Iranian Journal of Fuzzy Systems

Volume 16, Number 6, (2019), pp. 191-204

University of Sistan and Baluchestan

Robustified distance based fuzzy membership function for support vector machine classification

M. Mohammadi[1] and M. Sarmad[2]

[1,2] *Department of Statistics, Ferdowsi University of Mashhad, Iran*

manda.mohammadi@stu.um.ac.ir,sarmad@um.ac.ir

Abstract

Fuzzification of support vector machine has been utilized to deal with outlier and noise problem. This importance is achieved, by the means of fuzzy membership function, which is generally built based on the distance of the points to the class centroid. The focus of this research is twofold. Firstly, by taking the advantage of robust statistics in the fuzzy SVM, more emphasis on reducing the impact of outliers on the generalizability of SVM has been placed. Moreover, the variety of membership function for the elliptical data has been designated, based on the classic and robust Mahalanobis distance. Minimum covariance determinant and orthogonalised Gnanadesikan Kettenring estimators are employed in the structure of the robust–fuzzy SVM. By implementing the new membership function, the disadvantages of the traditional fuzzy membership function has been rectified. Simulated and real benchmarking data set confirm the effectiveness of the proposed methods. Compared with the traditional SVM and fuzzy SVM, these methods give a better performance on reducing the effects of outliers and significantly improves the classification accuracy and generalization.

Keywords: Support vector machine, Noise/outlier, Robust statistics, Fuzzy membership function, Minimum covariance determinant estimator, Orthogonalised Gnanadesikan Kettenring estimator.

2010 MSC: 68T10, 62H30, 62H86.

1 Introduction

Support vector machine (SVM) is a supervised classification technique which is derived from statistical learning theory. Due to the sensitivity of SVM to the outliers, the built decision boundary severely deviated from the optimal situation [17, 5]. To remedy this problem, various techniques have been proposed in the literature. A central SVM is proposed by Zhang [23] in which the class centers are utilized in the formulation of SVM. However, class center might be deficient, if some observations are located further from the centroid. As a substitute method to rectify the effect of outliers, a univariate robust measure namely the class median has been used by Kou [8]. On the other hand, robust kernels based on M-estimator for SVM has been proposed by Chen [3]. Such that, the influence of an outlier on both margin and training error is reduced by using the robust kernel. By considering the adaptive margin and "average" algorithm in Song et al. [17], other scheme of robust SVM is proposed to overcome the over fitting problem.

In addition to the above mentioned methods, "*fuzzification*" is another approach to prevent the adverse effect of outliers on the SVM accuracy. "*Membership function*" as a certain component of the fuzzy logic, try to assign the different weighting coefficient to each training sample that causes their effect on the position of SVM separating hyperplane can be controlled [9].

In addition to the mentioned properties, it is worthy of note that, choosing the proper membership function is an important step in the fuzzy SVM. According to Lin and Wang [9], one of the commonly chosen criterion to consider a point to be an outlier, is the distance to the center of the corresponding class. "Euclidean" distance is a common measure which is normally used in this respect. But this distance is not a proper metric for the data with correlated features. As the other alternative method, the "Mahalanobis" distance is proposed to measure the distance of each

Corresponding Author: M. Sarmad

Received: April 2017; Revised: December 2018; Accepted: April 2019.

point to the class center, while simultaneously the covariance of data is taking into account.
Note that, in both distances the arithmetic mean of each class is the primary choice for the class center. Unfortunately, this measurement is not robust to the outliers. A wide variety of robust estimators to substitute, are available in the literature (see [11, 12]), but the focus of this paper will be on the Minimum Covariance Determinant (MCD) [14] and Orthogonalised Gnanadesikan Kettenring (OGK) estimator [4], that are briefly explained in Section 2.
In this paper, "fuzzification" and "robustification" concepts have been associated which leads to highly diminish the sensitivity of SVM against the presence of outliers. So, the fuzzy membership functions using both distances based on the robust estimators is employed. Robust fuzzy SVM based on the MCD and OGK estimators is proposed, which is resulting to the increment of the SVM accuracy, as well as generalization ability.
Due to the different nature of these distances and their robust version, the precise inspection is required to make distinction to specify the optimal conditions for the proper use of the relevant method. It should be noted that, unlike the classic SVM, the degree of importance of data in different perspective of fuzzy SVM are not similar. The difference in weighting of the variety of fuzzy SVM membership functions with classic SVM is apparent in the Figure 1. With respect to Figure 1(a), it can be deduced that all observations are considered similarly with the same weight and no differentiation has been made among them which is obvious from the uniformity of the color of the graph.
Moreover, what can be seen form all of the fuzzy membership function depicted in the Figure 1 is that, the different parts of the chart vary in terms of degree of importance, in such a way that the points near the center are brighter and the darker the color of the regions, the more distant the center.
It is also obvious from 1(b) and 1(c), the Euclidean based membership function is not a suitable method for the data with correlated features, as assigning circle shape weight to this kind of data. In other words, equal weight has been considered for non semi–distance data to the center.
The other interesting fact from Figure 1(b) and 1(d) is that, for non robust distances (and consequently the membership functions), the presence of outlying observations has affected the center of observation, so that its location is drawn from the central part of the data towards the remote observation. This leads to assign a greater weight to the outliers and less weight to the other observations. In this case according to Barnett and Lewis [1], the masking phenomena has been occurred. But, fortunately the sensitivity of classical estimators to the presence of outliers can be resolve by using the robust version of them, as clear from Figure 1(c), 1(e) and 1(f). With respect to this figures, due to non movement of the center towards the outlying points, the sensitivity of the robust version of fuzzy membership to the outliers has been diminished.
On the other hand, in Figure 1(d), Mahalanobis based membership function fails to assign the proper weight to the data due to being covariance based structure of this distance which is doubly being affected by the outlying points. The utilization of robust Mahalanobis distance will result to insensitivity of the center to the presence of outliers and consequently assign weight to them according to their ellipsoidal shape (see Figures 1(e) and 1(f)). The details of the robust fuzzy membership functions is given in section 3. The layout of this paper is as follows. In the next section some preliminary concepts are recalled. The robust fuzzy membership function is discussed in Section 3. In Section 4, we apply our methods to a simple but illustrative toy-example and real data, then the experimental results are presented. The performance of different robust based membership functions are also compared and presented in Section 4. Finally, a conclusion is given in Section 5.

2 Some preliminary concepts

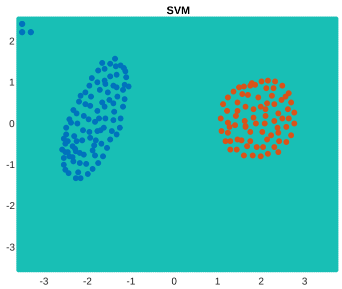
In this section, a brief review of the theory of SVM and fuzzy SVM is given. Some preliminary concepts for different fuzzy membership function based on the data with uncorrelated and correlated features has been explained. The other concepts to construct more robust version of fuzzy SVM, namely, MCD and OGK estimators are presented subsequently.

2.1 Support vector machine

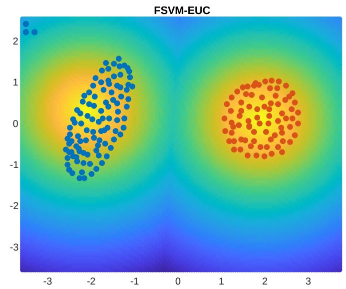
Support vector machine is the binary classification algorithm, aimed at maximizing the margin around the separating hyperplane and minimizing training error, simultaneously. The maximization of the margin leads to gain higher generalization ability. The designation of the binary classification problem is depicted as follows. Let S be a set of n training samples,

$$S = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_n, y_n)\},$$

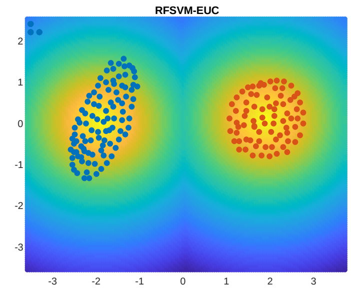
where $\mathbf{x}_i \in \mathbb{R}^p$ are p dimensional input vectors and $y_i \in \{-1, +1\}$ are the corresponding class labels, for $i = 1, \ldots, n$. The hyperplane can be represented as $\mathbf{w}^T\mathbf{x}_i - b = 0$. The hyperplane normal vector represents by $\mathbf{w}$ and b is a scalar which is determining the offset of the hyperplane from the origin. The optimal hyperplane can be obtained by the



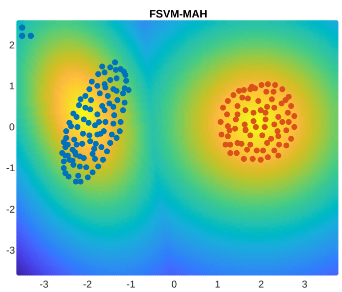
((a)) Classical SVM



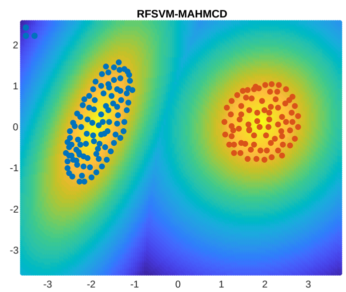
((b)) Fuzzy SVM using Euclidean distance



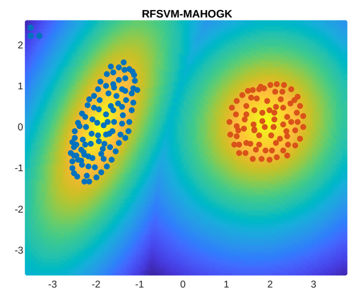
((c)) Robust fuzzy SVM using Euclidean distance



((d)) Fuzzy SVM using Mahalanobis distance



((e)) Robust fuzzy SVM using MCD based Mahalanobis distance



((f)) Robust fuzzy SVM using OGK based Mahalanobis distance

Figure 1: Difference of classical SVM, fuzzy SVM and variety of robust fuzzy SVM in terms of assigning weight to the observations

following primal optimization problem,

$$\min_{\mathbf{w},\boldsymbol{\xi}} \quad \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\boldsymbol{\xi}_i \tag{1}$$

subject to,

$$y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 - \boldsymbol{\xi}_i; \qquad \boldsymbol{\xi}_i \geq 0; \quad 1 \leq i \leq n.$$

For the non linearly separable data, a slack variable $\boldsymbol{\xi}$ has been added to the model. In order to define the constraint violation, the punishment parameter C, has to be determined in advance.
Finding the optimal hyperplane in (1) is feasible by transforming the primal into the dual form of the quadratic programming (QP) problem. The equivalent dual problem can be solved based on the Karush-Kuhn-Tucker (KKT) condition and Lagrange multipliers,

$$\max_{\boldsymbol{\alpha}} \quad \sum_{i=1}^{n}\boldsymbol{\alpha}i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\boldsymbol{\alpha}_i\boldsymbol{\alpha}_j y_i y_j \mathbf{x}_i^T\mathbf{x}_j \tag{2}$$

subject to,

$$\sum_{i=1}^{n}\boldsymbol{\alpha}_i y_i = 0; \qquad 0 \leq \boldsymbol{\alpha}_i \leq C; \quad 1 \leq i \leq n,$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$ is the vector of non-negative Lagrange multipliers. By solving the above quadratic optimization problem, $\boldsymbol{\alpha}_i$ and consequently, $\mathbf{w}^* = \sum_{i=1}^{n}\boldsymbol{\alpha}_i y_i \mathbf{x}_i$ is obtained. Based on the KKT condition, the bias term

$$b^* = y_i - \sum_{i=1}^{n}\boldsymbol{\alpha}_i^* y_i \mathbf{x}_i$$

can also be computed for any support vectors (the observations that their corresponding α_i are greater than zero). The sample point $\mathbf{x}_i$ is classified based on the sign of their classification function as follows,

$$f(\mathbf{x}) = \text{sign}(\mathbf{w}^T(\mathbf{x}_i) + b). \tag{3}$$

For the non-linear separable data in the feature space, the kernel function $K(\mathbf{x}_i, \mathbf{x}_j) = \boldsymbol{\phi}(\mathbf{x}_i).\boldsymbol{\phi}(\mathbf{x}_j)$ is used to find an optimal hyperplane in the higher dimensional space, where $\boldsymbol{\phi}(\mathbf{x}_i)$ is the non-linear mapping function.

2.2 Fuzzy support vector machine

Although, SVM is a powerful tool for the pattern classification; but its crisp approach leading to drops down the efficiency of this method in many applications [20, 17, 7]. The final decision of SVM depends on the small part of the training set, called "support vectors". This dependency will result in the sensitivity of this method to the noise and outliers [2, 5, 23]. On the other hand, in the practical problem, some training sample such as outliers can not definitely considered to belong to a specific class. So, it would be a crucial task to distinguish the normal data points from the outliers.

Fuzzy system proposed by Lotfi A. Zadeh [22], is built based on the fuzzy set theory and fuzzy logic. This method is useful to handle the complex real-world problems such as uncertainty and imprecision. Incorporation of the fuzzy set theory and classification techniques in handling the uncertainty and vagueness is highly beneficial which leads to enhancing the generalization ability of the classifiers. Recently, fuzzification of support vector machine algorithm have been rapidly developed, such as the work of Lin and Wang [9], Tang and Qu [18], Yang et al. [21], Tsujinishi and Abe [19] and Inoue and Abe [6].

Fuzzy SVM can be considered as an extension of support vector machine. It is capable of discarding the noisy data, by assigning low weight to them. The so called weight can be determined by the fuzzy membership function. Usually, fuzzy membership function is designed based on the distance of the points from the corresponding class center. A certain membership value $\boldsymbol{\mu}_i$ is given to each training sample $\mathbf{x}_i$, satisfying $\sigma \leq \boldsymbol{\mu}_i \leq 1$ with a sufficiently small constant $\sigma > 0$. Accordingly, the classical training sample S, is transformed to the fuzzy version,

$$S_f = \{(\mathbf{x}_1, y_1, \boldsymbol{\mu}_1), (\mathbf{x}_2, y_2, \boldsymbol{\mu}_2), \ldots, (\mathbf{x}_n, y_n, \boldsymbol{\mu}_n)\}.$$

The robust–fuzzy optimal separating hyperplane can be obtained based on the standard formulation of SVM, where fuzzification and robustification are latent in the membership $\boldsymbol{\mu}$; such that, each observation is gained the weight according to its distance to the robust center of the relevant class. It should be noted that, the fuzzification and robustification are carried out simultaneously and there is not any priority or delay in these process. The new membership function has to be embedded in the standard formulation of SVM.

$$\min_{\mathbf{w},\boldsymbol{\xi}} \ \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n} \boldsymbol{\xi}_i \boldsymbol{\mu}_i,$$

subject to,

$$y_i(\mathbf{w}^T\mathbf{x}_i + b) \geq 1 - \boldsymbol{\xi}_i; \qquad \boldsymbol{\xi}_i \geq 0; \ \ 1 \leq i \leq n.$$

The multiplication factor $\boldsymbol{\xi}_i\boldsymbol{\mu}_i$, can also be considered as a measure of error with different weight. Because it is combination of two term namely, the slack variable $\boldsymbol{\xi}_i$ and the attitude of a point towards the corresponding class $\boldsymbol{\mu}_i$ (Note that, the robustification of the membership function is discussed in Section 3). Similar to the traditional SVM, the following primal optimization problem can transform to the dual form and can be solved easily.

$$\max_{\boldsymbol{\alpha}} \ \sum_{i=1}^{n} \boldsymbol{\alpha} i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \boldsymbol{\alpha}_i \boldsymbol{\alpha}_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j$$

subject to,

$$\sum_{i=1}^{n} \boldsymbol{\alpha}_i y_i = 0; \qquad 0 \leq \boldsymbol{\alpha}_i \leq C\boldsymbol{\mu}_i; \ \ 1 \leq i \leq n,$$

The only difference of the fuzzy SVM with the classical SVM, is the upper bound of $\boldsymbol{\alpha}_i$, which is multiplication of the cost parameter C and the fuzzy membership function.

The low membership is assigned to the outlying observations and consequently, their effect on the classification has been reduced. On the other hand, the normal points have been gained a higher degree of membership.

2.3 Mahalanobis distance versus Euclidean distance

As already mentioned, in order to fuzzify the SVM, membership function has been employed, which is usually constructed based on the distance of each point to the class centroid. The most frequently used distance metrics in this respect, are Euclidean and Mahalanobis distance. But, which one to choose, depends on the shape of the classes.
Euclidean distance is wildly used for the data with uncorrelated features. However, the correlated features occur very often in many real world applications. So, the desired distance metric is the one which is inherently includes the covariance in its structure.
As the standard version of Euclidean distance, the Mahalanobis distance [10] is introduced. The aforementioned distance is suitable for the geometrical ellipsoidal shape data set. It is one common measure of distance, in the multivariate statistics. This distance, takes into account the correlation among variables, due to be considered the inverse of variance–covariance matrix in the computation. It can be defined as

$$MD = \sqrt{(\mathbf{x}-\bar{\mathbf{x}})^T \mathbf{S}^{-1}(\mathbf{x}-\bar{\mathbf{x}})}, \tag{4}$$

where $\mathbf{x}$ is the p–dimensional vector in $\mathbf{R}^p$, $\bar{\mathbf{x}}$ is the sample mean vector and $\mathbf{S}^{-1}$ denotes the inverse of variance-covariance matrix.
As previously mentioned, in the formulation of the membership function, the distance of each sample to the center of classes is considered. Based on this distance, the furthest points which are known as the outlying point, can be found. Actually this consideration has been done, for the sake of assigning each data points, different amount, to show the degree of importance of them.
However, the empirical mean as the most common measure of location, is highly affected by the outlying point [14]. Such that, a single data point located far away from the majority of data, has the ability to change this quantity, arbitrarily. Thus, in order to tackle the outlier problem, the utilization of the robust estimators has been offered. In such a way that, in the body of membership function, the classic center has been substituted by the robust centroid that is more resistant to the outliers.

2.4 Minimum covariance determinant (MCD)

One of the most common robust estimation of location and scale is the minimum covariance determinant which is proposed by Rousseeuw [16]. The MCD estimation aims to find the most compact set which contains approximately half of the data. This half–set, corresponds to $h = \lfloor \frac{(n+p+1)}{2} \rfloor$ observations. In order to find the optimum subset, all the possible subsets (C_h^n) of size h should be considered. The subset whose covariance matrix has the smallest determinant is the best choice to be the optimal subset

$$H_0 = \underset{H}{\operatorname{argmin}} \ \det(\operatorname{cov}(\mathbf{x}_i | i \in H)).$$

Let $\mathbf{X}$ denote a data matrix with n observations $(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ taken from p-dimensional space. The MCD estimator of location $\hat{\mu}_{MCD} = \frac{1}{h} \sum_{i \in H_0} \mathbf{x}_i$, is the average of the aforementioned subset and MCD estimator of the scale is the covariance matrix of the subset times a factor which is the multiplication of the consistency and finite sample correction [13]. It is worth mentioning that, the MCD estimator can be computed in a reasonable time using the FAST–MCD algorithm of Rousseeuw [15].

2.5 Orthogonalised gnanadesikan kettenring (OGK)

The orthogonalized Gnanadesikan–Kettenring (OGK) estimator is another recommended robust estimator derived by Maronna and Zamar [12]. OGK is the modification of the Gnanadesikan–Kettenring robust covariance estimate, that is initially proposed by Gnanadesikan and Kettenring [4]. It is approximately affine equivariant and positive definite estimator of scale. According to Gnanadesikan and Kettenring [4], this estimator begins with the robust estimate of covariance between two random variables X and Y. The pairwise covariance estimator of two variable is

$$\operatorname{cov}(X,Y) = \frac{1}{4}\{s(X+Y)^2 - s(X-Y)^2\}, \tag{5}$$

in which, s presents the standard division. The robust estimate will be obtained, if the robust estimate of variance is replaced with s. Let $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n \in \mathbf{R}^p$ be a data set. The general form of this estimator is obtained according to the following steps.

1. Let $m(.)$ and $s(.)$ be robust univariate location and scale estimators.
2. Construct $\mathbf{y}_i = \mathbf{D}^{-1}\mathbf{x}_i$ for $i = 1, 2, \dots, n$, based on the diagonal matrix $\mathbf{D}$, composed of variances of data ($\mathbf{D} = \text{diag}(s(X_1), s(X_2), \dots, s(X_p))$).
3. The correlation matrix $\mathbf{U}$ of $\mathbf{Y} = (Y_1, \dots, Y_p)$ is designed as

$$u_{jk} = \begin{cases} \frac{1}{4}[s(Y_j + Y_k)^2 - s(Y_j - Y_k)^2] & \text{if } j \neq k \\ 1 & \text{if } j = k \end{cases}$$

   The $\mathbf{U}$ matrix is symmetric but it is not positive definite.
4. The correlation matrix $\mathbf{U}$ is decomposed to the matrix contains its eigenvectors $\mathbf{E}$.
5. Project the data on the obtained eigenvectors of matrix $\mathbf{U}$, $\mathbf{Z} = \mathbf{YE}$.
6. Compute the robust variance of $\mathbf{Z}$, which is called $\mathbf{\Lambda} = \text{diag}(s^2(Z_1), \dots, s^2(Z_p))$. Set the $\hat{\boldsymbol{\mu}}(\mathbf{Y}) = \mathbf{Em}$, where $\mathbf{m} = (m(Z_1) \dots, m(Z_p))^T$ and compute the positive definite matrix $\hat{\mathbf{\Sigma}}(\mathbf{Y}) = \mathbf{E}\Lambda\mathbf{E}^T$
7. The robust estimators of location and scale are obtained by transforming back the initial data $\mathbf{X}$, i.e. $\hat{\boldsymbol{\mu}}_{OGK} = \mathbf{D}\hat{\boldsymbol{\mu}}(\mathbf{Y})$ and $\hat{\mathbf{\Sigma}}_{OGK} = \mathbf{D}\hat{\mathbf{\Sigma}}(\mathbf{Y})\mathbf{D}^T$.

3 Robust–Fuzzy SVM

In the fuzzy SVM, the attitude of each training point can be specified by the fuzzy membership function. Choosing a proper membership function is the fundamental step in fuzzy SVM. The membership function is indicator of the importance of the training samples to the decision surface. Such that, as the membership being bigger, the sample is more important than the others. So that, the different points have different impact on the learning process.
The membership is designated based on the distance between the sample points and the corresponding center. The following fuzzy membership function $\boldsymbol{\mu}_i$ with the above mentioned properties, is proposed by Lin and Wang [9].

$$\boldsymbol{\mu}_{EUC}(\mathbf{x}_i) = \begin{cases} 1 - \frac{\|\bar{\mathbf{x}}_+ - \mathbf{x}_i\|}{(r_+ + \delta)} & \text{if} \quad \mathbf{x}_i \in K_+ \\ 1 - \frac{\|\bar{\mathbf{x}}_- - \mathbf{x}_i\|}{(r_- + \delta)} & \text{if} \quad \mathbf{x}_i \in K_-. \end{cases}$$

Where, δ is an arbitrary small number to prevent the membership function to be zero. The negative and positive classes are shown by $K_\pm$. The class centers are defined as $\bar{\mathbf{x}}_\pm = \frac{1}{I_\pm}\sum_{i \in I_\pm} \mathbf{x}_i$, where $I_\pm = \sum_{i=1}^{n} I_{\mathbf{x}_\pm}(\mathbf{x}_i)$ and

$$I_{\mathbf{x}_\pm}(\mathbf{x}_i) = \begin{cases} 1 & \text{if} \quad \mathbf{x}_i \in K_\pm \\ 0 & \text{if} \quad \mathbf{x}_i \notin K_\pm, \end{cases}$$

in which $I_{\mathbf{x}_-}(\mathbf{x}_i)$ and $I_{\mathbf{x}_+}(\mathbf{x}_i)$ are presented the number of observation in each class. The radius of each class is depicted by $r_\pm = \max \|\bar{\mathbf{x}}_\pm - \mathbf{x}_i\|$. As the first proposed robust membership function, we need to robustify the center in the membership function by the robust estimators of location such as MCD.

$$\boldsymbol{\mu}_{EUCMCD}(\mathbf{x}_i) = \begin{cases} 1 - \frac{\|\bar{\mathbf{x}}_{MCD_+} - \mathbf{x}_i\|}{(r_+ + \delta)} & \text{if} \quad \mathbf{x}_i \in K_+ \\ 1 - \frac{\|\bar{\mathbf{x}}_{MCD_-} - \mathbf{x}_i\|}{(r_- + \delta)} & \text{if} \quad \mathbf{x}_i \in K_-. \end{cases}$$

The corresponding radius is robustified in the same manner as center, just by substituting the robust center. It should be noted that, due to the covariance based structure of the MCD estimator, it can not be computed for the high dimensional data. So, in this case, it is suggested to use any dimension reduction method such as principle component analysis to adjust the data to make them implementable for the fuzzy SVM.

3.1 Mahalanobis based membership function

Generally, real world situations often contain many problem with dependent data. In such cases, the Mahalanobis distance is an alternative to the Euclidean distance, that inversely weights the distance between the center of two data groups by the variances. The Mahalanobis based fuzzy membership function is depicted by,

$$\boldsymbol{\mu}_{MAH}(\mathbf{x}_i) = \begin{cases} 1 - \frac{\sqrt{(\mathbf{x}_i - \bar{\mathbf{x}}_+)^T \mathbf{S}_+^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}_+)}}{(r_+ + \delta)} & \text{if} \quad \mathbf{x}_i \in K_+ \\ 1 - \frac{\sqrt{(\mathbf{x}_i - \bar{\mathbf{x}}_-)^T \mathbf{S}_-^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}_-)}}{(r_- + \delta)} & \text{if} \quad \mathbf{x}_i \in K_-, \end{cases}$$

where, $r_{\pm} = \max(MD_{i\pm})$ are the Mahalanobis based class radius. The role of membership function is to specify the degree of importance of each point based on their position. That's why the radius of classes was utilized in the Euclidian based membership function. Similarly, in this method, the furthest point with the largest Mahalanobis is considered as the radius of each classes.

3.2 MCD based membership function

Due to be employed the conventional estimator of location and scale in the Mahalanobis distance; it is very sensitive to the outliers. So, as an alternative method, the MCD estimator of location and scale, is utilized. The robust–fuzzy membership function, can be presented as:

$$\boldsymbol{\mu}_{MAHMCD}(\mathbf{x}_i) = \begin{cases} 1 - \frac{\sqrt{(\mathbf{x}_i - \bar{\mathbf{x}}_{MCD_+})^T \mathbf{S}_{MCD_+}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}_{MCD_+})}}{(r_{MCD_+} + \delta)} & \text{if} \quad \mathbf{x}_i \in K_+ \\ 1 - \frac{\sqrt{(\mathbf{x}_i - \bar{\mathbf{x}}_{MCD_-})^T \mathbf{S}_{MCD_-}^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}_{MCD_-})}}{(r_{MCD_-} + \delta)} & \text{if} \quad \mathbf{x}_i \in K_-. \end{cases}$$

$\bar{\mathbf{x}}_{MCD_{\pm}}$ and $\mathbf{S}_{MCD_{\pm}}$ are the MCD estimation of location and scale of positive and negative class, respectively. The class radius based on the robust Mahalanobis using MCD estimators is presented by $r_{\pm} = \max(RMD_{i_{\pm}})$, where RMD is the robust mahalanobis distance.
It is worth taking into consideration that, due to implementation of the robust measure of location and scale, which is not unduly affected by the presence of outlying point, the SVM performance would not be affected.

3.3 OGK based membership function

The fuzzy membership function based on the OGK estimator, can be tuned similar to the MCD based membership function. To this end, just need to replace the MCD estimate of location and scale with the OGK estimators.

3.4 Algorithm outline of the robust fuzzy SVM

For more clarification, the algorithmic schemes of the robust fuzzy SVM is summarized as follows. It should be noted that, no differentiation has been made between the variety of the proposed robust fuzzy membership function in the following algorithm.

Algorithm 1: Algorithm outline for the robust fuzzy SVM

1 **Input:** A data matrix $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^n$, labels vector $\mathbf{Y} = \{\mathbf{y}_i\}_{i=1}^n$ and δ
2 **Output:** A robust fuzzy SVM
3 **Initialize:** $\boldsymbol{\mu}_{Robust_i} = \{\}$ // A robust fuzzy membership function parametrized by $\bar{\mathbf{x}}_{Robust_{\pm}}$,$\mathbf{S}_{Robust_{\pm}}^{-1}$ and $r_{Robust_{\pm}}$, where they are respectively center, covariance matrix and radius of negative and positive classes.
4 **for** i=1 : n
5 **do** $\boldsymbol{\mu}_{Robust_i} = \boldsymbol{\mu}_{Robust_i}(\mathbf{x}_i)$
6 **end for**
7 Train a SVM model using the robust fuzzy membership function

4 Experiment and Analysis

In this section a couple of linearly separable artificial data generated. Data with uncorrelated and correlated features are used to check the performance of the fuzzy membership function based on the classic and robust distances. In each scenario, the suitable distance has been utilized. It should be noted that, in final part of this section, some artificial non-linearly separable and real data have been employed as well.
For the experimental analysis, MATLAB 2016b as well as `svmtrain` toolbox are used. The built in function of MATLAB called `mvnrnd` is utilized to generate the data from the normal distribution. The artificial data, are partitioned to training and testing set. Such that 70% is chosen for the training set and the remaining of the observation has been assigned to the testing set. Since, the classification model has been produced in the training phase, we have only contaminated the training set to check the robustness of the proposed models. It should be noted that, the testing set is free of outliers. Designing the classifiers which are robust to the presence of outliers and have the ability of learn the model even in

such case is our aim.
For the real data, 10 folds cross-validation has been used in order to determine the best parameters of the model. This strategy is based on performing 10 iterations, such that each data set is randomly divided into 10 portion. In each iteration nine portion among ten is chosen as the training set and the remained one is used as the testing set. Each portion is used once as testing set. Then, in order to find the performance of the models, the average of all performance among 10 iteration is computed.
It is worth to note that, cross-validation was carried out only on the real data set. But, for the artificial data set, it was attempted to add outliers to the training data, so the random partitioning or cross-validation is not possible. The clean test set is generated initially and the performance of the model is computed based on this clean test set.
The performance of the model for both of the above mentioned procedures, including the division of data into the train and test set for the artificial data, as well as cross-validation method, for the real data are computed based on 100 times iterations and the averaging results has been reported in Tables 1- 5.

4.1 Artificial data set

In order to check the performance of the proposed methods, different structures regarding to whether there is dependency between data features or not, has been considered. Data with uncorrelated and correlated features are respectively considered in the following sub-sections. This structure is straightforwardly related to the covariance of the data. In the following examples, we have just generated the data based on the arbitrary simulation set-up. Our intention was just presenting the dependency and in-dependency of the data, which is very common in real world application. So, the diagonal and non-diagonal covariance matrix with different elements has been chosen in this respect. Moreover, the proposed methods are checked on some artificial non-linear separable data set as well.

4.1.1 Data set with uncorrelated features

Totally for each class, 400 observations were generated from a Gaussian distribution

$$N(\mathbf{x}|\mu_\pm, \Sigma_\pm) = \frac{1}{\sqrt{(2\pi)^d|\Sigma_\pm|}} \exp[-\frac{1}{2}(\mathbf{x}-\mu_\pm)^T\Sigma_\pm^{-1}(\mathbf{x}-\mu_\pm)]. \tag{6}$$

The mean vector and covariance matrix of positive and negative class are shown by $\boldsymbol{\mu}_+$, $\boldsymbol{\mu}_-$ and $\boldsymbol{\Sigma}_\pm$. Note that, the mean vectors of classes are different, but the covariance matrices are chosen similar. The off-diagonal elements of this matrix is the covariance between i^{th} and j^{th} elements of a random vectors and the diagonal elements are the variances.

$$\boldsymbol{\mu}_+ = \begin{bmatrix}2\\2\end{bmatrix} \qquad \boldsymbol{\mu}_- = \begin{bmatrix}-2\\-1\end{bmatrix} \qquad \text{and} \qquad \boldsymbol{\Sigma}_\pm = \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}.$$

As the outlying points, a small portion of data, is generated from $MVN(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*)$ with negative label. 10 percent of the observations are randomly substituted by the simulated bivariate normal observations that are generated as the outlying points.

$$\boldsymbol{\mu}^* = \begin{bmatrix}1\\5\end{bmatrix} \qquad \text{and} \qquad \boldsymbol{\Sigma}^* = \begin{bmatrix}2.5 & 1\\1 & 1\end{bmatrix}.$$

Figure 2 presents the simulated data, where negative and positive class has been showed by “circle” and “square” respectively. The outlying points can be recognized as their distance to the center of data mass. The clean testing data set is depicted by black coloured circle and square as negative and positive classes.
As argued in the literature, the fuzzy SVM assign different membership value to each training sample, which is resulting in decreasing the effect of outliers and obtaining better classifier models. In Table 1, the classification accuracy of SVM, fuzzy SVM based on the classic Euclidean distance and robust Euclidean distance, alternative fuzzy membership function using the classic Mahalanobis distance and the robust Mahalanobis distance based on the MCD and OGK is depicted. From now on, to simplify notation, we write FSVM–EUC, RFSVM–EUC, FSVM–MAH, RFSVM–MAHMCD and RFSVM–MAHOGK instead of fuzzy membership function using classic and robust Euclidean distance, classic and robust Mahalanobis distance, respectively.

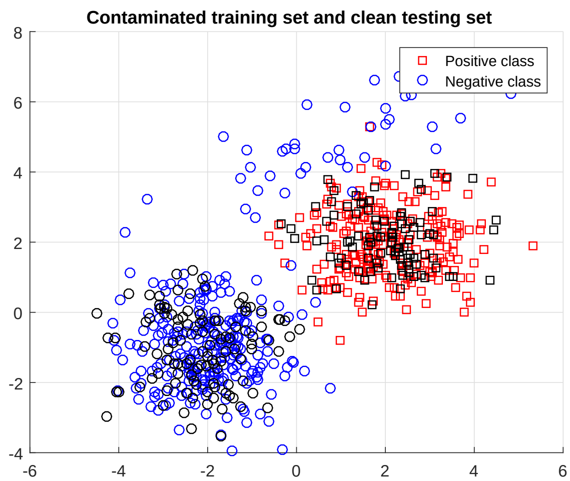


Figure 2: Scatter plot of the generated training data with outlying points and testing data without outliers

Table 1: Classification accuracy (in %) estimated at 95% confidence intervals of the SVM, FSVM–EUC, RFSVM–EUC, FSVM–MAH, RFSVM–MAHMCD and RFSVM–MAHOGK for data with uncorrelated features and different value of C parameter, where range from 2 to 2^8.

C	SVM	FSVM–EUC	RFSVM–EUC	FSVM–MAH	RFSVM–MAHMCD	RFSVM–MAHOGK
2	95.71 ±0.18	96.19 ±0.10	**97.14 ±0.02**	96.19 ±0.04	96.67 ±0.08	96.67 ±0.30
2^2	94.28 ±0.05	95.71 ±0.05	**97.62 ±0.01**	95.24 ±0.10	96.19 ±0.13	97.14 ±0.04
2^3	95.24 ±0.11	96.67 ± 0.03	**97.62 ±0.03**	95.71 ±0.80	97.14 ±0.05	96.19 ±0.07
2^4	95.71 ± 0.04	97.14 ±0.07	**98.09 ±0.01**	97.14 ±0.12	**98.09 ±0.03**	97.14 ±0.09
2^5	95.71 ±0.02	97.14 ±0.04	**98.57 ±0.02**	96.67 ±0.08	98.09 ± 0.04	97.14 ±0.11
2^6	96.67 ±0.20	97.14 ±0.01	97.14 ±0.04	97.14 ±0.03	97.14 ±0.06	**97.62 ±0.02**
2^7	92.86 ±0.07	93.81 ±0.05	**95.71 ±0.01**	95.24 ±0.05	95.24 ±0.03	95.24 ±0.03
2^8	94.28 ±0.15	95.71 ±0.10	**97.14 ±0.03**	94.76 ±0.03	96.67 ±0.05	96.67 ±0.03

It can be seen from Table 1 , that the accuracy of robust Euclidean distance is higher than the classic Euclidean distance and also the rest of the proposed membership functions. Both recommended robust Mahalanobis distance are outperformed the classic Euclidean and Mahalanobis distances, where in most of the cases, the MCD has better performance compared to the OGK.

4.1.2 Data set with correlated features

The second scenario, is designed for the data with correlated features. Similar to the previous simulation scheme 400 observations were generated from a bivariate Gaussian distribution as follows

$$\boldsymbol{\mu}_+ = \begin{bmatrix} 2 \\ 2 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}_+ = \begin{bmatrix} 1.5 & 0.9 \\ 0.9 & 1.3 \end{bmatrix}$$

$$\boldsymbol{\mu}_- = \begin{bmatrix} -3 \\ -3 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}_- = \begin{bmatrix} 1.3 & 0.9 \\ 0.9 & 1.4 \end{bmatrix}.$$

The percent of the outlying points is exactly similar to the first scenario. They are negative labelled bivariate normal distributed observations with the following mean and covariance.

$$\boldsymbol{\mu}^* = \begin{bmatrix} 2 \\ 4 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Sigma}^* = \begin{bmatrix} 0.2 & -0.1 \\ -0.1 & 0.2 \end{bmatrix}$$

Table 2: Classification accuracy (in %) estimated at 95% confidence intervals of the SVM, FSVM–EUC, FSVM–MAH, RFSVM–MAHMCD and RFSVM–MAHOGK for data with correlated features and different value of C parameter, where range from 2 to 2^8.

C	SVM	FSVM–EUC	FSVM–MAH	RFSVM–MAHMCD	RFSVM–MAHOGK
2	96.19 ±0.05	97.62 ±0.04	96.67 ±0.16	97.62 ±0.03	**98.09 ± 0.03**
2^2	94.28 ±0.07	96.19 ±0.05	95.24 ± 0.80	96.19 ±0.12	**96.67 ±0.03**
2^3	96.19 ±0.04	97.62 ±0.06	96.67 ±0.03	**98.09 ±0.01**	**98.09 ±0.02**
2^4	96.19 ±0.02	98.09 ±0.30	96.67 ±0.12	**98.57 ±0.04**	**98.57 ±0.01**
2^5	95.24 ±0.03	97.14 ±0.08	96.19 ±0.10	97.62 ±0.05	**98.09 ±0.04**
2^6	96.67 ±0.17	98.09 ±0.03	97.14 ±0.04	**99.05 ±0.03**	**99.05 ±0.01**
2^7	95.11 ±0.03	96.19 ±0.18	95.24 ±0.12	97.14 ±0.06	**98.09 ±0.01**
2^8	93.81 ±0.15	97.14 ±0.03	94.76 ±0.05	**97.62 ±0.02**	97.14 ±0.05

With respect to Table 2, the robust fuzzy membership function based on the Mahalanobis distance using OGK estimator, is outperformed the other competitors. In the comparison of robust membership function based on the OGK and MCD estimators, in most of the cases OGK performance is better than MCD. While MCD based membership function can be considered as the second best method. Fuzzy membership function based on the classical Mahalanobis distance is better than the classic SVM.

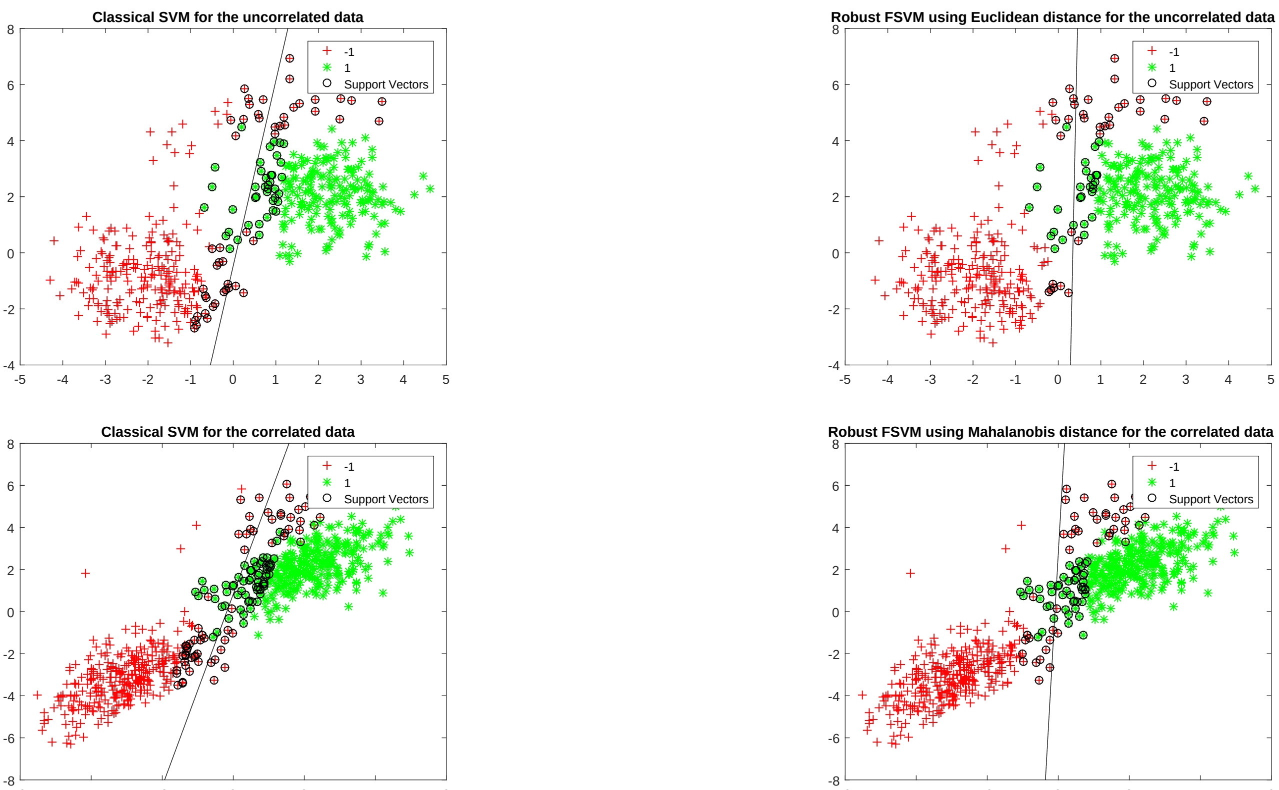
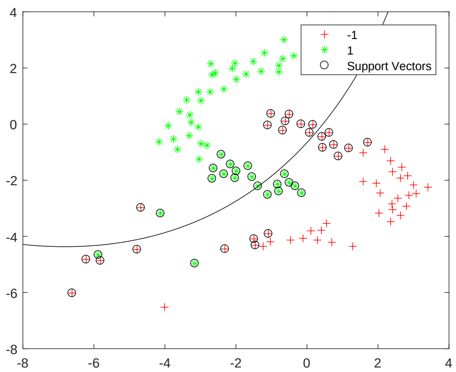


Figure 3: Separating boundary line using classical SVM and the proposed robust fuzzy SVM for the data with uncorrelated and correlated features

According to Figure 3, the sensitivity of SVM to the outliers has been shown for both structures of data, whether being with uncorrelated features or correlated. Existence of outliers in the data will lead to occurrence of the "masking" and "swamping" phenomena [1]. Such that, the true center will be misplaced to be further from the majority of data and be closer to the outlying points. Subsequently, some outliers which are closer to the center gain higher membership while some normal points gain less. Whereas, they are not worth to get such weights. So, this affect the hyperplane and consequently SVM accuracy. But, due to allocating each points different weight based on the robust fuzzy membership function, the place of hyperplane has not affected by the outlying points.

4.1.3 Non linearly separable data set

In order to check the performance of the proposed method, some well-known artificial data sets such as half–moon shape and XOR data set which are contaminated by some outlying points are used in this section. The Gaussian kernel with the parameter $\sigma = 5$ has been utilized to classify these data sets.



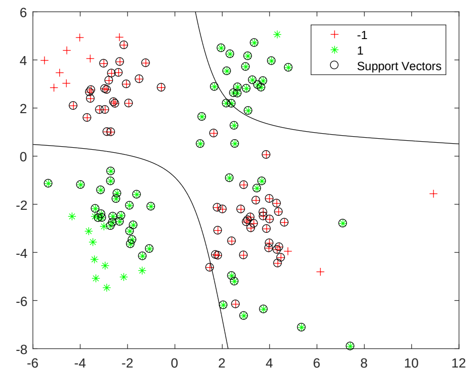


Figure 4: Classification of half–moon shape (left panel) and XOR (right panel) data using Gaussian kernel with the parameter $\sigma = 5$

Table 3: Classification accuracy (in %) estimated at 95% confidence intervals of the SVM, FSVM–EUC, RFSVM–EUC, FSVM–MAH, RFSVM–MAHMCD and RFSVM–MAHOGK

Data set	Classic SVM	FSVM–EUC	RFSVM–EUC	FSVM–MAH	RFSVM–MAHMCD	RFSVM–MAHOGK
Half–moon	87.79±0.40	89.83± 0.05	91.83 ±0.05	89.97± 0.10	**93.36 ± 0.01**	93.17± 0.03
XOR	61.00 ± 0.04	70.00 ±0.12	80.00 ± 0.05	91.66 ± 0.03	**98.33± 0.03**	96.67± 0.04

For the non-linear data, the robust fuzzy SVM based on the MCD has the best performance among the other methods. Note that, according to results obtained form the simulated data which is reported in Tables 1, 2 and 3, the higher accuracy of our proposed method among the other competitors, reflect the sufficiency of them in terms of generalization ability.

4.2 Real data set

In order to evaluate the performance of the proposed method some well known two classes benchmark data sets are used. The Diabetes, Heart and Bupa data set from UCI machine learning repository [1] and Biomed dat set from StatLib–Datasets Archive [2] are taken. It should be noted that, the Gaussian kernel has been used to classify these real world data sets. The chosen data set with their related characteristics are reported in Table 6. The performance of our proposed methods, SVM and different fuzzy SVM proposed by Lin [9] and Yang et al. [21] ($y' - SVM$) is reported in Table 4 and 5.

Table 4: Classification accuracy (in %) estimated at 95% confidence intervals obtained from testing set for the SVM, FSVM–EUC, RFSVM–EUC, FSVM–MAH, RFSVM–MAHMCD and RFSVM–MAHOGK

Data set	Classic SVM	FSVM–EUC	$y' - SVM$	RFSVM–EUC	FSVM–MAH	RFSVM–MAHMCD	RFSVM–MAHOGK
Diabetes	74.08 ± 0.10	72.06± 0.03	77.34	**82.26 ± 0.03**	74.11 ± 0.03	77.85 ± 0.07	77.09 ± 0.04
Biomed	71.00 ± 0.02	71.47± 0.03	89.06	**94.93 ± 0.03**	77.80 ± 0.05	89.86 ± 0.06	89.12 ± 0.02
Heart	83.11 ± 0.32	85.00 ± 0.06	85.19	86.75 ± 0.05	84.33 ± 0.04	**90.87± 0.05**	87.52 ± 0.01
Bupa	70.09± 0.01	71.63± 0.05	74.78	**77.87± 0.04**	72.52± 0.02	76.01 ± 0.14	74.93 ± 0.03

[1] http://archive.ics.uci.edu/ml/index.php
[2] http://lib.stat.cmu.edu/datasets/

Table 5: Classification accuracy (in %) estimated at 95% confidence intervals obtained from training set for the SVM, FSVM–EUC, RFSVM–EUC, FSVM–MAH, RFSVM–MAHMCD and RFSVM–MAHOGK

Data set	Classic SVM	FSVM–EUC	RFSVM–EUC	FSVM–MAH	RFSVM–MAHMCD	RFSVM–MAHOGK
Diabetes	75.01 ± 0.06	73.91± 0.05	84.02± 0.02	76.56±0.03	80.11± 0.01	79.68± 0.04
Biomed	73.03± 0.01	73.89± 0.04	96.12 ± 0.02	79.98 ± 0.05	90.80± 0.06	90.75 ± 0.05
Heart	85.82 ± 0.01	86.03 ± 0.03	88.63 ± 0.02	87.42 ± 0.05	91.05 ± 0.03	89.10 ± 0.07
Bupa	73.38± 0.02	73.59 ± 0.31	79.98 ± 0.04	75.07 ± 0.04	77.34 ± 0.01	75.84± 0.06

Note that, the reported classification accuracy in Tables, 1–5, is the average of 100 times repeated experiments accompanied by the 95% confidence intervals. It can be seen from Table 4 that the robust fuzzy membership function based on the Euclidean distance outperformed the other methods. So, RSVM-EUC is preferred rather than the rest of proposed methods for these three data sets. Also, the robust fuzzy membership based on MCD is the second ranked method.

Table 6: The benchmark data set

Data set	No. of observation	No. of features
Diabetes	768	8
Biomed	345	6
Heart	270	13
Bupa	209	4

4.3 Statistical comparison of results

The significance difference of the proposed methods over the existing classical SVM and the fuzzy SVM [9, 21] is obtained based on the paired t-test. The results of comparison of our proposed methods with the aforementioned method implies that, there are significant difference between RFSVM-EUC, RFSVM-MCD and RFSVM-OGK with classical SVM and the fuzzy SVM [9]. There is also evidence of significant difference between the mentioned method and fuzzy SVM of Yang[21]). Table 6 is devoted to the p-values of paired t-test related to the above mentioned methods. It should be

Table 7: P-values from paired t-test

Algorithm	SVM	FSVM	$y^{'} - SVM$
RFSVM-EUC	0.031	0.028	0.006
FSVM-MAH	**0.080**	**0.054**	**0.455**
RFSVM-MAHMCD	0.027	0.019	0.046
RFSVM-MAHOGK	0.038	0.029	0.041

noted that, at the 5% significance level, there is no difference between FSVM-MAH and the existing methods. The bold text is indicated the methods that are not statistically different.

5 Conclusions

The presence of outlier or noise can adversely affect the performance of support vector machine, which leads to weakening the generalization ability of this classifier. So, to overcome this problem, a variety of fuzzy membership function which is suited to different data structure is proposed in this paper. Our first method, tries to deactivate the outlier and noise effect by robustification of Euclidean distance. On the other hand, the Mahalanobis distance, which is usually employed for the elliptical shape data set, is used. Moreover, some robust estimators such as minimum covariance determinant and orthogonalised Gnanadesikan Kettenring are embedded in the formulation of Mahalanobis distance to make it resistance to the outliers. Therefore, by robustification of both distances employed in the fuzzy membership function, we have bilaterally reduced the impact of noise and outliers on the SVM accuracy. The performance of all above mentioned method with the classical SVM and fuzzy SVM has been compared. For the data with uncorrelated features, fuzzy membership function based on the robust Euclidean distance is outperformed the other method. The

MCD based Mahalanobis membership function ranked second behind the robust Euclidean. Additionally, for the data with correlated features, the performance of the membership function based on the robust Mahalanobis distance using OGK estimators is the best, which is followed by the MCD based membership function. Due to dependency of the results to the data structure, it is recommended to apply all the proposed fuzzy membership and choose the one with higher accuracy.

Acknowledgement

This paper is a part of the Ph.D thesis of Mandana Mohammadi (NO. 37943).

References

[1] V. Barnett, T. Lewis, *Outliers in statistical data*, Wiley, 1974.

[2] B. E. Boser, I. M. Guyon, V. N. Vapnik, *A training algorithm for optimal margin classifiers*, Proceedings of the Fifth Annual Workshop on Computational Learning Theory, ACM, (1992), 144–152.

[3] J. H. Chen, *M-estimator based robust kernels for support vector machines*, ICPR 2004, Proceedings of the 17th International Conference on Pattern Recognition, IEEE, **1** (2004), 168–171.

[4] R. Gnanadesikan, J. R. Kettenring, *Robust estimates, residuals, and outlier detection with multiresponse data*, Biometrics, **28** (1972), 81–124.

[5] I. Guyon, N. Matic, V. Vapnik, et al, *Discovering informative patterns and data cleaning*, 1996.

[6] T. Inoue, S. Abe, *Fuzzy support vector machines for pattern classification*, 2001. Proceedings. IJCNN'01. International Joint Conference on Neural Networks, IEEE, **2** (2001), 1449–1454.

[7] V. Kecman, *Learning and soft computing: Support vector machines, neural networks, and fuzzy logic models*, MIT press, 2001.

[8] Z. Kou, J. Xu, X. Zhang, L. Ji, *An improved support vector machine using class-median*, Proc of 8th Intl Conf on Neural Information Processing, **2** (2001), 883–887.

[9] C. F. Lin, S. D. Wang, *Fuzzy support vector machines*, IEEE Transactions on Neural Networks, **13**(2) (2002), 464–471.

[10] P. C. Mahalanobis, *On the generalized distance in statistics*, National Institute of Science of India, 1936.

[11] R. Maronna, R. D. Martin, V. Yohai, *Robust statistics*, John Wiley & Sons, Chichester, ISBN, **1**, 2006.

[12] R. A. Maronna, R. H. Zamar, *Robust estimates of location and dispersion for high-dimensional datasets*, Technometrics, **44**(4) (2002), 307–317.

[13] G. Pison, S. Van Aelst, G. Willems, *Small sample corrections for LTS and MCD*, Developments in Robust Statistics, Springer, **55** (2003), 330–343.

[14] P. J. Rousseeuw, *Least median of squares regression*, Journal of the American Statistical Association, **79**(388) (1984), 871–880.

[15] P. J. Rousseeuw, K. V. Driessen, *A fast algorithm for the minimum covariance determinant estimator*, Technometrics, **41**(3) (1999), 212–223.

[16] P. J. Rousseeuw, M. Hubert, *Robust statistics for outlier detection*, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, **1**(1) (2011), 73–79.

[17] Q. Song, W. Hu, W. Xie. *Robust support vector machine with bullet hole image classification*, IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), **32**(4) (2002), 440–448.

[18] H. Tang, L. S. Qu, *Fuzzy support vector machine with a new fuzzy membership function for pattern classification*, International Conference on Machine Learning and Cybernetics, 2008, IEEE, **2** (2008), 768–773.

[19] D. Tsujinishi, S. Abe, *Fuzzy least squares support vector machines for multiclass problems*, Neural Networks, **16**(5-6) (2003), 785–792.

[20] V. Vapnik, *The nature of statistical learning theory*, Springer Science & Business Media, 2013.

[21] C. Y. Yang, J. J. Chou, F. L. Lian, *Robust classifier learning with fuzzy class labels for large-margin support vector machines*, Neurocomputing, **99** (2013), 1–14.

[22] L. A. Zadeh, *Fuzzy sets*, Information and Control, **8**(3) (1965), 338–353.

[23] X. Zhang, *Using class-center vectors to build support vector machines*, Neural Networks for Signal Processing IX, 1999. Proceedings of the 1999 IEEE Signal Processing Society Workshop, IEEE, (1999), 3–11.

Archive of SID

Iranian Journal of Fuzzy Systems

Volume 16, Number 6 (2019), pp. 220

University of Sistan and Baluchestan

Robustified distance based fuzzy membership function for support vector machine classification

M. Mohammadi1 and M. Sarmad

تابع عضویت فازی مبتنی بر فاصله مقاوم برای ماشین بردار پشتیبان

چکیده. در این مقاله، روش فازی‌سازی ماشین بردار پشتیبان برای مقابله با تأثیر داده‌های پرت و دارای اغتشاش مورد استفاده قرار گرفته‌است. این مهم با استفاده از تابع عضویت فازی که ساختار آن غالباً بر اساس فاصله نقاط تا مرکز هر یک از کلاس‌ها است، ساخته می‌شود. تمرکز این پژوهش در قالب دو هدف اصلی ارائه می‌شود. ابتدا با بکارگیری آمار مقاوم در تئوری ماشین بردار پشتیبان فازی، تأکید بیشتر بر کاهش تأثیر داده‌های پرت بر قابلیت تعمیم‌پذیری ماشین بردار پشتیبان صورت گرفته‌است. علاوه‌براین، توابع عضویت متنوعی بر اساس فاصله ماهالانوبیس کلاسیک و مقاوم برای داده‌های بیضوی طراحی شده‌است. برآوردگرهای کمترین دترمینان کوواریانس و نانادیسکان- کتنرینگ متعامد در ساختار ماشین بردار پشتیبان مقاوم و فازی بکار رفته‌اند. با بکارگیری توابع عضویت جدید، معایب توابع عضویت فازی قدیمی برطرف شده‌است. اثربخشی روش‌های پیشنهادی با استفاده از مجموعه داده‌های شبیه‌سازی شده و واقعی تأیید گردید. در مقایسه با ماشین بردار پشتیبان قدیمی و فازی، روش‌های پیشنهادی عملکرد بهتری در کاهش اثرات داده‌های پرت دارند و به طور قابل توجهی صحت طبقه‌بندی و قابلیت تعمیم‌پذیری را بهبود می‌بخشند.