INFORMAÇÃO ESTRUTURAL DADA POR GRAFOS DE GABRIEL APLICADA À REGULARIZAÇÃO DE REDES NEURAIS RBF

Matheus N. Salgado*, Luiz C. B. Torres*, Frederico G. F. Coelho*, Antônio P. Braga*

*Programa de Pós-Graduação em Engenharia Elétrica
Universidade Federal de Minas Gerais
Av. Antônio Carlos 6627, 31270-901
Belo Horizonte, MG, Brasil

Emails: nogueiramaths@gmail.com, luizlitc@gmail.com, fredgfc@gmail.com, apbraga@ufmg.br

Abstract— The present work shows how the regularization of RBF can be performed using information from Gabriel's Graph data representation. The parameters of RBF radial functions are extracted directly from the structure of the graph and the regularization is performed removing radial functions, also using features from this data representation. In conclusion, the work finds ideal conditions for the regularization of RBF networks from the information obtained from Gabriel's Graph.

Keywords— Machine Learning, RBF, Regularizatio, Gabriel Graph, Classification.

Resumo— O presente trabalho mostra como a regularização de RBF pode ser realizada a partir de informações extraídas da representação dos dados via Grafo de Gabriel. Os parâmetros das funções radiais da RBF (centros e raios) são extraídos diretamente da estrutura do grafo e a regularização é feita através da eliminação de funções utilizando também características dessa representação. O trabalho determina as condições ideais para a regularização de redes RBF a partir das informações obtidas com o grafo.

Palavras-chave— Aprendizado de máquina, RBF, Regularização, Grafo de Gabriel, Classificação.

1 Introdução

Parte fundamental no desenvolvimento de qualquer classificador é a sua regularização. Uma vez que o modelo é obtido através de um conjunto de treinamento, é preciso extrair informações desse modelo que indiquem a sua generalização. O classificador com baixa generalização acaba se tornando especialista no conjunto de treinamento (*overfitting*), portanto com desempenho prejudicado para um conjunto de teste. Por outro lado, o classificador com generalização muito alta pouco aprendeu do conjunto de teste, portanto também tem o desempenho prejudicado. O objetivo da regularização é encontrar o grau de generalização que minimiza o erro de teste através de indicadores extraídos do modelo. Este trabalho investiga como a informação estrutural dada por Grafos de Gabriel (Zhang and King, 2002) pode ser aplicada ao processo de regularização de redes neurais RBF (*Radial Basis Function*) (Haykin, 1999).

No caso das redes RBF é mostrado que a magnitude do vetor de pesos da camada de saída indica generalização: quanto maior a norma, menor o grau de generalização (Haykin, 1999). Outro indicativo é a quantidade de funções radiais que projetam a camada escondida da rede. Este trabalho propõe a escolha dessas funções a partir de informações extraídas do Grafo de Gabriel.

Em seu artigo de 2002, Zhang apresenta um estudo sobre a relação entre as SVMs (Boser et al., 1992) e o Grafo de Gabriel, o que motivou o desenvolvimento do método CG-RBF (Torres et al., 2013) que utiliza o grafo para estimar os parâmetros de RBF e apresenta resultados promissores quando comparados a uma SVM (Boser et al., 1992). Além disso, trabalhos recentes têm demonstrado que classificadores eficientes podem ser projetados apenas utilizando características extraídas dessa estrutura de dados (Torres, Castro and Braga, 2015) (Torres, Castro, Coelho, Torres and Braga, 2015).

Outra forma de estimar centros e raios das funções radiais de uma RBF é através de heurísticas como K-médias e suas variações (Sing et al., 2003): *Fuzzy C-means* (FCM), redes *Self-Organizing Maps* (SOM) (Bouchired et al., 1998) e winner-takes-all (WTA). Já o método CG-RBF (Torres et al., 2013), mencionado anteriormente, não utiliza parâmetros para essa estimativa, uma vez que toda a informação é extraída diretamente das relações espaciais dos dados representada pelo Grafo de Gabriel.

Em sua abordagem, o CG-RBF realiza a regularização através da filtragem de ruído e escolha de funções radiais. Essa escolha é feita considerando, após a filtragem, os vértices do grafo que fazem conexão entre classes distintas, chamados Vetores de Suporte Estruturais (VSE)(Torres, Castro, Coelho, Torres and Braga, 2015). Essa estratégia parte do pressuposto de que os vértices que fazem essas conexões estão na margem de separação entre classes. Os VSE são dados críticos para os classificadores baseados no grafo de Gabriel, pois determinam sua superfície de decisão.

Para investigar o comportamento dos VSE na determinação dos parâmetros das funções radiais, o presente trabalho propõe uma comparação entre duas possíveis interpretações para os VSE: indicadores de ruído, portanto as funções radiais associadas a estes vértices são eliminadas para realizarmos a regularização; indicadores de margem, portanto as funções radiais associadas aos VSE são mantidas e as demais eliminadas. O objetivo aqui é então inferir a sua importância e função na regularização do modelo. O grafo realiza o papel de extrator de características para regularização de redes RBF.

O restante do trabalho está organizado da seguinte forma: na próxima seção, um breve referencial teórico sobre teoria de grafos e o Grafo de Gabriel. Na Seção 3, a primeira parte da metodologia aborda os VSE como indicadores de ruído e a segunda parte trata dos VSE como indicadores de margem. Na Seção 4, os resultados para as duas metodologias propostas são apresentados para uma base de dados sintética e também para bases reais. Por fim, na Seção 5, a conclusão.

2 Referencial Teórico

2.1 Teoria de Grafos

Um grafo $G(V,E)$ é uma estrutura matemática representada por um conjunto dos vértices V e um conjunto de pares ordenados E de V, chamados arestas do grafo G. Diversos problemas reais podem ser modelados através de grafos, como redes de computadores, perfis em redes sociais, ruas e esquinas de uma cidade (Bondy and Murty, 1976). Neste trabalho, um grafo especial é utilizado para modelar a estrutura de um conjunto de dados e extrair informações da relação espacial entre as amostras: o Grafo de Gabriel.

2.2 O Grafo de Gabriel

Considere um conjunto de dados $X = \{(x_i, d_i) | i = 1, ..., N\}$, em que $d_i \in D = \{+1, -1\}$ e $x_i \in R^n$, o grafo de Gabriel G_G de D com vértices $V = \{x_i \in X | i = 1, ..., N\}$ tem aresta E de vértices x_i e x_j se, e somente se, $\delta(x_i,x_j)^2 \leq [\delta(x_i,x_k)^2 + \delta(x_j,x_k)^2]$, $\forall x_k \epsilon V$ e $i \neq j \neq k$, em que $\delta(\cdot,\cdot)$ é a distância euclidiana.

Um importante subconjunto de V é o conjunto de vértices x_i e classe d_i que formam arestas com vértices x_j e classe $d_j \neq d_i$, chamado de Vetores de Suporte Estruturais (VSE) (Torres, Castro and Braga, 2015).

2.3 Treinamento da Rede RBF

Considere um conjunto de dados $X = \{x_i\} | i = 1, ..., N$, $x_i \in \Re^n$, associado aos rótulos $D = \{d_i\}$, $d_i \in \{+1, -1\}$. Considere ainda a rede RBF em que $\mathbf{H}$, Equação (1), é a projeção dos padrões de entrada pelas funções radiais $h_k(x_k, z_k)$, e $z_k = \{c_k, r_k\} | k = 1, ..., p$ são os parâmetros estimados de centro e raio, e p é o número de funções radiais. O vetor w corresponde aos pesos da camada de saída da rede.

$$\boldsymbol{H} = \begin{bmatrix} h_1(x_1,z_1) & h_2(x_1,z_2) & ... & h_p(x_1,z_p) \\ h_1(x_2,z_1) & h_2(x_2,z_2) & ... & h_p(x_2,z_p) \\ ... & ... & ... & ... \\ h_1(x_N,z_1) & h_2(x_N,z_2) & ... & h_p(x_N,z_p) \end{bmatrix} \quad (1)$$

O treinamento da rede é feito através da minimização da Equação (2), que representa o erro entre os rótulos com penalização para valores altos de w_j. A minimização segue o desenvolvimento mostrado nas Equações (3) - (5).

$$J = \sum_{i=1}^{N} \left(d_i - \hat{d}_i\right)^2 + \sum_{j=1}^{p} \lambda_j w_j^2 \quad (2)$$

$$\frac{\partial J}{w_j} = \frac{\partial J}{w_j} \sum_{i=1}^{N} \left(d_i - \hat{d}_i\right)^2 + \frac{\partial J}{w_j} \sum_{j=1}^{p} \lambda_j w_j^2 = 0 \quad (3)$$

$$2 \sum_{i=1}^{N} \left(\hat{d}_i - d_i\right) h(x_j, z_j) + 2\lambda_j w_j = 0 \quad (4)$$

$$\sum_{i=1}^{N} \hat{d}_i h(x_j, z_j) + \lambda_j w_j = \sum_{i=1}^{N} d_i h(x_j, z_j) \quad (5)$$

Em sua forma matricial, a Equação (5) é dada pela Equação (6), onde $\mathbf{H}$ é a matriz de projeção de entrada, $\hat{\mathbf{d}} = \mathbf{Hw}$, saída da rede, λ é um parâmetro de regularização, $\mathbf{I_p}$ é a matriz identidade de dimensão p e $\mathbf{d}$ é o vetor de rótulos.

$$\boldsymbol{H^T \hat{d}} + \lambda \boldsymbol{I_p w} = \boldsymbol{H^T d} \quad (6)$$

O vetor de pesos w da camada de saída da RBF será obtido através da Equação (7).

$$\boldsymbol{w} = \left(\boldsymbol{H^T H} + \lambda \boldsymbol{I_p}\right)^{-1} \boldsymbol{H^T D} \quad (7)$$

A classificação final é obtida através da simplificação mostrada na Equação (8), em que $sgn(\cdot)$ representa a função sinal.

$$\boldsymbol{\hat{d}^*} = sgn(\boldsymbol{Hw}) \quad (8)$$

O erro de classificação e_i de um dado x_i é obtido conforme a Equação (9), com valores possíveis de $\{0, 1\}$, devido à simplificação realizada.

$$e_i = \frac{(d_i - \hat{d}_i^*)^2}{4} \quad (9)$$

3 Metodologia

3.1 *Estimativa de Parâmetros das funções radiais da RBF*

As propriedades do Grafo de Gabriel podem ser utilizadas para estimar os centros e raios das funções radiais da RBF. As funções radiais são definidas aqui como uma função gaussiana $h(x,z) = \exp\left(\frac{\|x-c\|^2}{2\sigma^2}\right)$, onde $z = \{c, \sigma\}$.

Considere um conjunto de dados $X = \{x_i\} | i = 1, ..., N,\ x_i \in \Re^n$, associado aos rótulos $D = \{d_i\},\ d_i \in \{+1, -1\}$. $GG_X(V,E)$ é o Grafo de Gabriel resultante do conjunto de dados X. O centro c_i, parâmetro de $z_i = \{c_i, r_i\}$ da função radial $h_i(x_i, z_i)$, é estimado como o vértice $v_i \in GG_X(V,E)$. O raio σ_i é estimado como a média da metade do comprimento de toda aresta $e_j \in GG_X(V,E)$ associada ao vértice v_i. Um exemplo é mostrado na Figura 1.

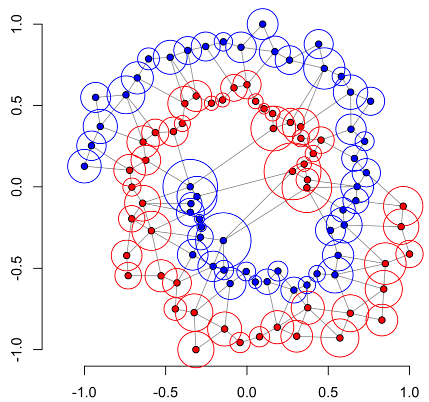

Figura 1: Centros (pontos) e raios (círculos) da RBF estimados através das propriedades do Grafo de Gabriel.

A avaliação dos padrões de entrada pelas funções radiais resultam na matriz de projeção H, Equação (1).

3.2 *Regularização via parâmetro λ*

A escolha do parâmetro de regularização λ, Equação (2), é feita através da minimização do erro de teste. A estimativa do erro de teste, por sua vez, é feita através de LOOCV (*Leave One Out Cross Validation*), conforme a Equação (10)

$$LOOCV = \frac{1}{N}\sum_{k=1}^{N}\left(\left[Hw^{(k)}\right]_k - d_k\right)^2, \qquad (10)$$

onde $[Hw]_k$ representa a estimativa $\hat{d}_k$ e $w^{(k)}$ é o vetor de pesos calculado através da Equação (7), retirando-se a linha k de H. Uma versão simplificada de LOOCV, mostrada na Equação (11), retorna o erro de validação cruzada em unidade de amostras rotuladas erroneamente por amostra. A Equação (11) foi obtida da Equação (10) da seguinte maneira: (i) caso a função $sgn(\cdot)$ retorne a mesma classe d_k, o resultado contribui com zero no somatório. (ii) caso $sgn(\cdot)$ seja diferente de d_k, o resultado será 2 ou -2, que contribui com $+1$ no somatório. Portanto, $LOOCV^*$ pode ser entendida como $\frac{Ne}{N}$, onde Ne é número de amostras rotuladas erroneamente.

$$LOOCV^* = \frac{1}{N}\sum_{k=1}^{N}\frac{\left(sgn(\left[Hw^{(k)}\right]_k) - d_k\right)^2}{4} \qquad (11)$$

A regularização é feita, então, penalizando os valores altos de w_i.

3.3 *Regularização por eliminação de funções radiais - Proposta I*

Além da regularização feita ajustando o valor de λ, este trabalho busca investigar o efeito de regularização provocado pela eliminação de algumas funções radiais. Essa eliminação é baseada nos vértices pertencentes ao conjunto chamado Vetores Suporte Estruturais (VSE), que são os vértices do grafo que têm conexão com vértices de classe oposta.

A estratégia é retirar a função radial h_i associada a um vértice v_i que tem conexão com a classe oposta - ou seja, um vértice $v_i \in \mathcal{S}$, sendo $\mathcal{S}$ o conjunto de vértices VSE. Foram projetados testes para investigar esse comportamento.

Os testes foram realizados na base de dados *spiral* para valores de desvio padrão *dp* entre [0.01,0.20; por 0.01], como são mostrados alguns exemplos na Figura 2. Quanto maior o valor de *dp*, maior a sobreposição das espirais e, consequentemente, maior o ruído nos dados. Para cada base de dados foi calculado o $LOOCV^*$ (Equação (11)) para 4 diferentes configurações: primeiro, sem regularização, w é calculado como na Equação (7) para $\lambda = 0$. Na segunda configuração foi escolhido o valor de λ que minimiza a Equação (10). Para a terceira e quarta configurações foram repetidas as escolhas de λ do primeiro e segundo, porém com a eliminação das funções radiais h_i associadas a vértices v_i que têm aresta com vértices de classes diferentes. A Tabela 1 resume as configurações descritas (Config 1 - Config 4), e os resultados estão apresentados na Figura 3.

A minimização da Equação (10) para obtenção de λ ótimo nas configurações 2 e 4, Tabela 1, foi feita através do método *Brent*. Além dos resultados de $LOOCV^*$ para cada configuração, também foi registrado o valor de λ ótimo para cada passo do teste.

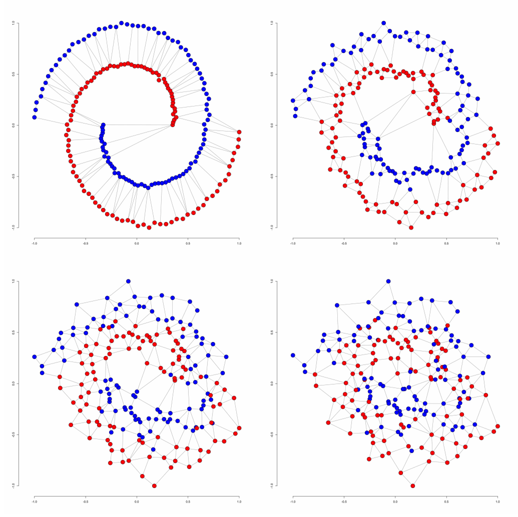

Figura 2: Base de dados *spiral* gerada com diferentes valores de desvio padrão: [0.01,0.06,0.13,0.2].

3.4 *Regularização por eliminação de funções radiais - Proposta II*

Na primeira proposta foi investigado o comportamento da tentativa de regularização através da eliminação de funções radiais associadas aos VSE. Na proposta II a intuição oposta é desenvolvida: os VSE como indicadores de margem. As funções radiais h_i associadas a vértices que fazem conexão com a classe oposta são mantidas e todas as demais são eliminadas. A estimativa dos raios e o treinamento da rede RBF são feitos como nas Seções 3.1 e 2.3, respectivamente.

No Grafo de Gabriel, um vértice que faz conexão com a classe oposta é crucial para a classificação (Torres, Castro and Braga, 2015): em problemas sem sobreposição de classes eles estão na margem de separação. Já em problemas com sobreposição de classes, são possíveis candidatos a ruído. Assim, intuitivamente, espera-se que a eliminação de funções radiais feita na Seção 3.3 tenha melhor desempenho para base de dados com sobreposição, e que a proposta da seção atual tenha melhor desempenho em bases de dados não sobrepostas.

Tabela 1: Retirada de funções radiais h_i e λ para cada configuração. As configurações 3 e 4 são referentes a proposta I, enquanto 5 e 6 são referentes a proposta II.

Configuração	Retirada de Funções Radiais h_i	λ
Config 1	Nenhuma	0
Config 2	Nenhuma	ótimo
Config 3	associadas aos VSE	0
Config 4	associadas aos VSE	ótimo
Config 5	não associadas aos VSE	0
Config 6	não associadas aos VSE	ótimo

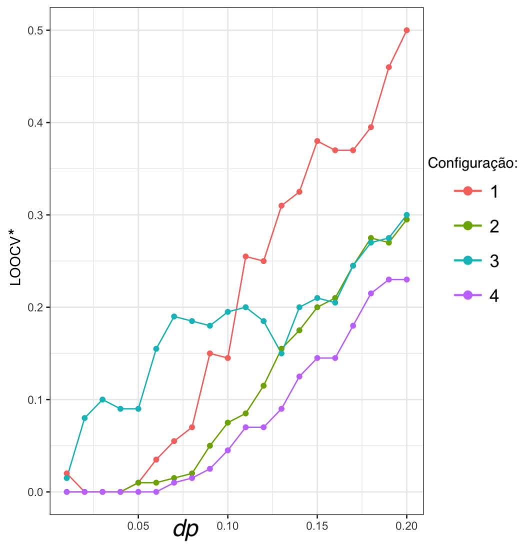

Figura 3: Resultados de testes feitos na base *spiral*. 1) $\lambda = 0$ sem eliminação de h_i; 2) λ escolhido min Equação (10) sem eliminação de h_i; 3) $\lambda = 0$ com eliminação de h_i; 4) λ escolhido min Equação (10) com eliminação de h_i.

O experimento descrito na Seção 3.3 foi repetido para o cenário atual acrescentando duas novas configurações para comparação: cofiguração 5, h_i associadas aos VSE são mantidas e as demais eliminadas, e $\lambda = 0$. A configuração 6 representa a mesma configuração anterior, porém para λ escolhido através da minimização da Equação (10). A Tabela 1 também resume as configurações 5 e 6.

Na seção seguinte os resultados dos testes são apresentados e discutidos. Além disso, também são apresentados testes em base de dados reais.

4 Resultados

Para melhor compreensão, os resultados estão organizados em três subseções: (1) Os resultados da proposta I de eliminação de funções radiais são comparados com não eliminação, e com a regularização via parâmetro λ; (2) São adicionados à discussão os resultados da proposta II comparados com os demais resultados obtidos; e (3) Experimentos em bases de dados reais são realizados e discutidos.

4.1 *Resultados - Proposta I*

Para cada valor de desvio padrão *dp* associado à distribuição que gerou a base *spiral* de teste, foi calculado o valor de LOOCV, para as 4 diferentes configurações (Tabela 1). Este resultado está apresentado na Figura 3.

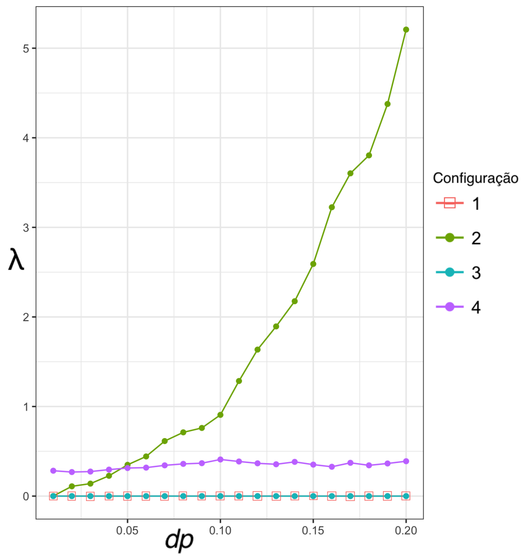

Figura 4: Valores escolhidos de λ para os testes realizados. 1) $\lambda = 0$ sem eliminação de h_i; 2) λ escolhido min Equação (10) sem eliminação de h_i; 3) $\lambda = 0$ com eliminação de h_i; 4) λ escolhido minimizando a Equação (10) com eliminação de h_i.

Apesar do pior desempenho para baixos valores de dp (que representa o ruído na base de dados), a configuração 3, que representa a eliminação de funções radiais e $\lambda = 0$, consegue bons resultados para valores mais altos de dp. Chega, inclusive, a ter melhores resultados de LOOCV do que a regularização sem eliminação de h_i. A combinação "regularização mais eliminação"de h_i obteve o melhor resultado para qualquer valor de dp.

As conclusões da Figura 3 devem estar acompanhadas das informações da Figura 4, que mostra os valores do λ escolhido em cada teste. Observa-se que para o grupo 2 (regularização sem eliminação de h_i) λ é crescente em relação a dp. Com essa mesma observação, concluimos também que a eliminação de funções radiais contribui para valores menores de λ que, na regularização, foi mantido aproximadamente constante em torno de 0.34.

Os resultados indicam que a eliminação de funções radiais pode ter caráter regulatório diferente da regulação pela magnitude de w.

4.2 Resultados - Proposta II

A proposta II procurou manter as funções radiais associadas aos VSE e eliminar as demais, considerando a intuição de que os VSE são indicadores de margem. Esse resultado está apresentado nas Figuras 5 e 6.

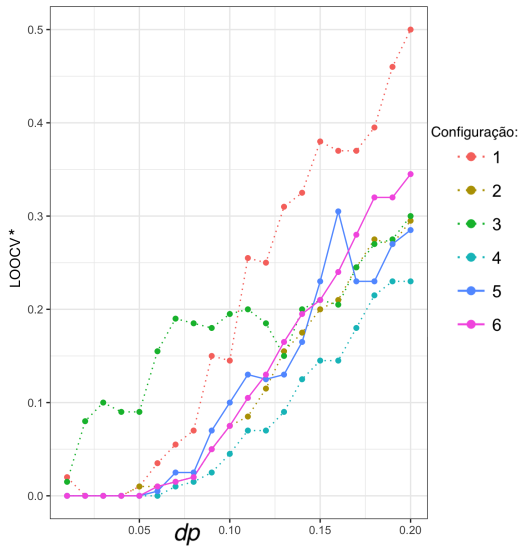

Figura 5: Dois novos grupos adicionados a Figura 3 para comparação: 5) h_i centradas nos VSE e $\lambda = 0$; 6) h_i centradas nos VSE e λ escolhido min Equação (10).

As configurações ímpares têm o parâmetro de regularização $\lambda = 0$. A configuração 5 apresenta desempenho superior ao da configuração 3 para baixos valores de dp, e desempenho similar para os demais valores. É válido lembrar que, na configuração 5, as funções radiais associadas aos VSE são escolhidas, ao contrário da configuração 3, que é associada aos demais vértices. Era esperado que a configuração 3 representasse bem as bases de dados com sobreposição e a configuração 5, as bases não sobrepostas. No entanto, isso não acontece. Os resultados da Figura 5 sugerem que a representatividade dos VSE é superior.

As configurações de números pares têm λ escolhido para minimizar a Equação (10). As configurações 2 e 6 apresentam comportamento parecido tanto no gráfico da Figura 5 quanto na Figura 6. Em ambos os casos o valor de λ cresce à medida que a sobreposição de dados aumenta. O mesmo não acontece para a configuração 4, que apresenta o melhor resultado de todos os grupos para qualquer valor de dp. Uma particularidade da configuração 4 é que ela elimina mais funções radiais para maiores valores de dp, ao contrário da configuração 6, que acrescenta. Os valores de λ na configuração 4 na Figura 6 realmente se destacam.

4.3 Bases de dados reais

Experimentos em base de dados reais também foram realizados para investigar o desempenho das metodologias propostas. 14 bases de dados reais binárias extraídas do *UCI* (Dheeru and Karra Taniskidou, 2017) foram utilizadas, exceto por *breastHess*, um problema de expressão genética (Hess et al., 2006). Mais informações sobre as bases de dados utilizadas são apresentadas na Tabela

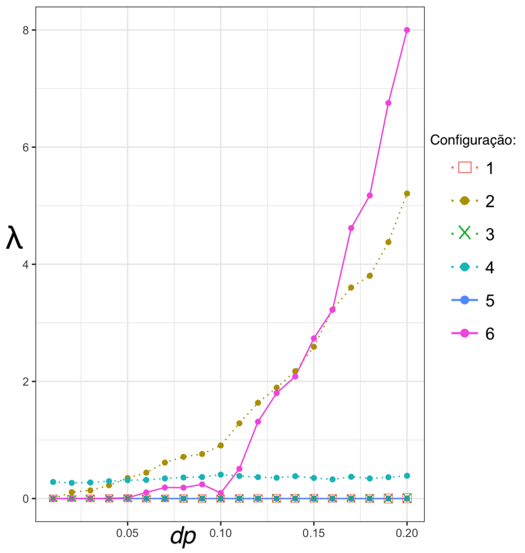

Figura 6: Valores escolhidos de λ para os testes realizados. (1-4) v. Figura 4. 5) h_i centradas nos VSE e $\lambda = 0$; 6) h_i centradas nos VSE e λ escolhido minimizando a Equação (10).

2. Os dados passaram por um pré-processamento de remoção de valores faltantes e reescala entre $\{-1, 1\}$. As bases de dados *segmentation* e *glass* não são originalmente binárias, mas foram reduzidas como mostrado em (Castro and Braga, 2013). A estimativa do erro de teste é feita da mesma forma que os experimentos anteriores: LOOCV, Equação (11). Os resultados são mostrados nas Tabelas 3 e 4 e avaliam a performance das 6 diferentes configurações propostas (Tabela 1) por LOOCV.

Para comparação entre a performance das diferentes configurações foi utilizado o teste de *Friedman*. Quanto menor o valor da média do $Rank(\mathcal{L})$, melhor a performance da configuração, como pode ser verificado na Tabela 5.

Os experimentos mostram que, para as bases de dados reais analisadas, o efeito da regularização através da eliminação de funções radiais só aparece se a regularização feita através da otimização de λ não for realizada. Tendo em vista o custo de um algoritmo de otimização, podemos concluir que a eliminação de funções radiais utilizando características do grafo de Gabriel para regularização é promissora.

5 Conclusão

O trabalho investiga características que levam à generalização em classificadores RBF usando informações extraídas da representação Grafo de Gabriel dos dados. Um conjunto especial de amostras é extraído do grafo e chamado de Vetores de Suporte Estruturais (VSE). A representação é utilizada não apenas para estimar os parâmetros das funções radiais da RBF, como também para eli-

Tabela 2: Informações sobre as 14 bases de dados reais em que os experimentos foram realizados.

Base de dados	abv	Amostras	Atributos
Segmentation	seg	210	18
Glass	gla	214	9
Appendicitis	app	106	7
Ionosphere	ion	351	33
Breastcancer	bre	683	9
Australian	aus	690	14
Diabetes	dia	768	8
BreastHess	bhe	133	30
Bupa	bup	345	6
Haberman	hab	306	3
Banknote	ban	1372	4
Fertility	fer	100	9
Parkinons	par	195	22
ILPD	ilp	579	10

Tabela 3: Resultado de LOOCV para as configurações 1, 2 e 3.

abv	Config1	Config2	Config3
seg	0.114	0.014	0.224
gla	0.061	0.033	0.107
app	0.255	0.142	0.226
ion	0.453	0.046	0.148
bre	0.095	0.031	0.063
aus	0.529	0.139	0.184
dia	0.431	0.236	0.271
bhe	0.241	0.173	0.241
bup	0.400	0.310	0.472
hab	0.402	0.291	0.278
ban	0.259	0.001	0.153
fer	0.350	0.120	0.120
par	0.056	0.046	0.210
ilp	0.420	0.297	0.314

Tabela 4: Resultado de LOOCV para as configurações 4, 5 e 6.

abv	Config4	Config5	Config6
seg	0.100	0.033	0.005
gla	0.037	0.033	0.033
app	0.113	0.217	0.151
ion	0.131	0.245	0.048
bre	0.038	0.082	0.034
aus	0.228	0.454	0.138
dia	0.286	0.392	0.236
bhe	0.226	0.308	0.173
bup	0.478	0.522	0.307
hab	0.265	0.301	0.284
ban	0.000	0.026	0.023
fer	0.120	0.390	0.120
par	0.169	0.174	0.041
ilp	0.287	0.356	0.297

Tabela 5: Média do $Rank(\mathcal{L})$ para as 6 diferentes configurações, resultado do teste de Friedman.

Configuração	Média $Rank(\mathcal{L})$
Config 1	5.2500
Config 2	2.0000
Config 3	4.2143
Config 4	2.8929
Config 5	4.7143
Config 6	1.9286

minar funções radiais, promovendo regularização. A primeira parte deste trabalho investiga a intuição de que os VSE indicam possíveis ruídos. A segunda parte investiga a noção inversa: de que os VSE indicam margem de separação. Além da eliminação de funções radiais específicas, os diferentes cenários são regularizados através da minimização da magnitude do vetor de pesos.

Era esperado que os VSE indicassem ruído para dados sobrepostos e indicassem margem para dados não sobrepostos. No entanto, o trabalho mostra que os VSE não podem ser considerados ruídos. Em bases de dados bem sobrepostas o resultado das configurações 3 e 5 são similares. Já em bases sem sobreposição, a configuração 5 tem desempenho muito superior à configuração 3. Isso indica que a representatividade dos VSE é grande. No entanto, as investigações mostram um fato curioso: quando combinados com a regularização por tamanho do vetor de pesos, o melhor desempenho foi observado para a configuração 4 (que exclui os VSE). Também observa-se que a eliminação de funções radiais mantém λ aproximadamente constante, enquanto para as configurações 2 e 6, é crescente. Por fim, percebe-se que, para problemas reais, a regularização através do ajuste de λ parece dominar sobre a regularização pela eliminação de funções radiais. Esses resultados ainda precisam ser investigados mais produndamente e serão tema de trabalhos futuros. Contudo fica claro que a metodologia é promissora.

Agradecimentos

Os autores agradecem à FAPEMIG, CAPES e CNPq pelo suporte dado a este trabalho.

Referências

Bondy, J. A. and Murty, U. S. R. (1976). Graph theory with applications.

Boser, B. E., Guyon, I. M. and Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers, *Proceedings of the fifth annual workshop on Computational learning theory*, ACM, pp. 144–152.

Bouchired, S., Ibnkahla, M., Roviras, D. and Castanie, F. (1998). Equalization of satellite mobile communication channels using combined self-organizing maps and rbf networks, *Acoustics, Speech and Signal Processing, 1998. Proceedings of the 1998 IEEE International Conference on*, Vol. 6, IEEE, pp. 3377–3379.

Castro, C. L. and Braga, A. P. (2013). Novel cost-sensitive approach to improve the multilayer perceptron performance on imbalanced data, *IEEE transactions on neural networks and learning systems* **24**(6): 888–899.

Dheeru, D. and Karra Taniskidou, E. (2017). UCI machine learning repository.

Haykin, S. (1999). Neural networks: A comprehensive foundation. 2º.

Hess, K. R., Anderson, K., Symmans, W. F., Valero, V., Ibrahim, N., Mejia, J. A., Booser, D., Theriault, R. L., Buzdar, A. U., Dempsey, P. J. et al. (2006). Pharmacogenomic predictor of sensitivity to preoperative chemotherapy with paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide in breast cancer, *Journal of clinical oncology* **24**(26): 4236–4244.

Sing, J., Basu, D., Nasipuri, M. and Kundu, M. (2003). Improved k-means algorithm in the design of rbf neural networks, *TENCON 2003. Conference on Convergent Technologies for the Asia-Pacific Region*, Vol. 2, IEEE, pp. 841–845.

Torres, L. C., Castro, C. L. and Braga, A. P. (2015). A parameterless mixture model for large margin classification, *Neural Networks (IJCNN), 2015 International Joint Conference on*, IEEE, pp. 1–6.

Torres, L. C., Lemos, A. P., Castro, C. L. and Braga, A. P. (2013). Projeto de redes rbf baseado na estrutura dos dados e em informações de margem, *Computational Intelligence, 2013 1st BRICS Countries Congress(BRICS-CCI) and 11th Brazilian Congress (CBIC) on*, pp. 1–7.

Torres, L., Castro, C., Coelho, F., Torres, F. S. and Braga, A. (2015). Distance-based large margin classifier suitable for integrated circuit implementation, *Electronics Letters* **51**(24): 1967–1969.

Zhang, W. and King, I. (2002). A study of the relationship between support vector machine and gabriel graph, *Neural Networks, 2002. IJCNN'02. Proceedings of the 2002 International Joint Conference on*, Vol. 1, IEEE, pp. 239–244.