Proceedings of the 6th INFORMS Workshop on Data Mining and Health Informatics (DM-HI 2011)
P. Qian, Y. Zhou, C. Rudin, eds.

Probabilistic Classification using Fuzzy Support Vector Machines

Marzieh Parandehgheibi
Operations Research Center
Massachusetts Institute of Technology
Cambridge, MA
marp@mit.edu

Abstract

In medical applications such as recognizing the type of a tumor as Malignant or Benign, a wrong diagnosis can be devastating. Methods like Fuzzy Support Vector Machines (FSVM) try to reduce the effect of misplaced training points by assigning a lower weight to the outliers. However, there are still uncertain points which are similar to both classes and assigning a class by the given information will cause errors. In this paper, we propose a two-phase classification method which probabilistically assigns the uncertain points to each of the classes. The proposed method is applied to the Breast Cancer Wisconsin (Diagnostic) Dataset which consists of 569 instances in 2 classes of Malignant and Benign. This method assigns certain instances to their appropriate classes with probability of one, and the uncertain instances to each of the classes with associated probabilities. Therefore, based on the degree of uncertainty, doctors can suggest further examinations before making the final diagnosis.

Keywords: Support Vector Machine(SVM), Fuzzy Classification, Training Data, Uncertain Points.

1 Introduction

Support Vector Machine (SVM) is one of the most powerful pattern classification methods known in recent years [1, 2]. In a two-class data set, SVM first maps the data set to a higher dimension to make the classes seperable. Then, it finds the seperating hyperplane while maximizing the margin between two classes. However, in many cases, the classes may have overlaps, i.e., they are not separable. In such cases, SVM allows some points to be on the wrong side with a specific cost, though it still maximizes the margin between the classes.

Moreover, there exist many applications, in which all points may not be exactly assigned to either of the classes, i.e., some points might be noisy. Generally, SVM assigns equal weights to all of the points. Thus, they have the same importance in determining the margin. Lin and Wang, in [4], addressed this deficiency by introducing a modification of SVM called "Fuzzy Support Vector Machine(FSVM)". By defining a fuzzy membership function, they assigned a fuzzy value or weight for each training point, and then ran SVM to reduce the effects of outliers and noisy points. FSVM has attracted a lot of attention on both improving and applying this method to different data sets. Diffent fuzzy membership funztions have been considered to improve the performance of FSVM [8], and an iterative Fuzzy Support Machine (IFSVM) method is introduced in [5]. Moreover, Abe and Inoue showed that FSVM can be generalized from two-class data sets to multiclass data sets [6].

Another pattern classification method is fuzzy classification, which is based on the truthness of value known as *membership value*. For each attribute of a class, a membership function is defined, and the range of membership values for every instance is between $[0, 1]$, where 1 shows the absolute truth and 0 shows the absolute false. For each class, an if-then rule is defined which shows if the membership values of an instance are in the determined ranges, then the instance would be assinged to that class.

Despite the great performance of FSVM and fuzzy classification, there is still a shortcoming. In medical applications such as disease recognition, a wrong diagnosis can have devastating effects on a person's life. To decrease the possibility of error, we propose a two-phase classification method which probabilistically assigns the uncertain points to each of the classes. First, FSVM is applied to the whole training data such that most of the uncertain points will be placed in the margin. Moreover, the certain points are assigned to appropriate classes. Next, a fuzzy membership function and an appropriate rule are defined to classify the points that were located in the margin. This will result in assigning uncertain points to each of the classes with a specific probability. The proposed method is applied to the Breast Cancer Wisconsin (Diagnostic) Dataset which consists of 569 instances in 2 classes of Malignant and Benign. This method assigns certain instances to their appropriate classes with probability of one, and the uncertain instances to each of the classes with associated probabilities. Therefore, based on the probability values, doctors can suggest further examinations before making the final diagnosis.

The organization of the rest of this paper is as follows. In Section 2, we present the idea behind Fuzzy Support Vector Machine method. In Section 3, we discuss the Fuzzy classification, and propose our probabilistic method. Finally, we provide simulation results in Section 4 and conclusions in Section 5.

2 Fuzzy SVM

In this section, the classic SVM is explained followed by FSVM. Suppose training data consists of N pairs $(x_1, y_1), \cdots, (x_N, y_N)$, where $x_i \in R^p$ and $y_i \in \{-1, 1\}$. Define a hyperplane by

$$x : f(x) = x^T\beta + \beta_0 = 0, \tag{1}$$

where β is a unit vector: $\|\beta\| = 1$. A classification rule induced by $f(x)$ is

$$G(x) = sign[x^T\beta + \beta_0]. \tag{2}$$

To deal with the overlap in the classes, SVM maximizes the margin between the training points for class 1 and -1 (M), but allows for some points to be on the wrong side of the margin. Defining the slack variables $\zeta = (\zeta_1, \zeta_2, ..., \zeta_N)$, the constraint can be modified as follows:

$$y_i(x_i^T\beta_+\beta_0) \geq M - \zeta_i, \quad \forall i = 1, \cdots, N, \tag{3}$$

$$\sum_{i=1}^{N} \zeta_i \leq \text{constant}, \quad \zeta_i \geq 0 \quad \forall i = 1, \cdots, N. \tag{4}$$

Defining $M = \frac{1}{\|\beta\|}$, we will have :

$$\min \qquad \|\beta\| + \sum_{i=1}^{N} \zeta_i \tag{5}$$

$$\text{s.t.} \qquad y_i(x_i^T\beta_+\beta_0) \geq 1 - \zeta_i, \quad \forall i = 1, \cdots, N \tag{6}$$

$$\zeta_i \geq 0, \quad \forall i = 1, \cdots, N \tag{7}$$

As explained earlier, in many real instances, all the data points do not have the same certaintity. Therefore, the uncertain points should get a lower weight, and have less contribution in determining the marginal region. In this paper, we consider a Gaussian function for the weights.

Suppose out of N points, N_1 points are in class 1 and N_2 remaining points are in class 2. Define the weight for each point as following:

$$W(x_i) = \prod_{j=1}^{p} exp^{\frac{-(x_{ij}-\mu_{jk})^2}{2\sigma_{jk}^2}}, \quad \forall x_i \in \text{Class } k, \tag{8}$$

where μ_{jk} and σ_{jk} refer to the mean and standard deviation of jth feature of all points in the class k, respectively. Moreover, x_{ij} indicates the jth feature value of ith point.

Then, we normalize the weights such that the total sum of the weights is equal to N, which is the sum of error costs for the classic SVM. In (9), $W_n(x_i)$ indicates the normalized weight.

$$W_n(x_i) = \frac{N}{\sum_{i=1}^{N} W(x_i)} W(x_i). \tag{9}$$

Finally, the wights show up in the objective function:

$$min \qquad \|\beta\| + \sum_{i=1}^{N} W_n(x_i)\zeta_i. \tag{10}$$

Points near to the center of each class have a higher weight than those farer. Therefore, near points will be classified certainly, and the points which are in the middle of the two classes, called uncertain points, will be located in the margin. In the next section, we discuss how to classify the marginal points probabilistically.

3 Fuzzy Classification of Marginal Points

In this part, we apply a fuzzy classification on the marginal points. Here, we use a fuzzy rule-based classification method, which has been applied to many data sets [9], [10]. The method used to generate the fuzzy rule is based on the mean and the standard deviation of each attribute [7]. Similar to the Gaussian weights in the previous section, a Gaussian fuzzy membership function A_{ik} is defined for every test point y_i located in the margin as

$$A_{ik} = \prod_{j=1}^{p} exp^{\frac{-(y_{ij}-\mu_{jk})^2}{2\sigma_{jk}^2}}, \quad \forall k \in 1, 2, \tag{11}$$

where μ_{jk} and σ_{jk} are the mean and standard deviation of training points of class k located in the margin, respectively. This membersip shows the closeness of element y_i to the center of kth class. To measure the related closeness of a point to both centers, a "membership probability" is defined for each marginal point as follows:

$$P_{i,C1} = \frac{A_{i,C1}}{A_{i,C1} + A_{i,C2}} \quad \text{and} \quad P_{i,C2} = 1 - P_{i,C1}. \tag{12}$$

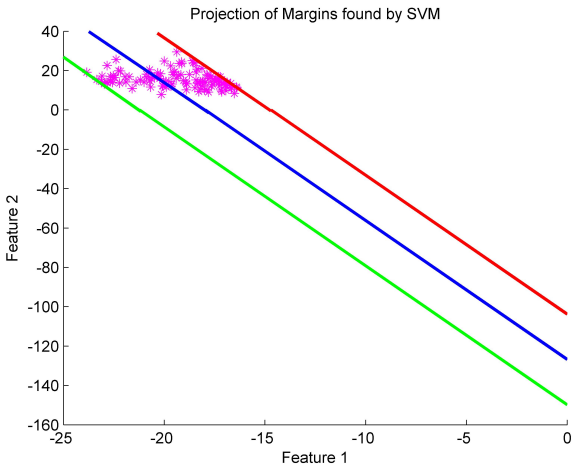
Points with probability more than 90% in class, will be assigned to that class. Otherwise, the given information is not sufficient to make a decision. Applying this probabilistic FSVM (PFSVM) on the Breast Cancer Wisconsin (Diagnostic) Dataset consisting of 569 instances, we will show that the probability of making a wrong decision is less than 1.23%, and the wrongly classified points in the FSVM will be determined as unknown classes, and will need additional information.

4 Simulation Results

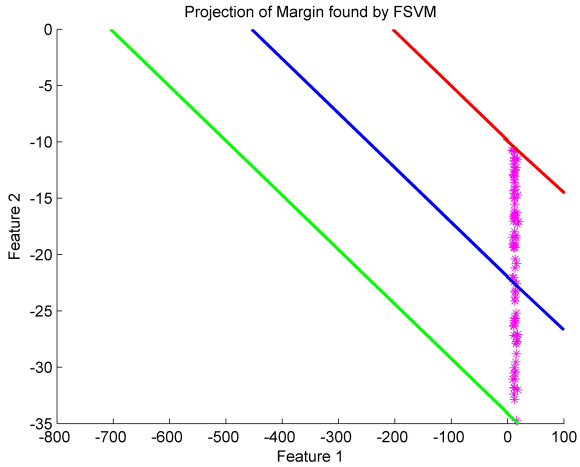
In this section, we examine the performance of PFSVM, and compare it with previous known methods. The data set we use is the Wisconsin breast cancer diagnostic dataset [11], which consists of 569 instances in two classes of

Malignan (M) and Benign (B) with 32 features per instance. First, the number of features is reduced from 32 to 23 by saving just one feature out of every set of features with correlation more than 0.95. Then by a 10-fold cross validation method the set of training and test data is determined. The two-phase probabilistic classifier is applied to the the new dataset.

In the first phase, we have obtained the margins using FSVM in which each training point gets a Gaussian weight. We then find the training points that are located inside the margin. Since there are 23 features, we project both margin and the points inside it onto a plane. Figure 1 compares the size of margins obtained by SVM and FSVM methods.



(a) SVM - Width of Margin: 0.895



(b) FSVM - Width of Margin: 1.931

Figure 1: Comparison of margins in SVM and FSVM

Moreover, observe that, on average, more than 80% of errors are located in the margin. Figure 2 illustrates an example. We can also observe that in all the cases, a Manignan cancer is classified in the Benign group which is more dangerous than misclassifying a benign type. Therefore, we double the weight of an error of type Manign.

By doubling the cost of misclassification of an "M" type instance, the errors occur in the margin with probability more than 98%. For comparison, we obtain the classifiers using both SVM and Fuzzy methods , and then apply it onto the test dataset. We also apply the probabilistic method explained in section 3 to the whole dataset, instead of marginal points found by FSVM. Table 1 shows the comparison.

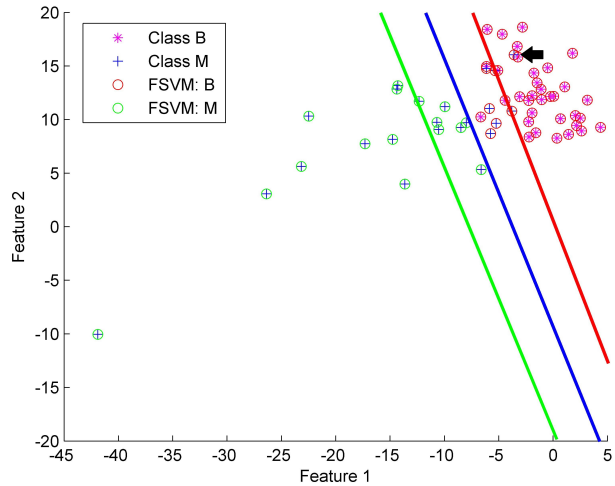


Figure 2: Errors located inside and outside of margin

Method \Run #	1	2	3	4	5	6	7	8	9	10	$percent_{ave}$
SVM_{err}	1	1	5	3	4	1	2	4	0	1	3.86
$FSVM_{err}$	4	4	5	7	7	3	2	1	3	4	7.02
$FUZZY_{err}$	3	3	5	8	4	6	5	3	3	1	7.19
$PFSVM_{err}$	1+1	1	0	0+2	3+1	0+2	0	1	2+1	0+2	1.63
$PFSVM_{undet}$	1	1	1	2	0	1	2	1	0	0	1.58

Table 1: Comparison of different classification methods

Table 1 shows that the error of probabilistic method is significantly smaller than the other methods. Note that the error of FSVM is higher than SVM which is due to the fact that the decreament of points' weights allows more errors and increases the margin. However, this increament of margin ensures us that the points outside of the margin are classified correctly.

Method \Run #	1	2	3	4	5	6	7	8	9	10	$percent_{ave}$
SVM_{err}	4	2	4	2	1	2	3	1	2	3	4.29
$FSVM_{err}$	6	2	3	2	1	2	4	1	4	5	5.36
$FUZZY_{err}$	7	2	5	4	1	3	6	3	4	4	6.96
$PFSVM_{err}$	0	2	1	0	0	2	0	0	0+1	1	1.23
$PFSVM_{undet}$	1	0	3	1	0	1	1	1	3	1	2.14

Table 2: Comparison of different classification methods - double cost

Table 2, shows the result of doubling the cost of errors in diagnosing Malignant cancer, i.e. if a cancer is Malignant, and diagnosed as Benign, this error has a higher cost than the reverse error. We can see that the classification rate is higher than 98.77% (error is 1.23%). Moreover, the sum of *error* and *undetermined* percentages is 3.37, which is less than all errors in both tables. Note that by running a probabilistic classification on the pure SVM, FSVM or Fuzzy classification, the sum of *error* and *undetermined* percentages will not be smaller than the current error. Therefore, neither of these classification methods can do better than the new probabilistic method.

5 Conclusion

In this paper, we considered the problem of disease diagnosis. We showed that in some cases, a deterministic classification method is not a proper method since the information may not be enough for making decision and mistakes can be devastating. Therefore, we came up with a probabilistic method and showed that it gives a smaller error rate than classic SVM and fuzzy classification. By applying our method to Breast Cancer Wisconsin (Diagnostic) Dataset, we showed that this method assigns a certain class to the cases that have strong evidence of being a Malignant or Benign type of cancer, where it assigns a probability to the cases which do not have enough information of being in a certain type.

References

[1] Burges, C., 1998, "A Tutorial on Support Vector Machines for Pattern Recognition", Data Mining and Knowledge Discovery, 2(2), 121167.

[2] Hastie, T., Tibshirani, R., Friedman, J., 2009, "The Elements of Statistical Learning: Data Mining, Inference, and Prediction", 2nd Edition, Springer-Verlag, New York.

[3] Schaefer, G., Zavisek, M., Drastuch, A., 2007, "Breat Cancer Classification Using Statistical Features and Fuzzy Classification of Thermograms", Proc. of IEEE International Conference on Fuzzy Systems, July 23-26, London, UK, 1-5.

[4] Lin, Ch., Wang, Sh., 2002, "Fuzzy Support Vector Machines", IEEE Transactions on Neural Networks, 13(2), 464-471.

[5] Shilton, A., Lai, D. , 2007, "Iterative Fuzzy Support Vector Machine Classification", Proc. of IEEE International Conference on Fuzzy Systems, July 23-26, London, UK, 1-6.

[6] Abe, Sh. , Inoue, T., 2002, "Fuzzy support vector machines for multiclass problems", Proc. of European Symposium on Artificial Neural Networks, 24-26 April, Bruges, Belgium, 113-118.

[7] Ravi, J., Ajith, A., 2004, "A Comparative Study of Fuzzy Classification Methods on Breast Cancer Data", Australasian Physical & Engineering Sciences in Medicine, 27(4), 213-218.

[8] Jiang, X. , Yi, Z. , Lv, J. C., 2006, "Fuzzy SVM with a new fuzzy membership function", Neural Computing & Applications, 15(3), 268-276.

[9] Ishibuchi, H., Nakashima, T., 1999, "Performance evaluation of fuzzy classifier systems for multi-dimensional pattern classification problems", IEEE Transactions on Systems, Man and Cybernetics, 29(2), 601-618.

[10] Ishibuchi, H., Nakashima, T., 1999, "Improving the performance of fuzzy classifier systems for pattern classification problems with continuous attributes", IEEE Transcations on Industrial Electronics, 46(6), 1057-1068.

[11] Merz, J., Murphy, P.M., 1996, UCI repository of machine learning databases. http://www.optical-network.com/topology.php .