

Modelo geométrico de margem larga baseado em propriedades estruturais de Grafos de Gabriel e em distância geodésica

Matheus N. Salgado, Luiz C. B. Torres*, e Antônio P. Braga

Universidade Federal de Minas Gerais,
Av. Antônio Carlos 6627, 31270-901, Belo Horizonte, MG, Brasil
{nogueiramaths, luizlitc}@gmail.com, apbraga@ufmg.br
<http://litc.cpdee.ufmg.br/>

Resumo. Uma nova abordagem na identificação de amostras da margem de separação em classificadores de margem larga é apresentada nesse trabalho. A partir de uma representação geométrica dos dados pelo Grafo de Gabriel, é definida uma métrica de grafo não paramétrica para leitura das informações contidas na representação estrutural do grafo. O trabalho mostra que a métrica sugerida tem distribuição normal para base de dados reais e pode ser utilizada para melhorar o desempenho de classificadores estruturais construídos a partir do Grafo de Gabriel através da filtragem de amostras que causam sobreposição.

Palavras-chave: Aprendizado de Máquina, Classificadores de Margem Larga, Grafo de Gabriel

1 Introdução

Um problema fundamental no desenvolvimento de classificadores de margem larga é a identificação de amostras junto à margem de separação, em relação às quais o classificador será construído, e a margem do separador correspondente, maximizada. Em modelos baseados em vetores de suporte, como as SVM's [1], estes são obtidos como resultado da otimização de funções-objetivo baseadas em estimativas de erro e de margem de separação. A formulação por Multiplicadores de Lagrange [2] resulta na identificação dos vetores de suporte com base na sensibilidade da função objetivo em relação à posição espacial das amostras. Em contraste com a formulação baseada em métodos de otimização, a construção de classificadores baseada em informações estruturais é fundamentada no princípio de que as informações de separação são intrínsecas às relações espaciais das amostras de entrada. Tem sido demonstrado que classificadores estruturais eficientes podem ser construídos a partir de Grafos de Gabriel [[3],[4],[5]]. O problema fundamental desta abordagem reside na leitura correta das informações contidas na representação estrutural do grafo.

* Bolsista do CNPq-Brasil (N°150254/2016-4)

Superfícies de separação resultantes de SVM's possuem margem máxima para problemas linearmente separáveis, no entanto, quando há sobreposição das amostras, a sua formulação requer a utilização de variáveis de folga [2]. Na prática, a utilização de variáveis de folga resulta na eliminação da sobreposição na região de separação, já que as amostras na região da margem deixam de influenciar no valor da função objetivo. Em SVM's a eliminação da sobreposição é formulada como parte do problema de otimização, sendo ela determinada pelo parâmetro de regularização [2]. O mesmo problema surge em modelos estruturais baseados em grafos, já que a não eliminação de amostras na região de separação pode resultar em modelos super-ajustados aos dados *overfitting*.

Em trabalhos anteriores [6] a eliminação das amostras na região de margem foi realizada por meio de filtragem baseada no grau de cada vértice. Este trabalho, por sua vez, apresenta uma abordagem para filtragem das amostras com base em propriedades intrínsecas à grafos e, particularmente, aos Grafos de Gabriel. O Grafo de Gabriel é um subgrafo da Triangulação de Delaunay [7] que, em sua construção, resulta em um grafo sempre conexo. Isso significa que há um caminho entre qualquer par de vértices e que a distância entre eles, também chamada de distância geodésica [8], pode ser calculada.

A partir dessas propriedades pode ser calculada a distância média entre diferentes classes: o quão uma amostra está distante das amostras de outras classes. As amostras da região de sobreposição estão mais próximas de outras classes do que as demais. Assim, é definida uma métrica de grafo não paramétrica que representa, indiretamente, a proximidade das amostras com a região de sobreposição. A filtragem então é realizada utilizando essa métrica: os padrões de menor valor em relação à ela são retirados, pois são considerados da região de sobreposição. É interessante ressaltar que essa métrica depende apenas das relações espaciais entre as amostras na construção do Grafo de Gabriel.

O restante do artigo encontra-se organizado da seguinte forma: A seção 2 apresenta um referencial teórico para construção do Grafo de Gabriel e do classificador baseado nessa representação dos dados. Também é apresentado o problema da sobreposição de classes para este classificador. Em seguida, na seção 3, a métrica de grafo sugerida para a filtragem é definida e visualizada para bases sintéticas. Na seção 4, experimentos em 17 bases reais são descritos para avaliação da métrica de grafo e comparação do desempenho da filtragem nela baseada. Por fim, na seção 5, as conclusões.

2 Referencial Teórico

2.1 Construção do Grafo de Gabriel

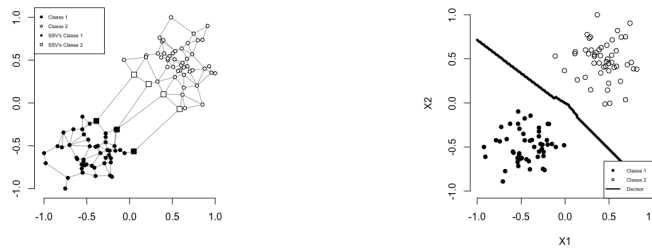
A estrutura é extraída através do modelo de Grafo de Gabriel [[7],[4]] dos dados. Considere um conjunto de dados $D = \{(x_i, y_i) | i = 1, \dots, N\}$, em que $y_i \in \{+1, -1\}$ e $x_i \in R^n$, o grafo G_G de D com vértices $V = \{x_i \in D | i = 1, \dots, N\}$ tem aresta E de vértices x_i e x_j se, e somente se, $\delta(x_i, x_j)^2 \leq [\delta(x_i, x_k)^2 + \delta(x_j, x_k)^2]$, $\forall x_k \in V$ e $i \neq j \neq k$, em que $\delta(\cdot, \cdot)$ é a distância euclidiana ao quadrado.

2.2 Classificador

O que define o classificador são as suas Arestas de Suporte (AS). AS's são escolhidas dentre as arestas do Grafo de Gabriel em que seus vértices x_i e x_j são de classes y_i e y_j diferentes. O conjunto dos vértices das AS's forma o conjunto de vetores suporte estruturais (VSE's) [3]. O classificador final é construído através do hiperplano formado pelo ponto médio mais próximo de uma amostra x que se deseja classificar; esse classificador é uma versão reduzida do classificador proposto por [3].

2.3 O Problema da Sobreposição de Classes

O que determina o classificador final são os VSE's das AS's [3]. Quando as classes estão distantes uma da outra, as AS's descrevem bem a fronteira de separação dos dados, como é mostrado na Figura 1. No entanto, quando há sobreposição de dados de diferentes classes, as AS's não descrevem uma fronteira suave de separação, causando *overfitting*, como é mostrado na Figura 2.



(a) Grafo de Gabriel dos dados originais e VSE's.

(b) Decisor final do classificador.

Fig. 1: Classificação com Grafo de Gabriel para dados sem classes sobrepostas.

A Figura 1a mostra um conjunto de dados modelado pelo Grafo de Gabriel e indica seus VSE's. Os dados da classe 1 vem da distribuição $\mathcal{N}(\mu = 3, \sigma = 0, 5)$ e da classe 2, da distribuição $\mathcal{N}(\mu = 6, \sigma = 0, 5)$. A Figura 1b mostra o decisor gerado a partir dos VSE's. A Figura 2a é análoga à 1a para as distribuições $\mathcal{N}(\mu = 3, \sigma = 2, 5)$ e $\mathcal{N}(\mu = 6, \sigma = 2, 5)$, classe 1 e 2, respectivamente. A Figura 2b é análoga à 1b para os dados de 2a.

No caso de dados com classes sobrepostas, escolher AS's como as arestas com vértices de classes distintas, simplesmente, causa *overfitting*. Este artigo apresenta uma métrica de grafo que revela características da fronteira entre os dados a partir de sua estrutura para eliminar a sobreposição de classes e, assim, escolher AS's que resultem em um classificador mais suave e eficiente.



(a) Grafo de Gabriel dos dados originais e VSE's. (b) Decisor final do classificador.

Fig. 2: Classificação com Grafo de Gabriel para dados com classes sobrepostas.

3 Detecção de Amostras na Margem de Separação entre Duas Classes

3.1 A Distância Entre Classes

A eliminação da sobreposição de amostras na margem de separação é formulada aqui a partir de propriedades intrínsecas ao Grafo de Gabriel, representação estrutural dos dados. A relação de distância relativa entre classes ao longo do grafo é calculada através da métrica de grafo $D_C(\cdot)$, Definição 1.

Considere $G_G = (V, E)$ um Grafo de Gabriel rotulado de ordem N . Cada rótulo define a classe de cada vértice $v \in V$. Cada aresta $e \in E$ tem um peso associado w igual a distância euclidiana das amostras que geraram aquela aresta. A Distância entre Classes $D_C : V \rightarrow \mathfrak{R} \geq 0$ de um vértice $v \in V$ é definida como a média da distância entre v e todos os vértices de classes distintas da sua própria.

Definição 1. *Sejam os rótulos de G_G $y_i = \{0, 1\}$, $i = \{1, 2, \dots, N\}$. U_0 e U_1 são subconjuntos de V formados pelos vértices de classe $y_i = 0$ e $y_i = 1$ e tamanho n_0 e n_1 , respectivamente.*

$$D'_C(u_i) = \left(\frac{\sum_{v \in U_0} d(u_i, v)}{n_0} \right)^{y_i} \left(\frac{\sum_{v \in U_1} d(u_i, v)}{n_1} \right)^{1-y_i}, \forall i. \quad (1)$$

$$D_C(u_i) = \frac{D'_C(u_i)}{\max \{D'_C(v)\}}, \forall v \in V. \quad (2)$$

O operador $d(\cdot, \cdot)$ indica a distância geodésica [8]. A métrica admite valores entre 0 e 1: $D_C(u_1) = 0$ quando os vértices da outra classe estão no mesmo ponto que u_1 e $D_C(u_1) = 1$ quando u_1 é o vértice mais distante.

As amostras da região da margem de separação podem ser identificadas através de $D_C(\cdot)$. A Figura 3 representa um grafo sintético dividido em duas classes com pontos dispostos como dois hexágonos regulares. O valor especificado em cada vértice indica a distância média entre aquele ponto e os pontos da

outra classe, como definido na Equação (2). Assim, percebe-se que os menores valores dessa métrica estão na fronteira entre as classes e vão aumentando quando se afastam dela.

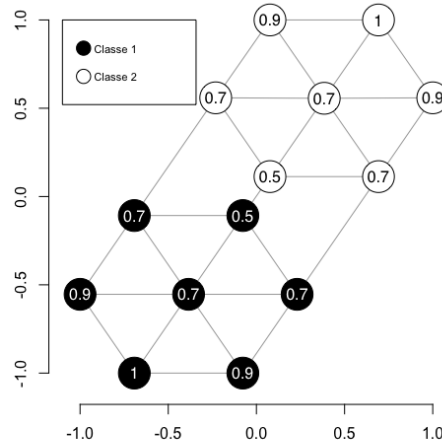


Fig. 3: $D_C(\cdot)$ calculado para cada vértice do grafo sintético.

3.2 Distância Entre Classes de Fronteira Não Linear

A métrica de grafo $D_C(\cdot)$, definida na Equação 2, pode ser calculada para qualquer conjunto de amostras modeladas via Grafo de Gabriel e o seu resultado avaliado para diferentes formatos de margem.

A base de dados sintética *benchmark fullmoon* tem margem de separação não linear, além de alguns dados sobrepostos, como mostrado no lado esquerdo da Figura 4. Seja \mathcal{W} o conjunto de vértices do Grafo de Gabriel construído através das amostras de *fullmoon*, mostrado no lado direito da Figura 4. A interpolação dos dados de $D_C(\mathcal{W})$ gera uma superfície que mostra que a métrica identifica amostras de uma margem de separação não linear, resultando em valores menores para essas amostras. O gráfico resultante dessa interpolação está apresentado na Figura 5 e mostra a distribuição aproximada da métrica para qualquer ponto entre as amostras. O vale formado pela superfície acompanha a margem de separação não linear entre as classes. Outra visualização desse resultado é mostrada pelas curvas de nível e mapa de calor da superfície aproximada, Figura 6a e 6b, respectivamente.

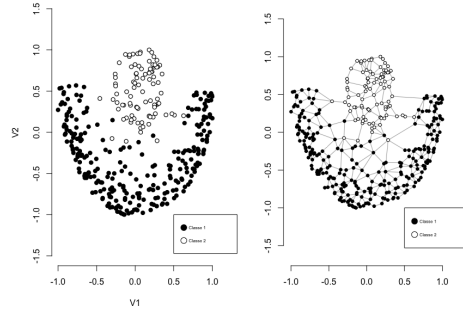


Fig. 4: Esquerda: Base de dados sintética *benchmark fullmoon*. Direita: Base de dados sintética *benchmark fullmoon* modelada pelo Grafo de Gabriel.

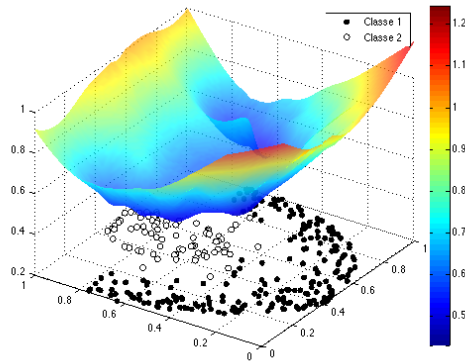
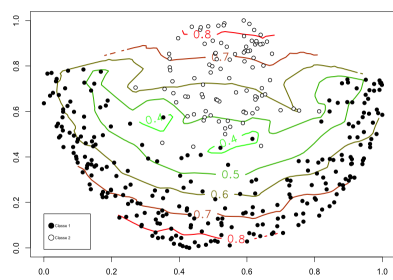
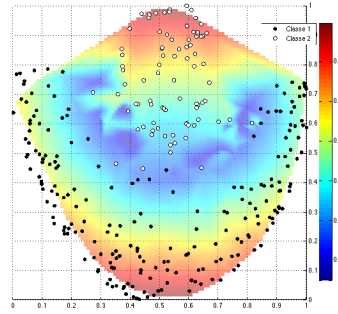


Fig. 5: Gráfico da Superfície de $D_C(\mathcal{W})$, base sintética *benchmark fullmoon*.

O valor de $D_C(\cdot)$ de uma amostra pode assumir qualquer valor entre 0 e 1, dependendo apenas da sua localização na representação estrutural. A distribuição da métrica para todas as amostras é mostrada pelo histograma da Figura 7, juntamente com uma curva gaussiana de média e desvio padrão estimados dos valores de $D_C(\mathcal{W})$. O comportamento normal de $D_C(\cdot)$ para a base *fullmoon* foi confirmado através do teste de Kolmogorov-Smirnov [9] que resultou em um p-valor de 0,7492, que implica na falha em rejeitar a hipótese nula de que as amostras seguem uma distribuição normal, para uma intervalo de confiança de 95%. Essa é uma característica interessante que será explorada em base de dados reais.



(a) Gráfico de Contorno de $D_C(\mathcal{W})$.



(b) Mapa de Calor de $D_C(\mathcal{W})$.

Fig. 6: Resultado da métrica de grafo $D_C(\mathcal{W})$, base sintética *benchmark fullmoon*.

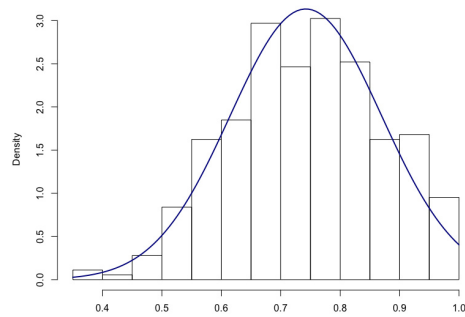


Fig. 7: Histograma de $D_C(\mathcal{W})$ normalizado, base sintética *benchmark fullmoon*.

4 Experimentos e Resultados em Bases Reais

Foram feitos dois experimentos em bases reais: um para verificar se a característica normal da distribuição de $D_C(\cdot)$ encontrada em bases sintéticas se repetia em bases reais; e outro para verificar a eficiência da filtragem baseada em $D_C(\cdot)$ se comparada à filtragem baseada no grau do vértice [6] e a um classificador sem filtragem.

Ambos experimentos foram realizados em 17 bases de dados reais extraídas do *UCI* [10], exceto por "bupa", extraída do *KEEL* [11], e por dois problemas de expressão gênica: "Golub" [12] e "breastCancerHess" [13]. Os dados passaram por um pré-processamento de remoção de valores faltantes e reescala entre $\{-1, 1\}$. Todas as bases de dados são originalmente binárias, exceto por "segmentation" e "glass", que foram reduzidas a binárias como mostrado em [14].

4.1 Verificação Empírica da Normalidade de $D_C(\cdot)$ em Bases Reais

Um Grafo de Gabriel foi construído para cada uma das 17 base de dados. A métrica de grafo $D_C(\cdot)$, Equação 2, foi avaliada para cada estrutura de dados. O teste de Kolmogorov-Smirnov [9] foi utilizado para verificar se as medidas de $D_C(\cdot)$ vêm de uma distribuição normal. A Tabela 1 mostra os resultados do teste: p-valor maior que 0,05 indica a falha em rejeitar a hipótese nula de que os dados seguem uma distribuição normal para um intervalo de confiança de 95%. 9 das 17 bases reais testadas têm distribuição normal.

4.2 Classificador com Filtragem baseada em $D_C(\cdot)$

O classificador baseado em Grafo de Gabriel foi utilizado para classificar os dados das 17 bases citadas. O experimento foi realizado empregando validação cruzada com 10 *folds*. A performance média AUC é mostrada na Tabela 2.

O classificador foi testado para 3 diferentes modelos: o primeiro, sem filtragem, consiste na construção do Grafo de Gabriel para todas as amostras. No segundo modelo foi utilizada a filtragem dos dados baseada no grau do vértice [6]. Por fim, no terceiro modelo, foi utilizada a filtragem baseada na métrica de grafo apresentada neste trabalho, $D_C(\cdot)$.

Seja \mathcal{V}_i o conjunto de vértices da base de dados de $id = i$ modelada via Grafo de Gabriel, sendo $i = 1, 2, \dots, 17$, Tabela 1. $D_C(\mathcal{V}_i)$ é o vetor resultante da avaliação de $D_C(v)$ para todo $v \in \mathcal{V}_i$ e retira-se todo $u \in \mathcal{V}_i$ que satisfaz a Inequação 3 em cada iteração da validação cruzada, em que $\overline{D_C(\mathcal{V})}$ é a média aritmética, $dv(\cdot)$ é o desvio padrão e $k = 0,4$. A performance média AUC para os três modelos é mostrada na Tabela [2].

$$D_C(u) < \overline{D_C(\mathcal{V}_i)} - k * dv(D_C(\mathcal{V}_i)). \quad (3)$$

O valor de $k = 0,4$ foi escolhido empiricamente após testes do classificador nas 17 bases para os valores de $k = 0; 0,1; 0,2; \dots, 2,0$. A Figura 8 mostra a

Table 1: Tabela resultado do teste Kolmogorov-Smirnov de normalidade para $D_C()$ aplicado à bases reais. Os valores maiores que 0,05 estão em negrito e indicam as bases que têm distribuição normal.

Base	Id	D	pvalue
<i>segmentation</i>	1	0,08	1,20E-01
<i>glass</i>	2	0,03	9,84E-01
<i>ionosphere</i>	3	0,09	5,72E-03
<i>sonar</i>	4	0,04	8,80E-01
<i>breastcancer</i>	5	0,11	1,93E-07
<i>australianStatlog</i>	6	0,08	1,45E-04
<i>diabetes</i>	7	0,06	9,17E-03
<i>breastCancerHess</i>	8	0,07	5,40E-01
<i>bupa</i>	9	0,07	5,47E-02
<i>banknote</i>	10	0,07	1,17E-06
<i>fertility</i>	11	0,13	6,96E-02
<i>parkinsons</i>	12	0,12	9,05E-03
<i>climate</i>	13	0,05	1,96E-01
<i>ILPD</i>	14	0,06	3,06E-02
<i>germanCredit</i>	15	0,03	2,02E-01
<i>heart</i>	16	0,10	1,36E-02
<i>golub</i>	17	0,08	7,44E-01

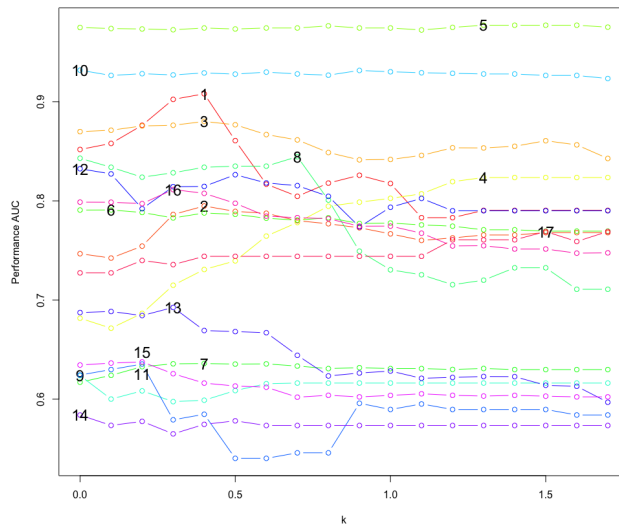


Fig. 8: Gráfico da performance AUC do classificador para diferentes valores de k. Cada base de dados está identificada por sua id correspondente, colocada no seu ponto de máximo.

performance AUC em cada uma das 17 bases, identificadas por sua *id* no seu ponto máximo.

Os três modelos de classificação foram comparados através do teste de *Friedman*, indicado em [15] para comparação de múltiplos modelos. O teste ranqueia os 3 modelos, como mostrado na Tabela 2. Quanto menor o valor da Média do Rank $R(\mathcal{L})$, melhor classificado é o modelo. Para verificar a significância das diferenças entre os valores de $R(\mathcal{L})$, encontrados pelo teste de *Friedman*, foi utilizado o método *Bonferroni* [15], que utiliza a distribuição *t-Student* ajustada para múltiplas comparações. O resultado dessa comparação é mostrado na Figura 9, que indica que o classificador com filtragem baseada em $D_C(\cdot)$ é estatisticamente diferente e, portanto, mais eficiente do que o classificador sem filtragem. O mesmo não pode ser verificado para o classificador com filtragem baseada no grau do vértice.

Table 2: Resultados: média de AUC e desvio padrão para três modelos de classificadores. Os melhores resultados estão apresentados em negrito. N_t/N_a indicam, respectivamente, o número total de amostras da base e o número de atributos.

Base	Classificador Sem Filtro	Classificador com Filtragem baseada no Grau do Vértice	Classificador com Filtragem baseada em $D_C(\cdot)$	N_t/N_a
<i>segmentation</i>	0,790 ± 0,193	0,785 ± 0,154	0,908 ± 0,054	210/18
<i>glass</i>	0,768 ± 0,178	0,921 ± 0,106	0,795 ± 0,121	214/9
<i>ionosphere</i>	0,838 ± 0,058	0,877 ± 0,083	0,880 ± 0,065	351/33
<i>sonar</i>	0,824 ± 0,065	0,728 ± 0,084	0,731 ± 0,106	208/60
<i>breastcancer</i>	0,972 ± 0,017	0,972 ± 0,015	0,974 ± 0,016	683/9
<i>australian</i>	0,769 ± 0,033	0,801 ± 0,019	0,788 ± 0,034	690/14
<i>diabetes</i>	0,630 ± 0,067	0,727 ± 0,037	0,636 ± 0,059	768/8
<i>breastHess</i>	0,686 ± 0,149	0,827 ± 0,120	0,834 ± 0,133	133/30
<i>bupa</i>	0,616 ± 0,094	0,561 ± 0,066	0,599 ± 0,110	345/6
<i>banknote</i>	0,875 ± 0,032	0,932 ± 0,026	0,929 ± 0,019	1372/4
<i>fertility</i>	0,534 ± 0,235	0,556 ± 0,226	0,585 ± 0,246	100/9
<i>parkinsons</i>	0,790 ± 0,084	0,782 ± 0,101	0,815 ± 0,073	195/22
<i>climate</i>	0,598 ± 0,098	0,731 ± 0,114	0,669 ± 0,130	540/18
<i>ILPD</i>	0,573 ± 0,057	0,600 ± 0,063	0,575 ± 0,052	579/10
<i>german</i>	0,604 ± 0,048	0,658 ± 0,068	0,616 ± 0,043	1000/24
<i>heart</i>	0,748 ± 0,136	0,800 ± 0,090	0,808 ± 0,084	270/13
<i>gobub</i>	0,769 ± 0,156	0,747 ± 0,121	0,744 ± 0,139	72/50
Média do Rank $R(\mathcal{L})$	2,5294	1,8235	1,6471	

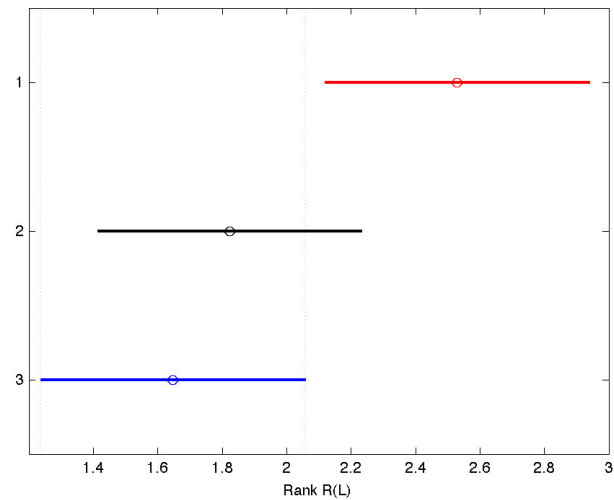


Fig. 9: Comparação entre os três modelos de classificadores através do ranqueamento do teste de *Friedman* e comparação pelo método *Bonferroni*. Em vermelho: classificador sem filtragem. Em preto: classificador com filtragem baseada no grau do vértice. Em azul: classificador com filtragem baseada em $D_C(\cdot)$.

5 Conclusão

Este trabalho apresenta uma abordagem para a identificação de amostras da margem em classificadores de margem larga utilizando apenas informações das relações espaciais entre as amostras. Essas informações são interpretadas através da representação do Grafo de Gabriel do conjunto de dados analisado, uma vez que classificadores eficientes baseados nesse modelo têm sido demonstrados. Mais especificamente, este trabalho apresenta uma métrica de grafo para a leitura dessas informações, investiga o seu comportamento e demonstra a sua utilização na filtragem de amostras que causam sobreposição dos dados (e que interferem no desempenho do classificador).

A métrica de grafo é definida a partir da distância geodésica relativa entre diferentes classes. Essa métrica é utilizada aqui para extrair informações do conjunto de dados para identificação de amostras da margem de separação e, assim, identificar e eliminar amostras que causam a sobreposição.

Os experimentos mostraram duas características dessa métrica de grafo. A primeira é que ela pode apresentar distribuição normal para base de dados reais. Das 17 diferentes bases de dados testadas, 9 apresentaram essa característica, o que representa uma quantidade expressiva. A segunda é que a métrica de grafo sugerida, Equação 2, identifica realmente amostras pertencentes à margem de separação. Como é mostrado neste trabalho, essa métrica pode ser utilizada na filtragem, retirando as amostras que apresentam os menores valores da métrica,

sendo o único filtro estatisticamente mais eficiente que o classificador sem filtro, dentre os testados.

6 Agradecimentos

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

References

- [1] Haykin, S.: Neural networks: A comprehensive foundation. 2^o
- [2] Boser, B.E., Guyon, I.M., Vapnik, V.N.: A training algorithm for optimal margin classifiers. In: Proceedings of the fifth annual workshop on Computational learning theory, ACM (1992) 144–152
- [3] Torres, L., Castro, C., Coelho, F., Torres, F.S., Braga, A.: Distance-based large margin classifier suitable for integrated circuit implementation. *Electronics Letters* **51**(24) (2015) 1967–1969
- [4] Torres, L.C., Coelho, F., Castro, C.L., Braga, A.P.: A graph of gabriel approach for large margin classifiers, LA-CCI-The Latin American Congress on Computational Intelligence Co-located with ARGENCON (2014)
- [5] Torres, L.C., Castro, C.L., Braga, A.P.: A parameterless mixture model for large margin classification. In: Neural Networks (IJCNN), 2015 International Joint Conference on, IEEE (2015) 1–6
- [6] Torres, L.: Classificador por arestas de suporte (CLAS): Métodos de aprendizado baseados em grafos de Gabriel. Tese de doutorado, UFMG (2016)
- [7] Zhang, W., King, I.: A study of the relationship between support vector machine and gabriel graph. In: Neural Networks, 2002. IJCNN'02. Proceedings of the 2002 International Joint Conference on. Volume 1., IEEE (2002) 239–244
- [8] Buckley, F., Harary, F.: Distance in graphs. Addison-Wesley (1990)
- [9] Wilcox, R.: Kolmogorov–smirnov test. *Encyclopedia of biostatistics* (2005)
- [10] Lichman, M.: UCI machine learning repository (2013)
- [11] Alcalá-Fdez, J., Fernández, A., Luengo, J., Derrac, J., García, S., Sánchez, L., Herrera, F.: Keel data-mining software tool: data set repository, integration of algorithms and experimental analysis framework. *Journal of Multiple-Valued Logic & Soft Computing* **17** (2011)
- [12] Golub, T.R., Slonim, D.K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J.P., Coller, H., Loh, M.L., Downing, J.R., Caligiuri, M.A., et al.: Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *science* **286**(5439) (1999) 531–537
- [13] 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.: Pharmacogenomic predictor of sensitivity to preoperative chemotherapy with paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide in breast cancer. *Journal of clinical oncology* **24**(26) (2006) 4236–4244
- [14] Castro, C.L., Braga, A.P.: Novel cost-sensitive approach to improve the multi-layer perceptron performance on imbalanced data. *IEEE transactions on neural networks and learning systems* **24**(6) (2013) 888–899
- [15] Demšar, J.: Statistical comparisons of classifiers over multiple data sets. *Journal of Machine learning research* **7**(Jan) (2006) 1–30