A Fuzzy C-means-based Approach for Selecting Reference Points in Minimal Learning Machines

Conference Paper *in* Communications in Computer and Information Science · July 2018

DOI: 10.1007/978-3-319-95312-0_34

4 authors, including:

Jose Alberth Vasconcelos Florêncio
Universidade Federal do Ceará
3 PUBLICATIONS **5** CITATIONS
SEE PROFILE

Madson Luiz Dantas Dias
Universidade Federal do Ceará
17 PUBLICATIONS **32** CITATIONS
SEE PROFILE

Ajalmar Rêgo da Rocha Neto
Instituto Federal de Educação, Ciência e Tecnologia do Ceará
68 PUBLICATIONS **313** CITATIONS
SEE PROFILE

A Fuzzy C-means-based Approach for Selecting Reference Points in Minimal Learning Machines

José A. V. Florêncio(✉), Madson L. D. Dias, Ajalmar R. da Rocha Neto(✉), and Amauri H. de Souza Júnior

Department of Teleinformatics, Federal Institute of Ceará, Treze de Maio Avenue, Fortaleza, Ceará 2081, Brazil
{jose.florencio,madson.dias}@ppgcc.ifce.edu.br,
{ajalmar,amauriholanda}@ifce.edu.br
http://ppgcc.ifce.edu.br/

Abstract. This paper introduces a new approach to select reference points of minimal learning machines (MLM) for classification tasks. The proposal is based on the Fuzzy C-means algorithm and consists of selecting data samples from regions where no overlapping between classes exists. Such an idea has been empirically shown capable of achieving simpler decision boundaries in comparison to the standard MLM, and thus less susceptible to overfitting. Experiments were performed using UCI data sets. The proposal was able to both reduce the number of reference points and achieve competitive performance when compared to conventional approaches for selecting reference points.

Keywords: Machine learning · Minimal learning machines
Fuzzy-C means

1 Introduction

The Minimal Learning Machine (MLM, [1]) is a supervised learning algorithm that has recently been applied to a diverse range of problems, such as fault detection [2], ranking of documents [3], and robot navigation [4].

The basic operation of MLM consists in a linear mapping between the geometric configurations of points in the input space and the respective points in the output space. The geometric configuration is captured by two distance matrices (input and output), computed between the training/learning points and a subset of it whose elements are called reference points (RPs). The learning step in the MLM consists of fitting a linear regression model between these two distance matrices. In the test phase, given an input, the MLM predicts its output by first computing distances in the input space and then using the learned regression

The authors would like to thank the IFCE for supporting their research.

G. A. Barreto and R. Coelho (Eds.): NAFIPS 2018, CCIS 831, pp. 398–407, 2018.
https://doi.org/10.1007/978-3-319-95312-0_34

model to predict distances in the output space. Those distances are then used to provide an estimate to the output.

The determination of the RPs, including its quantity, is fundamental to the quality of the surface boundary generated by the MLM model. In this regard, the original formulation of the MLM training algorithm establishes a random reference points choice, leaving just the number of points definition determined by the user. The random reference points choice ignores the data set disposition.

Clustering algorithms, such fuzzy C-means [5], are unsupervised machine learning methods that aim to separate objects into groups, based on the characteristics that these objects have. The basic idea is to bring objects with similar characteristics together into one group. The similarity of these objects is defined according to pre-established criteria.

The use of Fuzzy C-Means together with classification methods occurs mainly for the reduction of datasets with a large number of samples. In addition, some works use this algorithm and its variations in several types of applications, such as the classification of epilepsy risk [6], feature selection [7] and image segmentation [8].

In this work, it is proposed to use the algorithm Fuzzy C-Means for the selection of reference points of Minimum Learning Machines applied to problems of patterns recognition. Simulation with real-world datasets was performed to validate the proposal. The proposal was able to significantly reduce the number of reference points, as well as maintain its capacity of generalization equivalent or superior when compared to the approaches proposed in the original article [1].

The remainder of the paper is organized as follows. Section 2 briefly describes the MLM. Section 3 emphasizes the need for novel methods for selecting RPs. Section 4 introduces the Fuzzy C-Means MLM. Section 5 reports the empirical assessment of the proposal and the conclusions are outlined in Sect. 6.

2 Minimal Learning Machine

The Minimal Learning Machine is a supervised method whose training step consists of fitting a multiresponse linear regression model between distances computed from the input and output spaces. Output prediction for new incoming inputs is achieved by estimating distances in the output spaces using the underlying linear model followed by a search/optimization procedure in the space of possible outputs.

Basic Formulation. Let us define the learning problem as the problem of approximating a smooth continuous *target function* $f : \mathcal{X} \rightarrow \mathcal{Y}$ from the data $\mathcal{D} = \{(\mathbf{x}_n, \mathbf{y}_n = f(\mathbf{x}_n))\}_{n=1}^{N}$, where $\mathbf{x}_n \in \mathcal{X}$ and $\mathbf{y}_n \in \mathcal{Y}$. We call $\mathcal{X}$ and $\mathcal{Y}$ the input and output spaces, respectively. Henceforth, we assume $\mathcal{X} = \mathbb{R}^D$ and $\mathcal{Y} = \mathbb{R}^S$.

The MLM aims to approximate the target function f through the use of surrogate functions $\delta_k : \mathcal{Y} \rightarrow \mathbb{R}_+$ and $d_k : \mathcal{X} \rightarrow \mathbb{R}_+$. The surrogate functions are distance functions taken from fixed points $\mathcal{R} = \{(\mathbf{m}_k, \mathbf{t}_k = f(\mathbf{m}_k)) \in \mathcal{D}\}_{k=1}^{K}$, also

called reference points. More precisely, we have $d_k(\mathbf{x}) = d(\mathbf{x}, \mathbf{m}_k)$ and $\delta_k(\mathbf{y}) = \delta(\mathbf{y}, \mathbf{t}_k)$, with both $d(\cdot,\cdot)$ and $\delta(\cdot,\cdot)$ given by the Euclidean distance function. In addition, we refer to the set $\{\mathbf{m}_k\}_{k=1}^K$ as input reference points, and $\{\mathbf{t}_k\}_{k=1}^K$ as the corresponding output reference points.

We assume the existence of a mapping between the spaces induced by the distance functions δ and d. Formally, we have $g_k : \prod_{j=1}^K d_j(\mathcal{X}) \to \delta_k(\mathcal{Y})$, or equivalently $g_k : \mathbb{R}_+^K \to \mathbb{R}_+$. Considering now the data $\mathcal{D}$, we collect the distances taken in the input space between the data points and the input reference points in a matrix $\mathbf{D} \in \mathbb{R}_+^{N\times K}$. Similarly, take the pointwise distance matrix in the output space between the N data points (outputs $\mathbf{y}_n$) and the output reference points to be represented by $\boldsymbol{\Delta} \in \mathbb{R}_+^{N\times K}$. Using the data, we are interested in finding the mapping g_k using the model $\Delta_{n,k} = g_k(\mathbf{D}_{n,\cdot}) + \epsilon_n$ for all $n = 1, \ldots, N$. The term ϵ_n represents the residuals whereas $\mathbf{D}_{n,\cdot}$ denotes the n-th row of the matrix $\mathbf{D}$; by the same token, $\Delta_{n,k}$ stands for the element in the n-th row and k-th column of $\boldsymbol{\Delta}$.

The MLM assumes that the mappings g_k can be sufficiently well approximated by linear models. In doing so, we have that distances in the output space can be approximated by a linear combination of distances in the input space, i.e., $\Delta_{n,k} = \mathbf{D}_{n,\cdot}\mathbf{b}_k + \epsilon_n$, where $\mathbf{b}_k \in \mathbb{R}^K$ represents the coefficients of the linear mapping g_k. Putting all the mappings together for all data points, we represent the so-called *distance regression* model of the MLM in a matrix form given by

$$\boldsymbol{\Delta} = \mathbf{DB} + \boldsymbol{\varepsilon}, \tag{1}$$

where the matrix $\mathbf{B} \in \mathbb{R}^{K\times K}$ comprises the K vectors of coefficients $\mathbf{b}_k$ in its columns.

Given that, the MLM computes a function $h_{\mathbf{B}}(\mathbf{x}) : \mathcal{X} \to \mathcal{Y}$ given by:

$$h_{\mathbf{B}}(\mathbf{x}) = \arg\min_{\mathbf{y}} \sum_{k=1}^{K} \left[\delta_k^2(\mathbf{y}) - \left(\sum_{i=1}^{K} d_i(\mathbf{x}) B_{i,k} \right)^2 \right]^2, \tag{2}$$

where $\delta_k(\mathbf{y}) = \|\mathbf{y} - \mathbf{t}_k\|$ represents the Euclidean distance between $\mathbf{y}$ and the k-th output reference point $\mathbf{t}_k$; similarly, $d_i(\mathbf{x}) = \|\mathbf{x} - \mathbf{m}_i\|$ denotes the Euclidean distance between $\mathbf{x}$ and the i-th input reference point $\mathbf{m}_i$; K denotes the number of reference points.

Learning Algorithm. The learning algorithm of the Minimal Learning Machine simply requires the (i) selection of the set reference points $\{(\mathbf{m}_k, \mathbf{t}_k)\}$; and (ii) determination of the parameters $\mathbf{B}$. With regard to the selection of reference points, in the original proposal, the MLM assigns the reference points randomly from the available data points for learning. This paper focus on alternatives to such random assignment.

Since the reference points are taken from the data, we have that $K \leq N$. The number of reference points K controls the model capacity, thus it can be used to avoid overfitting. Under the normal conditions where the number of selected reference points is smaller than the number of training points (i.e., $K < N$), the matrix $\mathbf{B}$ can be approximated by the usual least squares estimate

$$\hat{\mathbf{B}} = (\mathbf{D}^T\mathbf{D})^{-1}\mathbf{D}^T\boldsymbol{\Delta}, \tag{3}$$

where $\mathbf{D}$ and $\boldsymbol{\Delta}$ are the pairwise distance matrices between the data and the reference points in the input and output space respectively.

Out-of-Sample Prediction. Predicting the outputs for new input data mainly refers to solving the minimization problem embedded in Eq. (2). For an out-of-sample input point $\mathbf{x}$ whose distances from the K input reference points $\{\mathbf{m}_k\}_{k=1}^K$ are computed, i.e., $d_1(\mathbf{x}) \dots d_K(\mathbf{x})$, we then estimate the distances between its unknown output $\mathbf{y}$ and the output reference points using the linear model between distances, that is

$$\hat{\delta}_k(\mathbf{y}) = \sum_{i=1}^{K} d_i(\mathbf{x})\hat{B}_{i,k}, \qquad \forall k = 1, \dots, K. \tag{4}$$

Together the estimates $\hat{\delta}_1(\mathbf{y}) \dots \hat{\delta}_K(\mathbf{y})$ can be used to locate $\mathbf{y}$ in the $\mathcal{Y}$-space. The location of $\mathbf{y}$ can be estimated from the minimizer given in Eq. (2) and rewritten here to emphasize the dependence of $\mathbf{y}$:

$$\hat{\mathbf{y}} = \arg\min_{\mathbf{y}} \sum_{k=1}^{K} \left((\mathbf{y} - \mathbf{t}_k)^T(\mathbf{y} - \mathbf{t}_k) - \hat{\delta}_k^2(\mathbf{y}) \right)^2. \tag{5}$$

It is worth mentioning that $\hat{\delta}_k(\mathbf{y})$ is not a function of $\mathbf{y}$ but rather a point estimate of the actual distance function $\delta_k(\mathbf{y}) = \|\mathbf{y} - \mathbf{t}_k\|_2$. Thus, from an optimization perspective, $\hat{\delta}_k(\mathbf{y})$ must be treated as a constant.

For the classification case, where outputs $\mathbf{y}_n$ are represented using the 1-of-S encoding scheme[1]. It was showed in [9] that under the assumption that the classes are balanced, the optimal solution to Eq. (5) is given by

$$h_{\mathbf{B}}(\mathbf{x}) = \hat{\mathbf{y}} = \mathbf{t}_{k^*}, \tag{6}$$

where $k^* = \arg\min_k \hat{\delta}_k(\mathbf{y})$. It means that output predictions for new incoming data can be carried out by simply selecting the output of the nearest reference point in the output space, estimated using the linear model $\hat{\mathbf{B}}$. This method was named Nearest Neighbor MLM (NN-MLM).

3 Empirical Analysis of RP Selection

Similarly to the support vector machines (SVM, [10]) and the relevance vector machines (RVM, [11]), the out-of-sample phase of the MLM requires the determination of a subset of points taken from the training set. However, unlike the

[1] A S-level qualitative variable is represented by a vector of S binary variables or bits, only one of which is *on* at a time. Thus, the j-th component of an output vector $\mathbf{y}$ is set to 1 if it belongs to class j and 0 otherwise.

SVM and RVM, the choice of such points is not a by-product of the training step. In the original MLM proposal, the choice of such points is random, leaving just the number of points determined by the user. However, such an approach is not a warranty of model effectiveness. In other words, a model generated from a random reference points selection can provide bad decision boundaries that can be overfitting or underfitting. In Fig. 1 we show examples of decision boundaries generated by underfitting, overfitting and appropriate-fitting models.

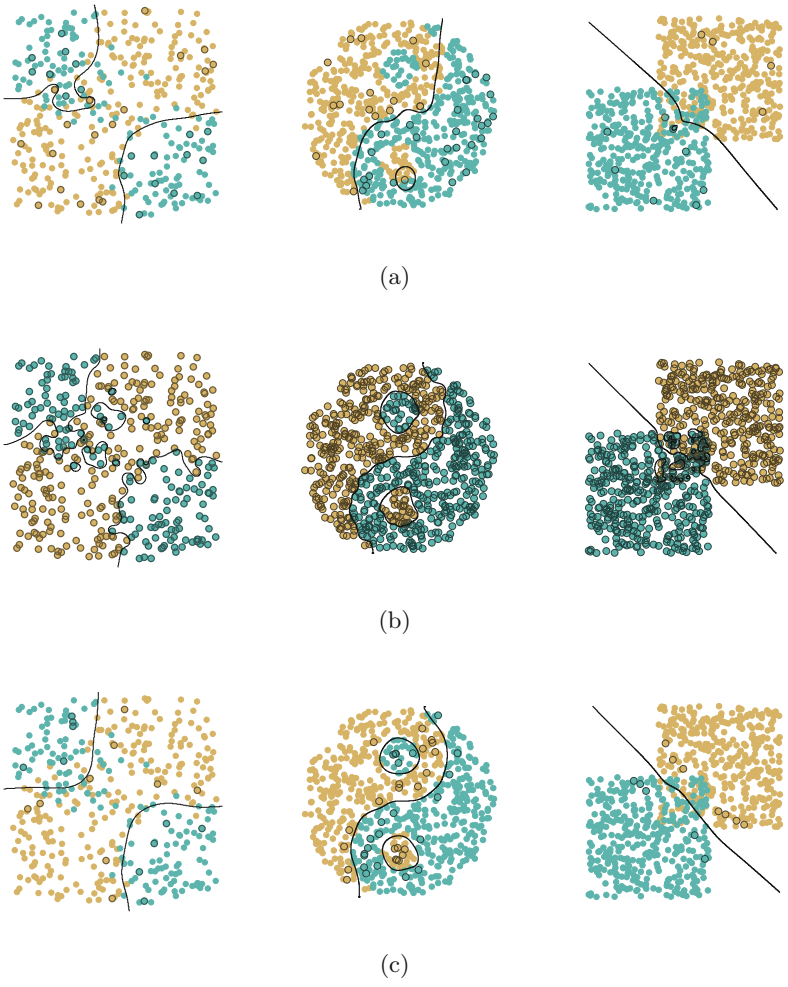

Fig. 1. Examples of decision boundaries generated by (a) underfitting, (b) overfitting and (c) appropriate-fitting models.

As can be seen in Fig. 1(a), using a small value for K and making the choice at random can lead to an unrepresentative subset, disregarding possibly important regions. Conversely, with the use of all points in the data set as reference points, a rather complex decision surface is created, as can be seen in Fig. 1(b). The ideal choice is to make reference points not in "confusing" regions, in other words,

regions of data overlap. In Fig. 1(c) the ideal choice of a subset of reference points is presented. It is interesting to note that the number of points in Fig. 1(a) and (c) is the same.

4 Proposal: Fuzzy C-means MLM (FCM-MLM)

Our proposal, called fuzzy C-means MLM (FCM-MLM), uses the fuzzy C-means [5] algorithm as the main tool for select reference points (RPs) in minimal learning machines. As in the original proposal, FCM-MLM requires only a single hyper-parameter K, which in our case denotes the maximum number of RPs. The number of RPs obtained by the FCM-MLM will always be less than or equal to the parameter K (i.e. $|\mathcal{R}| \leq K$). This is possible due to the execution of an extra step for removing points that are in heterogeneous regions—regions that contain data of different classes. This is carried out by eliminating subsets of data (derived from fuzzy C-means) that contain patterns of different classes.

Briefly, FCM-MLM comprises three steps. In the first step, the fuzzy C-means algorithm runs under all the data set. After that, the next step is to perform the removal of prototypes resulting from the execution of fuzzy C-means that group patterns from different classes. Finally, the set of homogeneous prototypes is selected as RPs. The main idea of the proposal is to ensure that reference points are well distributed over the input space and, jointly, to ensure that such points are not located in class-overlapping regions. Algorithm 1 presents the pseudocode of the FCM-MLM method.

Algorithm 1. FCM-MLM

Input: Initial RPs number (K), data set inputs ($\mathcal{X}$) and outputs ($\mathcal{Y}$)

Output: Regression model ($\hat{\mathbf{B}}$), set of PRs inputs ($\mathcal{R}$) and outputs ($\mathcal{T}$)

1: Apply the fuzzy C-means algorithm in the whole data set

$$\{\boldsymbol{c}_k\}_{k=1}^{K}, \{\mu_{nk}\}_{n=1,k=1}^{N,K} \leftarrow C\text{–}\textsc{Means}(\mathcal{X}, K)$$

where $\mu_{nk} \in [0, 1]$ are the cluster membership values.

2: Create K subsets with label of the data

$$\mathcal{Y}^k \leftarrow \{\boldsymbol{y}_n \in \mathcal{Y} \mid k = \underset{1 \leq i \leq K}{\arg\max}\ \mu_{ni}\}, \qquad 1 \leq k \leq K$$

3: Create the RP set only with the closest patterns of homogeneous centroids

$$\mathcal{R} \leftarrow \bigcup_k \left\{ \underset{\boldsymbol{x}_n \in \mathcal{X}}{\arg\min} \|\boldsymbol{x}_n - \boldsymbol{c}_k\| \right\}, \quad \forall k : |\mathcal{Y}^k| = 1$$

The set $\mathcal{T}$ are given by the corresponding output of the elements in $\mathcal{R}$.

4: Compute the distance matrices $\mathbf{D}_x$ using $\mathcal{X}$ and $\mathcal{R}$; and $\boldsymbol{\Delta}_y$ using $\mathcal{Y}$ and $\mathcal{T}$

5: Compute $\hat{\mathbf{B}}$ using the Eq. (3)

6: **return** $\mathcal{R}$, $\mathcal{T}$, $\hat{\mathbf{B}}$

5 Simulations and Discussion

For a qualitative analysis, we have applied FCM-MLM, RN-MLM and FL-MLM to solve an artificial problem. The problem, well-known Ripley (RIP) dataset problem consists of two classes where the data for each class have been generated by a mixture of two Gaussian distributions (Fig. 2).

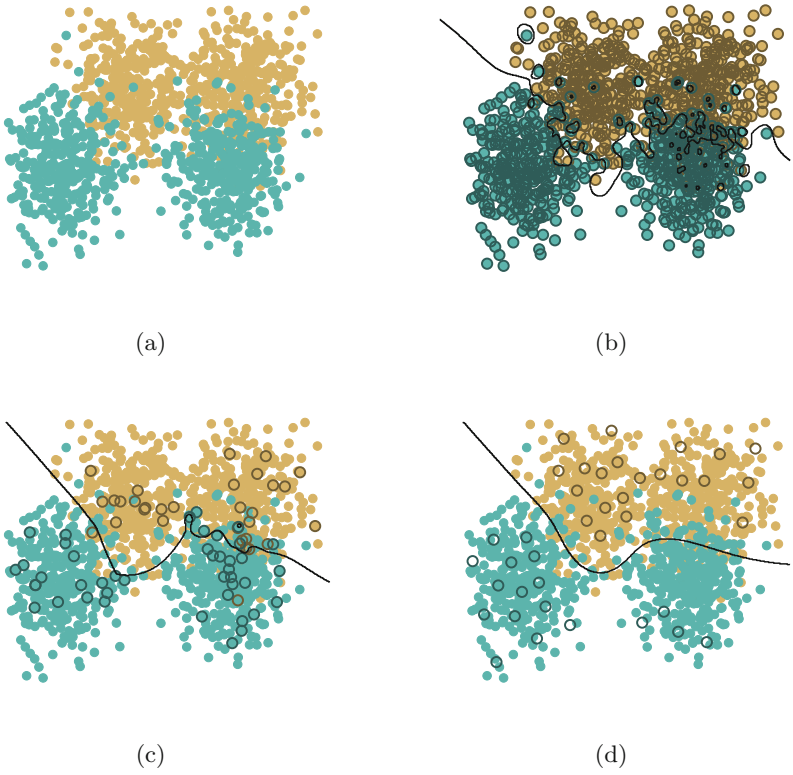

Fig. 2. (a) RIP data set and decision boundaries generated by (b) FL-MLM, (c) RN-MLM, and (d) FCM-MLM.

Based on the Fig. 2, we can infer that FCM-MLM produced better decision boundary when compared to the other algorithms. In the Fig. 2(c) and (d), one can see that the number of RPs for FCM-MLM is lower than the number of RPs for RN-MLM. Moreover, the decision boundary generated from the FCM-MLM is more smoothed than the other models.

Experiments with real-world benchmarking data sets were also carried out in this work. We used UCI data sets [12]; Heart (HEA), Haberman's Survival (HAB), Vertebral Column Pathologies (VCP), Breast Cancer Wisconsin (BCW), Statlog Australian Credit Approval (AUS), Pima Indians Diabets

(PID) and Human Immunodeficiency Virus protease cleavage (HIV). In addition, three well-known artificial data sets were also used in our simulations, Two Moon (TMN), Ripley (RIP) and Banana (BNA). Some description, abbreviation, number of patterns (*# patterns*) and number of features (*# features*) about the aforementioned datasets are presented in the Table 1.

Table 1. General description for datasets used in this work.

Dataset	*Abbreviation*	*# patterns*	*# train*	*# test*	*# features*
Heart	HEA	270	216	54	13
Haberman's Survival	HAB	306	245	61	3
Vertebral Column	VCP	310	248	62	6
Breast Cancer W.	BCW	688	550	138	9
Australian Credit A.	AUS	690	552	138	14
Pima Indians Diabetes	PID	768	614	154	8
Two Moon	TMN	1001	801	200	2
Ripley	RIP	1250	1000	250	2
HIV-1 Protease Cleavage	HIV	3272	2617	655	8
Banana	BNA	5300	4240	1060	2

The performance of our proposal is compared to two variants of the MLM, regarding the selection of RPs. The first variant is the full MLM (FL-MLM), in which the set of reference points is equal to the training set (i.e., $K = N$). The second variant is the random MLM (RN-MLM), where we randomly select K reference points from the training data. It corresponds to the original proposal. A combination of the k-fold cross-validation and holdout methods was used in the experiments. The holdout method with a 80% training and 20% test division was used to estimate the performance metrics. In Table 2 we report the performance metrics of each RP selection method.

The adjustment of the parameter K for the FCM-MLM and the RN-MLM model was performed using grid search combined with 10-fold cross-validation. The RPs were selected in the range of 5–100% (with a step size of 5%) of the available training samples. The classification error was used to choose the best value of K. Each experiment was performed for 30 independent runs.

In order to verify the possible equivalence between the classifier accuracies, we perform a statistical hypothesis tests. Such tests aim to establish the limits beyond which two samples should no longer be considered to be taken from the same population, but as belonging to two different populations. That being said, we adopted a non-parametric test, named Friedman, which does not rely on any assumptions about the form of distribution that is taken to have generated the accuracy values. Besides, it can be used when comparing three or more classifiers [13]. Given the null and the alternative hypothesis, that all algorithms are equivalent or not, respectively; whether the test provided a significance level

value less than or equal to the chosen significance level (in our case, 0.01), the test suggests that the observed data is inconsistent with the null hypothesis and, thus, the null hypothesis should be rejected. Information about superiority and inferiority can also be infered by Friedman statistical hypothesis test.

Table 2. Performance comparison – Accuracy (ACC) and reduction percentage in comparison with the training set (RED) – with the FCM-MLM, RN-MLM and FL-MLM; and results of statistical tests. The symbols ✓ and ✗ with respect to the Friedman statistical test means equivalence and no equivalence, respectively.

Data set	*Metric*	FCM-MLM		RN-MLM			FL-MLM	
HEA	ACC	71.36	6.03	70.00	6.29	✓	**72.16** ± 6.29	✓ ✗
	RED	**59.44**	16.50	33.12	18.10			
HAB	ACC	**72.13**	4.47	71.97	4.27	✓	68.09 ± 4.98	✗ ✗
	RED	**88.01**	11.58	80.20	14.22			
VCP	ACC	**84.78**	4.48	82.58	4.39	✗	82.15 ± 4.23	✗ ✓
	RED	**80.60**	5.99	56.51	26.86			
BCW	ACC	**97.00**	1.31	96.98	1.40	✓	96.96 ± 1.27	✓ ✓
	RED	**86.91**	4.65	62.61	22.71			
AUS	ACC	69.13	3.06	**72.15**	3.05	✗	70.97 ± 3.59	✓ ✓
	RED	**60.65**	5.13	55.59	22.22			
PID	ACC	73.44	2.86	**74.59**	2.58	✗	73.16 ± 2.38	✓ ✗
	RED	**84.61**	8.11	75.92	16.10			
TMN	ACC	**99.87**	0.22	99.82	0.28	✓	99.87 ± 0.22	✓ ✓
	RED	**63.63**	23.71	61.92	20.72			
RIP	ACC	**89.81**	1.88	89.75	1.77	✓	88.32 ± 1.61	✗ ✗
	RED	**87.30**	11.01	76.64	18.83			
HIV	ACC	**86.68**	1.30	86.50	1.30	✓	85.99 ± 1.14	✗ ✗
	RED	**96.73**	1.39	75.32	23.16			
BNA	ACC	88.16	0.90	**89.87**	0.81	✗	87.58 ± 0.89	✓ ✗
	RED	78.19	23.94	**89.33**	2.54			

By analyzing the Table 2 one can conclude that the performances of the FCM-MLM were equivalent or even superior to those achieved by the RN-MLM and FL-MLM for each data sets evaluated. Moreover, one can also see that our proposal achieves sparse solutions, i.e., the FCM-MLM produces a reduced set of RPs.

6 Conclusions

In this paper, we propose an algorithm to select the reference points of the MLM for classification tasks based on the fuzzy C-means algorithm. Three strategies of

MLM reference point selection are evaluated. Our proposal called FCM-MLM is able to obtain the RP subset for MLMs. The experimental results indicate that the FCM-MLM represents a good alternative to the random selection, providing a competitive classifier while maintaining its simplicity.

References

1. de Souza Júnior, A.H., Corona, F., De Barreto, G.A., Miché, Y., Lendasse, A.: Minimal learning machine: a novel supervised distance-based approach for regression and classification. Neurocomputing **164**, 34–44 (2015)
2. Coelho, D.N., De Barreto, G.A., Medeiros, C.M., 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, 9–12 December 2014, pp. 42–48. IEEE (2014)
3. Alencar, A.S.C, Caldas, W.L., Gomes, J.P., de Souza, A.H., Aguilar, P.A., Rodrigues, C., Wellington, F., 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, 4–7 November 2015, pp. 305–309. IEEE (2015)
4. Marinho, L.B., Almeida, J.S., Souza, J.W.M., de Albuquerque, V.H.C., Rebouças Filho, P.P.: A novel mobile robot localization approach based on topological maps using classification with reject option in omnidirectional images. Expert Syst. Appl. **72**, 1–17 (2017)
5. Bezdek, J.C.: Pattern Recognition with Fuzzy Objective Function Algorithms. Advanced Applications in Pattern Recognition, 2nd edn. Plenum Press, New York (1987)
6. Prabhakar, S.K., Rajaguru, H.: PCA and K-means clustering for classification of epilepsy risk levels from EEG signals—a comparitive study between them. In: 2015 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS), pp. 83–86, November 2015
7. 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.) ECML PKDD 2014. LNCS (LNAI), vol. 8726, pp. 306–321. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-662-44845-8_20
8. Wang, L., Pan, C.: Robust level set image segmentation via a local correntropy-based k-means clustering. Pattern Recogn. **47**(5), 1917–1925 (2014)
9. Mesquita, D.P.P., Gomes, J.P.P., Souza Junior, A.H.: Ensemble of efficient minimal learning machines for classification and regression. Neural Process. Lett. **46**, 751–766 (2017)
10. Vapnik, V.N.: The Nature of Statistical Learning Theory. Springer, New York (1995). https://doi.org/10.1007/978-1-4757-2440-0
11. Tipping, M.E.: Sparse Bayesian learning and the relevance vector machine. J. Mach. Learn. Res. **1**, 211–244 (2001)
12. Lichman, M.: UCI machine learning repository (2013)
13. Demšar, J.: Statistical comparisons of classifiers over multiple data sets. J. Mach. Learn. Res. **7**, 1–30 (2006)