

# A Maximal Margin Classification Algorithm Based on Data Field

Zhao Long<sup>1,\*</sup>, Wang ShuLiang<sup>2</sup> and Lin Yi<sup>1</sup>

<sup>1</sup>State Key Laboratory of software Engineering, Wuhan University, Wuhan 270000, China; <sup>2</sup>School of Software, Beijing Institute of Technology, Beijing, 100081, China

**Abstract:** This paper puts forward a new maximal margin classification algorithm based on general data field (MM-GDF). This method transforms the linear inseparable problem into finding the nearest points in the general data field (GDF). GDF is inspired by the physical field. Different dimensions represent the different properties. Not all attributes play a decisive role in the classification process. Therefore, how to find decisive data points and attributes is a critical issue. This research builds a general data field structure in high dimension data sets. The role of data point is expanded from local to global by GDF. We calculate the weights of data points and features by the potential value in data field space. We put it into practice. Experiments show that MM-GDF method is better than the exiting common methods.

**Keywords:** Generalized data field, local hyper plane, maximal margin, nearest neighbors, SVM.

## 1. INTRODUCTION

The local hyper plane distance nearest neighbor's classification builds an on-linear decision surface with maximal local margin in the input space, with in variance inferred from the local neighborhood rather than the prior knowledge, so that it performs very well in many applications MM-GDF is a new nearest neighbor classifiers. The distance metric of MM-GDF was calculated in general data space. A new classification hyper plane was constructed which has been general used in SVM. SVM has attracted substantial interest in machine learning and pattern recognition since the mid-1990s [1, 2]. SVM try to find an optimal separating hyper plane with the maximal margin. This procedure can be cast into a constrained quadratic optimization problem which has a unique solution. In MM-GDF, we still find the  $k$  nearest neighbor points to construct the classification hyper plane. The  $k$  nearest neighbor (KNN) classification is a popular method for classification [3-6]. Independent distance metric is used in 1NN method [7]. The performance of KNN relies on appropriate distance metric of the input data points. Sometime the performance of KNN is even better than 1NN [8]. Although KNN is simple and the parameter is only one, the value of  $K$  is generally small [10].

Recently, a series of improved methods have been present [11-15]. HKNN method performs very well among the manifold-related methods in some applications [16]. The extension and improvement of the HKNN algorithm named adaptive local hyper plane (ALH) algorithm is present by Yang Tao [13, 17].

The attributes metrics and data preprocessing are basic problems in the research of SVM. Maximal margin algorithm doesn't need any particular statistical distribution of

the training data [9]. It requires the tuning of only one free parameter. The goal of building the maximal margin classification model is maximizing the distinctions between the input data points by a decision surface. The method presented in this paper has been compared to the other maximal margin classification methods, including SVM, HKNN, and ALH. The MM-GDF method improves the hyper plane selection strategy. It calculates data sets attribute weights in GDF. The data point in space exist a certain potential value with the other nearby points. If the qualities of data points are equal, the distribution of the data-intensive areas will have a higher potential value. The potential function obtains the maximum point in the vicinity. The potential function may reflect the distribution of data intensive. This distribution will be used as to estimate the overall distribution. The rest of this paper is organized as follows: Section 2 presents the concepts of data field and general data field. The steps of MM-GDF are present in Section3. The competitive experiment and analysis of the results are proposed in Section 4.

## 2. BACKGROUND

In order to overcome the weaknesses of gravity long range nuclear, data field space is introduced. Inspired by the ideas of the physical field, the interaction between the data points is introduced as data field potential value.

### 2.1. Definition of Data Field

The data field equipotential surface is used to describe the greater amount potential pitch spaces. The less potential field equipotential lines means the distance radiation attenuation of data points are more obvious.  $D = \{x_1, x_2, \dots, x_p\}$  is a data set which have  $p$  features.  $\Omega \subseteq R^p$  is the data space of  $D$ .  $x$  is a potential data point which has a certain quality in  $\Omega$ ,  $x \in \Omega$ . There is an action field around  $x$ . Every point of this field will be affected by the other points. Thereby a potential data field is defined on the entire data

space. The potential value of any data point can be expressed as

$$\hat{\varphi}(x) = \frac{1}{n\sigma} \sum_{i=1}^n m_i \times K\left(\frac{x-x_i}{\sigma}\right) \quad (1)$$

where  $K(x)$  is the potential function,  $\sigma$  is an impact factor and  $m_i$  is the mass of data object  $x_i$ .  $K(x)$  is the unit potential function that meets  $\int_{K(x)} dx = 1, \int xK(x) dx = 0$ .  $\sigma$  is an impact factor that control the interaction process between the points. The quality of the object  $x_i$  is  $m_i$  that satisfy the normalization condition  $\sum_{i=1}^n m_i = 1$ .

This function is different with the kernel function. The selection algorithm of current kernel function has certain degree randomness, and lacks of theoretical foundation. The introduction of the data field theory can be a good solution to this problem. The potential value is the strength of field source, such as the quality of the particle or point charge electricity.

### 2.2. General Data Field

Data field further extends to the general data field through the impact factor becoming different in multidimensional space. Assumed that the density function of the data set has been known, the potential accuracy of the estimate of the function depends primarily on factors and units of potential function.

Dataset  $D$  consists  $n$  data points  $\{x_1^{(p)}, x_2^{(p)} \dots x_i^{(p)}, x_n^{(p)}\}$  that show  $n$  data points in the data space of  $\Omega$ . Each data point  $x_i^{(p)}$  is treated as a particle with certain nature  $Nat_i$ .  $x_i^{(p)}$  has a potential mass. In  $\Omega$ , every data points are mutually interacted.  $x_i^{(p)}$  affects other points, and it is further affected by all the other points. All the fields from different local points  $x_i^{(p)}$  ( $i=1, 2, \dots, n$ ) are superposed in global data space  $\Omega$ . This superposition enables the data field to characterize the interaction among different points. The superposed interaction virtually creates a data field in the whole data space  $\Omega$ . The potential value of an arbitrary point  $x$  in the data field is defined as

$$\varphi(x) = \sum_{i=1}^n Nat_i \times K\left(\frac{x - X_i}{\sigma}\right) \quad (2)$$

where,  $K(x)$  is the unit potential function that satisfies  $\int K(x) dx = 1, \int xK(x) dx = 0$ .  $\sigma$  is an impact factor.  $Nat_i$  ( $\sum_{i=1}^n Nat_i = 1, Nat_i \geq 0$ ) is the nature of  $x_i^{(p)}$ , we can take the value of the data point or the quality of the data point as the nature.

However, the impact factor  $\sigma$  is anisotropic. For the existent multi-dimensional data field, the potential function as Equation (1) shows that the impact factor  $\sigma$  is stable. That is, the isotropic  $\sigma$  is supposed to have the same value when measured in different directions. As Equation (2), the impact factor  $\sigma$  is different in multi-dimensions. The potential value has different properties along different axes. When measured in different directions, a different potential value is obtained. The Equation (2) is not appropriate to some extent. Thus, in

order to get a better estimation of the potential function, the general data field is defined as:

$$\hat{\varphi}(x) = \frac{1}{|H|} \sum_{i=1}^n Nat_i \times K(H^{-1}(x - X_i)) \quad (3)$$

where  $K(x)$  is a multivariate potential function and  $H$  is a positive definite  $P \times P$  matrix that is a non-singular constant matrix. Let  $H = \sigma A$ , where  $\sigma > 0, |A| = 1$ . Therefore, the Equation (3) can also be written as

$$\hat{\varphi}(x) = \frac{1}{\sigma^P |A|} \sum_{i=1}^n Nat_i \times K\left(A^{-1}\left(\frac{x - X_i}{\sigma}\right)\right) \quad (4)$$

In some cases, each data object is considered to be the same in terms of mass, Equation (4) is simplified as

$$\hat{\varphi}(x) = \frac{1}{n|H|} \sum_{i=1}^n K(H^{-1}(x - X_i)) \quad (5)$$

### 3. MAXIMAL MARGIN CLASSIFICATION ALGORITHM BASED ON DATA FIELD

The basic idea of MM-GDF algorithm is evaluating the local hyper plane by implying potential values of the various features. Data field theory is applied in high-dimensional data processing and analysis. It has a breakthrough accuracy of data processing in practical application. This algorithm aims high-dimensional data processing in SVM and the multi-view mining theory of SVM. Based a wide range of sample sub-space division rules, it builds the data field structure of the appropriate data sets. According to the scalability of source data in different features, data field theory is extended to general data field. The GDF reveals the intrinsic link between data attributes and data classification results. GDF makes the SVM not only is able to complete the classification of high-dimensional data, but also has high interpretability.

Recently, ALH classification method is present by Tao & Kecman. They use optimizing the local margin and hyper plane in the original and weighted input space to classify data. MM-GDF is different from ALH. MM-GDF has two hyper parameters: T and K, the weights of feature are considered in GDF.

#### 3.1. General Data Field used in Classification

We extend the maximal margin method to high-dimensional data space according analyzing the linkages between data attributes. We obtain the distribution of data points and the corresponding potential values on the different dimensions. The same value data points have different potential values in GDF. The respective dimension of the data in a multidimensional space should have more anisotropic. Through GDF, we depict the relationships of data points by potential value. Each data point can be calculated out of its

corresponding potential value  $\varphi(x) = \sum_{i=1}^m Nat_i \times K(H^{-1}(x - X_i))$ . The potential value of point  $x$  is  $\varphi(x)$ ; if the difference between the data dimensions isn't considered.

However, the feature weights are important to the classifier. In the actual classification study, we can see that the weights of one data points in different dimensions are not entirely equal. We use two potential values to measure fea-

Table 1. Characteristic of data sets.

Data	Instances	Attributes	Classes
Cancer	198	32	3
Derm	366	33	6
Glass	214	9	6
Heart	270	13	2
Pro	997	20	3
Iris	150	4	3
Sonar	208	60	2
Wine	178	13	3
Vote	232	16	3

ture’s contribution to the current categories, respectively within class potential value  $S_w$ , it represent the distance and similarity of each sample in the same class. The similarity within class is greater when the within-class potential value is smaller. The inter-class potential value  $S_b$  is used to measure the difference of the original characteristic of a class with another. The great value of  $S_b$  means the characteristics are more dispersed, this feature play great role in the distinction between classes. When the distributions of data points are evenly, the potential values of the location points approximately equal, the corresponding feature importance is close to 0. Conversely, when the distributions of data points are asymmetrical, the potential values of the location data points are also asymmetrical; the importance of this feature is great. According to the above ideas, we transform the original data set to the data set in data field.

$$\varphi(x) = \sum_{i=1}^m Nat_i \times \sum_{j=1}^n K(H_x^{-1}x - H_{X_i}^{-1}x_{i,j}) \tag{6}$$

where  $x$  is a data point,  $i$  is the number of record,  $j$  is the feature number,  $H_x^{-1}$  is the importance measure of  $x$  dimension,  $H_{x_i}^{-1}$  is the importance measure of  $x_i$  dimension. The weight of dimension is calculated by summing the corresponding potential value of data point on this dimension. And then a new data space is constructed. This data space not only reflects the differences between the data points, also considers the anisotropy differences between the different dimensions.

### 3.2. Specific Steps of MM-GDF

The mathematical model knowledge cannot represent the reality information in dealing with the complex and ever-changing reality data. Some important part may be ignored, and a large number of useless noise data points may exist. The unbalanced properties and potential value of data points can represent some implicit knowledge. Specifically, the first step is selected  $K$  sequence nearest points in Maximal Margin algorithm. Although HKNN expressions explicitly, but the  $K$  value is essential in the classification process. When selected prototypes cannot be choose correctly, the

accuracy of the classification will be very low. Therefore choose of classified prototypes are essential to classifier. The parameter selection is particularly important in maximum margin classification method. This paper present a new ALH algorithm based on general data field, the range of parameters is fixed, it still makes a very good classification results. It reduces the human factor in classification accuracy. The steps of MM-GDF classification algorithm are as follows:

**Algorithm MM-GDF:**  $q, x, K, T$  be a query sample,  $q$  is the test samples,  $x$  is the data points of the samples,  $K$  and  $T$  are the parameter,  $K$  is the neighborhood size.

- Step 1: Calculating the potential of  $x$ , map  $x$  to the general data space  $\Omega$ ;
- Step 2: For every  $q$ , select  $k$  nearest data points, the value of  $K$  is 1:10;
- Step 3: For each  $q$ , we define the local hyper plane with  $J_c^*(q) = \sum_{j=1}^D w_j (V_j \alpha^* + \bar{p}_j - q_j)^2 + \lambda (\alpha^*)^T (\alpha^*)$ ;
- Step 4: Compute the  $k$ -local hyper plane distance  $Dist(x_i, q) = \sqrt{\sum_{j=1}^D w_j (x_{ij} - q_j)^2}$  with  $w_j = \hat{\varphi}(x) = \frac{1}{n|H|} \sum_{i=1}^n K(H^{-1}(x - X_i))$ ;
- Step 5: Classify the query  $q$  to the class label  $f(q) = \operatorname{argmin}_c J_c^*(q)$ .

### 4. EXPERIMENTS AND RESULTS

We used 9 datasets of the UCI Machine Learning Repository. The denotations of the datasets are as follows Table1:

Therefore, the previous performance study of MM-GDF simply used a single loop of cross-validation procedure, where classifiers with different hyper-parameters are trained and tested and the performance with the best hyper-parameters is reported. Our purpose in this work is not only to compare MM-GDF methods with traditional classifiers, but also to assess the true accuracy of MM-GDF methods

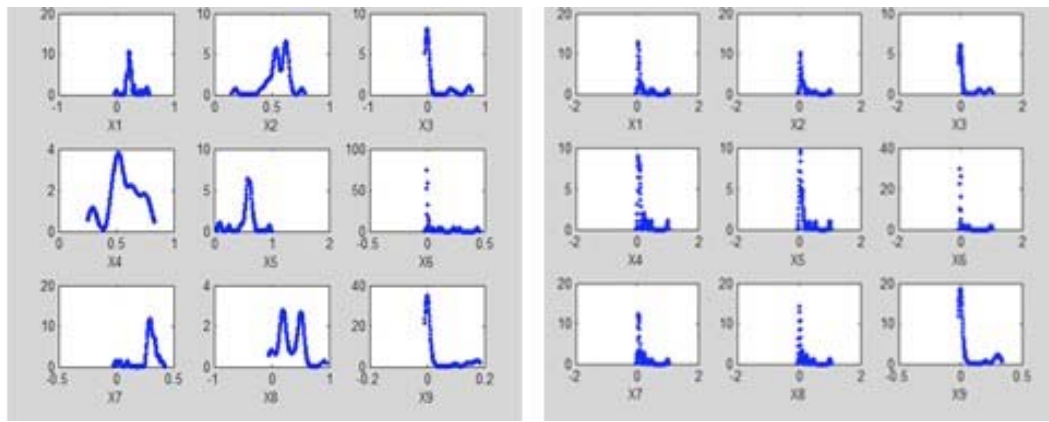


Fig. (1). Frequency histogram of glass’s origin data points and its potential values.

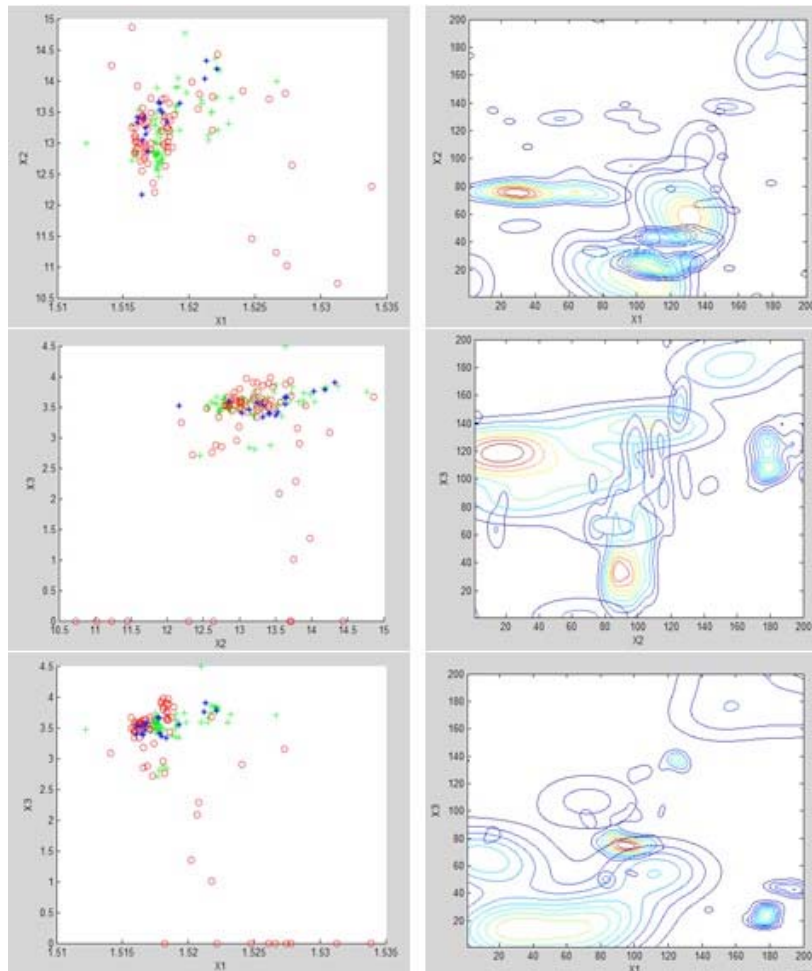


Fig. (2). Two-dimensional scatter plots and GDF equipotential line chart.

and competing classifiers. Therefore, the hyper-parameters should be determined without seeing the correct labels of the test data sets. In this work, we test the performance of classification algorithms by further separating training data into two parts, which are respectively used for hyper-parameter tuning and validation. We employ the ten-fold cross-validation procedure for each of the data sets in Table 1.

Through the above importance measure function, the scatter diagram and potential function values image of glass data set are given as follows:

$X_1, X_2 \dots X_9$  are the features of glass data set. Horizontal axe of sub graph is the range of each feature. Vertical axis is the frequency of the corresponding value.

From the comparison between origin data and potential data it can be seen that relative to the origin frequency distribution, the general data field frequency graph is more balanced and discrete. We take three categories data in glass data sets, and draw the scatter plots and data field image of  $X_1, X_2$  and  $X_3$ :

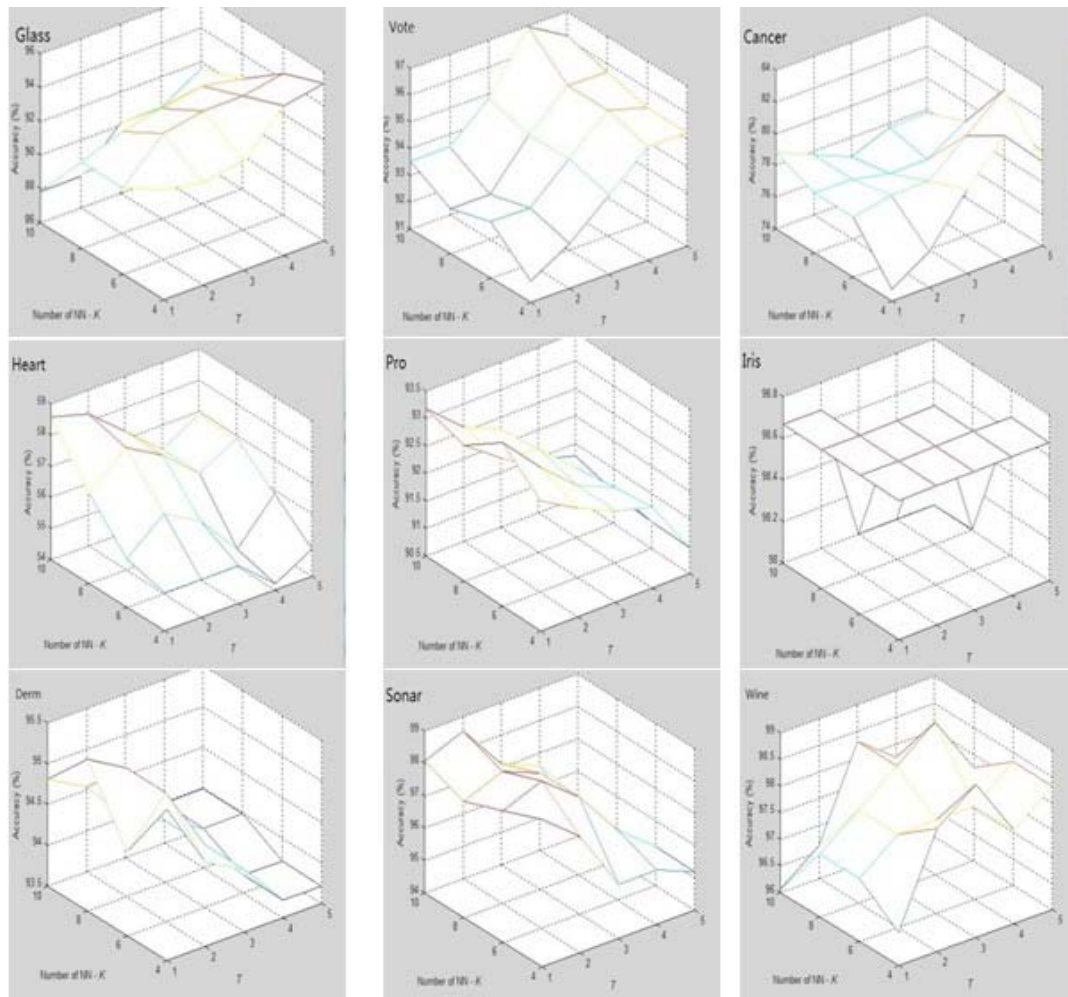


Fig. (3). Classification accuracies and parameters selection of ALH-GDF.

Table 2. Classification accuracies of 7 methods on 9 data sets.

Data set	k-NN	SVM	NFL	HKNN	CNN	ALH	MM-GDF
Cancer	79.8	79.8	74.2	76.3	69.7	81.8	<b>82.3</b>
Derm	96.7	<b>98.1</b>	95.9	97.8	95.9	97.27	95.35
Glass	72.4	72.9	68.2	74.3	72.4	74.3	<b>95.3</b>
Heart	<b>59.9</b>	58.6	52.5	54.2	49.2	56.9	58.59
Pro	89.5	91.4	90.3	91.1	90.8	91.57	<b>93.48</b>
Iris	96.7	98.0	89.3	97.3	92.0	96.67	<b>98.67</b>
Sonar	87.5	88.9	88.9	92.3	89.4	90.38	<b>98.56</b>
Wine	97.8	<b>98.9</b>	96.1	98.3	95.5	<b>98.9</b>	<b>98.9</b>
Vote	92.7	<b>97.0</b>	95.3	96.1	93.1	96.55	<b>97.0</b>
Average	85.89	87.07	83.41	86.41	83.11	87.15	<b>90.91</b>

X1, X2 and X3 are the features of glass data set. Each color of the left sub graph means one class. Every point in left sub graph corresponds one data point. The scatter plots can't describe the distribution rule of the data points. The lines in right sub graphs are the GDF equipotential lines. All data points are revolving around data field. The equipotential

lines are denser when the potential values are large. Conversely, the sparse equipotential lines mean the small corresponding potential values. The processed data points based on GDF are more discrete than the original data. The combinations of data points within same class are stronger. The positive role of GDF for classification is proved.

We compare GDF with several competing classifiers on 9 real data sets to prove the effectiveness of MM-GDF: (1) KNN; (2) SVM—Gaussian kernel; (3) NFL—nearest feature line; (4) HKNN; (5) CNN—center-based nearest neighbor; (6) ALH-Adaptive local hyper plane classification [17]; (7) MM-GDF. In MM-GDF method, the value of  $K$  is taken from 1 to 10, the range of  $T$  is 1 to 5, while the value of the parameter  $\sigma$  is given in accordance with the definition of the data field. The accuracies of MM-GDF are shown in Fig. (3) when the different parameters  $K$  and  $T$  are selected.

The leave-one-out cross-validation procedure is used to estimate the hyper parameters estimation and calculate the accuracies for all methods over all data sets. The results of all methods are shown in Table 2.

From the Table 2, we can see that MM-GDF is overall the best algorithm. The accuracy of MM-GDF is best on 7 data sets. It indicates that the optimal values of  $K$  and  $T$  can always be adaptively specified by MM-GDF during each run of the experiment. Moreover, it is interesting to see that the accuracy rate of the MM-GDF is much higher than the other classifiers, which indicates consistency and stability of the proposed model.

## CONCLUSION

In this paper, a new classification method MM-GDF is proposed. This method improves the maximal margin method by calculating the feature weight and data potential value in general data field. The sums of potential values used to calculate the weights of the between-group and within-group. The feature weight and the weighted Euclidean distance in general data field are then used for  $K$  data point's selection and local hyper plane construction. MM-GDF works very well on 9 UCI data sets. MM-GDF does not need any additional parameter. It is more practical. Further research is putting the algorithm on high dimension data sets and makes it more effective.

## CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

## ACKNOWLEDGEMENTS

This work was supported by National Natural Science Fund of China (61472039, 61173061, and 71201120).

## REFERENCES

- [1] V.N. Vapnik, 1995, The Nature of Statistical Learning Theory, Springer-Verlag, New York.
- [2] C.J.C. Burges, 1998, A tutorial on support vector machines for pattern recognition, Data Mining and Knowledge Discovery 2 (2), 121-167.
- [3] C. Domeniconi, J. Peng, D. 2002, Gunopulos, Locally adaptive metric nearest neighbor classification, IEEE Trans. Pattern Anal. Mach. Intell. 24 (9), 1281-1285.
- [4] B. Li, Y.W. Chen, Y.Q. Chen, 2008, The nearest neighbor algorithm of local probability centers, IEEE Trans. Syst. Man Cybern. 38 (1), 141-153.
- [5] S.N. Karl, Q.N. Truong, 2009, An adaptable k-nearest neighbors algorithm for MMSE image interpolation, IEEE Trans. Image Process. 18 (9), 1976-1987.
- [6] M.H. Tristan, R. Stephane, 2009, Tailored aggregation for classification, IEEE Trans. Pattern Anal. Mach. Intell. 31 (11), 2098-2105.
- [7] L. Devroye, L. Györfi, G. Lugosi, 1996, A Probabilistic Theory of Pattern Recognition, Springer-Verlag, pp. 71-75.
- [8] E. Blanzieri, F. Melgan, 2008, Nearest neighbor classification of remote sensing images with the maximal margin principle, IEEE Trans. Geosci. Remote Sensing 46 (6), 1804-1811.
- [9] G. Guo, C.R. Dyer, 2005, Learning from examples in the small sample case: face expression recognition, IEEE Trans. Syst. Man Cybern. 35, 477-488.
- [10] B. Zhang, S.N. Srihari, 2004, Fast k-nearest neighbor classification using cluster-based trees, IEEE Trans. Pattern Anal. Mach. Intell. 26, 525-528.
- [11] P. Vincent, Y. Bengio, 2002, K-local hyperplane and convex distance nearest neighbor algorithms, Advances in Neural Information Processing Systems (NIPS), vol. 14, MIT Press, Cambridge, MA, pp. 985-992.
- [12] Okun, 2007, K-local hyperplane distance nearest neighbor algorithm and protein fold recognition, Pattern Recognition Image Anal. 17 (4), 621-630.
- [13] V. Kecman, T. Yang, 2009, Adaptive local hyperplane for regression tasks, in: Proceedings of the International Joint Conference on Neural Networks, pp. 1566-1570.
- [14] Q.B. GAO, Z.Z. Wang, 2007, Center-based nearest neighbor classifier, Pattern Recognition 40, 346-349.
- [15] L. Nanni, 2006, A novel ensemble of classifiers for protein folds recognition, Neurocomputing 69, 2434-2437.
- [16] L. Nanni, A. Lumini, 2006, An ensemble of K-local hyperplanes for predicting protein-protein interactions, Bioinformatics 22, 1207-1210.
- [17] Yang T, Kecman V. 2008, Adaptive local hyperplane action. Neurocomputing, 71: 3001-3004.

Received: September 16, 2014

Revised: December 23, 2014

Accepted: December 31, 2014

© Long *et al.*; Licensee Bentham Open.

This is an open access article licensed under the terms of the (<https://creativecommons.org/licenses/by/4.0/legalcode>), which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.