza nenhum processo recursivo pode obter resultados muito próximos a outros métodos robustos da literatura e corrobora para a sua eficácia.

4.6 Cenário II - Modelos de Classificação clássicos

Nesta seção iremos fazer uma análise comparativa do método proposto com métodos clássicos de classificação conhecidos na literatura. Foram testados dois modelos Gaussianos RBF (Berthold & Diamond, 1995) e *Bayes* (Christian et al., 2014), dois que não usam kernel Gaussianos, o MLP (Multi Layer Perceptron) (Haykin, 2005) e o ELM (Extream Learning Machine) (Huang et al., 2011) e dois modelos conhecidos por sua robustez em problemas de classificação, SVM (Vapnik, 1995) e LS-SVM (Suykens et al., 2002) ambos com uso de kernels radiais. Como todos os modelos foram do pacote *caret*, que seleciona a melhor saída ajustada para a entrada selecionada, houve a necessidade de ajuste de parâmetros feitos pelo próprio pacote. O objetivo foi criar um *benchmarking* entre os métodos e demonstrar nossa hipótese na qual o método utilizando ou não a retirada de ruídos a depender do problema estudado, pode obter resultados relevantes. Esta hipótese vem da

4.6 Cenário II - Modelos de Classificação clássicos

análise comparativa apresentada na Tabela 2.1 entre uma rede RBF e uma MLP convencional. Todos os resultados foram apresentados nas Tabelas 4.13 à 4.21.

4.6.1 Experimento I

Para este experimento foram testadas as bases menos desbalanceadas ou em que o desbalanceamento não afeta diretamente o problema da sobreposição descrito na seção 3.5. Foram feitas seis análises separadas para facilitar a interpretação dos resultados como demonstrado a seguir:

1^a Análise

Na Tabela 4.13, pode ser observado que o método proposto, sem retirada de ruído obteve resultados muito próximos dos demais testados. Para ambas as bases o melhor resultado foi obtido pelo método LS-SVM.

Tabela 4.13: *Benchmarking* com classificadores clássicos

Modelo	<i>Append</i> - 106/7 N_{v+}/N_{v+} 21/85				<i>Austral</i> - 690/14 N_{v+}/N_{v-} 307/383			
	AUC				AUC			
	Treina	σ_N	Teste	σ_N	Treina	σ_N	Teste	σ_N
<i>Dom</i>	0,8043	0,0262	0,7672	0,1770	0,8765	0,0060	0,8722	0,0460
<i>Dom-SR</i>	0,9889	0,0351	0,7290	0,2141	0,9711	0,0054	0,8621	0,0462
LSSVM-Radial	0,8087	0,0153	0,7740	0,1635	0,8933	0,0079	0,8694	0,0379
SVM-Radial	0,7589	0,0360	0,7436	0,2123	0,8841	0,0123	0,8586	0,0405
ELM	0,6989	0,0876	0,7085	0,1980	0,7659	0,0599	0,7438	0,0742
RBF-Clássico	0,9355	0,0509	0,7622	0,1694	0,9397	0,0048	0,8541	0,0346
<i>Bayes</i>	0,8110	0,0202	0,7607	0,1672	0,8455	0,0085	0,7977	0,0627
MLP	0,7968	0,0137	0,7622	0,1694	0,8954	0,0133	0,8551	0,0425

Os resultados mostraram que para ambas as bases a retirada de ruídos, influenciou na performance do método piorando seu resultado, entretanto para a base *Australian*, menos desbalanceada, a retirada de ruídos, embora não tenha ajudado na melhora de performance do método, obteve resultados médios maiores que alguns métodos clássicos. Ainda para esta base, o método proposto obteve resultados iguais ao classificador LS-SVM, o que demonstra que a hipótese colocada pelo método demonstra eficiência. O que pode ser observado também para a base *Appendicitis*.

4.6 Cenário II - Modelos de Classificação clássicos

2ª Análise

Na Tabela 4.14 pode ser observado que o método foi bem eficaz. Aqui, no entanto para cada base, o método respondeu diferente. Para a base *Banknote* a retirada de ruídos obteve o mesmo resultado do método sem a filtragem com resultados iguais ao SVM-Radial e melhores que outros métodos clássicos. Já para a base *Blood* a retirada de ruídos diminuiu a performance, embora tenha saído melhor, em média, que os demais métodos. Para esta base, o resultado do método proposto sem a filtragem, obteve as melhores AUC de teste entre todos os testados

Tabela 4.14: *Benchmarking* com classificadores clássicos

Modelo	<i>Bankl</i> - 1372/4 N_{v+}/N_{v-} 610/762				<i>Blood</i> -748/4 N_{v+}/N_{v-} 178/570			
	AUC				AUC			
	Treina	σ_N	Teste	σ_N	Treina	σ_N	Teste	σ_N
<i>Dom</i>	1,0000	0,0000	1,0000	0,0000	0,6776	0,0091	0,7217	0,1171
<i>Dom-SR</i>	1,0000	0,0000	1,0000	0,0000	0,9697	0,0165	0,6763	0,1327
LSSVM-Radial	0,9981	0,0030	0,9973	0,0030	0,6355	0,0096	0,6308	0,1089
SVM-Radial	1,0000	0,0000	1,0000	0,0000	0,6516	0,0126	0,6638	0,1597
ELM	0,9789	0,0014	0,9790	0,0143	0,5421	0,0174	0,5471	0,0520
RBF-Clássico	1,0000	0,0000	0,9993	0,0022	0,7977	0,0100	0,6229	0,1432
<i>Bayes</i>	0,9197	0,0052	0,9187	0,0224	0,5577	0,0025	0,5475	0,1193
MLP	1,0000	0,0000	0,9991	0,0027	0,6580	0,0085	0,6515	0,1557

Aqui, novamente o método demonstrou que a abordagem proposta permite classificar as amostras de teste com muita assertividade. Tanto na filtragem quanto sem, o método proposto ficou igual ou melhor que outras abordagens, mostrando sua robustez.

3ª Análise

Na Tabela 4.15, pode ser observado que o método proposto foi bem eficaz, mas com abordagens diferentes. Para a base *Breast-Cancer* a filtragem melhorou o resultado e aproximou de outros métodos estudados, embora os melhores resultados tenham sido obtidos pelo RBF-Clássico. Para a base *Breast-Hess* o método sem filtragem obteve resultados muito próximos e melhores em média que alguns métodos estudados. Para esta base, o melhor resultado foi obtido pelo método SVM-Radial.

4.6 Cenário II - Modelos de Classificação clássicos

Tabela 4.15: *Benchmarking* com classificadores clássicos

Modelo	<i>Br-Cancer</i> - 683/9 N_{v+}/N_{v-} - 444/239				<i>Br-Hess</i> - 133/30 N_{v+}/N_{v-} - 99/34			
	AUC				AUC			
	Treina	σ_N	Teste	σ_N	Treina	σ_N	Teste	σ_N
<i>Dom</i>	0,9712	0,0029	0,9687	0,0206	0,7934	0,0242	0,7197	0,1747
<i>Dom-SR</i>	0,9933	0,0025	0,9703	0,0187	0,9284	0,0407	0,6793	0,0989
LSSVM-Radial	0,9733	0,0028	0,9707	0,0217	0,8425	0,0400	0,6792	0,1442
SVM-Radial	0,9946	0,0029	0,9691	0,0195	0,8052	0,0382	0,8314	0,1827
ELM	0,9333	0,0311	0,9444	0,0470	0,5196	0,0620	0,5159	0,0503
RBF-Clássico	0,9975	0,0080	0,9743	0,0195	0,9967	0,0069	0,6204	0,1318
<i>Bayes</i>	0,9651	0,0039	0,9564	0,0185	0,9086	0,0183	0,8139	0,1144
MLP	0,9861	0,0020	0,9645	0,0210	0,9669	0,0181	0,7250	0,1654

Os resultados demonstraram que a metodologia proposta foi bem assertiva para as bases estudadas, sendo que para a *Breas-tCancer* a filtragem melhorou a performance ficando próximo do resultados de métodos clássicos. Já para a *Breast-Hess* a filtragem comprometeu a performance, muito provavelmente, pela característica desta base ser um pouco desbalanceada. Nesta situação a metodologia sem filtragem apresentou resultados satisfatórios quando comparados com outros métodos, sendo melhores que alguns e piores que outros.

4^a Análise

Na Tabela 4.16, demonstra os resultados encontrados. Para ambas as bases, o método proposto sem filtragem obteve bons resultados quando comparados com os métodos clássicos. Para a base *Bupa*, os resultados foram melhores que os demais, sendo o LS-SVM o que obteve melhor resultado dentre os classificadores. Para a base *Diabetes* o melhor resultado foi obtido pelo modelo MLP.

Tabela 4.16: *Benchmarking* com classificadores clássicos

Modelo	<i>Bupa</i> - 345/6 N_{v+}/N_{v-} - 145/200				<i>Diabetes</i> - 768/8 N_{v+}/N_{v-} - 500/268			
	AUC				AUC			
	Treina	σ_N	Teste	σ_N	Treina	σ_N	Teste	σ_N
<i>Dom</i>	0,7541	0,0137	0,6725	0,0722	0,7585	0,0096	0,7187	0,0566
<i>Dom-SR</i>	0,8959	0,0282	0,6430	0,0711	0,9496	0,0078	0,7100	0,0549
LSSVM-Radial	0,7850	0,0440	0,6665	0,0747	0,7627	0,0097	0,7206	0,0606
SVM-Radial	0,7804	0,0346	0,6487	0,0506	0,7582	0,0135	0,7140	0,0481
ELM	0,5818	0,0522	0,5390	0,0879	0,6885	0,0444	0,6874	0,0726
RBF-Clássico	0,9916	0,0080	0,6082	0,0664	0,9397	0,0308	0,7145	0,0513
<i>Bayes</i>	0,7665	0,0126	0,6587	0,0826	0,7354	0,0109	0,7141	0,0646
MLP	0,7604	0,0435	0,6616	0,0723	0,7514	0,0063	0,7340	0,0393

4.6 Cenário II - Modelos de Classificação clássicos

Os resultados demonstraram que o método proposto com filtragem não melhorou a performance, muito provavelmente por serem bases com pouca sobreposição. O método proposto foi bem eficaz e apresentou valores muito próximos dos demais métodos conhecidos provando sua robustez.

5^a Análise

Na Tabela 4.17 pode ser observado que o método com filtragem não só melhorou a performance em relação ao outro como para a base *German*, obteve resultados médios superiores aos demais estudados. Para a base *Heart*, o melhor resultado foi obtido pelo método RBF-Clássico.

Tabela 4.17: *Benchmarking* com classificadores clássicos

Modelo	<i>German</i> - 1000/24 N_{v+}/N_{v-} 700/300				<i>Heart</i> - 270/13 N_{v+}/N_{v-} 150/120			
	AUC				AUC			
	Treina	σ_N	Teste	σ_N	Treina	σ_N	Teste	σ_N
<i>Dom</i>	0,6526	0,0236	0,6241	0,0485	0,8390	0,0148	0,8145	0,0880
<i>Dom-SR</i>	0,7093	0,0267	0,6803	0,0442	0,9572	0,0219	0,8187	0,0814
LSSVM-Radial	0,7061	0,0236	0,6552	0,0422	0,8955	0,0188	0,8097	0,0794
SVM-Radial	0,7664	0,0121	0,6500	0,0441	0,8720	0,0170	0,8268	0,0875
ELM	0,5000	0,0000	0,5000	0,0000	0,7269	0,0469	0,7564	0,1067
RBF-Clássico	0,9991	0,0016	0,5750	0,0502	0,9468	0,0097	0,8342	0,0817
<i>Bayes</i>	0,6685	0,0631	0,6267	0,0923	0,8615	0,0106	0,8037	0,0681
MLP	0,7823	0,0324	0,6502	0,0546	0,9196	0,0320	0,8234	0,8049

O método sem filtragem obteve resultados menos eficazes que o que filtra. Ambas as bases têm características de sobreposição que ao serem filtradas permitem uma maior acurácia demonstrada em 4.22 e AUC de teste. Os resultados demonstraram ainda que a metodologia proposta foi capaz de aproximar os resultados dos métodos clássicos, ficando inclusive com resultados médios melhores para a base *German*, demonstrando que o objetivo foi alcançado.

6^a Análise

Na Tabela 4.18 podem ser observados os resultados para as bases *IndianLiver-Patients* e *Ionosphere*. Nesta análise o método proposto alcançou valores muito próximos dos demais e pouco superiores sendo mais eficaz que os demais. Para a base *Indian* o método com filtragem foi melhor e para a base *Ionosphere* o método sem filtragem obteve melhores resultados.

4.6 Cenário II - Modelos de Classificação clássicos

Tabela 4.18: *Benchmarking* com classificadores clássicos

Modelo	<i>ILP</i> - 579/10 N_{v+}/N_{v-} - 414/165				<i>Ionos</i> - 351/33 N_{v+}/N_{v-} - 225/126			
	AUC				AUC			
	Treina	σ_N	Teste	σ_N	Treina	σ_N	Teste	σ_N
<i>Dom</i>	0,6442	0,0148	0,5731	0,0641	0,9596	0,0060	0,9430	0,0518
<i>Dom-SR</i>	0,6497	0,0739	0,5732	0,0341	0,9884	0,0058	0,9369	0,0391
LSSVM-Radial	0,5992	0,0273	0,5581	0,0444	0,9405	0,0301	0,9354	0,0764
SVM-Radial	0,5072	0,0160	0,4953	0,0109	0,9708	0,0086	0,9359	0,0513
ELM	0,4995	0,0017	0,5000	0,0000	0,7384	0,0518	0,7174	0,0926
RBF-Clássico	0,9510	0,1381	0,5569	0,0408	0,9972	0,0074	0,9355	0,0588
<i>Bayes</i>	0,7227	0,0088	0,5723	0,0599	0,9141	0,0062	0,8983	0,0602
MLP	0,6183	0,0738	0,5779	0,0835	0,9874	0,0038	0,8810	0,0529

Os resultados demonstram que também para estas bases o método foi bem eficaz apresentando valores muito próximos e ligeiramente superiores no caso da base *Ionosphere*. Aqui foi observado ainda que a filtragem permitiu um mínimo ganho de performance na resposta da base *Indian* o que demonstra que mesmo com o pequeno desbalanceamento, a retirada de amostras sobrepostas não piorou o resultado. Para a base *Ionosphere* o método proposto demonstrou que é tão eficiente quanto os demais estudados.

4.6.2 Experimento II

Neste experimento foram testadas as bases muito desbalanceadas ou em que o desbalanceamento entre as classes descrito na seção 3.5.3 é superior à 65 %. Foi observado o efeito da filtragem de ruído da aplicação da técnica de *undersampling* descrita em (Haibo & Garcia, 2009). Foram feitas três análises demonstradas a seguir:

1^a Análise

Na Tabela 4.19 pode ser observado que o modelo proposto com filtragem obteve uma melhoria da performance para ambas as bases, apresentando valores médios melhores que os demais para a base *Fertility*. No caso da base *Climate*, embora tenha obtido valores maiores que alguns métodos estudados, o melhor resultado foi alcançado pelo método MLP.

4.6 Cenário II - Modelos de Classificação clássicos

Tabela 4.19: *Benchmarking* com classificadores clássicos

Modelo	<i>Climate</i> - 540/18 N_{v+}/N_{v-} - 494/46				<i>Fertility</i> - 100/9 N_{v+}/N_{v-} - 12/88			
	AUC				AUC			
	Treina	σ_N	Teste	σ_N	Treina	σ_N	Teste	σ_N
<i>Dom</i>	0,5012	0,0037	0,5000	0,0000	0,5045	0,0144	0,5000	0,0000
<i>Dom-SR</i>	0,6676	0,0754	0,6618	0,1231	0,6550	0,0919	0,6681	0,2727
LSSVM-Radial	0,7178	0,0715	0,5516	0,0725	0,8300	0,0973	0,5215	0,1725
SVM-Radial	0,6475	0,0872	0,5100	0,0316	0,5000	0,0000	0,5000	0,0000
ELM	0,5000	0,0000	0,5000	0,0000	0,5000	0,0000	0,5000	0,0000
RBF-Clássico	0,5000	0,0000	0,5000	0,0000	0,8136	0,2028	0,4944	0,0176
<i>Bayes</i>	0,7916	0,0270	0,7136	0,1406	0,5141	0,0227	0,5000	0,0000
MLP	0,9268	0,0244	0,8496	0,0820	0,6111	0,1615	0,4889	0,0234

Os resultados demonstram que a técnica de *undersampling* de fato melhorou a performance do método conforme sugere (Haibo & Garcia, 2009). Entretanto a acurácia foi um pouco comprometida conforme pode ser visto em 4.22. Este problema pode ter ocorrido devido ao tamanho da base de treinamento ser muito próxima da base de teste separada. Este efeito ocorreu para ambas as bases testadas. No caso da base *Fertility* o método obteve resultados acima dos métodos de classificação clássicos o que demonstra que a técnica de *undersampling* e conjunto com a filtragem podem deixar o método bastante robusto.

2^a Análise

Na Tabela 4.20 pode ser observado que para a base *Glass* a técnica não foi tão eficiente quanto para a base *Haberman*.

Tabela 4.20: *Benchmarking* com classificadores clássicos

Modelo	<i>Glass</i> - 214/9 N_{v+}/N_{v-} - 29/185				<i>Haberman</i> - 306/3 N_{v+}/N_{v-} - 225/81			
	AUC				AUC			
	Treina	σ_N	Teste	σ_N	Treina	σ_N	Teste	σ_N
<i>Dom</i>	0,9557	0,0172	0,9066	0,1027	0,6895	0,0162	0,5605	0,0695
<i>Dom-SR</i>	0,9828	0,0120	0,8841	0,1079	0,9315	0,0301	0,5679	0,0994
LSSVM-Radial	0,9826	0,0062	0,8246	0,1064	0,6467	0,0193	0,5786	0,0695
SVM-Radial	0,8771	0,0145	0,8800	0,1026	0,6031	0,0404	0,5288	0,0449
ELM	0,8795	0,0334	0,8091	0,1949	0,5691	0,0444	0,5201	0,0843
RBF-Clássico	0,9866	0,0133	0,9121	0,1039	0,9340	0,0366	0,5526	0,0674
<i>Bayes</i>	0,9802	0,0072	0,9162	0,1032	0,5891	0,0324	0,5528	0,0552
MLP	0,9702	0,0148	0,9114	0,1029	0,6313	0,0169	0,6076	0,0931

Para estas bases o modelo proposto ficou com AUC de teste próximas de outros modelos sendo que para a *Glass* o melhor resultado foi para o modelo RB-Clássico

4.6 Cenário II - Modelos de Classificação clássicos

enquanto para a *Haberman* o melhor resultado foi obtido pelo método MLP. Os resultados demonstraram que a técnica de *undersampling* não alcançou o objetivo para a base *Glass*. Neste caso outra técnica conforme descrita em (Castro & Braga, 2013) poderia melhorar os resultados. Entretanto, mesmo com uma base muito desbalanceada, o método sem filtragem obteve resultados muito próximos de alguns métodos clássicos e até superior a outros o que significa que pode ser aplicado em bases deste tipo sem qualquer objeção. Já para a base *Haberman*, onde a técnica alcançou seu objetivo, a filtragem foi capaz de aproximar os resultados do método proposto dos demais métodos estudados, muito provavelmente pela base possuir menos sobreposição apesar do grande desbalanceamento.

3ª Análise

Na Tabela 4.21 podem ser observados os resultados das últimas bases testadas. É possível observar que um comportamento muito parecido com os do teste anterior. Para a base *Parkinson* modelo proposto sem filtragem foi mais eficiente que o com filtragem, mas obteve valores inferiores aos demais estudados. Para a base *WPBC* o modelo com filtragem apresentou um ganho de performance em relação ao sem filtragem, conseguiu resultados melhores que outros métodos, porém o melhor resultado foi obtido pelo método MLP.

Tabela 4.21: *Benchmarking* com classificadores clássicos

Modelo	<i>Parkins</i> - 195/22 N_{v+}/N_{v-} 147/48				<i>WPBC</i> - 194/33 N_{v+}/N_{v-} 46/148			
	AUC				AUC			
	Treina	σ_N	Teste	σ_N	Treina	σ_N	Teste	σ_N
<i>Dom</i>	0,8278	0,0349	0,7925	0,1514	0,6231	0,0207	0,5484	0,0944
<i>Dom-SR</i>	0,9269	0,0337	0,6888	0,0953	0,6884	0,0459	0,5941	0,1139
LSSVM-Radial	0,9216	0,0555	0,8139	0,1196	0,7833	0,0455	0,6113	0,1352
SVM-Radial	0,8191	0,0258	0,7742	0,1574	0,5908	0,0794	0,5067	0,0344
ELM	0,5661	0,0947	0,5482	0,1164	0,5000	0,0000	0,5000	0,0000
RBF-Clássico	1,0000	0,0000	0,9111	0,0959	0,8762	0,1325	0,5229	0,0701
<i>Bayes</i>	0,8205	0,0187	0,7810	0,1040	0,7371	0,0672	0,5920	0,1010
MLP	0,9517	0,0366	0,8439	0,1273	0,9043	0,0470	0,6121	0,1332

Os resultados demonstraram mais uma vez que a técnica de *undersampling* é eficaz para algumas bases e menos para outras. No caso da base *Parkinsons* o uso da técnica não melhorou a performance sendo talvez inadequada para este problema. Talvez para este caso outra técnica como descrita em (Castro & Braga,

4.6 Cenário II - Modelos de Classificação clássicos

2013) surta um efeito melhor. Já para a base *WPBC* a técnica provou sua eficiência permitindo resultados eficazes próximos de outros estudados. Todavia para ambas as bases, o modelo proposto foi eficiente quando comparado com outros métodos de classificação clássicos estudados.

4.6.3 Análise de Significância dos experimentos

Nesta seção aplicaremos os testes estatísticos de significância para a análise comparativa dos métodos de classificação, utilizando o teste não paramétrico de Wilcoxon (Demsar, 2006), conforme já descrito. Na Tabela 4.22 podem ser observados os resultados para as acurácias de teste. Esta nova métrica foi usada para uma nova abordagem analítica dos resultados. Os melhores resultados do método proposto e dos demais métodos foram destacados (negrito).

Tabela 4.22: *Benchmarking* de Acurácia de teste dos classificadores clássicos

Base	<i>Dom</i>	<i>Dom-SR</i>	LS-SVM	SVM	ELM	RBF	<i>Bayes</i>	MLP
<i>Append</i>	0,8673	0,8673	0,8773	0,8782	0,8482	0,8582	0,8291	0,8582
<i>Austral</i>	0,8710	0,8623	0,8623	0,8536	0,7391	0,8464	0,8101	0,8420
<i>Bank</i>	1,0000	1,0000	0,9971	1,0000	0,9759	0,9993	0,9235	0,9993
<i>Blood</i>	0,6400	0,6200	0,5900	0,6200	0,5000	0,5600	0,5700	0,6200
<i>Br-Cancer</i>	0,9678	0,9679	0,9707	0,9605	0,9561	0,9722	0,9590	0,9634
<i>Br-Hess</i>	0,8187	0,8115	0,7830	0,8341	0,7374	0,7747	0,8049	0,7533
<i>Bupa</i>	0,6869	0,6827	0,6958	0,6812	0,5821	0,6376	0,6495	0,6813
<i>Climate</i>	0,9148	0,6426	0,9259	0,9167	0,9148	0,9148	0,9432	0,9444
<i>Diabetes</i>	0,7682	0,7643	0,7734	0,7635	0,7292	0,7513	0,7539	0,7643
<i>Fertility</i>	0,8800	0,6400	0,8944	0,8800	0,8800	0,8700	0,8800	0,8600
<i>German</i>	0,7320	0,7000	0,7580	0,7570	0,7000	0,7150	0,7140	0,7190
<i>Glass</i>	0,9628	0,9530	0,9701	0,9626	0,9065	0,9719	0,9673	0,9675
<i>Habermam</i>	0,7056	0,7381	0,7384	0,7254	0,7216	0,6862	0,7285	0,7482
<i>Heart</i>	0,8148	0,8222	0,8185	0,8370	0,7259	0,8318	0,8074	0,8074
<i>ILP</i>	0,6994	0,7099	0,7098	0,7082	0,7151	0,7097	0,6338	0,7065
<i>Ionos</i>	0,9515	0,9431	0,9344	0,9429	0,7409	0,9431	0,9148	0,9061
<i>Parkin</i>	0,8613	0,8511	0,8863	0,8971	0,7274	0,9379	0,6040	0,8929
<i>WPBC</i>	0,7718	0,7895	0,7984	0,7626	0,7626	0,7574	0,6534	0,7721

Discussão dos Resultados

Estes resultados demonstraram que o método proposto obteve acurácias muito próximas e para algumas bases até melhores que métodos clássicos da literatura. É possível observar ainda que em média, o métodos LS-SVM e SVM-Radial, obtiveram as melhores acurácias de teste. Estes métodos são conhecidos pela sua

eficácia e robustez na solução de problemas de classificação. A análise mostrou ainda o mesmo comportamento ocorrido nos modelos Gaussianos para as bases *Climate* e *Fertility*, ou seja, a técnica de *undersampling*, apresentou ganhos nas AUC de teste conforme descrito na Tabela 4.19 mas comprometeu a acurácia (em vermelho), muito provavelmente pelo tamanho amostrado do conjunto de treinamento em relação ao conjunto testado, conforme explicado anteriormente. Para o caso destas bases como analisado no *benchmarking* dos testes Gaussianos, uma nova técnica como apresentada em (Castro & Braga, 2013) deva ser utilizada para validar uma possível melhoria de performance.

Teste de Wilcoxon

O teste de significância foi feito entre o método proposto *Dom* e o método *LS-SVM* que obtiveram em média, os melhores resultados. A hipótese colocada é que o método proposto seja significativo com um nível igual à 0,05, em relação ao classificador de melhor resultado. O resultado do teste demonstrou um valor de $z_w = -1,3283$ com $\bar{w}_w = 85,5$ e $\sigma_w = 22,96$. O p-valor de Wilcoxon $p_w = 0,18352$ não é significativo para um $p_w \leq 0,05$. Isto equivale dizer que o método proposto em comparação com o método LS-SVM obteve a mesma significância ou em outras palavras, neste caso, se equivalem estatisticamente. Como consequência, o fato do método proposto não necessitar de arbitragem de parâmetros e de não utilizar técnicas recursivas o elege como um bom classificador perante o LS-SVM do pacote *Caret* e o demais testados.

4.7 Conclusão do capítulo

Neste capítulo demonstramos nossa hipótese nos levou a resultados satisfatórios quando comparados com os de modelos largamente usados na literatura. Nos classificadores Gaussianos a metodologia proposta obteve Acurácias muito próximas do método *Affinity propagation* que possui técnica recursiva para encontrar os melhores centros através de processos de otimização demonstrado na Tabela 4.12, já para os Classificadores clássicos a abordagem se aproximou da metodologia LS-SVM que também que no pacote *Caret* utiliza processos de validação cruzada e de minimização de erros para encontrar os melhores resultados, demonstrados na

4.7 Conclusão do capítulo

Tabela 4.22. O teste não paramétrico de Wilcoxon demonstrou que a abordagem não possui diferenças estatísticas relevantes quando comparada com os modelos *F-Affinity propagation* e LS-SVM o que a valida com um bom classificador sendo um diferencial o fato de não se necessitar de ajustes dos parâmetros c e σ das Gaussianas.

Capítulo 5

Aplicação

Neste capítulo apresentamos uma aplicação para o método em um problema de classificação de despacho de usinas. O problema foi retirado do artigo: Despacho econômico em sistema de seis barras com modulação de usinas Térmicas e Eólicas, (Queiroz & Junior, 2016). Ao final pretendemos demonstrar que a abordagem proposta pode ser utilizada em problemas práticos, complexos e com grande demanda para os dias atuais.

5.1 Explicando o problema

Devido às crises energéticas que o país atravessou nos últimos anos, o modelo atual adotado pelo Brasil baseado em usinas Hidrelétricas e térmicas tem sido colocado em questão. Neste contexto surge com mais força a ideia de se criar uma matriz energética com a contribuição de fontes renováveis como: biomassa, solar e eólica, para atender à crescente demanda de energia. Esta diversificação de matriz tem sido incentivada com algum sucesso em países europeus e nos EUA com o surgimento de geração distribuída que utiliza pequenas gerações próximas às cargas. Por outro lado, devido à nossa capacidade hídrica, a utilização de pequenas geradoras hidrelétricas ainda é uma realidade no Brasil. Estas unidades, conhecidas como usinas fio d'água, por não possuírem reservatório ou este ser muito pequeno, possuem um custo operacional mais baixo, embora a sua oferta de geração nem sempre seja mais barata que a de outras como as citadas acima: (eólicas e térmicas mais modernas e solares). Tal constatação, pode ser obser-

5.1 Explicando o problema

vada na Tabela 5.1 da EPE (Empresa de Pesquisa Energética), órgão responsável pela regulamentação de leilões e definição dos preços, onde estão inseridos os custos de ofertas de energias dos leilões ocorridos entre o período de 2004 à 2016 (www.epe.gov.br, 2017).

Tabela 5.1: Preços de Leilões da EPE entre 2004 e 2016

Tipo de Tecnologia	Preço em R\$ por MW/h			Número Unidades
	min	médio	máx	
UHT	58,35	105,44	209,50	28
PCH	108,04	155,66	207,64	59
UTE	52,69	132,97	279,00	181
UEOL	87,50	126,93	212,33	602
UFV	108,04	270,02	305,51	96

Onde o tipo de Tecnologia se refere às usinas: UHT (Usina Hidroelétrica Tradicional), PCH (Pequena Central Hidroelétrica), UTE (Usina Termoelétrica), UEOL (Usina Eólica) e UFV (Usina Foto-Voltaica). Todavia, em se tratando de leilões, há várias questões a serem consideradas para a oferta dos preços, dentre elas: a hidrografia dos anos anteriores, a previsão de hidrografia para anos futuros, o tipo de tecnologia disponível, a regulamentação governamental vigente entre outras possíveis (www.epe.gov.br, 2017), como resultado, os preços ofertados não seguem um padrão. Em outras palavras, os custos de Térmicas modernas, conhecidas por terem um preço mais elevado, podem ser menores que os de usinas Foto-Voltaicas e Eólicas, que tecnologicamente, estão na vanguarda das geradoras no país. Por outro lado, as usinas Eólicas e Foto-Voltaicas, por ocuparem uma área menor que as usinas tradicionais, geralmente têm sido concebidas como GD (Geração distribuída) que para a legislação brasileira tem carga máxima de 30 MW. Neste contexto, os preços destacados na Tabela 5.1, nos levam à algumas observações: primeiro, o número pequeno de unidades Hidrelétricas tradicionais (UHT) é menor devido ao fato das concessões serem de 30 anos o que evita o comprometimento de novos contratos; Segundo, o alto custo das usinas Foto-Voltaicas (UFV), muito provavelmente se devem a um percentual pouco explorado no país; Terceiro, o número elevado de unidades de geração Eólica (UEOL) muito provavelmente se deve à quantidade de unidades classificadas com GD e quarto, o custo dde geração de Termoelétricas, supostamente mais elevado, se apresenta abaixo apenas das Foto-Voltaicas confirmando o fato de serem mais dispendiosas o que fez com que nos últimos dois anos o governo Brasileiro criasse as bandeiras

tarifárias a serem impostas sobre as contas. Destas bandeiras divididas entre: verde, amarela e vermelha, apenas a verde, não possui o repasse da utilização das usinas Termoelétricas.

5.2 Modelagem do Artigo estudado

No contexto do problema colocado na seção anterior, uma boa estratégia para minimizar custos de geração e ao mesmo tempo suprir a demanda crescente por energia seria diversificar a matriz energética. Todavia, tal ação não é trivial. Há de se considerar vários aspectos, dentre eles o ROI (*Return Over Investment*) dos investidores que pode ser mais lento. Assim, a abordagem mais eficiente no estudo de viabilidade é a utilização de modelos que tentem apontar as questões de imprevisibilidade como: Insolação e temperatura para o caso de UFV, ocorrência de ventos viáveis para o caso das Eólicas e a vazão dos rios para o caso das PCH. Estes modelos, permitem os estudos matemáticos de viabilidade para mitigar os riscos, permitindo assim que o foco dos investidores se direcionem para o que pode ser medido, ou seja, o tipo de tecnologia a ser utilizada, a mão de obra disponível, possíveis subsídios do governo entre outras questões de legislação. O artigo estudado (Queiroz & Junior, 2016), trata desta modelagem. Através do estudo de um modelo contendo um sistema com seis barras, duas usinas térmicas e duas GD Eólicas colocadas próximas á carga, tenta analisar como o acréscimo das Eólicas pode contribuir para a diminuição do despacho de geração das usinas Termoelétricas, mais caras durante um período diário de variação de demanda. A principal questão a ser levantada é que devido à incerteza do vento, as usinas Eólicas não são despacháveis, neste sentido o despacho será ditado pelas Termoe-létricas que possuem um custo de geração mais elevado. A Figura 5.1, demonstra o sistema estudado do artigo. O Artigo contém outras duas tabelas importantes para o estudo. A Tabela 5.2 demonstra os custos médios de geração por centavo de dólar para as unidades Térmicas definidas como (GT) e Eólicas definidas como (GW).

5.2 Modelagem do Artigo estudado

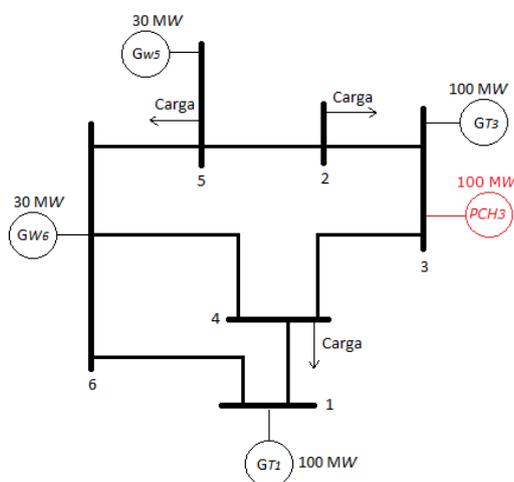


Figura 5.1: Sistema genérico de seis barras

Os resultados do artigo foram obtidos através da utilização de um algoritmo que objetiva o menor custo de despacho em função do fluxo de carga gerado pela dicotomia Geração/Carga. Para as usinas Eólicas foi considerado um cenário de (Abuella & Hatziaioniu, 2015) com o vento variando entre 5m/s e 15m/s sendo a velocidade de corte igual a 25m/s. Foi ainda considerado um valor de penalidade para ambas as usinas Eólicas, descrito na Tabela 5.2 para compensar a ausência de vento.

Tabela 5.2: Preços de geração em US\$

Usina	Custo Variável US\$ MW/h	Custo Fixo US\$ MW/h	Custo Marginal US\$ MW/h	Geração em MW	
				min	máx
GT1	0,012	12	105	5	100
GT3	0,0096	9,6	96	5	100
GW5	0,032	8	0	0	30
GW6	0,016	6	0	0	30

Para as usinas Termoelétricas foram considerados os custos: marginal, que se refere ao preço mínimo de geração, o fixo que á acrescido ao marginal no momento que entra em operação, e o variável depende do combustível, ou seja, que tende a se elevar ao longo do tempo. Os valores utilizados foram retirados de (Allen et al., 2014). Já a Tabela 5.3 que demonstra os tipos de carga distribuídas durante uma medição diária. Estes valores foram obtidos através dos dados coletados em (Queiroz & Junior, 2016) que simula dados de barras com valores reais de linhas

5.3 Aplicação do método proposto

de transmissão com modelagens conforme descritas em (Zimmerman, 1995).

Tabela 5.3: Curva de Carga diária

Intervalo	Fator de carga	Tipo	Horário
1	0,1	Leve	00:00 às 5:00
2	0,2 a 0,6	Média	5:01 às 10:00 e 21:00 às 23:59
3	0,7 a 1,0	Nominal	10:01 às 13:45 e 17:45 às 20:59
4	1,01 a 1,3	Pesada	13:46 às 17:44

Na Figura 5.2 podem ser vistas as curva das cargas de cada barra do sistema estudado e descrito na Figura 5.1 e os horários descritos na Tabela 5.3, considerando um dia útil.

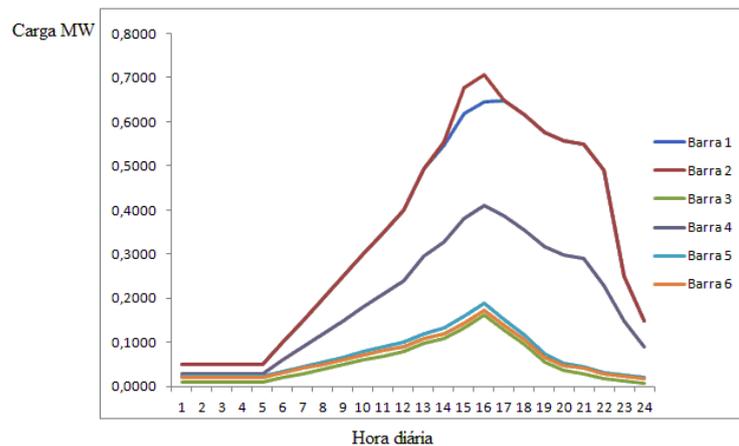


Figura 5.2: Curva de carga do sistema

5.3 Aplicação do método proposto

O artigo mostra que a natureza do problema do despacho é de classificação. Posto isto, a partir dos valores de velocidade de vento, custo de geração, consumo de carga e tipo de usina, propõem-se aplicar o método com o objetivo de determinar o despacho. Para tal, algumas premissas são aceitas: primeiro, devido à incerteza do vento, o despacho será sempre das UTE e/ou da PCH. Segundo, os valores de geração, são fictícios, servindo como parâmetro de estudo para se definir o custo de despacho mais baixo, ou seja, não há uma relação direta dos valores com os estipulados na Tabela 5.1 de valores de leilões da EPE. Terceiro, foi considerado

dois cenários, sendo um com chuva onde haverá o despacho com PCH e o outro com seca onde o despacho será ditado pelas UTE. Quinto, foi considerado três perfis de carga, sendo o primeiro com carga em dia útil, o segundo aos sábados e o terceiro aos domingos e feriados. Sexto, foi colocado uma janela de medição diária com o tempo de um mês sendo analisado o mês de janeiro. Por fim, como o desempenho dos métodos demonstrados pelos testes de Wilcoxon descritos nas Tabelas (4.12 e 4.22) indicaram resultados melhores para a abordagem *F-Affinity propagation* para os métodos gaussianos e LS-SVM para os classificadores, foram aplicados apenas estes métodos em conjunto com a metodologia proposta para o tratamento do problema.

5.4 Bases testadas

Para os experimentos foram utilizadas seis bases montadas a partir dos dados apresentados em (Queiroz & Junior, 2016) cujas variáveis são: dias do mês de janeiro, tipo de usina, velocidade do vento e carga nas barras. Além destes dados foi inserida também a variável chuva para simular o despacho da PCH. A Tabela 5.4 descreve as bases com os dados coletados para os cenários estudados (chuva e estiagem) e para os três testes: dias úteis (Desp. Carga), para finais de semana e feriados (Desp. F semana) e para o custo (Desp. Custo), conforme descritos na seção anterior.

Tabela 5.4: Bases para os cenários de despacho

Nome da base	N_d	N	N_{v+}	N_{v-}
Desp. Chuva Carga	5	744	450	294
Desp. Chuva F semana	5	744	217	527
Desp. Chuva Custo	5	744	19	725
Desp. Seca Carga	5	744	731	13
Desp. Seca F semana	5	744	216	528
Desp. Seca Custo	5	744	647	97

Onde N como já descrito é o número de amostras, N_d o número de dimensão da base e N_{v+} e N_{v-} o número de amostras (vértices do grafo de Gabriel) de cada classe. Para os rótulos foram definidos os três tipos de despachos de cada

cenário considerando a incerteza do vento modelado para as usinas UEOL com ventos entre $5m/s$ e $15m/s$ e a presença e ausência de chuva modelado para a usina PCH. Assim, Os resultados permitirão fazer uma análise comparativa e uma discussão de qual usina pode ser despachada em cada cenário, como pode ser visto na próxima seção.

5.5 Experimentos

Conforme descrito nas seções anteriores, foram estudados dois cenários. Para o cenário chuvoso teremos o despacho tanto das Termoelétricas quanto da Pequena Central Hidroelétrica. Os custos de geração estão descritos na Tabela 5.2 para as UTE e UEOL e foi acrescentado o valor de US\$0,10544 MW/h para a usina PCH no sistema. Como informado, devido à incerteza do vento, as Eólicas não são despacháveis. Já para o cenário de estiagem a PCH por serem fio d'água pode sofrer restrição de despacho, em outras palavras, o despacho será ditado apenas pelas Termoelétricas. Para ambos os cenários pretende-se analisar o comportamento do despacho em relação ao custo, carga e dias da semana.

5.5.1 Cenário I - Período com chuva

Neste cenário foi analisado o mês de janeiro com chuva. Esta situação permite o despacho da usina PCH que possui uma composição de custo fixo, e das Termoelétricas que possuem uma composição de custo variável no tempo por dependerem do custo de combustível. Já as Eólicas, embora tenham para este estudo um custo menor, devem ser compensadas pelo período de ausência de vento o que significa que seu custo entra independente de seu funcionamento.

1º Teste - Despacho por carga em dias úteis

Nas Tabelas 5.5 e 5.6 podem ser observados os resultados obtidos pelos três modelos considerando o despacho das usinas pela demanda, ou carga consumida/gerada. Neste cenário, as GDs Eólicas mesmo com vento abundante, o que não ocorre devido à incerteza, não conseguiriam suprir a carga, ou seja, haverá uma elevação no custo devido ao despacho das Termoelétricas.

Tabela 5.5: Despacho por Carga considerando dia útil

Dados	Modelo	AUC				Acurácia	
		Treina	σ_N	Teste	σ_N	Teste	σ_N
Despacho Carga	<i>Dom</i>	0,9967	0,0006	0,9901	0,0089	0,9933	0,0095
	LS-SVM	0,9955	0,0007	0,9938	0,0110	0,9958	0,0111
	<i>F-Affinity</i>	0,9849	0,0020	0,9803	0,0221	0,9811	0,0193

Tabela 5.6: Despacho por Usina

Tipo Usina	Despacho
UEOL - UTE	28%
UTE	26%
UEOL - PCH	16%
PCH	5%

Discussão dos Resultados

O modelo proposto obteve resultados muito próximos ao LS-SVM e pouco melhor que a metodologia *F-Affinity propagation* sendo capaz de classificar o despacho com eficácia. Todavia o ponto mais relevante é observar o resultado descrito na Tabela 5.6 onde é demonstrado o resultado da classificação do despacho, considerando que em 26% do tempo ocorreu despacho de Termoeletricas e em 5% despacho com PCH.

2º Teste - Despacho por carga em finais de semana e feriados.

Nas Tabelas 5.7 e 5.8 foram anotados os resultados dos testes para o despacho de carga em finais de semana e no feriado do mês de janeiro. Nesta situação, a carga por ser mais leve, deve diminuir o despacho das Termoeletricas uma vez que há a contribuição d PCH.

Tabela 5.7: Despacho por Carga - finais de semana e feriado

Dados	Modelo	AUC				Acurácia	
		Treina	σ_N	Teste	σ_N	Teste	σ_N
Despacho Final de Semana	<i>Dom</i>	0,9274	0,0053	0,8956	0,0302	0,9001	0,0123
	LS-SVM	0,9021	0,0201	0,8972	0,0654	0,9140	0,0280
	<i>F-Affinity</i>	0,9127	0,0056	0,8979	0,0403	0,9207	0,0225

Tabela 5.8: Despacho por Usina

Tipo Usina	Despacho
UEOL - UTE	24%
UTE	23%
UEOL - PCH	19%
PCH	9%

Discussão dos Resultados

Pode ser observado na Tabela 5.7 que para este caso o método proposto obteve resultados muito próximos que o LS-SVM e da metodologia F-*Affinity propagation* que obteve a melhor classificação. Contudo, esta diferença foi muito pequena, podendo dizer que a metodologia proposta foi eficaz conseguindo resultados de classificação próximos de métodos conhecidos na literatura. Na Tabela 5.8 foi possível perceber ainda uma diminuição na quantidade de vezes que o despacho ocorreu com as usinas Termoelétricas foi compensado pelo aumento de 16% para 19% do despacho da PCH em conjunto com as GD Eólicas. Este resultado reflete que para a situação de carga mais leve não é necessário o aumento de geração nas unidades Termoelétricas. Resultado também se repete na ausência de vento com um aumento de 5% para 9%.

3º Teste - Despacho por custo de geração.

Nas Tabelas 5.9 e 5.10, foram anotados os resultados dos testes considerando o custos dos despachos com Termoelétricas e PCH e da gerção das GDs Eólicas. Para o cenário de chuva, a PCH contribuirá com o custo de geração, as GD's também terão a composição do custo, considerando uma parcela de custo variável descrito na Tabela 5.2, devido à incerteza do vento. Já as Termoelétricas terão o custo variável de acordo com as cargas ao longo do mês.

Tabela 5.9: Despacho por Custo

Dados	Modelo	AUC				Acurácia	
		Treina	σ_N	Teste	σ_N	Teste	σ_N
Despacho Custo	<i>Dom</i>	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000
	LS-SVM	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000
	F- <i>Affinity</i>	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000

Tabela 5.10: Despacho por Usina

Tipo Usina	Despacho
UEOL - UTE	32%
UTE	30,67%
UEOL - PCH	25,33%
PCH	13,33%

Discussão dos Resultados

Na Tabela 5.9 pode ser observado um equilíbrio nos resultados. Este equilíbrio muito provavelmente se deu devido à natureza dos dados serem "fortemente" separáveis. De fato, a metodologia proposta foi capaz de classificar os valores com a mesma eficácia dos métodos LS-SVM e F-*Affinity propagation*. Pode ser observado ainda na Tabela 5.10 que o número de Termoelétricas despachadas diminuiu enquanto houve um pequeno acréscimo de PCH no percentual de despacho por custo. Este resultado permite concluir que no cenário de chuva os custos com a PCH tornam-se mais atraentes aumentando o despacho da unidade. por outro lado, há uma queda de despachos de UTE que tendem a ser mais elevados ao longo do dia. Como foi possível perceber um equilíbrio entre os dois casos estudados, onde ocorreu uma pequena modificação de despacho das Termoelétricas, foi feito um novo teste onde foi considerado a demanda nos finais de semana e feriado do mês de janeiro. Nesta situação, é sabido que a demanda por ser menor poderá modificar as características do despacho.

5.5.2 Cenário II - Período com estiagem

Para este experimento, a usina PCH, fio d'água deve ter o despacho reduzido ou anulado. Em períodos de seca ou estiagem este tipo de usina, por não terem reservatório ou estes serem muito pequenos, não devem ser despachadas por segurança e neste contexto, as usinas Termoelétricas assumem tal função. Isto significa um aumento de custo de geração que acabam sendo repassados pelas bandeiras tarifárias, aos consumidores.

1º Teste - Despacho por carga em dias úteis

Nas Tabelas 5.11 e 5.12 podem ser observados os resultados obtidos pelos três

modelos considerando o despacho das usinas pela demanda, ou carga consumida/gerada. Neste cenário, por não termos a PCH o despacho será ditado pelas UTE. As GDs Eólicas como já explicado no cenário de chuva, não conseguem suprir a carga, ou seja, haverá uma elevação de geração das UTE ao longo do dia.

Tabela 5.11: Despacho por Carga considerando dia útil

Dados	Modelo	AUC				Acurácia	
		Treina	σ_N	Teste	σ_N	Teste	σ_N
Despacho Carga	<i>Dom</i>	0,9875	0,0281	0,9983	0,0017	0,9935	0,0165
	LS-SVM	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000
	<i>F-Affinity</i>	1,0000	0,0000	1,0000	0,0000	1,0000	0,0000

Tabela 5.12: Despacho por Usina

Usina	Despacho
UEOL - UTE	62,67%
UTE	36%
UEOL - PCH	1,33%
PCH	0%

Discussão dos Resultados

Na Tabela 5.11 foram colocados os resultados que ficaram muito próximos. Os resultados demonstraram que o método proposto foi bastante eficaz na classificação, embora tenha ficado atrás dos demais, os valores foram muito próximos. Os métodos: gaussiano *F-Affinity propagation* e de classificação clássico LS-SVM obtiveram melhor desempenho. Já na Tabela 5.12, observa-se que para o cenário de estiagem, a usina PCH foi despachada 1,33% no período, o que demonstra que a metodologia foi capaz de interpretar o cenário em que ela não pode entrar em operação. Como consequência o despacho é ditado pelas Termoelétricas, todavia, a parcela principal do despacho é feita em conjunto com as GD (Eólicas). Outro ponto a ser observado é quando se compara com o cenário de chuva, Tabela 5.6 onde pode ser observado que com a estiagem houve um aumento de 10% do despacho das UTE, enquanto que dobrou a contribuição das unidades UEOL em conjunto com as Termoelétricas saindo de 28% para 62,67%. Este resultado já era esperado e na ausência de chuva, mas com vento, as Eólicas ajudem a aliviar a carga das UTE.

2º Teste - Despacho por carga em finais de semana e feriados.

Na Tabelas 5.13 e 5.14 foram anotados os resultados para o despacho considerando a carga em finais de semana e no feirado. Os testes demonstraram que o método proposto obteve valores próximos dos demais, sendo que o LS-SVM foi o melhor classificador.

Tabela 5.13: Despacho por Carga - finais de semana e feriado

Dados	Modelo	AUC				Acurácia	
		Treina	σ_N	Teste	σ_N	Teste	σ_N
Despacho final de semana	<i>Dom</i>	0,9171	0,0037	0,8886	0,0391	0,8991	0,0301
	LS-SVM	0,9226	0,0085	0,8987	0,0394	0,9085	0,0309
	F- <i>Affinity</i>	0,8247	0,0316	0,8217	0,0560	0,8750	0,0306

Tabela 5.14: Despacho por Usina

Usina	Despacho
UEOL - UTE	70,67%
UTE	28%
UEOL - PCH	0%
PCH	0%

Discussão dos Resultados

Os resultados demonstraram que o método proposto foi eficaz acima do Gaussiano F-*Affinity propagation*, mas ficando abaixo do LS-SVM. Já a Tabela 5.14 demonstra ainda que ocorreu a classificação correta dos despachos. A usina PCH não foi despachada, como esperado, por se tratar de um mês com pouca vazão de água. Como consequência o despacho continua sendo feito pelas Termoelétricas, entretanto, ao se comparar com o cenário anterior na Tabela 5.8, a parcela principal do despacho é feita em conjunto com as GD (Eólicas) ocorrendo um aumento de 24% para 70,67%, quase triplicou, compensado com a diminuição de 36% para 28% de UTE para a situação de ser ter vento com certa abundância. Observa-se ainda uma elevação de 5% de contribuição das UTE. Em dias de carga mais leve, o sistema fica mais aliviado e em consequência disto há diminuição da carga, contudo, na ausência do vento as Eólicas não contribuem necessitando sempre do despacho das UTE. Estes resultados corroboraram com o descrito anteriormente que em períodos de seca o aumento de geração das Termoelétricas independente

da carga ser leve ou pesada causa um aumento de custos que é repassado via bandeira tarifária e que seria maior sem a contribuição das GDs Eólicas.

3º Teste - Despacho por custo de geração.

Nas Tabelas 5.15 e 5.16 foram inseridos os resultados do cenário considerando o custo. Para o cenário de seca, o custo tende a se elevar devido ao despacho só das unidades UTE mais caras ao longo do tempo. Pode ser observado que o método proposto obteve resultados muito próximos, embora menores que os demais. A metodologia *F-Affinity propagation* foi a que obteve os os melhores resultados.

Tabela 5.15: Despacho por Custo

Dados	Modelo	AUC				Acurácia	
		Treina	σ_N	Teste	σ_N	Teste	σ_N
Despacho Custo	<i>Dom</i>	0,9964	0,0031	0,9822	0,0306	0,9892	0,0106
	LS-SVM	0,9875	0,0053	0,9853	0,0232	0,9946	0,0070
	<i>F-Affinity</i>	0,9845	0,0027	0,9861	0,0236	0,9960	0,0065

Tabela 5.16: Despacho por Usina

Usina	Despacho
UEOL - UTE	68%
UTE	32%
UEOL - PCH	0%
PCH	0%

Discussão dos Resultados

Estes resultados demonstraram que o método proposto embora tenha ficado abaixo dos demais, conseguiu o objetivo de classificar o despacho considerando o menor preço. Na Tabela 5.16 pode ser observado que não ocorreu despacho da PCH como esperado. Por sua vez, com relação ao despacho das Termoelétricas, ocorreu um rearranjo de usinas despachadas. Os resultados mostram que 68% do tempo, o custode geração foi composto pela soma das Termoelétricas com as Eólicas e que em 32%, na ausência do vento, o despacho foi feito só por Termoelétricas. Quando comparado com o cenário anterior na Tabela 5.10, ocorreu um aumento de apenas 2% no custo das UTE enquanto que a contribuição das UEOL, como já mencionado acima, dobrou. Embora pareça que este aumento seja significativo,

há de se ressaltar que a ausência das GDs, neste caso, aumentaria ainda mais o custo das Termoelétricas que são mais caras ao longo do tempo. Esta elevação do custo, como já informado é de alguma forma repassado aos consumidores através das bandeiras tarifárias.

5.6 Conclusão do capítulo

Neste capítulo demonstramos que a metodologia proposta pode ser aplicada em um problema de classificação de despacho de usinas. Através dos dados coletados no artigo (Queiroz & Junior, 2016), vimos uma nova abordagem de classificação utilizando o grafo de Gabriel com conjunto Dominante. A metodologia proposta neste trabalho foi capaz de resultados assertivos quando comparados com a de classificadores conhecidos nos cenários diferentes propostos. Para o cenário de chuva foi observado uma parcela maior de contribuição da usina PCH e para o cenário de estiagem a maior contribuição das Termoelétricas. De fato, como pode ser observado comparando os cenários das Tabelas (5.6 e 5.12), Tabelas (5.8 e 5.14) e Tabelas (5.10 e 5.16), fica bem evidente que as usinas Eólicas podem contribuir para o alívio do sistema e ao mesmo tempo que contribuem para a diminuição do custo ao longo do tempo embora o vento seja sempre o fator mais desafiador.

Capítulo 6

Conclusões

A hipótese levantada foi que o uso do grafo de Gabriel e do sub-conjunto dominante do grafo poderia ser usado para encontrar, sem a necessidade de ajuste, os parâmetros c e σ das Gaussianas das RBFs usadas na rede neural em problemas de classificação. Como comprovado pelos experimentos, o uso de vértices dominantes do grafo como centros e a média da distância euclidiana entre estes vértices e suas adjacências como raio das Gaussianas demonstraram que a abordagem proposta encontrou resultados de AUC e Acurácia muito próximos de métodos largamente utilizados na literatura em problemas de classificação, validando a hipótese. Foi observado ainda que a abordagem proposta evita o *overfitting* na solução, uma vez que o número de RBFs ou neurônios é menor que o número de amostras do conjunto de treinamento, evitando a superposição de padrões de entrada e saída deste. Outro ponto positivo para a metodologia proposta é determinística e não utiliza recursos de otimização, o que a deixa mais simples e de fácil configuração. Quando comparada aos demais métodos estudados em ambos os cenários (métodos Gaussianos e classificadores clássicos), onde a arbitragem de parâmetros é necessária além processos de validação cruzada, processos de otimização e abordagens estocásticas são utilizados para se obter os melhores resultados, os *benchmarking* demonstraram que o objetivo de atuar como classificador para o método proposto (*Dom*) foi alcançado com valores bem consistentes, próximos dos demais.

A metodologia proposta também definiu uma forma de filtragem de ruídos, ou

retirada de amostras sobrepostas, na versão *Dom-SR*, cujo objetivo foi melhorar eficácia do método para problemas onde a sobreposição interfere na acurácia. Todavia, foi possível perceber que esta abordagem tem resultados melhores ou piores a depender do problema a ser tratado. As análises demonstraram que determinada base com um número substancial de sobreposição pode ser deteriorada caso haja maior desequilíbrio entre as amostras e que nestes casos, a filtragem pode comprometer negativamente o resultado. Outra analogia nesta linha permitiu observar que em algumas bases, mesmo com muito desequilíbrio, esta filtragem melhorou o resultado aplicando-se a técnica de *undersampling*, conforme descrito pela literatura, enquanto que para outras esta não se mostrou eficiente. Esta dicotomia nos leva a concluir que existem casos em que outras técnicas para o tratamento de desbalanceamento devam ser utilizadas. Uma outra questão observada foi o resultado de AUC de teste e acurácia para as bases: *Climate* e *Fertility* onde o desbalanceamento é considerável. Ocorreu para ambos os casos uma melhoria da AUC de teste com uma piora da acurácia de teste. Este antagonismo nos resultados demonstrou a necessidade de se rever o modelo do treinamento para estes dois problemas, pois é plausível que a diminuição da base treinada devido à retirada de sobreposição em conjunto com a técnica de *undersampling* tenha comprometido o conjunto de teste tendendo este a predizer perdas iguais aos erros de classificação, sendo que o conjunto de treinamento é bem maior não tendo proporcionalmente perdas iguais e como consequência aumentando o erro de classificação, abrindo a possibilidade de aplicação de outras técnicas de filtragem para bases desbalanceadas.

Os testes não paramétricos de Wilcoxon demonstraram que a modelagem proposta (*Dom*), foi significativa não apresentando estatisticamente diferenças em relação aos melhores modelos nos dois cenários. Em outras palavras, o modelo foi tão eficaz quanto o melhor Gaussiano *F-Affinity propagation* e quanto o melhor classificador clássico LS-SVM e neste sentido, o fato da abordagem ser simples, não necessitar de calibragem e ser determinística, representa uma substancial vantagem na comparação entre estes modelos estudados.

Quando aplicado a um problema proposto de despacho de usinas em um sistema de seis barras, foi possível observar que o método conseguiu classificar correta-

mente o despacho de usinas Termoelétricas e da PCH em um sistema de seis barras com duas GD Eólicas inseridas no sistema. Conforme esperado, as UEOL, com a presença do vento, contribuem com o alívio de carga permitindo a diminuição do despacho de Termoelétricas mais caras devido o custo variável do combustível. Foi possível perceber ainda que a PCH por ser fio d'água, não deve ser despachada em períodos de estiagem. Neste contexto, a metodologia proposta (*Dom*) se mostrou assertiva no objetivo de classificar os despachos das usinas nos cenários de chuva e estiagem quando comparada com os resultados das abordagens LS-SVM e *Affinity-propagation*, métodos conhecidos da literatura como ótimos classificadores.

Por fim, podemos concluir que esta nova abordagem cumpriu seu objetivo de classificar amostras apresentando um método de simples configuração e que não necessita de ajustes de parâmetros a priori, podendo ser utilizada para solucionar problemas conhecidos da literatura com níveis de complexidade e dimensões diferentes e servir também como instrumento de futuros *benchmarking* com outros modelos de classificadores.

6.1 Proposta de continuidade

A metodologia proposta permitiu observar dois pontos que são focos de uma continuidade com o objetivo de aprimorá-la. O primeiro seria a possibilidade de apresentar apenas um método em que a filtragem só seja aplicada quando necessária. Isto permitiria em um só código atuar em bases com sobreposição e desbalanceamento de maneira a ter um resultado mais apurado. Esta proposta, em resumo, nada mais seria do que aplicar o melhor resultado apresentado nas Tabelas 4.12 e 4.22 do método proposto em um só modelo. O segundo seria o estudo mais aprofundado dos problemas ocorridos nas bases: *Climate* e *Fertility*, ou outras de natureza parecida em que a técnica de *undersampling* melhorou a AUC de teste ao mesmo tempo que diminuiu a acurácia, efeito peculiar observado apenas para estas duas bases. Como observado, a proposição é que seja utilizadas técnicas diferentes como demonstrada em (Castro & Braga, 2013) que poderiam corrigir esta inconsistência dando ainda mais robustez ao método.

Referências

- Abuella, M. & C. J. Hatziadoniu (2015). The economic dispatch for integrated wind power systems using particle swarm optimization. 71
- Albert, A. (1972). *Regression and the Moore-Penrose pseudoinverse.*, Volume 3. Elsevier. 27
- Allen, J. W., B. F. W., & B. S. Gerald (2014). *Power Generation, Operation and Control.* John Wiley & Sons Inc. and Hoboken New Jersey, 2014. 71
- Asuncion, A. & D. J. Newman (2007). UCI machine learning repository. 4, 42
- Aurenhammer, F. & R. Klein (1989). *Voronoi Diagrams.* Institut für Grundlagen der Informationsverarbeitung Technische Universität Graz Klosterwiesgasse 32/2 A-8010 Graz, Austria. Fern Universität Hagen Praktische Informatik VI Elberfelder Straße 95 D-58084 Hagen, Germany. 12
- Bache, K. & M. Lichman (2013). Uci machine learning repository. 43
- Barroso, M. M. A. (2001). Conjuntos dominantes: uma ferramenta para tratamento de dados especiais. *Aula Magna* 7(6), 68–75. 17, 18, 29
- Berg, M. D., O. Cheong, M. V. Kreveld, & M. Overmars (2008). *Computational Geometry: Algorithms and Applications, Third Edition.* Springer-Verlag TELOS, Santa Clara, CA, USA, 3rd. 10, 13
- Berthold, M. R. & J. Diamond (1995). Boosting the performance of rbf networks with dynamic decay adjustment, in advances in neural information processing systems. In *Advances in neural information processing systems*, pp. 521–528. 43, 57

- Bishop, C. M. (1995). *Neural networks for pattern recognition*. 27, 46
- Bohme, T., F. Goring, & Harant (2001). Menger's theorem. *Journal of Graph Theory, May 2001* 37(1), 35–36. 8
- Braga, P. A., F. L. P. A. Carvalho, & B. T. Ludemir (2007). *Redes Neurais Artificiais Teoria e Aplicações*. Editora ABDR. 1, 3, 23, 24, 26, 30
- Carvalho, R. P. B., S. W. Lacerda, & P. A. Braga (2007). Rrs + ls-svm: a new strategy for a priori sample selection. *Neural Comput & Applic -2007- DOI 10.1007/s00521-007-0085-y* 16, 227–234. 36
- Castro, L. C. A. & P. A. Braga (2013). Novel cost-sensitive approach to improve the multilayer perceptron performance on imbalanced data. *IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS*. 24(6). 3, 39, 40, 44, 54, 57, 64, 66, 84
- Chiu, S. L. (1994). Fuzzy model identification based on cluster estimation. *Journal of Intelligent & fuzzy systems* 2(3), 267–278. 21, 26, 43, 46
- Christian, R., R. Nils, L. Karsten, S. Gero, & Z. Marc (2014). Classification and visualization. 43, 57
- Demsar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Faculty of Computer and Information Science Trzaska 25 Ljubljana, Slovenia*. 4, 42, 65
- Duda, R., P. Hartg, & D. G. Stork (2012). *Pattern classification*. john wiley & sons. 26
- Even, S. (1973). *Algorithmic combinatorics*. 17
- Frey, B. J. & D. Dueck (2007). Clustering by passing messages between data points. *science* 315(5814), 972–976. 21, 43, 46
- Gallier, J. (2008). *Notes on Convex Sets, Polytopes, Polyhedra, Combinatorial Topology, Voronoi Diagrams and Delaunay Triangulations*. Department of Computer and Information Science University of Pennsylvania Philadelphia, PA 19104, USA e-mail: jean@cis.upenn.edu. 12

- Haibo, H. & E. A. Garcia (2009). Learning from imbalanced data. *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*. 21(9). 3, 39, 40, 44, 53, 62, 63
- Haykin, S. (2005). *Neural Networks, A Comprehensive Foundation, Second Edition*. (Second Edition ed.). Pearson Education (Singapore) Pte. Ltd., Indian Branch, 482 F.I.E. Patparganj, Delhi 110092, India. 1, 3, 5, 22, 43, 57
- Haykin, S. (2009). *Neural networks and learning machines*. (third Edition ed.). Copyright textcopyright 2009 by Pearson Education, Inc., Upper Saddle River, New Jersey 07458. 1, 22, 23, 24
- Haynes, W. T. and Hedetniemi, S. & S. T. Slater (1998). *Fundamentals of Domination in Graphs*. Cpyring ©1998 by Marcel Dekker, INC. All Rights Reserved, 270 Madison Avenue, New York, Nesw York,1006. 14, 16
- Huang, G. B., D. H. Wang, & Y. Lan (2011). Extreme learning machines: a survey. *International Journal of Machine Learning and Cybernetics* 2(2), 107–122. 43, 57
- Huang, G. B., Q. Y. Zhu, & C. K. Siew (2006). Extreme learning machine: theory and applications. *Neurocomputing* 70(1), 489–501. 43
- Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika*, [S.l.]. 32(3), 241–254. 25
- Klein, R. (1989). *Lecture Notes in Computer Science, Concrete and Abstract Voronoi Diagrams. A Comprehensive Foundation, Second Edition*. Printing and binding: Druckhaus Beltz, Hemsbach/Bergstr. 214513140-543210 - Printed on acid-free paper, Springer-Verlag Berlin Heidelberg 1989. 9, 11
- Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE* 78(9), 1464–1480. 26
- Ore, O. (1962). *Theory of Graphs*. Camer. Mathematic Society, Colloq. Public, 38. Povidence, RI. 15
- Queiroz, O. M. & B. G. F. Junior (2016). Despacho econômico em sistema de seis barras com modulagem de usinas térmicas e eólicas. 4, 68, 70, 71, 73, 81

- Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics* 20, 53–65. 21, 43, 46
- Ruck, D. W., S. K. Rogers, M. Kabrisky, M. E. Oxley, & B. W. Suter (1990). The multilayer perceptron as an approximation to a bayes optimal discriminant function. *IEEE Transactions on Neural Networks* 1(4), 296–298. 3, 43
- Sheehan, J. S. (1986). Graph theory (encyclopedia of mathematics and its applications, 21). *Bulletin of the London Mathematical Society* 18(1), 83–84. 5, 6, 7, 8
- Sing, J. K., D. Basu, M. Nasipuri, & M. Kundu (2003). Improved k-means algorithm in the design of rbf neural networks. In *TENCON 2003. Conference on Convergent Technologies for the Asia-Pacific Region*, Volume 2, pp. 841–845. IEEE. 21, 26, 43, 46
- Sousa, H. V. (2015). *Um estudo sobre conjuntos cliques-dominantes em grafos*. Dissertação de mestrado, Universidade Estadual de Campinas. Instituto de Computação, Programa de Pós-Graduação em Ciência da Computação, Campinas, São Paulo. 14, 15
- Souza, P. V. C., L. C. B. Torres, A. P. Braga, & F. G. Coelho (2014). Análise de algoritmos de maximização de margem baseados em grafos de gabriel e fecho afim. 13
- Students & others working in the social sciences (2017). Social science statistics. 46
- Suykens, J. A. K., T. Gestel, J. Brabanter, B. Moor, & J. Vandewalle (2002). Least-squares support vector machines. *World Scientific: Singapore..* 43, 57
- Torres, L. C. B. (2012). *Uma nova Abordagem Baseada em Margem Para Seleção de Modelos Neurais*. Dissertação de mestrado, UFMG, Belo Horizonte. 1, 13, 28
- Torres, L. C. B. (2016). *Classificador por Arestas de Suporte (CLA): Métodos de Aprendizagem Baseados em Grafos de Gabriel*. Tese de doutorado, UFMG, Belo Horizonte. 1, 10, 12, 13, 27, 31, 36, 46

REFERÊNCIAS

- Tutte, W. T. (1998). *Graph theory as I have known it*. OUP Oxford. 7, 8
- Vapnik, V., C. C. (1995). Support-vector networks. *Machine learning, [S.l.] 20(3)*, 273–297. 26, 43, 57
- Wanderley, B. F. M. (2013). *Estudos em Estimaco de Densidade por Kernel: Mtodos de Seleo de Caractersticas e Estimaco do Parmetro Suavizador*. Tese de doutorado, UFMG, Belo Horizonte. 1, 22, 25, 26, 36
- www.epe.gov.br (2017). Empresa de pesquisa energtica. 69
- Zhang, W. & I. King (2002). *A study of the relationship between support vector machine and gabriel graph*. In *Neural Networks. IJCNN 02*. Proceedings of the 2002 International Joint Conference on, vol. 1. 13
- Zimmerman, R. D. (1995). *Comprehensive distribution power flow: Modeling, formulation, solution algorithms and analysis*. 72