

Um novo método baseado em protótipos para seleção de pontos de referência em máquinas de aprendizado mínimo

José A. V. Florêncio, Madson L. D. Dias e Ajalmar R. da Rocha Neto

Departamento de Teleinformática, Instituto Federal do Ceará
Av. Treze de Maio, 2081, Benfica CEP: 60040-215, Fortaleza, Ceará, Brazil
{jose.florencio,madson.dias}@ppgcc.ifce.edu.br e ajalmar@ifce.edu.br

Resumo Este trabalho introduz uma nova proposta de seleção de pontos de referência (*reference points*, RPs) para máquinas de aprendizado mínimo (*Minimal Learning Machine*, MLM) aplicadas a problemas de classificação com a utilização do algoritmo *k*-médias. O procedimento básico de treinamento do MLM está relacionado a seleção de um subconjunto de padrões de treinamento, chamados de pontos de referência. Tal subconjunto é utilizado para a realização de um mapeamento entre as configurações geométricas das amostras (pontos de entrada) e seus rótulos (pontos de saída) correspondentes. A proposição original sugere que a seleção dos pontos de referência seja feita de maneira aleatória, deixando a cargo do utilizador apenas a definição do número de pontos a serem utilizados. O método de busca em grade, combinado com validação cruzada pode ser utilizados para seleção automática da quantidade de pontos. Porém, esse tipo de método de seleção pode acarretar em uma escolha inadequada dos pontos de referência, fazendo com que amostras ruidosas ou mesmo discrepantes componham tal subconjunto. A fim de superar o problema na escolha de pontos de referência, o método *k*-médias assume um papel relevante na seleção de pontos de referência de MLM para problemas de classificação. A ideia principal é remover do conjunto de pontos de referência, amostras discrepantes, padrões não-relevantes ou aqueles que podem estar corrompidos com o ruído e, assim, permitir que o classificador possa alcançar acurácias mais elevadas. A proposta foi validada com a realização de experimentos com conjuntos de dados do mundo real. A técnica foi capaz de reduzir significativamente o número de pontos de referência, bem como manter a capacidade de generalização equivalente ou superior a versão convencional de seleção de pontos de referência, quando comparada a abordagens iniciais.

Palavras-chave: Aprendizagem de máquina, Máquinas de aprendizado mínimo, Pontos de Referência, *k*-médias.

1 Introdução

Máquina de aprendizado mínimo (*Minimal Learning Machine*, MLM) é uma proposta recente de um algoritmo de aprendizado supervisionado, que pode ser aplicado a problemas de classificação e regressão [11]. Esta técnica vem sendo aplicada a uma diversa gama de problemas do mundo real, tais como a detecção de falhas em motores [2] e o ranqueamento de documentos [1].

O funcionamento do MLM se baseia na presunção da existência de um mapeamento entre as configurações geométricas de pontos do espaço de entrada e as configurações geométricas dos respectivos pontos no espaço de saída. Tal mapeamento é representado por duas matrizes de distâncias (entrada e saída), computadas em relação a todos os pontos do conjunto de treinamento e um subconjunto do mesmo, chamados de pontos de referência (*reference points*, RPs). O aprendizado do MLM é realizado através da determinação de um modelo de regressão linear entre as duas matrizes de distância. Dessa forma, dado um ponto no espaço de entrada, pode-se calcular a localização deste ponto no espaço de saída, através do modelo de regressão aprendido [11].

A formulação original do algoritmo de treinamento de uma máquina de aprendizado mínimo estabelece que a escolha dos pontos de referência seja realizada de maneira aleatória, deixando a cargo do utilizador apenas a definição do número de pontos utilizados [11]. Dessa forma, a escolha do número de pontos necessários é realizada via tentativa e erro.

Métodos de agrupamento, tais como o k -médias [6], são técnicas de aprendizado não-supervisionado que visam de separar objetos em grupos, baseando-se nas características que esses objetos possuem. A ideia básica consiste em reunir em um mesmo grupo os objetos com características similares. A similaridade destes objetos é definida de acordo com critérios pré-estabelecidos.

A utilização do k -médias em conjunto com métodos de classificação de padrões se dá, principalmente na geração de conjuntos reduzidos de bases de dados com grande número de amostras. Além disso, alguns trabalhos utilizam esta técnica e variações dela em diversos tipos de aplicações, como na classificação risco de epilepsia [10], seleção de características [14] e segmentação de imagens [15].

Neste trabalho, propõe-se a utilização do algoritmo k -médias para a seleção de pontos de referência de Máquinas de Aprendizado Mínimo aplicadas a problemas de classificação de padrões. Para validação da proposta foram realizadas simulações com conjuntos de dados do mundo real. A proposta foi capaz de reduzir significativamente o número de pontos de referência, bem como manter a capacidade de generalização equivalente ou superior, quando comparada a abordagens propostas no artigo original [11].

O restante do artigo está organizado da seguinte forma. Na Seção 2 são apresentados os fundamentos do MLM. O método k -médias é descrito na Seção 4, enquanto a proposta de seleção de pontos de referência é apresentada na Seção 5. Em seguida, na Seção 6 são apresentados os resultados das simulações computacionais e, por fim, as conclusões são relatadas na Seção 7.

2 Máquinas de aprendizado mínimo

Considere um conjunto de dados com N , composto por um conjunto de padrões de entrada $X = \{\mathbf{x}_n\}_{n=1}^N$ e suas respectivas saídas $Y = \{\mathbf{y}_n\}_{n=1}^N$, de tal maneira que $\mathbf{x}_n \in \mathcal{X} = \mathbb{R}^D$ é o n -ésimo padrão de entrada e $\mathbf{y}_n \in \mathcal{Y} = \mathbb{R}^S$ é sua respectiva saída. Além disso, assume-se que \mathcal{X} e \mathcal{Y} são os espaços de entrada e saída, respectivamente. Considera-se também a existência de um mapeamento $f: \mathcal{X} \rightarrow \mathcal{Y}$ entre o espaço de entrada e o espaço de saída, de tal forma que

$$\mathbf{Y} = f(\mathbf{X}) + \mathbf{E}, \quad (1)$$

em que as colunas da matriz \mathbf{X} e \mathbf{Y} correspondem a D entradas e S saídas, respectivamente. Além disso, o número de linha das duas matrizes corresponde ao número de observações N e a matriz \mathbf{E} corresponde a matriz de resíduos.

Dito isso, a formulação de uma máquina de aprendizado mínimo (*Minimal Learning Machine*, MLM) pode ser dividida em duas etapas: regressão de distâncias e estimação da saída.

2.1 Regressão entre distâncias

O primeiro passo é a seleção de um subconjunto de dados, conhecido como conjunto de *pontos de referência*, retirado a partir do conjunto total de dados. As entradas deste subconjunto de dados é representado por $R = \{\mathbf{m}_k\}_{k=1}^K$ e suas respectivas saídas são representadas por $T = \{\mathbf{t}_k\}_{k=1}^K$, em que $R \subseteq X$ e $T \subseteq Y$.

É definido $\mathbf{D}_x \in \mathbb{R}^{N \times K}$ tal que a k -ésima coluna $\mathbf{d}(X, \mathbf{m}_k)$ contenha as distâncias $d(\mathbf{x}_n, \mathbf{m}_k)$ entre os N pontos de entrada \mathbf{x}_n e o k -ésimo ponto de referência \mathbf{m}_k .

Da mesma forma que \mathbf{D}_x , é definida a matriz $\mathbf{\Delta}_y \in \mathbb{R}^{N \times K}$ tal que a k -ésima coluna $\mathbf{\delta}(Y, \mathbf{t}_k)$ contem as distâncias $\delta(\mathbf{y}_n, \mathbf{t}_k)$ entre os N pontos de saída \mathbf{y}_n e a saída \mathbf{t}_k do k -ésimo ponto de referência.

Assumindo que o mapeamento entre a matriz de distâncias de entrada \mathbf{D}_x e a matriz correspondente de saída $\mathbf{\Delta}_y$ é linear, o modelo de regressão multiresposta pode ser escrito como:

$$\mathbf{\Delta}_y = \mathbf{D}_x \mathbf{B} + \mathbf{E}. \quad (2)$$

Quando o número de pontos de referência selecionados é menor que o número de pontos disponíveis, a matriz \mathbf{B} pode ser aproximada por:

$$\hat{\mathbf{B}} = (\mathbf{D}'_x \mathbf{D}_x)^{-1} \mathbf{D}'_x \mathbf{\Delta}_y. \quad (3)$$

Para um padrão de teste $\mathbf{x} \in \mathbb{R}^D$, cujas distâncias dos padrões de referência estão disponíveis no vetor $\mathbf{d}(\mathbf{x}, R) = [d(\mathbf{x}, \mathbf{m}_1) \dots d(\mathbf{x}, \mathbf{m}_K)]$, as distâncias estimadas correspondentes entre um padrão desconhecido \mathbf{y} e as saídas conhecidas $\{\mathbf{t}_k\}_{k=1}^K$ dos pontos de referência são

$$\hat{\mathbf{\delta}}(\mathbf{y}, T) = \mathbf{d}(\mathbf{x}, R) \hat{\mathbf{B}}. \quad (4)$$

O vetor $\hat{\mathbf{\delta}}(\mathbf{y}, T) = [\hat{\delta}(\mathbf{y}, \mathbf{t}_1) \dots \hat{\delta}(\mathbf{y}, \mathbf{t}_K)]$ fornece uma estimativa da configuração geométrica de \mathbf{y} e o subconjunto T , no espaço- \mathcal{Y} .

2.2 Estimação da saída

O problema de estimar a saída \mathbf{y} , dadas as saídas $\{\mathbf{t}_k\}_{k=1}^K$ de todos os pontos de referência e uma estimativa $\hat{\delta}(\mathbf{y}, T)$ das suas distâncias mútuas, pode ser entendido como um problema de mutileração [9] para estimar a posição no espaço- \mathcal{Y} .

Várias estratégias podem ser usadas para solucionar o problema de mutileração. Do ponto de vista geométrico, encontrar $\mathbf{y} \in \mathbb{R}^S$ é equivalente a solucionar o conjunto sobredeterminado de K equações não-lineares correspondentes a hiper-esferas S -dimensionais, centradas em \mathbf{t}_k e passando através de \mathbf{y} . Na Figura 1 é graficamente representado o problema para $S = 2$.

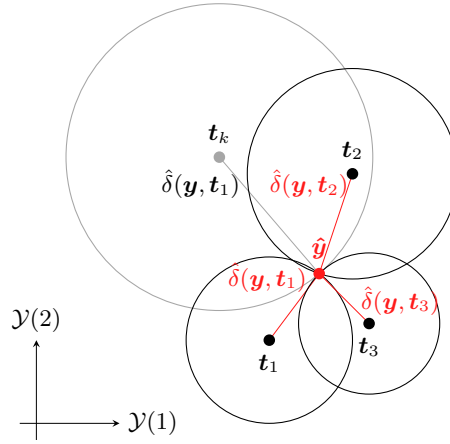


Figura 1. Estimação da saída ($S = 2$).

Dado o conjunto de $k = 1, \dots, K$ hiper-esferas, cada uma delas com raios iguais a $\hat{\delta}(\mathbf{y}, \mathbf{t}_k)$

$$(\mathbf{y} - \mathbf{t}_k)'(\mathbf{y} - \mathbf{t}_k) = \hat{\delta}^2(\mathbf{y}, \mathbf{t}_k), \quad (5)$$

a localização de \mathbf{y} é estimada a partir da minimização da função objetivo

$$J(\mathbf{y}) = \sum_{k=1}^K \left((\mathbf{y} - \mathbf{t}_k)'(\mathbf{y} - \mathbf{t}_k) - \hat{\delta}^2(\mathbf{y}, \mathbf{t}_k) \right)^2. \quad (6)$$

Devido à incerteza introduzida pelas estimativas $\hat{\delta}(\mathbf{y}, \mathbf{t}_k)$, uma solução ótima para a Equação (6) pode ser conseguida através de um método genérico de otimização.

A proposição original do MLM, utiliza o método de Levenberg-Marquardt [7] para solucionar o problema de otimização presente na Equação 6 [11].

2.3 MLM aplicado a problemas de classificação

Em problemas de classificação, considera-se um conjunto de dados com N de amostras $X = \{\mathbf{x}_n\}_{n=1}^N$, em que $\mathbf{x}_n \in \mathcal{X}$ e um conjunto correspondente de rótulos $L = \{l_n\}_{n=1}^N$, com $l_n \in \{C_s\}_{s=1}^S$, com C_s denota a s -ésima classe.

O MLM pode ser estendido a problemas de classificação de maneira direta, através da representação das S classes com a utilização do esquema de codificação 1-de- S . Em tal abordagem, uma variável qualitativa com S símbolos é representada por um vetor de S variáveis binárias (ou bits), com somente uma das componentes “ativa” a cada momento [8].

Na classificação de uma amostra \mathbf{x} com classe $l \in \{C_s\}_{s=1}^S$ desconhecida, a classe estimada \hat{l} associada a saída estimada $\hat{\mathbf{y}}$, encontrado pelo método tradicional do MLM é $\hat{l} = C_{s^*}$, onde

$$s^* = \arg \max_{s=1, \dots, S} \left\{ \hat{y}^{(s)} \right\}, \quad (7)$$

e $\hat{y}^{(s)}$ denota a s -ésima componente do vetor $\hat{\mathbf{y}}$.

2.4 MLM de vizinho mais próximo

O procedimento de estimação de saída do método MLM é a etapa de maior complexidade computacional, uma vez que existe a necessidade de solucionar um problema de otimização. Assim, a redução da complexidade computacional nesta etapa é particularmente valiosa. No que diz respeito ao caso de classificação, pode-se notar que a natureza finita e discreta dos resultados possíveis para a previsão de classe torna o procedimento de estimativa de saída mais fácil.

A fase de estimação da saída visa encontrar a melhor estimativa para a saída de uma determinada entrada \mathbf{x} , a partir das distâncias estimadas entre este ponto e o conjunto de pontos de referência $\{\hat{\delta}(\mathbf{y})\}_{k=1}^K$ no espaço de saída. Em tarefas de classificação, o conjunto de valores possíveis para $\hat{\mathbf{y}}$ é limitado ao número de classes. Dessa forma, uma alternativa à solução do problema de otimização apresentado na Equação 6 é adotar a saída (rótulo) associada ao ponto de referência mais próximo no espaço de saída [8].

A MLM de vizinho mais próximo (*Nearest Neighbor Minimal Learning Machine*, NN-MLM) estima a saída a partir do rótulo associado ao ponto de referência mais próximo no espaço de saída. A principal vantagem do NN-MLM é a não dependência de métodos baseados em gradiente, acelerando assim a fase de testes do MLM sem comprometer sua precisão. Dessa forma, a classe estimada \hat{l} , a partir de um padrão de entrada \mathbf{x} , é dada por $\hat{l} = l_{k^*}$, em que

$$k^* = \arg \min_{k=1, \dots, K} \left\{ \hat{\delta}^2(\mathbf{y}, \mathbf{t}_k) \right\}, \quad (8)$$

e l_{k^*} representa o rótulo do k -ésimo ponto de referência.

3 Análise empírica da seleção de pontos de referência

Bem como em máquinas de vetores suporte (*support vector machines*, SVM) [13] e máquinas de vetores de relevância (*relevance vector machines*, RVM) [12], na etapa de predição, o MLM requer um subconjunto de pontos selecionados a partir do conjunto de dados. Todavia, diferentemente dos classificadores SVM e RVM, a escolha de tais pontos não é um subproduto direto da sua etapa de treinamento. Neste caso, tais pontos (chamados de pontos de referência) são selecionados na etapa inicial do processo de treinamento. Na proposição original do MLM, a escolha destes pontos é realizada de maneira aleatória, deixando a cargo do usuário apenas a quantidade de pontos a serem utilizados.

Embora a escolha aleatória traga bons resultados numéricos na grande maioria dos casos, essa abordagem não traz garantias de eficácia. No caso em que o valor de K é pequeno, a escolha aleatória pode resultar em um conjunto de pontos de referência não representativo para a base de dados completa. Isso fará com que o modelo criado não seja complexo o bastante, configurando uma condição de subajustamento (*underfitting*). Por outro lado, a utilização de um valor grande para K , ou mesmo a utilização de todo o conjunto de treinamento como pontos de referência, faz com que o modelo criado seja muito complexo, tendo a ocorrência de um superajustamento (*overfitting*). Além disso, o modelo criado não tem nenhuma característica de esparsidade. Na Figura 2 são ilustrados exemplos de superfícies de decisão geradas por modelos subajustados, sobreajustados e modelos com ajustamento ideal.

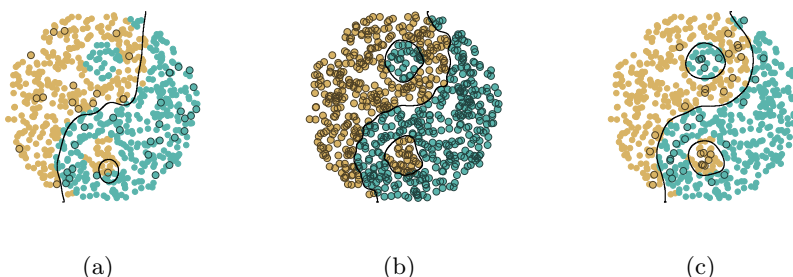


Figura 2. Superfícies de decisão (a) subajustada, (b) sobreajustada e (c) ideal.

Como pode ser observado na Figura 2(a), utilizar um valor pequeno para K e realizar a escolha de forma aleatória pode acarretar em um subconjunto não representativo, desconsiderando regiões possivelmente importantes. De forma contrária, com a utilização de todos os pontos do conjunto de dados como pontos de referência, cria-se uma superfície de decisão bastante complexa, pode ser observado na Figura 2(b). A escolha ideal é fazer com que os pontos de referência não se encontrem em regiões de confusão, ou seja, regiões de sobreposição de dados. Na Figura 2(c) é apresentada a escolha ideal de um subconjunto de pontos de referência. É interessante observar que o número de pontos nas figuras 2(a) e 2(c) é o mesmo.

4 O algoritmo k -médias

O algoritmo k -médias, proposto por J. MacQueen (1967) [6], é um algoritmo simples de quantização vetorial e agrupamento iterativo que particiona N pontos de dados em K subconjuntos disjuntos, denotados por $S = \{S_k\}_{k=1}^K$, de modo a minimizar o critério da soma dos quadrados. Formalmente, tal algoritmo minimiza a seguinte função objetivo:

$$J = \sum_{k=1}^K \sum_{\mathbf{x}_n \in S_k} \|\mathbf{x}_n - \boldsymbol{\mu}_k\|^2, \quad (9)$$

em que \mathbf{x}_n é o n -ésimo vetor pertencente ao conjunto S_k , $\boldsymbol{\mu}_k$ é centróide geométrico dos pontos de dados pertencentes a S_k .

O algoritmo mais comum usa uma técnica de refinamento iterativo. Também referido como algoritmo de Lloyd [5], este algoritmo consiste em um simples processo de reestimação em duas etapas:

- i. Atribuição: Os pontos de dados são atribuídos ao grupo cujo centróide é o mais próximo a cada ponto e,
- ii. Atualização: Cada centróide do grupo é recalculado para o centro (média) de todos os pontos de dados atribuídos a ele.

Estas duas etapas são alternadas até que um critério de parada seja alcançado, de modo que não haja mais alterações na atribuição de pontos de dados a algum grupo. Toda iteração requer $N \times K$ comparações, representando a complexidade do tempo de uma iteração. O Algoritmo 1 apresenta o pseudocódigo de uma possível implementação do método k -médias.

Algoritmo 1 k -médias

Entradas:

$\{\mathbf{x}_n \in \mathbb{R}^D\}_{n=1}^N$: conjunto de entradas
 K : número de grupos

Saídas:

$\{\boldsymbol{\mu}_k \in \mathbb{R}^D\}_{k=1}^K$: conjunto de centroides
 $\{c_n \in \mathbb{N} \cap [1, K]\}_{n=1}^N$: rótulos

- 1: Inicializar os K centroides dos grupos $\{\boldsymbol{\mu}_k\}_{k=1}^K \in \mathbb{R}^D$ aleatoriamente
 - 2: **enquanto** não convergir **faça**
 - 3: **para** $n \leftarrow 1$ **até** N **faça**
 - 4: $c_n \leftarrow \arg \min_k \|\mathbf{x}_n - \boldsymbol{\mu}_k\|^2, \quad 1 \leq k \leq K$
 - 5: **fim para**
 - 6: **para** $k \leftarrow 1$ **até** K **faça**
 - 7: $\boldsymbol{\mu}_k \leftarrow \frac{\sum_{n=1}^N 1\{c_n = k\} \mathbf{x}_n}{\sum_{n=1}^N 1\{c_n = k\}}$
 - 8: **fim para**
 - 9: **fim enquanto**
 - 10: **return** $\{\boldsymbol{\mu}_k\}_{k=1}^K, \{c_n\}_{n=1}^N$
-

5 Proposta: *overlap removed* MLM (OR-MLM)

A proposta *overlap removed* MLM (OR-MLM) utiliza do algoritmo k -médias como ferramenta principal na seleção de pontos de referência para máquinas de aprendizado mínimo. Assim como na proposição original, o OR-MLM necessita de um único parâmetro K , que neste caso é o número máximo de pontos de referência. O número de pontos de referência obtidos pelo OR-MLM será sempre menor ou igual ao parâmetro K (i.e. $|R| \leq K$). Isto é possível graças a execução de uma etapa de remoção de pontos de referência que estejam em regiões heterogêneas¹. Isto é feito através da eliminação dos subconjuntos de dados (oriundos que k -médias) que contenham padrões de classes diferentes.

De maneira sucinta, o OR-MLM é realizado em três etapas. Na primeira etapa o algoritmo k -médias é executado sob todo o conjunto de dados. Depois disso, a próxima etapa é realizar a remoção de protótipos advindos da execução do k -médias que agrupem padrões de diferentes classes. Por fim, o conjunto de protótipos homogêneos é selecionado como pontos de referência. A ideia principal da proposta é fazer com que os pontos de referência fiquem bem distribuídos sobre todo o conjunto de dados e, conjuntamente, fazer com que tais pontos não se encontrem em regiões de sobreposição. O Algoritmo 2 apresenta o pseudocódigo do método OR-MLM.

Algoritmo 2 KM-MLM

Entradas:

- X : conjunto de entradas
- L : conjunto de rótulos
- K : número inicial de pontos de referência

Saídas:

- R : conjunto de pontos de referência
 - L : conjunto de rótulos dos pontos de referência
 - $\hat{\mathbf{B}}$: modelo de regressão
- 1: Aplicar k -média sobre todo o conjunto de dados

$$\{\boldsymbol{\mu}_k\}_{k=1}^K, \{c_n\}_{n=1}^N \leftarrow k\text{-MÉDIAS}(X, K)$$

- 2: Criar k conjuntos com os rótulos dos dados

$$L_k \leftarrow \{\mathbf{l}_n \in L \mid c_n = k\}, \quad 1 \leq k \leq K$$

- 3: Criar conjunto de pontos de referência apenas com centroides homogêneos

$$R \leftarrow \bigcup_k \{\boldsymbol{\mu}_k\} \text{ e } T \leftarrow \bigcup_k \{L_k\}, \quad \forall k: |L_k| = 1$$

- 4: Calcular as matrizes de distâncias \mathbf{D}_x através de X e R ; e $\mathbf{\Delta}_y$ através de Y e T
 - 5: Calcular o modelo de regressão $\hat{\mathbf{B}}$ através da Equação (3)
 - 6: **return** $R, T, \hat{\mathbf{B}}$
-

¹ Consideram-se regiões heterogêneas, regiões que contenham dados de classes diferentes, isto é, locais com sobreposição de dados.

6 Simulações computacionais

A fim de avaliar o comportamento dos métodos de seleção, foram realizadas simulações com conjuntos de dados reais, disponíveis no repositório de aprendizado de máquinas da UCI [4]. Além disso, outros dois conjuntos de dados artificiais foram utilizados nas simulações. Informações sobre os conjuntos de dados utilizados neste trabalho são apresentados na Tabela 1.

Tabela 1. Conjuntos de dados utilizados.

Conjunto de dados	Abreviação	# Padrões	# Treino	# Teste	# Atributos
<i>Heart</i>	HEA	270	216	54	13
<i>Haberman's Survival</i>	HAB	306	245	61	3
<i>Vertebral Column</i>	VCP	310	248	62	6
<i>Ionosphere</i>	ION	351	281	70	34
<i>Breast Cancer W.</i>	BCW	688	550	138	9
<i>Australian Credit A.</i>	AUS	690	552	138	14
<i>Pima Indians Diabetes</i>	PID	768	614	154	8
<i>Two Moon*</i>	TMN	1001	801	200	2
<i>Ripley*</i>	RIP	1250	1000	250	2

* Conjuntos de dados artificiais.

O desempenho da proposta foi comparado com duas variações do MLM, no que dizem respeito a seleção de pontos de referência. A primeira variação é a *full* MLM (FL-MLM), na qual o conjunto de pontos de referência é igual ao conjunto de treinamento ($R = X$). Na segunda variação, *random* MLM (RN-MLM), são selecionados aleatoriamente K pontos de referência, tal como os autores do MLM reportam no artigo original.

Foi utilizada uma combinação dos métodos de validação cruzada *holdout* e *k-fold* nos experimentos realizados. O método *holdout* com uma divisão de 80%–treinamento, 20%–teste foi utilizado para estimação das métricas de comparação, que neste caso são a taxa de acerto, ou acurácia (ACC) e percentual de redução de pontos de referência em relação ao conjunto de treinamento (RED). Além da análise numérica dos resultados, também foi executado um teste estatístico de hipótese para verificar a possível equivalência entre os métodos propostos e os presentes no estado da arte.

O ajuste do parâmetro K para os métodos RN-MLM e OR-MLM (número de pontos de referência e número máximo de pontos de referência, respectivamente) foi realizado utilizando o método de busca em grade, combinado com validação cruzada *kfold* de 5 divisões. Os pontos de referência foram selecionados aleatoriamente numa gama de 5–100% (com taxa de variação de 5%) das amostras de treinamento disponíveis. O erro de classificação foi utilizado para a escolha do melhor valor de K .

Cada experimento foi realizado 30 vezes para cada par *conjunto de dados/método de seleção*, com base na seleção aleatória dos subconjuntos de dados de treinamento e teste. As estimativas globais das métricas de avaliação foram tomadas como uma média geral das 30 realizações para cada conjunto de dados.

Independente do conjunto de dados e do método de seleção, todos os dados são submetidos a uma etapa de normalização, fazendo com que valores medidos em escalas diferentes sejam ajustados para a mesma escala. Nos experimentos realizados utilizou-se a normalização de média zero, dessa forma, os dados estarão dispostos sob uma média 0 e desvio padrão 1.

Tabela 2. Métricas de desempenho – percentual da taxa de acerto (ACC) e percentual de redução de pontos de referência em relação ao conjunto de treinamento (RED) – para os métodos OR-MLM, RN-MLM e FL-MLM; e resultados de testes estatísticos*. Os símbolos ✓ e ✗ significam equivalência ou não equivalência, respectivamente.

<i>base</i>	<i>métrica</i>	OR-MLM	RN-MLM	FL-MLM
HEA	ACC	69,88 ± 6,43	70,00 ± 6,29 ✓	72,16 ± 6,29 ✗ ✗
	RED	44,94 ± 17,31	33,12 ± 18,10	
HAB	ACC	72,57 ± 4,60	71,97 ± 4,27 ✓	68,09 ± 4,98 ✗ ✗
	RED	92,50 ± 6,72	80,20 ± 14,22	
VCP	ACC	84,89 ± 4,29	82,58 ± 4,39 ✗	82,15 ± 4,23 ✗ ✗
	RED	75,51 ± 13,48	56,51 ± 26,86	
ION	ACC	93,33 ± 2,87	93,10 ± 2,73 ✓	94,14 ± 2,74 ✓ ✗
	RED	57,77 ± 12,40	42,18 ± 21,05	
BCW	ACC	96,84 ± 1,38	96,98 ± 1,40 ✓	96,96 ± 1,27 ✓ ✓
	RED	77,72 ± 17,29	62,61 ± 22,71	
AUS	ACC	70,53 ± 3,12	72,15 ± 3,05 ✗	70,97 ± 3,59 ✓ ✗
	RED	72,58 ± 13,53	55,59 ± 22,22	
PID	ACC	74,26 ± 2,30	74,59 ± 2,58 ✓	73,16 ± 2,38 ✗ ✗
	RED	88,82 ± 7,60	75,92 ± 16,10	
TMN	ACC	99,87 ± 0,22	99,82 ± 0,28 ✓	99,87 ± 0,22 ✓ ✓
	RED	71,34 ± 16,78	61,92 ± 20,72	
RIP	ACC	90,17 ± 2,04	89,75 ± 1,77 ✓	88,32 ± 1,61 ✗ ✗
	RED	90,27 ± 5,64	76,64 ± 18,83	

* O teste estatístico adotado foi o teste não paramétrico de Friedman [3].

Através da análise da Tabela 2 pode-se inferir que o método de seleção proposto se mostra suficientemente estável, em termos de taxa de acerto, já que o mesmo sempre obtém valores médios de acurácia superiores ou equivalentes aos outros métodos. Em termos de número de pontos de referência, o método proposto obteve resultados consideravelmente melhores que todos os outros métodos de seleção.

Para uma análise qualitativa, os métodos de seleção estudados foram aplicados a um conjunto de dados artificial bastante conhecido na literatura, chamado ripley (RIP). Neste experimento, o parâmetro K , do método RN-MLM foi selecionado através da busca em grade e validação cruzada *kfold* com 5 divisões sobre todo o conjunto de dados (1250 amostras). O mesmo valor de K foi utilizado no método OR-MLM. O conjunto de dados RIP, bem como os resultados deste experimento são apresentados na Figura 3.

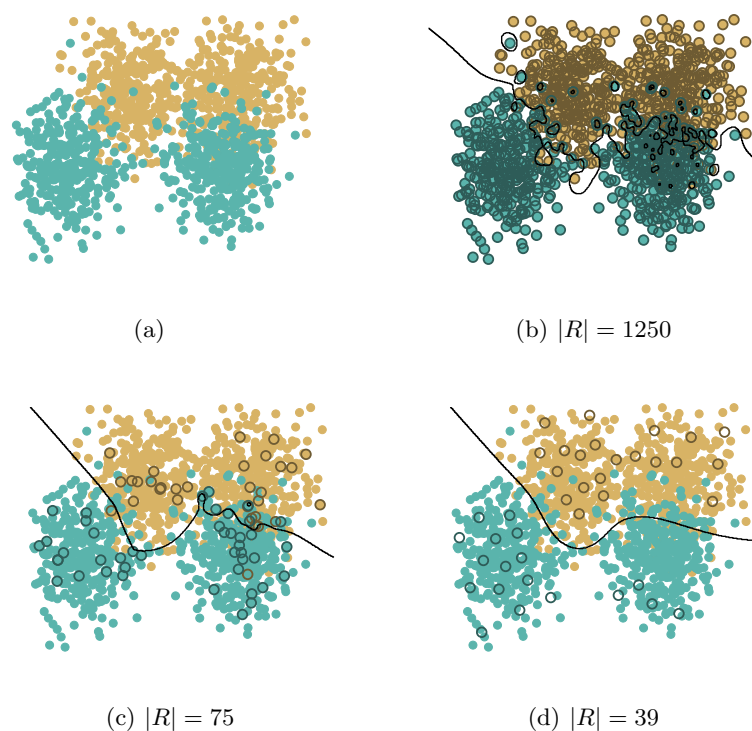


Figura 3. Análise qualitativa da proposta OR-MLM. (a) Conjunto de dados e superfícies de decisão geradas pelos métodos (b) FL-MLM, (c) RN-MLM e (d) OR-MLM, quando aplicadas ao conjunto de dados RIP.

Com base na Figura 3, pode-se inferir que o OR-MLM produziu uma melhor superfície de decisão quando comparados aos outros algoritmos. Na Figura 3(d), pode-se ver que o número de RPs para OR-MLM é menor do que o número de RPs para RN-MLM (Figura 3(c)). Além disso, o limite de decisão gerado a partir do OR-MLM é mais suave do que os outros modelos.

7 Conclusões

Neste trabalho foi proposto um método para seleção de pontos de referência para classificadores MLM, chamado *overlap removed* MLM. Para realizar este objetivo, utilizou-se uma abordagem baseada no algoritmo k -médias, que busca fazer com que os pontos de referência fiquem bem distribuídos sobre todo o conjunto de dados e, conjuntamente, fazer com que tais pontos não se encontrem em regiões de sobreposição. Com base nos resultados obtidos, pode-se inferir que a proposta atinge a expectativa desejada, uma vez que o número de pontos de referência escolhidos pela proposta é bem menor que o número de pontos obtidos por métodos habituais de seleção e ao mesmo tempo, a taxa de acerto da proposta se mantém similar as outras técnicas.

Referências

1. Alencar, A.S.C., Caldas, W.L., Gomes, J.P.P., de Souza, A.H., Aguilar, P.A.C., Rodrigues, C., Franco, W., de Castro, M.F., Andrade, R.M.C.: Mlm-rank: A ranking algorithm based on the minimal learning machine. In: 2015 Brazilian Conference on Intelligent Systems, BRACIS 2015, Natal, Brazil, November 4-7, 2015. pp. 305–309. IEEE (2015)
2. Coelho, D.N., Barreto, G.D.A., Medeiros, C.M.S., Santos, J.D.A.: Performance comparison of classifiers in the detection of short circuit incipient fault in a three-phase induction motor. In: 2014 IEEE Symposium on Computational Intelligence for Engineering Solutions, CIES 2014, Orlando, FL, USA, December 9-12, 2014. pp. 42–48. IEEE (2014)
3. Demšar, J.: Statistical comparisons of classifiers over multiple data sets. *The Journal of Machine Learning Research* 7, 1–30 (2006)
4. Lichman, M.: UCI machine learning repository (2013)
5. Lloyd, S.P.: Least squares quantization in PCM. *IEEE Trans. Information Theory* 28(2), 129–136 (1982), <https://doi.org/10.1109/TIT.1982.1056489>
6. MacQueen, J.: Some methods for classification and analysis of multivariate observations. In: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*. pp. 281–297. University of California Press, Berkeley, Calif. (1967), <http://projecteuclid.org/euclid.bsm/1200512992>
7. Marquardt, D.W.: An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics* 11(2), 431–441 (1963), <https://doi.org/10.1137/0111030>
8. Mesquita, D.P.P., Gomes, J.P.P., Souza Junior, A.H.: Ensemble of efficient minimal learning machines for classification and regression. *Neural Processing Letters* (Jan 2017), <https://doi.org/10.1007/s11063-017-9587-5>
9. Niewiadomska-Szynkiewicz, E., Marks, M.: Optimization schemes for wireless sensor network localization. *Applied Mathematics and Computer Science* 19(2), 291–302 (2009), <https://doi.org/10.2478/v10006-009-0025-3>
10. Prabhakar, S.K., Rajaguru, H.: Pca and k-means clustering for classification of epilepsy risk levels from eeg signals #8212; a comparative study between them. In: 2015 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS). pp. 83–86 (Nov 2015)
11. de Souza Júnior, A.H., Corona, F., Barreto, G.D.A., Miché, Y., Lendasse, A.: Minimal learning machine: A novel supervised distance-based approach for regression and classification. *Neurocomputing* 164, 34–44 (2015)
12. Tipping, M.E.: Sparse bayesian learning and the relevance vector machine. *Journal of Machine Learning Research* 1, 211–244 (2001), <http://www.ai.mit.edu/projects/jmlr/papers/volume1/tipping01a/abstract.html>
13. Vapnik, V.N.: *The Nature of Statistical Learning Theory*. Springer, New York (1995)
14. Wang, D., Nie, F., Huang, H.: Unsupervised feature selection via unified trace ratio formulation and k-means clustering (TRACK). In: Calders, T., Esposito, F., Hüllermeier, E., Meo, R. (eds.) *Machine Learning and Knowledge Discovery in Databases - European Conference, ECML PKDD 2014, Nancy, France, September 15-19, 2014. Proceedings, Part III. Lecture Notes in Computer Science*, vol. 8726, pp. 306–321. Springer (2014), <https://doi.org/10.1007/978-3-662-44845-8>
15. Wang, L., Pan, C.: Robust level set image segmentation via a local correntropy-based k-means clustering. *Pattern Recognition* 47(5), 1917–1925 (2014), <https://doi.org/10.1016/j.patcog.2013.11.014>