

# A Smoothing Optimization Approach Applied to the Supervised MDS Method

V. Xavier, N. Maculan, and J. Pessanha

**Abstract**—This paper presents an efficient approach to the Supervised MDS method. This method handles the problems of data visualization, supervised classification and bipartite ranking. In order to overcome the non-differentiable nature of the Supervised MDS method, the mathematical formulation proposed in this work is based on the hyperbolic smoothing technique. The performance of the algorithm is evaluated by computational experiments. The results show that the proposed methodology presented, in most cases, better results than the results available in the literature. Furthermore, the methodology presents a good performance in relation to the methods Logistic regression, Naive Bayes and Support Vector Machine.

**Index Terms**—Supervised classification, Bipartite ranking, Nonlinear optimization, Hyperbolic smoothing.

## I. INTRODUCTION

O método Supervised Multidimensional Scaling (SMDS) [1] aborda os problemas de redução de dimensionalidade, classificação supervisionada e *ranking* bipartido. O termo Multidimensional Scaling (MDS) designa uma família de métodos de redução de dimensionalidade, na qual as relações de distância existentes entre as observações em um espaço de alta dimensão são também idealmente preservadas ao máximo em um espaço métrico de mais baixa dimensão.

Os métodos de redução de dimensionalidade são muitas vezes utilizados com o objetivo de visualização de dados, em especial, quando as projeções são feitas para os espaços ou, casos em que é possível visualizar os dados com diagramas de dispersão. O método Supervised MDS pertence a classe de métodos de redução de dimensionalidade, entretanto tem como objetivos principais a classificação supervisionada e o *ranking* bipartido.

Em um cenário típico de classificação supervisionada buscase, a partir de um conjunto de atributos conhecidos, inferir a classe ou categoria de uma nova observação por meio de um modelo previamente ajustado com base em uma amostra ou conjunto de dados para os quais são conhecidos os respectivos atributos e classes. Já no problema geral de *ranking* ou ordenação busca-se no espaço de alta dimensão uma relação de ordem ou de preferência entre as observações. Uma aplicação típica e relevante de tal resultado está presente em sistemas de recomendação e ferramentas de busca e recuperação de informação, onde a ordenação induz uma relação de preferências. O

problema de *ranking* bipartido é um caso especial do problema geral de *ranking* em que as observações possuem duas classes. A informação das classes é utilizada na construção do modelo de aprendizagem fundamentado em uma função que induz o *ranking* nas observações. Trata-se de um problema relacionado com o problema de classificação supervisionada binária. Desse modo, alguns algoritmos de classificação supervisionada podem ser usados no contexto do problema do *ranking* bipartido [2].

As formulações matemáticas intrínsecas ao método Supervised MDS possuem a característica de serem não diferenciáveis. Para cada formulação não diferenciável, o presente trabalho descreve uma formulação matemática alternativa, fundamentada na Suavização Hiperbólica (Hyperbolic Smoothing) e, dessa forma, capaz de superar as dificuldades impostas pela não diferenciabilidade e de permitir a utilização de métodos de otimização poderosos, baseados na aproximação da série de Taylor de primeira ou de segunda ordem. Trata-se da mesma estratégia de suavização empregada no algoritmo HSMDS proposto em Xavier [3] e baseada em uma função diferenciável de classe  $C^\infty$ .

O artigo está organizado em seis seções. A seguir, na seção II, tem-se uma introdução ao método Supervised MDS. Na sequência, a seção III descreve a introdução da Suavização Hiperbólica. A seção IV apresenta a proposta de suavização do método Supervised MDS. Na seção V o desempenho da metodologia proposta é ilustrado por dois tipos de experimentos computacionais. Primeiro são apresentados experimentos propostos por Witten e Tibshirani para uma comparação direta com resultados publicados na literatura [1], em seguida é apresentada uma comparação do desempenho computacional em relação a outros algoritmos tradicionais para os problemas de classificação supervisionada e *ranking* bipartido. Por fim, na seção VI constam as principais conclusões do trabalho.

## II. O MÉTODO SUPERVISED MDS

O método Supervised MDS em uma primeira fase faz uso do conjunto de treinamento, incorporando a informação da classe na projeção das observações no espaço de baixa dimensão. Em uma segunda fase, as observações projetadas permitem classificar observações novas.

Considere uma matriz de distâncias euclidianas,  $D_{n \times n}$ , calculada a partir de um conjunto de  $n$  observações,  $\mathbf{z}_i, i = 1, \dots, n$ , cada uma com  $S$  atributos, i.e., cada observação pertence a um espaço de dimensão  $S$ . O processo de redução de dimensionalidade busca a representação de cada observação em um espaço de dimensão inferior  $d$ , onde  $d < S$ . Desta forma, o método Supervised MDS busca obter um novo conjunto de observações,  $\mathbf{x}_i, i = 1, \dots, n$ , pertencentes ao espaço

V. L. Xavier, Departamento de Estatística, Instituto de Matemática e Estatística, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brasil, viniciuslx@ime.uerj.br.

N. Maculan, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brasil, nelson.maculan@gmail.com.

J. F. M. Pessanha, Departamento de Estatística, Instituto de Matemática e Estatística, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, Brasil, professorjfm@hotmail.com.

de menor dimensão  $d$ , i.e.,  $\mathbf{x}_i \in R^d$ , de modo que gere uma matriz de distâncias, que se aproxime ao máximo da matriz de dissimilaridades  $D_{n \times n}$ , tendo com base a minimização da seguinte função de escalonamento para observações do conjunto de treinamento:

$$f(\mathbf{x}) = \left\{ \begin{array}{l} \frac{1}{2}(1-\alpha) \sum_{i=1}^n \sum_{j=1}^n (D_{ij} - \|\mathbf{x}_i - \mathbf{x}_j\|_2)^2 + \\ \alpha \sum_{i,j:y_j > y_i} (y_j - y_i) \sum_{l=1}^d \left( \frac{D_{ij}}{\sqrt{d}} - (x_{jl} - x_{il}) \right)^2 \end{array} \right\}^2 \quad (1)$$

em que  $y_i$  representa a classificação ou resultado associado à observação  $i$ ,  $y_i \in \{1, 2\}$  e  $\alpha \in [0, 1]$  é um parâmetro de ponderação que balanceia o primeiro termo da função objetivo com o segundo.

O parâmetro  $\alpha$  pode ser ajustado em função do erro de classificação fora da amostra utilizando validação cruzada. Deve ser observado que quando  $\alpha = 0$  a formulação (1) é equivalente à formulação *Least Squares MDS*.

O primeiro termo do somatório corresponde à formulação *Least Squares MDS* [4], enquanto o segundo termo contempla a componente associada às informações das classes. O segundo termo tende a afastar as observações de classes diferentes,  $y_i \neq y_j$ . Note que esse termo só ocorre para os valores nos quais  $y_j > y_i$ . Dado que as distâncias  $D_{ij}$  são sempre positivas, no processo de minimização cada parcela  $(x_{jl} - x_{il})$  tende a ser positiva e tal parcela é positiva, se e somente se,  $x_{jl} > x_{il}$ . Logo, no espaço de dimensão reduzida, as observações na classe 2 tenderão a ter coordenadas maiores do que as das observações na classe 1 em todas as componentes.

Embora a formulação (1) considere a informação da classe ela aborda somente o problema de redução de dimensionalidade e não tem a capacidade de generalização para uma observação nova.

O critério Supervised MDS para a classificação generaliza a formulação (1) através de dois novos procedimentos: a inclusão da observação nova na função objetivo e a utilização do vetor  $\mathbf{x}$ , resultante da minimização da formulação (1), como referência no espaço de baixa dimensão. O vetor  $\mathbf{x}$  corresponde às observações do conjunto de treinamento no espaço reduzido. Pretende-se estimar as coordenadas de uma observação nova no espaço reduzido,  $\mathbf{x}_{n+1}$ , e a classe desta nova observação  $y_{n+1}$ . Assim, o índice  $n+1$  faz referência a uma observação nova genérica e  $D_{n+1}$  representa o vetor de distâncias no espaço de alta dimensão entre cada observação do conjunto de teste e a observação nova.

A generalização da formulação (1) é feita com a suposição sobre a classe da observação nova. Assim, a seguinte formulação considera a hipótese da observação nova pertencer à classe 1, ou seja,  $y_{n+1} = 1$  :

$$h_1(\mathbf{x}_{n+1}) = \left\{ \begin{array}{l} (1-\alpha) \sum_{i=1}^n (D_{i,n+1} - \|\mathbf{x}_i - \mathbf{x}_{n+1}\|_2)^2 + \\ \alpha \sum_{i:y_i=2, 1 \leq i \leq n} \sum_{l=1}^d \left( \frac{D_{i,n+1}}{\sqrt{d}} - (x_{il} - x_{n+1,l}) \right)^2 \end{array} \right\}^2 \quad (2)$$

Adicionalmente, a formulação (3) considera a hipótese de a observação nova pertencer à classe 2. i.e.,  $y_{n+1} = 2$ :

$$h_2(\mathbf{x}_{n+1}) = \left\{ \begin{array}{l} (1-\alpha) \sum_{i=1}^n (D_{i,n+1} - \|\mathbf{x}_i - \mathbf{x}_{n+1}\|_2)^2 + \\ \alpha \sum_{i:y_i=1, 1 \leq i \leq n} \sum_{l=1}^d \left( \frac{D_{i,n+1}}{\sqrt{d}} - (x_{n+1,l} - x_{il}) \right)^2 \end{array} \right\}^2 \quad (3)$$

A classificação da nova observação inicia-se com a minimização das funções  $h_1(\mathbf{x}_{n+1})$  e  $h_2(\mathbf{x}_{n+1})$  em relação a  $\mathbf{x}_{n+1}$  para obter duas alternativas para as coordenadas da observação nova no espaço reduzido. Na sequência, a regra de classificação mais simples consiste em atribuir a observação nova à classe correspondente ao menor valor entre as duas funções minimizadas. Portanto, a classe imputada à nova observação corresponde àquela que produziu o menor resíduo no processo de redução de dimensionalidade:

$$y_{n+1} = \arg \min_{k \in \{1,2\}} \left\{ \min_{\mathbf{x}_{n+1}} h_k(D_{n+1}, \mathbf{x}_{n+1}) \right\} \quad (4)$$

A regra de classificação pode ser representada pela seguinte forma alternativa:

$$\begin{cases} \min_{\mathbf{x}_{n+1} \in \mathbb{R}^d} h_1(\mathbf{x}_{n+1}) - \min_{\mathbf{x}_{n+1} \in \mathbb{R}^d} h_2(\mathbf{x}_{n+1}) < 0 \rightarrow \hat{y}_{n+1} = 1 \\ \text{caso contrário } \hat{y}_{n+1} = 2 \end{cases} \quad (5)$$

A classificação pode ser feita de um modo mais geral, substituindo o valor zero na expressão (5) por um valor fixo,  $v$ , representando um limiar genérico.

#### A. Supervised MDS para Ranking Bipartido

O problema de *ranking* bipartido é um caso especial do problema de *ranking* e é diretamente relacionado com o problema de classificação supervisionada. No problema de *ranking* bipartido busca-se uma função com imagem no espaço unidimensional,  $f : R^S \rightarrow R^1$ . A ordenação dos valores obtidos pela função aplicada nas observações produz um *ranking*. De forma análoga ao problema de classificação, no problema de *ranking* bipartido deseja-se separar as classes através de um *ranking* ou ordenamento. No método Supervised MDS o *ranking* bipartido é construído pela ordenação das difereças:

$$\min_{\mathbf{x}_{n+1}} h_1(D_{n+1}, \mathbf{x}_{n+1}) - \min_{\mathbf{x}_{n+1}} h_2(D_{n+1}, \mathbf{x}_{n+1}) \quad (6)$$

Para qualquer par de observações  $\mathbf{x}_i$  e  $\mathbf{x}_j$ , tem-se a seguinte propriedade desejada para a função  $f$ :  $f(x_i) > f(x_j)$  se  $y_i > y_j$ . O número de pares onde essa propriedade não é violada é um indicador da qualidade do *ranking*. Assim, uma medida de avaliação do *ranking* é o erro do *ranking* bipartido, definido por:

$$\frac{1}{n_1 n_2} \sum_{i:y_i=1} \sum_{j:y_j=2} \left( 1_{f(\mathbf{x}_j) > f(\mathbf{x}_i)} + \frac{1}{2} 1_{f(\mathbf{y}_i) = f(\mathbf{y}_j)} \right) \quad (7)$$

### III. SUAVIZAÇÃO HIPERBÓLICA

Com o objetivo de obter uma alternativa diferenciável para as formulações (1), (2) e (3), o conceito da Suavização Hiperbólica (SH) é apresentada nesta seção. Atualmente existe uma grande diversidade de problemas em que a abordagem da SH tem sido aplicada com sucesso, incluindo problemas difíceis da classe NP-difícil, problemas não diferenciáveis e não convexos com um grande número de mínimos locais, por exemplo, *Multisource Fermat-Weber* [5] [6], *Hub Location* [7] e recobrimento de um corpo sólido [8]. Xavier et Xavier [9] apresentam uma revisão bibliográfica do conjunto de aplicações bem-sucedidas abordadas com a Suavização Hiperbólica. No contexto de métodos de redução de dimensionalidade, Xavier [3] apresenta em sua tese de doutorado uma nova metodologia supervisionada para redução de dimensionalidade baseada em protótipos. No contexto de classificação não supervisionada, o problema de *clustering* é abordado em [10] [11]. Dados dois pontos  $x_i$  e  $x_j$  e a distância euclidiana entre eles,  $u = \|\mathbf{x}_i - \mathbf{x}_j\|_2$ . A distância euclidiana é não diferenciável quando  $x_i = x_j$ . A função  $\theta(u, \gamma)$  usada para suavização é definida por:

$$\theta(u, \gamma) = \sqrt{u^2 + \gamma^2} \quad (8)$$

A função  $\theta(u, \gamma)$  tem as seguintes propriedades [3]:

- (a)  $\theta(u, \gamma) > u, \quad \forall \gamma > 0;$
- (b)  $\lim_{\gamma \rightarrow 0} \theta(u, \gamma) = u;$
- (c)  $\theta(u, \gamma)$  pertence à classe  $C^\infty$  de funções diferenciáveis;
- (d)  $\theta'(u, \gamma) = u/(u^2 + \gamma^2)^{1/2}$
- (e)  $\theta''(u, \gamma) = \gamma^2/(u^2 + \gamma^2)^{3/2}$
- (f)  $\theta''(0, \gamma) = 1/|\gamma|$
- (g)  $\lim_{\gamma \rightarrow 0} \theta''(0, \gamma) = \infty$

Note que a propriedade (b) implica que a função  $\theta(u, \gamma)$  é uma aproximação assintótica da distância euclidiana. A propriedade (c) permite a utilização de métodos de otimização poderosos baseados em aproximação da série de Taylor de primeira ou de segunda ordem. Devido à propriedade (e), a curvatura da função  $\theta(u, \gamma)$  é crescentemente atenuada na medida em que se aumenta o parâmetro  $\gamma$ . Pela propriedade (f), a curvatura assume o valor máximo quando  $u = 0$ . De forma similar, pelas propriedades (b) e (g), a curvatura da função  $\theta(u, \gamma)$  no ponto  $u = 0$  tende a infinito na medida em que  $\theta(u, \gamma)$  se aproxima de  $u$ .

### IV. HIPERBOLIC SMOOTHING SUPERVISED MDS

Voltando a atenção para as formulações expressas pelas funções (1), (2) e (3), nota-se que a não diferenciabilidade é uma propriedade comum nessas três formulações. Tem-se como proposta aplicar SH e substituir essas funções por um conjunto de funções diferenciáveis. Com este propósito em mente, problemas intrinsecamente não diferenciáveis são aproximados por alternativas completamente diferenciáveis. Por conseguinte, a função Supervised MDS (1) será substituída por:

$$\tilde{f}(\mathbf{x}, \gamma) = \left\{ \begin{array}{l} \frac{1}{2}(1 - \alpha) \sum_{i=1}^n \sum_{j=1}^n (D_{ij} - \theta(\|\mathbf{x}_i - \mathbf{x}_j\|_2, \gamma))^2 + \\ \alpha \sum_{i,j: y_j > y_i} (y_j - y_i) \sum_{l=1}^d \left( \frac{D_{ij}}{\sqrt{d}} - (x_{jl} - x_{il}) \right)^2 \end{array} \right\} \quad (9)$$

As funções (2) e (3) são substituídas respectivamente por:

$$\tilde{h}_1(\mathbf{x}_{n+1}, \gamma) = \left\{ \begin{array}{l} (1 - \alpha) \sum_{i=1}^n (D_{i,n+1} - \theta(\|\mathbf{x}_i - \mathbf{x}_j\|_2, \gamma))^2 + \\ \alpha \sum_{i: y_i=2, 1 \leq i \leq n} \sum_{l=1}^d \left( \frac{D_{i,n+1}}{\sqrt{d}} - (x_{il} - x_{n+1,l}) \right)^2 \end{array} \right\} \quad (10)$$

$$\tilde{h}_2(\mathbf{x}_{n+1}, \gamma) = \left\{ \begin{array}{l} (1 - \alpha) \sum_{i=1}^n (D_{i,n+1} - \theta(\|\mathbf{x}_i - \mathbf{x}_j\|_2, \gamma))^2 + \\ \alpha \sum_{i: y_i=1, 1 \leq i \leq n} \sum_{l=1}^d \left( \frac{D_{i,n+1}}{\sqrt{d}} - (x_{n+1,l} - x_{il}) \right)^2 \end{array} \right\} \quad (11)$$

Para resolver a formulação suavizada (9) (10) e (11), [3] propôs o algoritmo HSMDS. O algoritmo busca resolver uma sequência de problemas diferenciáveis, onde o parâmetro de suavização é gradualmente reduzido por um fator de redução. Assim, é gerada uma sequência de problemas suavizados que se aproximam gradativamente dos problemas originais (1) (2) e (3). Cada problema nessa sequência de aproximação é completamente diferenciável, permitindo assim a aplicação de métodos eficientes e robustos de otimização sem restrições, tais como o Gradiente Conjugado e os métodos da família Quasi-Newton.

O programa R [12] possui um conjunto de funções para otimização não linear, sendo utilizado com sucesso em diversas aplicações reais, por exemplo, no problema de alocação ótima de amostras [13]. A metodologia proposta foi implementada utilizando a linguagem estatística R. As tarefas de otimização foram realizadas por meio do método de Gradiente Conjugado implementado na rotina `optim` da biblioteca `stats`. Em relação à regra de parada, em cada problema irrestrito utilizou-se o máximo de 150 iterações do método Gradiente Conjugado.

As observações do conjunto de treinamento são projetadas no espaço reduzido considerando a minimização da equação (9). A redução de dimensionalidade de cada observação do conjunto de teste é independente uma das outras. As coordenadas destas observações no espaço de baixa dimensão são obtidas através da minimização das funções suavizadas (10) e (11). Por serem processos de otimização independentes, as observações são projetadas utilizando o recurso de computação paralela em CPU, por meio da biblioteca `doSNOW` [14].

### V. RESULTADOS E DISCUSSÃO

#### A. Comparação com o Método Supervised MDS

Witten e Tibshirani [1] apresentam duas tabelas de resultados computacionais obtidos em dois experimentos

independentes para dois objetivos distintos: classificação supervisionada e *ranking* bipartido. Com o objetivo de validação e comparação da metodologia proposta, alguns experimentos foram realizados seguindo exatamente os mesmos critérios descritos em [1]. Os dados sintéticos utilizados nos experimentos foram gerados pelos seguintes modelos:

1- Modelo *Constant*, assim chamado porque há uma média constante para as observações em cada classe.

$$\mathbf{z}_i \sim \begin{cases} N(\mathbf{0}, \mathbf{I}_{SXS}) & \text{se a observação } i \text{ for da classe 1} \\ N(-\mathbf{0}, \mathbf{I}_{SXS}) & \text{se a observação } i \text{ for da classe 2} \end{cases}$$

$\mathbf{I}_{SXS}$ , matriz de covariância tem a forma da matriz identidade, e vetor de média com todas as componentes iguais.

2- Modelo *two-sided*, assim chamado porque as observações da classe 2 formam dois grupos distintos.

$$\mathbf{z}_i \sim \begin{cases} N(\mathbf{0}, \mathbf{I}_{SXS}) & \text{se a observação } i \text{ for da classe 1} \\ N(-1, \mathbf{I}_{SXS}) \text{ ou } N(1, \mathbf{I}_{SXS}) & \text{com igual probabilidade} \\ & \text{se a observação } i \text{ for da classe 2} \end{cases}$$

3- Modelo linear, assim chamado porque há uma tendência linear nas observações, em função do índice da observação

$$\mathbf{z}_i \sim N\left(\frac{3i}{n}, \mathbf{I}_{SXS}\right)$$

As primeiras observações são da classe 1 e as seguintes da classe 2. A observação  $i$  tem vetor de média com todas as componentes iguais a  $3i/n$ . Novamente,  $n$  representa o número de observações,  $\mathbf{z}_i$ ,  $i = 1, \dots, n$ , cada uma com  $S$  atributos.

Os resultados das Tabelas I e II, foram gerados com os conjuntos de treinamento e de teste de tamanhos iguais. Todos os parâmetros de ajuste foram selecionados utilizando validação cruzada. Os erros de classificação e do *ranking* bipartido são calculados no conjunto de teste, com base em 50 simulações de conjuntos de treinamento e de conjuntos de teste. Os conjuntos de teste foram gerados com o mesmo modelo e com o mesmo número de observações do conjunto de treinamento.

Nas primeiras três colunas da Tabela I constam o modelo de geração da base de dados sintéticos utilizado, a dimensão do espaço e o número de observações. Nas quatro colunas subsequentes constam os erros de classificação para o método Supervised MDS, primeiro com o parâmetro  $\alpha = 1$ , para SMDS e a versão suavizada HS SMDS, em seguida tem-se os resultados obtidos por validação cruzada (*Cross Validation*, CV). Estão destacados em negrito o melhor resultado obtido considerando  $\alpha = 1$  e o melhor resultado considerando validação cruzada.

Os resultados obtidos pela metodologia proposta correspondem as colunas HS SMDS e HS SMDS CV. As demais colunas são exatamente iguais àquelas apresentadas por Witten e Tibshirani [1]. Os resultados obtidos mostram que o método proposto obteve um erro médio menor em 10 dos 12 casos analisados.

De modo análogo ao apresentado para o problema de classificação, é feita a apresentação dos resultados para o

TABELA I  
CLASSIFICAÇÃO SUPERVISIONADA

Modelo	S	n	SMDS	HS SMDS	SMDS	HS SMDS
			$\alpha = 1$	$\alpha = 1$	CV	CV
Two-sided	5	20	0,264	<b>0,2460</b>	0,295	<b>0,2800</b>
Two-sided	5	50	0,2288	<b>0,2088</b>	0,2432	<b>0,2328</b>
Two-sided	15	20	0,089	<b>0,0580</b>	0,115	<b>0,1060</b>
Two-sided	15	50	0,0664	<b>0,0480</b>	<b>0,0664</b>	0,0736
Linear	5	20	0,131	<b>0,1300</b>	0,142	<b>0,1220</b>
Linear	5	50	0,1324	<b>0,1232</b>	0,13	<b>0,1184</b>
Linear	15	20	0,115	<b>0,0700</b>	0,107	<b>0,0540</b>
Linear	15	50	<b>0,0752</b>	0,0824	0,0744	<b>0,0728</b>
Constant	5	20	0,289	<b>0,2580</b>	0,264	<b>0,2120</b>
Constant	5	50	<b>0,2172</b>	0,2216	<b>0,2024</b>	0,2048
Constant	15	20	0,249	<b>0,1980</b>	0,144	<b>0,0820</b>
Constant	15	50	0,1472	<b>0,1160</b>	0,0996	<b>0,0688</b>

TABELA II  
RANKING BIPARTIDO

Modelo	S	n	SMDS	HS SMDS	SMDS	HS SMDS
			$\alpha = 1$	$\alpha = 1$	CV	CV
Two-sided	5	20	0,2454	<b>0,1778</b>	0,2788	<b>0,2084</b>
Two-sided	5	50	0,193	<b>0,1534</b>	0,2069	<b>0,1790</b>
Two-sided	15	20	0,1162	<b>0,0128</b>	0,1462	<b>0,0386</b>
Two-sided	15	50	0,0646	<b>0,0116</b>	0,0646	<b>0,0256</b>
Linear	5	20	<b>0,058</b>	0,0588	0,0736	<b>0,045</b>
Linear	5	50	0,0565	<b>0,0499</b>	0,0538	<b>0,0485</b>
Linear	15	20	0,0296	<b>0,0182</b>	0,027	<b>0,0136</b>
Linear	15	50	0,0173	<b>0,0172</b>	0,0148	<b>0,0146</b>
Constant	5	20	0,203	<b>0,1868</b>	0,1904	<b>0,1574</b>
Constant	5	50	0,1593	<b>0,1267</b>	0,1432	<b>0,1104</b>
Constant	15	20	<b>0,1306</b>	0,1354	<b>0,048</b>	0,0496
Constant	15	50	0,0669	<b>0,0600</b>	0,0456	<b>0,0286</b>

problema de *ranking* bipartido. A Tabela II exibe os erros de *ranking* bipartido obtidos pela metodologia proposta (HS SMDS e HS SMDS CV) e os apresentados por Witten e Tibshirani [1].

Uma análise comparativa revela que o método proposto HS SMDS, considerando  $\alpha = 1$ , obteve o menor erro médio em 10 dos 12 casos analisados. Considerando os resultados com validação cruzada o método proposto HS SMDS obteve o menor erro médio em 11 dos 12 casos analisados.

### B. Comparação de Métodos de Classificação Supervisionada e Ranking Bipartido

Adicionalmente aos experimentos com dados sintéticos, realizou-se outro conjunto de experimentos usando dados disponíveis no UCI Machine Learning Repository [15] e dados de microarranjos de DNA, *microarrays*, disponíveis em [16], a seguir relacionados:

O conjunto de dados *SPECTF Heart Data Set* (SPECTF.train) [15] é composto por dados de imagens cardíacas de tomografia computadorizada. Cada paciente é classificado em duas categorias: normal e anormal. O conjunto tem 80 observações com 44 atributos,  $R^{44}$ .

O conjunto de dados Parkinsons [17] (*parkinsons.data*) é composto por uma série de medições de voz de 31 pessoas, sendo 23 com a doença de Parkinson. O conjunto tem 195 observações com 22 atributos,  $R^{22}$ .

O conjunto de dados *Connectionist Bench Sonar, Mines vs. Rocks* (*sonar.alldata*) [15] é composto por dados de sinais de sonar. O conjunto tem 111 observações provenientes da aplicação do sonar em um cilindro de metal simulando uma mina terrestre e 97 observações provenientes da aplicação do sonar sobre rochas. Cada observação com 60 atributos,  $R^{60}$ .

O conjunto Alon [18] é composto de dados de microarranjos de DNA do tecido do cólon, com 40 pacientes com câncer e 22 normais, cada amostra com 2000 atributos,  $R^{2000}$ .

O conjunto Golub [19] é composto de dados de microarranjos de DNA de 25 pacientes com leucemia mieloide aguda e 47 com leucemia linfoblástica aguda, cada amostra com 7129 atributos,  $R^{7129}$ .

Como o intuito de avaliar o desempenho do método proposta em relação a outros métodos consagrados pela literatura foram selecionados os seguintes métodos: Regressão Logística (RL), Naive Bayes (NB) e Support Vector Machine (SVM), com kernel linear e polinomial de grau dois. Para cada método foi utilizada a respectiva biblioteca: *klaR* [20], *stats* [12] e *e1071* [21]. Para o ajuste dos parâmetros do Kernel foi utilizada a biblioteca *kernelab* [22].

O método de validação cruzada *k-fold*, com valor de *k* igual a 10, foi utilizado para avaliar o desempenho de cada algoritmo. Para eliminar a diferença entre as amostras dos dados na validação cruzada todos os métodos foram testados com as mesmas *k* amostras. Além disso, a validação cruzada foi feita com balanceamento da proporção de cada classe nas *k* amostras.

Na Tabela III consta o erro de classificação para cada um dos métodos selecionados. Destaca-se que o método proposto obteve o menor valor de erro em quatro dos cinco conjunto de dados testados, sendo que no conjunto Golub o método SVM com Kernel linear e polinomial também obteve o menor valor de erro. Na Tabela IV consta o erro de *ranking* bipartido para cada um dos métodos selecionados. Destaca-se que o método proposto obteve um menor valor de erro em dois dos cinco conjunto de dados testados.

No método Supervised MDS o número de variáveis do problema de otimização para a redução de dimensionalidade é dado por  $nd$ , onde  $n$  é o número de observações e  $d$  é a dimensão do espaço em que as observações são projetadas. Sendo assim, o número de variáveis não depende do número  $S$  de atributos de entrada. Dessa forma o algoritmo proposto pode ser aplicado na classificação e no *ranking* bipartido em problemas com alta dimensão e poucas amostras, característica comum em dados de microarrays e presente nos conjuntos Alon e Golub.

## VI. CONCLUSÕES

Comparado com outros métodos tradicionais para os problemas de classificação supervisionada e *ranking* bipartido o método Supervised MDS produz bons resultados e em alguns casos resultados até superiores, sendo assim pode ser considerado um método muito eficaz e promissor.

TABELA III  
CLASSIFICAÇÃO

Dados	S	n	HS SMDS	RL	NB	SVM	
						Linear	Polinomial
Specfict	44	80	<b>0.1750</b>	0.3625	0.2750	0.2750	0.2750
Parkson	22	195	0.1589	0.1490	0.2648	<b>0.1180</b>	<b>0.1180</b>
Sonar	60	208	<b>0.2500</b>	0.3211	0.2539	0.2820	0.2820
Alon	2000	66	<b>0.1290</b>	0.5476	0.1452	0.1785	0.1785
Golub	7129	72	<b>0.0276</b>	0.4009	0.2923	<b>0.0276</b>	<b>0.0276</b>

TABELA IV  
RANKING BIPARTIDO

Dados	S	n	HS SMDS	RL	NB	SVM	
						Linear	Polinomial
Specfict	44	80	<b>0.1687</b>	0.2250	0.2250	<b>0.1687</b>	<b>0.1687</b>
Parkson	22	195	0.1186	0.1154	0.1154	<b>0.1125</b>	<b>0.1125</b>
Sonar	60	208	0.1944	<b>0.1420</b>	<b>0.1420</b>	0.1896	0.1896
Alon	2000	66	<b>0.0958</b>	0.3708	0.3708	0.1916	0.1916
Golub	7129	72	0.0280	0.2097	0.2097	<b>0.0120</b>	<b>0.0120</b>

Neste artigo é proposta uma nova abordagem para os problemas de redução de dimensionalidade, classificação supervisionada e *ranking* bipartido tendo como base a utilização da Suavização Hiperbólica nas formulações (1), (2) e (3).

A minimização das distorções entre a projeção no espaço reduzido e no espaço original dos dados não é uma tarefa trivial, tendo a abordagem metodológica proposta se mostrado eficaz na busca de um mínimo local profundo. Com base no conjunto dos resultados obtidos, pode-se verificar a confiabilidade e a eficácia da metodologia proposta para ambos os problemas de classificação supervisionada e *ranking* bipartido.

Em comparação com resultados publicados na literatura, o algoritmo HS SMDS apresentou um desempenho bem superior na maioria das comparações, tanto com validação cruzada e também com parâmetro fixo. Além disso, o algoritmo HS SMDS também apresentou um resultado bom comparativamente com os métodos Regressão Logística, Naive Bayes e SVM com kernel linear e polinomial. Assim, pode-se cogitar que a proposta apresentada neste artigo seja uma boa alternativa frente a apresentada em [1]. Além disso, a proposta apresentada tem como diferencial a utilização de processamento paralelo nas projeções das observações do conjunto de teste.

Tendo em vista uma ampla difusão da metodologia proposta, está sendo desenvolvida uma biblioteca no software livre R.

## REFERÊNCIAS

- [1] D. M. Witten and R. Tibshirani, "Supervised multidimensional scaling for visualization, classification, and bipartite ranking," *Computational Statistics & Data Analysis*, vol. 55, no. 1, pp. 789–801, 2011.
- [2] S. Agarwal, "A study of the bipartite ranking problem in machine learning," Tech. Rep., 2005.
- [3] V. L. Xavier, "Uma abordagem eficiente para métodos não lineares de redução de dimensionalidade e uma nova metodologia supervisionada para redução de dimensionalidade baseada em protótipos," Ph.D. dissertation, Universidade Federal do Rio de Janeiro, 2016.
- [4] A. J. Kearsley, R. A. Tapia, and M. W. Trosset, "The solution of the metric stress and sstress problems in multidimensional scaling using newton's method," Tech. Rep., 1994.

- [5] V. L. Xavier, F. M. França, A. E. Xavier, and P. M. Lima, "A hyperbolic smoothing approach to the multisource weber problem," *Journal of Global Optimization*, vol. 60, no. 1, pp. 49–58, 2014.
- [6] V. L. Xavier and A. E. Xavier, "Accelerated hyperbolic smoothing method for solving the multisource fermat weber and k-median problems," *Knowledge-Based Systems*, 2019.
- [7] A. E. Xavier, C. M. Gesteira, and V. L. Xavier, "Solving the continuous multiple allocation p-hub median problem by the hyperbolic smoothing approach," *Optimization*, vol. 64, no. 12, pp. 2631–2647, 2015.
- [8] D. C. Lubke, V. L. Xavier, H. M. Venceslau, and A. E. Xavier, "Flying elephants method applied to the problem of covering solid bodies with spheres," *International Journal of Metaheuristics*, vol. 7, no. 1, pp. 30–42, 2018.
- [9] A. E. Xavier and V. L. Xavier, "Flying elephants: a general method for solving non-differentiable problems," *Journal of Heuristics*, vol. 22, no. 4, pp. 649–664, 2016.
- [10] V. L. Xavier, "Resolução do problema de agrupamento segundo o critério de minimização da soma de distâncias," Master's thesis, M. Sc. Thesis—COPPE—UFRJ, Rio de Janeiro, 2012.
- [11] A. E. Xavier and V. L. Xavier, "Solving the minimum sum-of-squares clustering problem by hyperbolic smoothing and partition into boundary and gravitational regions," *Pattern Recognition*, vol. 44, no. 1, pp. 70 – 77, 2011. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0031320310003432>
- [12] R. C. Team, "R: A language and environment for statistical computing, r foundation for statistical computing, austria, 2015," 2018.
- [13] J. A. M. Brito and A. E. Xavier, "A nonlinear optimization algorithm applied to optimal allocation problem," *IEEE Latin America Transactions*, vol. 14, no. 12, pp. 4762–4767, 2016.
- [14] R. Analytics and S. Weston, "dosnow: Foreach parallel adaptor for the snow package," *R package version*, vol. 1, no. 3, 2009.
- [15] D. Dua and C. Graff, "UCI machine learning repository," 2017. [Online]. Available: <http://archive.ics.uci.edu/ml>
- [16] C. A. Jeffery IB, Higgins DG, "Higgins laboratory: Comparison and evaluation of microarray feature selection methods," <http://www.bioinf.ucd.ie/people/ian/>.
- [17] M. A. Little, P. E. McSharry, S. J. Roberts, D. A. Costello, and I. M. Moroz, "Exploiting nonlinear recurrence and fractal scaling properties for voice disorder detection," *Biomedical engineering online*, vol. 6, no. 1, p. 23, 2007.
- [18] U. Alon, N. Barkai, D. A. Notterman, K. Gish, S. Ybarra, D. Mack, and A. J. Levine, "Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays," *Proceedings of the National Academy of Sciences*, vol. 96, no. 12, pp. 6745–6750, 1999.
- [19] T. R. Golub, D. K. Slonim, P. Tamayo, C. Huard, M. Gaasenbeek, J. P. Mesirov, H. Coller, M. L. Loh, J. R. Downing, M. A. Caligiuri *et al.*, "Molecular classification of cancer: class discovery and class prediction by gene expression monitoring," *science*, vol. 286, no. 5439, pp. 531–537, 1999.
- [20] C. Roever, N. Raabe, K. Luebke, U. Ligges, G. Szepannek, M. Zentgraf, M. U. Ligges, and S. SVMlight, "The klar package," *Department of Statistics, University of Dortmund*, 2006.
- [21] E. Dimitriadou, K. Hornik, F. Leisch, D. Meyer, and A. Weingessel, "Misc functions of the department of statistics (e1071), tu wien," *R package*, vol. 1, pp. 5–24, 2008.
- [22] A. Karatzoglou, A. Smola, K. Hornik, and A. Zeileis, "kernlab-an s4 package for kernel methods in r," *Journal of statistical software*, vol. 11, no. 9, pp. 1–20, 2004.



**Nelson Maculan** Possui graduação em Engenharia de Minas e Metalurgia pela Universidade Federal de Ouro Preto (1965), mestrado (D.E.A.) Matemática Estatística - Université de Paris VI (Pierre et Marie Curie) (1967), doutorado em Engenharia de Produção pela Universidade Federal do Rio de Janeiro (1975) e Diplôme d'Habilitation à Diriger des Recherches (H.D.R) en Sciences de la Gestion (1988), Université Paris-Dauphine (Paris IX). Atualmente é Professor Emérito da Universidade Federal do Rio de Janeiro. Tem experiência na área de Ciência da Computação, com ênfase em Matemática da Computação, atuando principalmente nos seguintes temas: otimização combinatória, programação inteira, programação linear, geração de colunas e otimização global.



**José Francisco Moreira Pessanha** é bacharel em Estatística (1992) pela Escola Nacional de Ciências Estatísticas (ENCE) e engenheiro electricista (1994) formado pela Universidade do Estado do Rio de Janeiro (UERJ). Obteve o grau de M.Sc. em Engenharia Elétrica pela Universidade Federal do Rio de Janeiro (COPPE/UFRJ) em 1999 e D.Sc. pela Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio) em 2006. Dr. Pessanha é pesquisador do Centro de Pesquisas de Energia Elétrica (CEPEL) e professor adjunto do curso de graduação em estatística do Instituto de Matemática e Estatística da UERJ. Em 2016 realizou pós doutorado no INESC TEC, Porto. Tem experiência na aplicação de métodos estatísticos e de otimização em sistemas elétricos de potência, em especial em previsão de carga, previsão probabilística da geração eólica, confiabilidade de sistemas elétricos e tarifação de sistemas de distribuição.



**Vinicius Layter Xavier** Possui graduação em Estatística (2008) pela Universidade do Estado do Rio de Janeiro e Mestrado (2012) e Doutorado (2016) em Engenharia de Sistemas pela Universidade Federal do Rio de Janeiro. Atualmente é professor do Departamento de Estatística da Universidade do Estado do Rio de Janeiro. Possui experiência em record linkage, clustering, classificação supervisionada, otimização e métodos de redução de dimensionalidade.