

A New Fuzzy Membership Computation Method for Fuzzy Support Vector Machines

Trung Le, Dat Tran, Wanli Ma and Dharmendra Sharma

Faculty of Information Sciences and Engineering

University of Canberra, Australia

Email: {trung.le, dat.tran, wanli.ma, dharmendra.sharma}@canberra.edu.au

Abstract—Support vector machine (SVM) considers all data points with the same importance in classification problems, therefore SVM is very sensitive to noisy data or outliers. Current fuzzy approach to two-class SVM introduces a fuzzy membership to each data point in order to reduce the sensitivity of less important data, however computing fuzzy memberships is still a challenge. It has been found that the performance of fuzzy SVM highly depends on the computation of fuzzy memberships, hence in this paper, we propose a new method to compute fuzzy memberships and we also extend the fuzzy approach for two-class SVM to one-class SVM. Experiments performed on a number of popular data sets to evaluation the proposed fuzzy SVMs show promising classification results.

I. INTRODUCTION

Current support vector machine (SVM) is sensitive to noise and outliers because SVM assumes all data points have the same importance in classification problems. In order to address this problem, a fuzzy approach was introduced to reduce the sensitivity of less important data [6]. This approach assigns a fuzzy membership value as a weight to each training data point and uses this weight to control the importance of the corresponding data point. A similarity measure function to compute fuzzy memberships were introduced in [9]. However, they had to assume that outliers should be somewhat separate from the normal data. A recent investigation [7] has investigated the effect of the trade-off parameter C to the model of conventional two-class SVM and introduced a triangular membership function to set higher grades to the data points in regions containing data of both classes. However this method could be applied with some assumptions involved and hence computing fuzzy memberships is still a challenge. It has been found that the performance of fuzzy SVM highly depends on the determination of fuzzy memberships therefore in this paper, we propose a new method to compute fuzzy memberships for the current fuzzy two-class SVM and we also extend the fuzzy approach as well as the computation method to one-class and multi-class SVMs. In our viewpoint, regions in feature space containing data points of different classes are more important than other regions containing data points of one class only. Fuzzy memberships of data points in those mixed regions should have the highest membership value and data points in other regions should have lower fuzzy membership values. We propose to use a fuzzy clustering technique to determine clusters in mixed regions and fuzzy memberships for other data points are determined by their closest cluster

accordingly. Experiments performed on a number of public data sets using fuzzy c -means clustering technique for fuzzy SVM show that the proposed fuzzy membership computation method can provide promising classification rates.

The paper is organised as follows. Section 2 reviews current SVMs. Section 3 reviews fuzzy SVM for two-class classification. Section 4 presents our proposed fuzzy SVMs and explains how to compute fuzzy memberships in Section 5. Section 6 presents experimental results on popular data sets and a conclusion is given in Section 7.

II. CURRENT SVMs

A. One-Class Support Vector Machine (OCSVM)

A hyperplane is determined to separate all normal data and at the same time maximise the margin between the normal data and the hyperplane [14]. OCSVM can be modelled as follows

$$\min_{w, \rho} \left(\frac{1}{2} \|w\|^2 - \rho + \frac{1}{\nu s} \sum_{i=1}^s \xi_i \right) \quad (1)$$

subject to

$$\begin{aligned} w^T \phi(x_i) &\geq \rho - \xi_i & i = 1, \dots, s \\ \xi_i &\geq 0, & i = 1, \dots, s \end{aligned} \quad (2)$$

where w is the normal vector of the hyperplane, ρ is the margin, ν is a positive constant, x_i , $i = 1, \dots, s$ are data points, ξ_i , $i = 1, \dots, s$ are slack variables, and $\phi(\cdot)$ is a kernel function.

The decision function is $f(x) = \text{sign}(w^T \phi(x) - \rho)$. The unknown x is a normal data point if $f(x) = +1$ or an abnormal data point if $f(x) = -1$.

B. SVM Classification (SVMC)

SVMC was originally proposed to deal with the balanced data sets [15]. However, by selecting two appropriately proportional trade-off parameters, it can be used to deal with imbalanced datasets. Let x_i , $i = 1, \dots, m_1$ be normal data points with label $y_i = +1$ and x_i , $i = m_1 + 1, \dots, s$ be abnormal data points with label $y_i = -1$, and $m_2 = s - m_1$. SVMC can be modelled as follows

$$\min_{w, b} \left(\frac{1}{2} \|w\|^2 + C_1 \sum_{i=1}^{m_1} \xi_i + C_2 \sum_{i=m_1+1}^s \xi_i \right) \quad (3)$$

subject to

$$\begin{aligned} y_i[w^T\phi(x_i) + b] &\geq 1 - \xi_i & i = 1, \dots, s \\ \xi_i &\geq 0, & i = 1, \dots, s \end{aligned} \quad (4)$$

where C_1 , C_2 and b are real numbers. The decision function is $f(x) = \text{sign}(w^T\phi(x) + b)$.

C. Support Vector Data Description (SVDD)

SVDD [15] aims at drawing an optimal hypersphere containing normal data. Abnormal data are outside this hypersphere. The optimisation problem is as follows

$$\min_{R, \xi} \left(R^2 + C_1 \sum_{i=1}^{m_1} \xi_i + C_2 \sum_{i=m_1+1}^s \xi_i \right) \quad (5)$$

subject to

$$\begin{aligned} \|\phi(x_i) - c\|^2 &\leq R^2 + \xi_i & i = 1, \dots, m_i \\ \|\phi(x_i) - c\|^2 &\geq R^2 - \xi_i & i = m_i + 1, \dots, s \\ \xi_i &\geq 0, & i = 1, \dots, s \end{aligned} \quad (6)$$

where R and c are radius and centre of the hypersphere, respectively. The decision function is $f(x) = \text{sign}(R^2 - \|\phi(x) - c\|^2)$.

D. Small Sphere & Large Margin (SSLM)

The SSLM approach combines the ideas of OCSVM and conventional two-class SVM [18] in minimising a hypersphere containing all normal data and simultaneously maximising the margin which is the distance from outliers (abnormal data) to the surface of the optimal hypersphere. This SSLM approach can be formulated by the following optimisation problem:

$$\min_{R, c, \xi, \rho} \left(R^2 - \nu\rho^2 + \frac{1}{\nu_1 m_1} \sum_{i=1}^{m_1} \xi_i + \frac{1}{\nu_2 m_2} \sum_{i=m_1+1}^s \xi_i \right) \quad (7)$$

subject to

$$\begin{aligned} \|\phi(x_i) - c\|^2 &\leq R^2 + \xi_i & i = 1, \dots, m_i \\ \|\phi(x_i) - c\|^2 &\geq R^2 + \rho^2 - \xi_i & i = m_i + 1, \dots, s \\ \xi_i &\geq 0, & i = 1, \dots, s \end{aligned} \quad (8)$$

where ν , ν_1 and ν_2 are three positive constants, ρ^2 is outside margin (distance from abnormal data to the surface of the hypersphere).

It can be seen that minimising the cost function (7) will make the radius R as small as possible and the margin ρ^2 as large as possible. Therefore this approach is called Small Sphere and Large Margin (SSLM). The hypersphere only surrounds the positive class (normal data) and SSLM aims to find a large margin between this hypersphere and the abnormal data points. The decision function is $f(x) = \text{sign}(R^2 - \|\phi(x) - c\|^2)$.

III. FUZZY APPROACH TO TWO-CLASS SVM

Although SVM is a powerful tool for solving classification problems, there are still some limitations. Each data point belongs to either one class or the other and all training points of a class are treated uniformly. In many real-world applications, it is often that some data points are more important than others in the classification problem. To overcome this problem, a fuzzy approach to SVM introduces fuzzy membership λ_i assigned to each data point x_i as seen in the following optimal hyperplane problem [10]

$$\min_{w, b} \left(\frac{1}{2} \|w\|^2 + C \sum_{i=1}^s \lambda_i \xi_i \right) \quad (9)$$

subject to

$$\begin{aligned} y_i[w^T\phi(x_i) + b] &\geq 1 - \xi_i & i = 1, \dots, s \\ \xi_i &\geq 0, & i = 1, \dots, s \end{aligned} \quad (10)$$

where λ_i is fuzzy membership of data point x_i . This fuzzy SVM is applied to sequential learning and fuzzy membership is considered as a function of time $\lambda_i = f(t_i)$, where t_i is the time the data point x_i arrived in the system. In classification problems, Lee *et al.* [9] introduced a similarity measure function to compute fuzzy memberships. However, they had to assume that outliers should be somewhat separate from the normal data. Recently Hsu *et al.* [7] have investigated the effect of the trade-off parameter C to the model of conventional two-class SVM and introduced a triangular membership function to set higher grades to the data points in mixed regions.

IV. NEW FUZZY SVMs

A fuzzy approach to both SVDD and SSLM will be presented in this section. We will present a general form for SVDD and SSLM then our fuzzy approach to this general form. We then extend this fuzzy approach to OCSVM and SVMC.

A. A General Form for SVDD and SSLM

The equations (6) and (8) can be rewritten as follows

$$\begin{aligned} y_i \|\phi(x_i) - c\|^2 &\leq y_i R^2 + \xi_i, & i = 1, \dots, s \\ y_i \|\phi(x_i) - c\|^2 &\leq y_i R^2 + z_i \rho^2 + \xi_i, & i = 1, \dots, s \end{aligned} \quad (11)$$

where $z_i = \min\{0, y_i\}$. The optimisation problem for both SVDD and SSLM can be written as follows

$$\min_{R, c, \xi, \rho} \left(R^2 - \nu\rho^2 + \frac{1}{\nu_1 m_1} \sum_{i=1}^{m_1} \xi_i + \frac{1}{\nu_2 m_2} \sum_{i=m_1+1}^s \xi_i \right) \quad (12)$$

subject to

$$\begin{aligned} y_i \|\phi(x_i) - c\|^2 &\leq y_i R^2 + z_i \rho^2 + \xi_i & i = 1, \dots, s \\ \xi_i &\geq 0, & i = 1, \dots, s \\ z_i &= \min\{0, y_i\}, & i = 1, \dots, s \end{aligned} \quad (13)$$

It is seen that if $\nu = 0$, $z = 0$, $\frac{1}{\nu_1 m_1} = C_1$, $\frac{1}{\nu_2 m_2} = C_2$ then (12) and (13) become (5) and (6) for SVDD. If $z_i = \min\{0, y_i\}$, $i = 1, \dots, s$, we have the equations (7) and (8) for SSLM.

B. Fuzzy SVDD and Fuzzy SSLM

The fuzzy optimisation problem for both SVDD and SSLM is proposed as follows

$$\min_{R,c,\xi,\rho} \left(R^2 - \nu\rho^2 + \frac{1}{\nu_1 m_1} \sum_{i=1}^{m_1} \lambda_i \xi_i + \frac{1}{\nu_2 m_2} \sum_{i=m_1+1}^s \lambda_i \xi_i \right) \quad (14)$$

subject to

$$\begin{aligned} y_i \|\phi(x_i) - c\|^2 &\leq y_i R^2 + z_i \rho^2 + \xi_i & i = 1, \dots, s \\ \xi_i &\geq 0, & i = 1, \dots, s \\ z_i &= \min\{0, y_i\}, & i = 1, \dots, s \end{aligned} \quad (15)$$

where weights $\lambda_i \in [0, 1]$, $i = 1, \dots, s$ are regarded as fuzzy memberships. The corresponding Lagrange is written as follows

$$\begin{aligned} L(R, c, \xi, \alpha, \beta) &= R^2 - \nu\rho^2 \\ &+ \frac{1}{\nu_1 m_1} \sum_{i=1}^{m_1} \lambda_i \xi_i + \frac{1}{\nu_2 m_2} \sum_{i=m_1+1}^s \lambda_i \xi_i \\ &+ \sum_{j=1}^s \alpha_j (y_j \|\phi(x_j) - c\|^2 - y_j R^2 - z_j \rho^2 - \xi_j) \\ &- \sum_{k=1}^s \beta_k \xi_k \end{aligned} \quad (16)$$

where $\alpha_i \geq 0$ and $\beta_i \geq 0$ are Lagrange multipliers. Setting derivatives of $L(R, c, \xi, \alpha, \beta)$ with respect to primal variables to 0, we obtain the dual form

$$\min_{\alpha} \left(\sum_{i=1}^s \sum_{j=1}^s \alpha_i \alpha_j y_i y_j K(x_i, x_j) - \sum_{i=1}^s \alpha_i y_i K(x_i, x_i) \right) \quad (17)$$

subject to

$$\begin{aligned} 0 &\leq \alpha_i \leq \frac{\lambda_i}{\nu_1 m_1}, & i = 1, \dots, m_1 \\ 0 &\leq \alpha_j \leq \frac{\lambda_j}{\nu_2 m_2}, & j = m_1 + 1, \dots, s \\ \sum_{i=1}^s \alpha_i y_i &= 1, & \sum_{i=1}^s \alpha_i = 2\nu + 1 \end{aligned} \quad (18)$$

The dual form is also a quadratic optimization problem and has the same form as the dual of the ν -SVM [13], thus it can be solved with the ν -SVM solver in the LIBSVM software [5]. The parameters R , c and ρ are calculated as follows

$$\begin{aligned} R^2 &= \frac{1}{n_1} P_1 \\ \|c\|^2 &= \sum_{i=1}^s \sum_{j=1}^s \alpha_i \alpha_j y_i y_j K(x_i, x_j) \\ \rho^2 &= \frac{1}{n_2} P_2 - \frac{1}{n_1} P_1 \end{aligned} \quad (19)$$

where

$$\begin{aligned} n_1 &= |S_1|, & n_2 &= |S_2| \\ S_1 &= \left\{ x_i \mid 0 < \alpha_i < \frac{\lambda_i}{\nu_1 m_1}, \quad 1 \leq i \leq m_1 \right\} \\ S_2 &= \left\{ x_j \mid 0 < \alpha_j < \frac{\lambda_j}{\nu_2 m_2}, \quad m_1 + 1 \leq j \leq s \right\} \end{aligned}$$

$$\begin{aligned} P_1 &= \sum_{x_i \in S_1} \left[K(x_i, x_i) + \|c\|^2 - 2 \sum_{k=1}^s y_k \alpha_k K(x_k, x_i) \right] \\ P_2 &= \sum_{x_i \in S_2} \left[K(x_i, x_i) + \|c\|^2 - 2 \sum_{k=1}^s y_k \alpha_k K(x_k, x_i) \right] \end{aligned}$$

To classify an unknown data point x , the following decision function is used

$$f(x) = \text{sign} \left(R^2 - \|c\|^2 - K(x, x) + 2 \sum_{k=1}^s \alpha_k y_k K(x, x_k) \right) \quad (21)$$

The unknown x is a normal data point if $f(x) = +1$ or an abnormal data point if $f(x) = -1$

C. Fuzzy OCSVM

Similarly, fuzzy OCSVM can be modeled as follows

$$\min_{w,\rho} \left(\frac{1}{2} \|w\|^2 - \rho + C \sum_{i=1}^s \lambda_i \xi_i \right) \quad (22)$$

subject to

$$\begin{aligned} w^T \phi(x_i) &\geq \rho - \xi_i & i = 1, \dots, s \\ \xi_i &\geq 0, & i = 1, \dots, s \end{aligned} \quad (23)$$

We can derive the dual form as follows

$$\min_{\alpha} \left(\frac{1}{2} \sum_{i=1}^s \sum_{j=1}^s \alpha_i \alpha_j y_i y_j K(x_i, x_j) \right) \quad (24)$$

subject to

$$\begin{aligned} 0 &\leq \alpha_i \leq \lambda_i C_1, & i = 1, \dots, m_1 \\ 0 &\leq \alpha_j \leq \lambda_j C_2, & j = m_1 + 1, \dots, s \\ \sum_{i=1}^s \alpha_i &= 1, & i = 1, \dots, s \end{aligned} \quad (25)$$

D. Fuzzy SVMC

Fuzzy SVMC is modeled as follows

$$\min_{w,b} \left(\frac{1}{2} \|w\|^2 + C_1 \sum_{i=1}^{m_1} \lambda_i \xi_i + C_2 \sum_{i=m_1+1}^s \lambda_i \xi_i \right) \quad (26)$$

subject to

$$\begin{aligned} y_i [w^T \phi(x_i) + b] &\geq 1 - \xi_i & i = 1, \dots, s \\ \xi_i &\geq 0, & i = 1, \dots, s \end{aligned} \quad (27)$$

where weights $\lambda_i \in [0, 1]$, $i = 1, \dots, s$ are regarded as fuzzy memberships. The corresponding dual form is as follows

$$\min_{\alpha} \left(\frac{1}{2} \sum_{i=1}^s \sum_{j=1}^s \alpha_i \alpha_j y_i y_j K(x_i, x_j) - \sum_{i=1}^s \alpha_i \right) \quad (28)$$

subject to

$$\begin{aligned} 0 &\leq \alpha_i \leq \lambda_i C_1, & i = 1, \dots, m_1 \\ 0 &\leq \alpha_j \leq \lambda_j C_2, & j = m_1 + 1, \dots, s \\ \sum_{i=1}^s y_i \alpha_i &= 0, & i = 1, \dots, s \end{aligned} \quad (29)$$

V. NEW FUZZY MEMBERSHIP COMPUTATION METHOD

The normal and abnormal data points are normally mixed and the task of fuzzy SVM is to construct a hyperplane or hypersphere in feature space to separate normal data from abnormal data. Hence we assume that the data points in the mixed regions are important and they should have the highest fuzzy membership value. Other data points are less important and should have lower fuzzy membership values.

A. Fuzzy Clustering Membership

We use fuzzy clustering techniques to determine clusters in the mixed regions. These clusters contain both normal and abnormal data points. Fuzzy memberships of these data points are set to 1 and fuzzy memberships of other data points are determined by their closest cluster accordingly. Although clustering is performed in the input space, according to most of current kernel functions, relative distances between data points are preserved so we can apply the clustering results obtained in the input space to the feature space. The following algorithm is used to determine fuzzy memberships for all data points:

Fuzzy Membership Determination Algorithm

- 1) Select a clustering algorithm
- 2) Perform clustering on the training data set
- 3) Determine a subset containing clusters that contain both normal and abnormal data. Denote this subset as MIXEDCLUS.
- 4) For each data point $x \in \text{MIXEDCLUS}$, set its fuzzy membership to 1
- 5) For each data point $x \notin \text{MIXEDCLUS}$, do the following
 - Find out the cluster whose center is closest to x
 - Calculate fuzzy membership of x with this cluster

B. The Role of Fuzzy Memberships

The term $\sum_i \lambda_i \xi_i$ is regarded as a weighted sum of empirical errors to be minimised in fuzzy SVMs. If a misclassified point x_i is not in a mixed cluster, its fuzzy membership λ_i is small and hence its error ξ_i can be large as long as $\lambda_i \xi_i$ is still minimised. On the other hand, if it is in a mixed cluster, its fuzzy membership is 1 and hence its error ξ_i must be small such that $\lambda_i \xi_i$ remains minimised. This means that the decision boundary tends to move to mixed region to reduce empirical errors in this region. Figure 1 demonstrates the difference between decision boundaries for SVM and fuzzy SVM in two class classification problem. Positive (normal) data points are represented by plus (+) signs and negative (abnormal) data points—minus (-) signs. The dashed and solid circles are the decision boundaries for SVM and fuzzy SVM, respectively. There are some misclassified normal points located on the left hand side and the dashed circle is near those data points to reduce distances (empirical errors) from them to it. In fuzzy SVM, those points will have small fuzzy memberships and hence their distances to the solid circle, i.e. the decision boundary for fuzzy SVM, can be longer and the fuzzy SVM can reduce empirical errors in the mixed region represented by

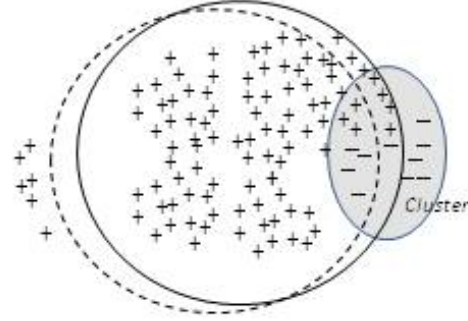


Fig. 1. Decision boundaries for SVM (dashed circle) and fuzzy SVM (solid circle) in a two class (+ and -) classification problem.

TABLE I

NUMBER OF DATA POINTS IN 14 DATA SETS. #POS: TOTAL POSITIVE DATA POINTS, #NEG: TOTAL NEGATIVE DATA POINTS, m_1 : POSITIVE DATA POINTS FOR TRAINING, m_2 : NEGATIVE DATA POINTS FOR TRAINING, AND d : DIMENSION. THE REMAINING DATA POINTS ($\#pos - m_1$) AND ($\#neg - m_2$) WERE USED FOR TESTING.

Data set	#pos	#neg	m_1	m_2	d
Australian	383	307	245	12	14
B. Cancer	444	239	355	18	10
Biomed	67	127	53	2	5
Bioinformatics	221	117	176	9	20
Diabetes	500	268	400	21	8
DNA	464	485	371	19	180
Fourclass	307	555	245	15	2
Glass	70	76	56	2	9
Ionosphere	255	126	180	9	34
Liver	200	145	116	6	6
Spectf	254	95	203	10	44
SVMGuide3	296	947	236	12	22
Vehicle	212	217	169	8	18
Wine	59	71	47	2	13

the grey cluster. In this case, fuzzy SVM has less misclassified points than SVM. The misclassified normal data points on the left-hand side can be considered as outliers due to their long distances to the normal data set and fuzzy SVM reduces its sensitivity to these outliers.

VI. EXPERIMENTAL RESULTS

We performed experiments on 14 well-known data sets related to machine fault detection and bioinformatics [16]. Table 1 presents total training data (#pos) and test data (#neg) available in the data sets. These data sets were balanced so we created at random imbalanced subsets such that the ratio of number of normal data points (m_1) and number of abnormal data points (m_2) was 19:1. Creating subsets were repeated 10 times and the average classification rate for 10 times was determined. The classification rate acc is measured as [8]

$$acc = \sqrt{acc^+ acc^-} \quad (30)$$

where acc^+ and acc^- are the classification accuracy on normal and abnormal data, respectively.

TABLE II
CLASSIFICATION RESULTS (IN %) ON 14 DATA SETS FOR OCSVM, FUZZY OCSVM, SVMC AND FUZZY SVMC.

Data set	OCSVM	FOCSVM	SVMC	FSVMC
Australian	78.1	78.8	43.9	63.5
B-Cancer	96.2	96.2	97.0	97.7
Biomed	76.3	76.3	40.5	44.3
Bioinformatics	62.3	64.0	50.2	51.0
Diabetes	67.8	67.9	51.7	57.0
DNA	62.3	63.1	30.2	39.6
Fourclass	91.1	91.1	62.7	62.7
Glass	79.7	79.7	57.6	59.8
Ionosphere	91.1	91.8	86.3	89.7
Liver	60.9	63.5	43.7	52.4
Spectf	58.2	59.7	9.1	33.3
SVMGuide3	48.4	51.9	45.7	50.6
Vehicle	51.0	54.0	38.6	51.5
Wine	92.9	92.9	24.1	84.9

The popular RBF kernel function $K(x, x') = e^{-\gamma \|x-x'\|^2}$ was used in our experiments to compare with the SSLM approach. We used the same parameter settings suggested in [20], the parameter γ was searched in $\{\sigma_0^2/16, \sigma_0^2/8, \sigma_0^2/4, \sigma_0^2/2, \sigma_0^2, 2\sigma_0^2, 4\sigma_0^2, 8\sigma_0^2, 16\sigma_0^2\}$, where σ_0^2 is the mean norm of the training data. For SVMs and fuzzy SVMs, the penalty parameters C , C_1 and C_2 were searched over the grid $\{0.01, 0.05, 0.1, 0.5, 1, 5, 10, 50, 100, 500\}$, such that the ratio C_2/C_1 belonged to

$$\left\{ \frac{1}{4} \times \frac{m_1}{m_2}, \frac{1}{2} \times \frac{m_1}{m_2}, \frac{m_1}{m_2}, 2 \times \frac{m_1}{m_2}, 4 \times \frac{m_1}{m_2} \right\} \quad (31)$$

For OCSVM, the parameter ν was searched in $\{0.01k, 0.1k\}$, where k was an integer ranging from 1 to 9. For SSLM, the parameter ν was searched in $\{10, 30, 50, 70, 90\}$, while ν_1 and ν_2 were selected from $\{0.001, 0.01\}$. Fuzzy c -means technique [3] was chosen for clustering data sets. Number of clusters was searched in $\{1, 2, 4, 6, 8, 10\}$.

Table 2 presents classification results for OCSVM, fuzzy OCSVM, SVMC and fuzzy SVMC. Table 3 presents classification results for SVDD, fuzzy SVDD, SSLM and fuzzy SSLM. From these tables, we can see that the proposed fuzzy SVMs achieved higher classification rates than their corresponding non-fuzzy SVMs for most of the data sets. Each of these results was averaged on 10 results performed on 10 different training and test data sets created randomly (note in each time the same training and test data sets were used for both SVM and its corresponding fuzzy SVM).

VII. CONCLUSION

We have presented a new method to determine the level of importance for all data points, a method to compute fuzzy memberships, and extend the fuzzy approach for two-class SVM to one-class SVM for novelty detection. An explanation of fuzzy membership role and how fuzzy SVM can achieve better classification results are provided. Experimental results on different data sets showed promising classification results for the proposed fuzzy SVM approach.

TABLE III
CLASSIFICATION RESULTS (IN %) ON 14 DATASETS FOR SVDD, FUZZY SVDD, SSLM AND FUZZY SSLM.

Data set	SVDD	FSVDD	SSLM	FSSLM
Australian	86.8	89.1	68.8	77.6
B-Cancer	98.6	98.6	97.2	97.2
Biomed	71.4	71.4	63.8	64.1
Bioinformatics	85.3	87.9	44.8	51.3
Diabetes	71.6	72.8	62.5	63.1
DNA	91.2	92.7	66.5	69.5
Fourclass	99.2	99.2	95.7	95.7
Glass	59.8	61.3	66.2	76.8
Ionosphere	94.8	94.9	92.1	92.7
Liver	60.3	65.2	60.9	61.8
Spectf	75.3	75.3	53.7	53.8
SVMGuide3	48.4	51.9	45.6	50.5
Vehicle	52.2	55.5	42.9	48.1
Wine	99.3	99.3	94.3	94.3

REFERENCES

- [1] Bishop, C.M. (1994). Novelty detection and neural network validation *IEEE Proceedings-Vision, Image, and Signal Processing*
- [2] Barnett V. and Lewis, T. (1978). *Outliers in statistical data* 3rd ed. Wiley
- [3] Bezdek, J.C. (1981). *Pattern Recognition with Fuzzy Objective Function Algorithms*, Plenum Press, New York and London.
- [4] Campbell, C. & Bennet, K.P. (2001). A linear programming approach to novelty detection *Advances in neural information processing*
- [5] Chang, C.-C. and Lin, C.-J. (2001). LIBSVM: A Library for Support Vector Machines. The LIBSVM software package is available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- [6] Chuang, C., Jeng, S. Su, J., and Hsiao, C. (2002). Robust support vector regression networks for function approximation with outliers, *IEEE Trans. Neural Networks*, 13, 1322-1330
- [7] Hsu, C.-C., Han, M.-F., Chang, S.-H., Chung, H.-Y. (2009). Fuzzy Support Vector Machines with the Uncertainty of Parameter C. *Expert Systems With Application*, 44, 6654-6658.
- [8] Kubat, M. and Matwin, S. (1997). Addressing the curse of imbalanced training set: One-sided selection *Proc. 14th International Conference on Machine Learning*
- [9] Lee, G.H., Taur, J. S. and Tao, C.W. (2006). A Robust Fuzzy Support Vector Machine for Two-class Pattern Classification *International Journal of Fuzzy Systems*, 8, 2, 76-87
- [10] Lin, C.-F. and Wang, S.-D. (2002). Fuzzy Support Vector Machines. *IEEE Transactions on Neural Networks*, 13, 2, 464-472.
- [11] Lin, Y., Lee, Y. and Wahba, G. (2002). Support vector machine for classification in nonstandard situations, *Machine Learning*, 15, 1115-1148.
- [12] Moya, M.M., Koch, M.W., and Hostetler, L.D. (1991). One-class classifier networks for target recognition applications *Proceedings of world congress on neural networks*, 797-801.
- [13] Roberts, S. and Tarassenko, L. (1994). A Probabilistic Resource Allocation Network for Novelty Detection *Neural Computation*, 6, 270-284.
- [14] Schlkopf and Smola, A. J. (2002). *Learning with kernels* Cambridge: The MIT Press
- [15] Tax, D. M. J. and Duin, R. P. W. (2004). Support vector data description *Machine Learning*, 54, 45-56
- [16] Tax, D. M. J. (2009). Data sets downloadable at <http://ict.ewi.tudelft.nl/~davidt/occ/index.html>
- [17] Towel, G.G. (2000). Local expert autoassociator for anomaly detection *Proc. 17th International Conference on Machine Learning*
- [18] Vapnik, V. (1995). The nature of statistic learning theory. *New York: Springer Verlag*
- [19] Vert, J. and Vert, J.P. (2006). Consistency and convergence rates of one class SVM and related algorithm *Journal of Machine Learning Research*, 7, 817-854.
- [20] Yu, M. and Ye, J. (2009). A Small Sphere and Large Margin Approach for Novelty Detection Using Training Data with Outliers *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 31, 2088-2092