# 불완전 데이터의 패턴 분석을 위한 miSVMs

이기영[1,*], 김대원[2], 이도헌[2], 이광형[1,2]
한국과학기술원 전자전산학과[1], 바이오시스템학과[2]
kylee@bisl.kaist.ac.kr

# A New Support Vector Machines for Classifying Uncertain Data

Kiyoung Lee*, Dae-Won Kim, Doheon Lee, and Kwang H. Lee
Department of EECS and Department of BioSystems, KAIST

요 약

Conventional support vector machines (SVMs) find optimal hyperplanes that have maximal margins by treating all data equivalently. In the real world, however, the data within a data set may differ in degree of uncertainty or importance due to noise, inaccuracies or missing values in the data. Hence, if all data are treated as equivalent, without considering such differences, the optimal hyperplanes identified are likely to be less optimal. In this paper, to more accurately identify the optimal hyperplane in a given uncertain data set, we propose a membership-induced distance from a hyperplane using membership values, and formulate three kinds of membership-induced SVMs.

## 1. Introduction

Despite many benefits of SVMs [1][2], conventional SVMs lack a mechanism for reflecting variations in the uncertainty or importance of data in a data set, and hence treat all data as equivalent. In many real world problems, however, data may have different degree of uncertainty due to noise, inaccuracies or missing values in the data set.

One approach, called Fuzzy SVMs (FSVMs), to including differences in importance in input data within the SVM formalism was proposed by Lin and Wang [3]. They introduced fuzzy membership values $\mu_i$ into the compensation part for misclassification in the objective function of conventional SVMs, and modified the error term in the conventional SVMs as

$$O_f = \frac{1}{2}\,\mathbf{w}\cdot\mathbf{w} + C\sum_{i=1}^{n}\mu_i\xi_i, \qquad (1)$$

, where $\xi_i$ is a slack variable [1]. However, FSVMs are limited in that they reflect all the differences in an input data set; only margin error data are treated differently in this approach, even though all the data may have differing degrees of uncertainty or importance.

To resolve the above problems in FSVMs and find more reasonable and more optimal OHPs from uncertain data, in the present study we propose three kinds of SVMs that take into account all differences in uncertainty or importance by means of a membership-induced distance measure.

## 2. Membership-induced SVMs

### 2.1 Motivation and approach

Conventional SVMs cannot reflect the different uncertainties of below two types of data in finding the OHP. Consider, two data sets (Figs. 1a and 1b) whose *uncertain* input vector $\mathbf{x}_u$ has moved to $P_u$ from original point $P_o$ due to noise in $\mathbf{x}_u$. If $\mathbf{x}_u$ has not moved, then the optimal solution is $OHP_o$.

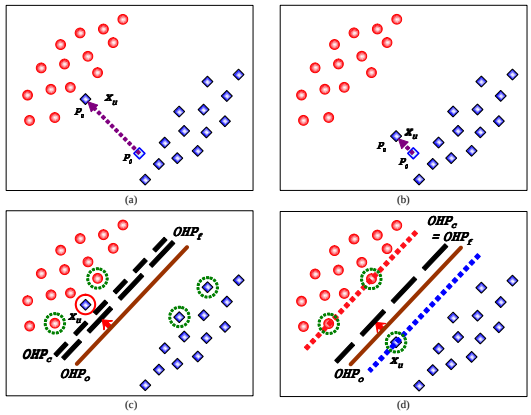


Fig. 1. Uncertainty handling in conventional SVMs and FSVMs.

**Regarding $\mathbf{x}_u$ as a margin error:** If the $\mathbf{x}_u$ in Fig. 1a is regarded as a margin error, then the slack variable $\xi_u$ for $\mathbf{x}_u$ is nonzero, and thus the data set is regarded as linearly nonseparable. In this case, conventional SVMs identify $OHP_c$ in Fig. 1c as the OHP, even though $OHP_o$ or similar hyperplanes are more optimal. FSVMs partially solve this problem by decreasing the effect of the margin error through the use of fuzzy membership. FSVMs would identify a hyperplane such as $OHP_f$ as the OHP. There is, however, no golden rule governing the location of $OHP_f$ is located in Fig. 1c because we do not know the original position of the input vector $\mathbf{x}_u$. In other words, the FSVMs find only one of many feasible candidate OHPs between $OHP_c$ and $OHP_o$.



Copyright(c) 2005 NuriMedia Co.,Ltd

**Regarding $\mathbf{x}_u$ as a support vector on the margin:** If the $\mathbf{x}_u$ in Fig. 1b is regarded as a support vector on the margin, then all the slack variables for the data including the input vector $\mathbf{x}_u$ are zero, and thus the data set is regarded as linearly separable. This is a more serious problem for FSVMs because, in this case, FSVMs find the same OHP as conventional SVMs (shown in Fig. 1d), even though the OHP may be closer to $OHP_o$. That is, neither the conventional SVMs nor FSVMs have a mechanism to reflect the uncertainty in $\mathbf{x}_u$, and hence neither will find the optimal solution.

In this paper, we define membership-induced distances using fuzzy membership values for all data to reflect all the uncertainties in a given set of data. By maximizing the membership-induced distance, we find *membership-induced OHPs* that are more reliable and more optimal hyperplanes than those identified by conventional SVMs and FSVMs.

## 2.2 Membership-induced Distance

Suppose we have an uncertain data set $S = \{\mathbf{s}_i : (y_i, \mathbf{x}_i, \mu_i) \mid i = 1, ..., n\}$ in which each datum $\mathbf{s}_i$ has three components: a label ($y_i \in \{-1, +1\}$); an input vector ($\mathbf{x}_i \in R^N$); and a fuzzy membership value ($\mu_i, 0 < \mu_i \leq 1$), where $\mu_i$ is the degree of certainty of the datum $\mathbf{s}_i$. We define a membership-induced distance (*mi-distance*) between $\mathbf{s}_i$ and a hyperplane.

*Definition 1* (Membership-induced Distance) Let $m$ ($\in \{R^+ \cup \{0\}\}$) be a *fuzziness control parameter*, and $(\mathbf{w}, b)$ be a hyperplane. The membership-induced distance between datum $\mathbf{s}_i$ and the hyperplane, denoted by $\delta_{mi}(i)$, is then defined as

$$\delta_{mi}(i) = \frac{|\mathbf{w} \cdot \mathbf{x}_i + b|}{\mu_i^m \parallel \mathbf{w} \parallel}, \quad i = 1, ..., n. \tag{2}$$

Note that $\delta_{mi}(i)$ decreases with increasing $\mu_i^m$. Hence, data with larger values of $\mu_i^m$ more strongly influence the OHP. Furthermore, when $m = 0$, the mi-distance equals the Euclidean distance, and thus our SVMs based on the mi-distance act like conventional SVMs. When $m \to \infty$, the mi-distance for uncertain data approaches infinity, and thus all uncertain data are neglected in the calculation for finding the OHP.

## 2.3 Formulation for Membership-induced OHPs

### 2.3.1 Formulation of MISVM for the linearly separable case

We first define the membership-induced margin (*mi-margin*), denoted as $\delta_{mi}^m$, of a fuzzy set $S$ with respect to a hyperplane $(\mathbf{w}, b)$ as the minimum mi-distance while classifying the data set $S$ correctly. By maximizing the mi-margin $\delta_{mi}^m$, we find the *membership-induced OHP* (MIOHP).

Let us suppose that an uncertain data set $S$ is linearly separable. For this case, $\delta_{mi}^m$ can be written as

$$\delta_{mi}^m = \min_{i=1,...,n} \frac{1}{\mu_i^m} |\mathbf{w} \cdot \mathbf{x}_i + b| \ / \ \parallel \mathbf{w} \parallel \tag{3}$$

Because scalar multiplication does not affect identical equation, we can normalize the hyperplane $(\mathbf{w}, b)$ to satisfy $\min_{i=1,...,n} \frac{1}{\mu_i^m} |\mathbf{w} \cdot \mathbf{x}_i + b| = 1$. Then, we can maximize the mi-margin by minimizing the objective function $O_{mi}^l$

$$O_{mi}^l = \frac{1}{2} \mathbf{w} \cdot \mathbf{w} \tag{4}$$

$$\textit{subject to} \quad \frac{1}{\mu_i^m} y_i (\mathbf{w} \cdot \mathbf{x}_i + b) - 1 \geq 0, \ i = 1, ..., n. \tag{5}$$

The problem of minimizing Eq.4 subject to Eq.5 is a convex optimization problem, thus by introducing Lagrange multipliers $\alpha_i$, and by using KKT conditions[1], we obtain the dual representation of the optimization problem: *maximize* $D(\alpha)_{mi}^l$

$$D(\alpha)_{mi}^l = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{1}{\mu_i^m} \frac{1}{\mu_j^m} \alpha_i \alpha_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j, \tag{6}$$

$$\textit{subject to} \quad \sum_{i=1}^n \frac{1}{\mu_i^m} \alpha_i y_i = 0, \quad 0 \leq \alpha_i, \quad i = 1, ..., n. \tag{7}$$

This dual representation is a constrained quadratic problem; hence we can derive the $\check{\alpha}_i$ that satisfy the Eq.6. Furthermore, we can see that if the fuzziness control parameter $m$ equals zero, then this dual representation is equivalent to the separable case in conventional SVMs. Thus, our MISVMs are a general extension of conventional SVMs. The complementary KKT condition of this case is defined as

$$\check{\alpha}_i \{ \frac{1}{\mu_i^m} y_i (\mathbf{w} \cdot \mathbf{x}_i + b) - 1 \} = 0, \ i = 1, ..., n. \tag{8}$$

In this case, the data $\mathbf{x}_i$ corresponding to $\check{\alpha}_i > 0$ are the support vectors that lie on the mi-margin.

After solving Eq.6, the offset $b_o$ of the $MIOHP_l$ can be obtained using the complementary KKT condition, Eq.8.

The decision function for a test input vector $\mathbf{x}_t$ is

$$f(\mathbf{x}_t)_{mi}^l = sgn(\sum_{i=1}^n \frac{1}{\mu_i^m} \check{\alpha}_i y_i \mathbf{x}_u \cdot \mathbf{x}_i + b_o). \tag{9}$$

### 2.3.2 Formulation for the soft mi-margin

If an uncertain data set $S$ is linearly nonseparable, we first handle this case using slack variables $\zeta_i$ for each datum. The $\zeta_i$ for each datum is the amount of mi-distance from the limit (the corresponding mi-margin) to a opposite class. Since the exponential-weighted fuzzy membership $\mu_i^m$ of each datum $\mathbf{s}_i$ determines the degree of certainty or importance on a MIOHP, the term $\mu_i^{vm} \zeta_i$ of the datum $\mathbf{s}_i$ is a measure of the weighted mi-margin error for that datum, where $v$ ($\geq 1$) is a constant. Thus, the MIOHP for this case can be found by minimizing the objective function $O_{mi}^c$,

$$O_{mi}^c = \frac{1}{2} \mathbf{w} \cdot + C \sum_{i=1}^n \mu_i^{vm} \zeta_i \tag{10}$$

$$\textit{subject to} \quad \frac{1}{\mu_i^m} y_i (\mathbf{w} \cdot \mathbf{x}_i + b) - 1 + \zeta_i \geq 0, \ (\zeta_i \geq 0, \ i = 1, ..., n). \tag{11}$$

, where $C$ ($\in R^+$) is a regulation parameter.

Similar to the linearly sparable case, we can obtain the dual representation of the optimization problem: *maximize* $D(\alpha)_{mi}^c$

$$D(\alpha)_{mi}^c = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \frac{1}{\mu_i^m} \frac{1}{\mu_j^m} \alpha_i \alpha_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j, \tag{12}$$



Copyright(c) 2005 NuriMedia Co.,Ltd

TABLE I
PREDICTION ACCURACY OF FSVMS AND MISVMS WHEN $\gamma = 1.4$

| C | FSVMs | MISVMs | | | | |
|---|---|---|---|---|---|---|
| | | $m$ =0.0 | 0.1 | 0.5 | 1.0 | 4.0 |
| 1 | **37.33** | 60.00 | 58.44 | 62.40 | 61.60 | 61.07 |
| 2 | 37.60 | 70.40 | 71.47 | 71.73 | 72.27 | 63.20 |
| 5 | 37.60 | 71.47 | 76.27 | 74.67 | 73.60 | 60.80 |
| 50 | 39.67 | 64.67 | 76.33 | **77.00** | 67.00 | 59.33 |
| 500 | 49.67 | 57.33 | 71.33 | 67.33 | 58.00 | 59.33 |
| 2000 | **59.33** | 57.33 | 71.00 | 65.67 | 56.67 | 59.33 |
| 10000 | 42.67 | 34.00 | **63.33** | 52.00 | 36.00 | 50.00 |

$$subject\ to \quad \sum_{i=1}^{n} \frac{1}{\mu_i^m}\alpha_i y_i = 0, \quad 0 \le \alpha_i \le \mu_i^{vm} C, \quad i = 1, ..., n. \tag{13}$$

, and we can derive the complementary KKT conditions:

$$\check{\alpha}_i\{\frac{1}{\mu_i^m} y_i(\mathbf{w}\cdot\mathbf{x}_i + b) - 1 + \zeta_i\} = 0, \quad i = 1, ..., n. \tag{14}$$

$$(\mu_i^{vm} C - \check{\alpha}_i)\zeta_i = 0, \quad i = 1, ..., n. \tag{15}$$

The decision function for a test vector $\mathbf{x}_t$ is

$$f(\mathbf{x}_t)_{mi}^{l} = sgn(\sum_{i=1}^{n} \frac{1}{\mu_i^m}\check{\alpha}_i y_i \mathbf{x}_t \cdot \mathbf{x}_i + b_o), \tag{16}$$

2.3.3 Formulation for kernelization

As seen in Eqs.12, and 16, the dual form of the objective function and the decision function of MISVMs are represented entirely in terms of inner products of pairs of input vectors. Thus we can kernelize the MISVMs. The kernelized version of the decision function in Eq.16 for MISVMs is

$$f(\mathbf{x}_t)_{mi}^{k} = sgn(\sum_{i=1}^{n} \frac{1}{\mu_i^m}\check{\alpha}_i y_i k(\mathbf{x}_t, \mathbf{x}_i) + b_o). \tag{17}$$

## 3. Experiments

To investigate the success of these attempts, we conducted a variety of tests in which FSVMs and MISVMs were applied to Iris Plant data [4]. The Iris plant data set has three classes; one of these classes is linearly separable from the others, but the other two classes are not linearly separable from each other. The original Iris Plant data set does not contain any uncertainty. Thus, to create a training data set with uncertainty, we randomly created missing attributes $\gamma$. We randomly divided the Iris plant data set into two sets, and carried out a cross-validation with these sets after creating missing attributes.

Table I shows selected results obtained using the FSVMs and MISVMs for a data set generated using $\gamma = 1.4$, a relatively high value that gave an average missing rate of attributes of approximately 0.46. The accuracies of the FSVMs ranged from 37.33% to 59.33%. Thus, 59.33% was the optimal case of the FSVMs for the data set. By comparison, our proposed MISVMs gave substantially better results for all selected $C$ values considered, with accuracies ranging from 63.33% to 77.00%. For each $C$, the improvement achieved by MISVMs compared to FSVMs ranged from 19.67% up to 102.34%. For the optimal case, the accuracy of the MISVMs was 77.00%, a 29.78% improvement on the result achieved using FSVMs.

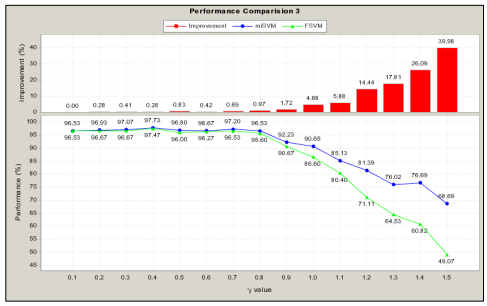


Fig. 2. Comparison of the accuracies of FSVMs and MISVMs with $\gamma$.

We performed an experiment in which we chose 15 values of $\gamma$ in the range of 0.1 to 1.5, and randomly generated five data sets for each value of $\gamma$. We then performed a cross-validation for each data set, found the prediction accuracies for the optimal case, and averaged the accuracies for each value of $\gamma$. The results of this experiment are depicted in Fig. 2. In general, the averaged prediction accuracies of the FSVM and MISVMs decrease with increasing $\gamma$. However, the gradient of this decrease in the averaged prediction accuracies is steeper for the FSVMs than for the MISVMs, indicating that, for the present data sets, the MISVMs give increasingly superior results compared to the FSVMs as the uncertainty is increased (see upper part of Fig. 2). Thus, the results of this experiment demonstrate the superiority of MISVMs over FSVMs in handling uncertainty when the degree of uncertainty is large.

## 4. Concluding Remarks

In this paper we have discussed the problems of uncertainty handling in SVMs. To resolve the problems and to more precisely reflect the uncertainties in a given data set, in the present work we introduced a fuzziness control parameter and proposed a membership-induced distance measure. Using this measure, we developed three kinds of MISVMs. Comparisons of the prediction performance of MISVMs with the performances of conventional SVMs and FSVMs showed that the proposed MISVMs approach better reflects uncertainties in a data set compared to the FSVMs method.

### REFERENCES

[1] B. Schölkopf and A.J. Smola, Learning with Kernels, MIT Press, Cambridge, MA, 2002.
[2] C.J.C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition", *Data Mining and Knowledge Discovery*, vol. 2, pp. 121-1678, 1998.
[3] Chun-Fu Lin and Sheng-De Wang, "Fuzzy Support Vector Machines", *IEEE Trans. Neural Networks*, vol. 13, no. 2, pp. 464-471, 2002.
[4] C.L. Blake and C.J. Merz, UCI repository of machiune learning database , *http://www.ics.uci.edu/ mlearn/MLRepository.html/*, 1998.



Copyright(c) 2005 NuriMedia Co.,Ltd