ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 113-118

Fuzzy Support Vector Machines for Multiclass Problems

Shigeo Abe and Takuya Inoue
Graduate School of Science and Technology, Kobe University

Rokkodai, Nada, Kobe, Japan

Abstract. Since support vector machines for pattern classification are based on two-class classification problems, unclassifiable regions exist when extended to $n\,(>2)$-class problems. In our previous work, to solve this problem, we developed fuzzy support vector machines for one-to-$(n-1)$ classification. In this paper, we extend our method to pairwise classification. Namely, using the decision functions obtained by training the support vector machines for classes i and $j\,(j \neq i, j = 1, \ldots, n)$, for class i we define a truncated polyhedral pyramidal membership function. The membership functions are defined so that, for the data in the classifiable regions, the classification results are the same for the two methods. Thus, the generalization ability of the fuzzy support vector machine is the same with or better than that of the support vector machine for pairwise classification. We evaluate our method for four benchmark data sets and demonstrate the superiority of our method.

1 Introduction

Support vector machines outperform conventional classifiers especially when the number of training data is small and there is no overlap between classes [1, pp. 47–61]. For the conventional support vector machines, an n-class problem is converted into n two-class problems and for the ith two-class problem, class i is separated from the remaining classes. By this formulation, however, unclassifiable regions exist. To solve this problem, Kreßel [2] converts the n-class problem into $n(n-1)/2$ two-class problems which cover all pairs of classes. This method is called pairwise classification. By this method also unclassifiable regions remain. To resolve unclassified regions for the pairwise classification, Platt et al. [3] proposed decision-tree-based pairwise classification. Unclassifiable regions are resolved but decision boundaries are changed as the order of tree formation is changed. To solve this problem we proposed fuzzy support vector machines for one-to-$(n-1)$ classification [4].

In this paper, we extend our method to pairwise classification. Namely, using the decision functions obtained by training the support vector machines

ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 113-118

for pairs of classes, for each class we define a truncated polyhedral pyramidal membership function. The membership functions are defined so that, for the data in the classifiable regions, the classification results are the same with pairwise classification.

In Section 2, we explain two-class support vector machines, and in Section 3 we discuss fuzzy support vector machines for pairwise classification. In Section 4 we compare performance of the fuzzy support vector machine with that of the support vector machine for pairwise classification.

2 Two-class Support Vector Machines

Let m-dimensional inputs $\mathbf{x}_i\,(i = 1, \ldots, M)$ belong to Class 1 or 2 and the associated labels be $y_i = 1$ for Class 1 and -1 for Class 2. If these data are linearly separable, we can determine the decision function: $D(\mathbf{x}) = \mathbf{w}^t\,\mathbf{x} + b$, where $\mathbf{w}$ is an m-dimensional vector, b is a scalar, and

$$y_i\,D(\mathbf{x}_i) \geq 1 \quad \text{for} \quad i = 1, \ldots, M. \tag{1}$$

The distance between the separating hyperplane $D(\mathbf{x}) = 0$ and the training datum nearest to the hyperplane is called the margin. The hyperplane $D(\mathbf{x}) = 0$ with the maximum margin is called the optimal separating hyperplane.

Now consider determining the optimal separating hyperplane. The Euclidean distance from a training datum $\mathbf{x}$ to the separating hyperplane is given by $|D(\mathbf{x})|/\|\mathbf{w}\|$. Thus assuming the margin δ, all the training data must satisfy

$$\frac{y_k D(\mathbf{x}_k)}{\|\mathbf{w}\|} \geq \delta \qquad \text{for} \quad k = 1, \ldots, M. \tag{2}$$

If $\mathbf{w}$ is a solution, $a\mathbf{w}$ is also a solution where a is a scalar. Thus, we impose the following constraint:

$$\delta\,\|\mathbf{w}\| = 1. \tag{3}$$

From (2) and (3), to find the optimal separating hyperplane, we need to find $\mathbf{w}$ with the minimum Euclidean norm that satisfies (1).

The data that satisfy the equality in (1) are called support vectors.

Now the optimal separating hyperplane can be obtained by minimizing

$$\frac{1}{2}\,\|\mathbf{w}\|^2 \tag{4}$$

with respect to $\mathbf{w}$ and b subject to the constraints:

$$y_i\,(\mathbf{w}^t\,\mathbf{x}_i + b) \geq 1 \qquad \text{for} \qquad i = 1, \ldots, M. \tag{5}$$

We can solve (4) and (5) converting them into the dual problem. The above formulation can be extended to nonseparable cases.

ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 113-118

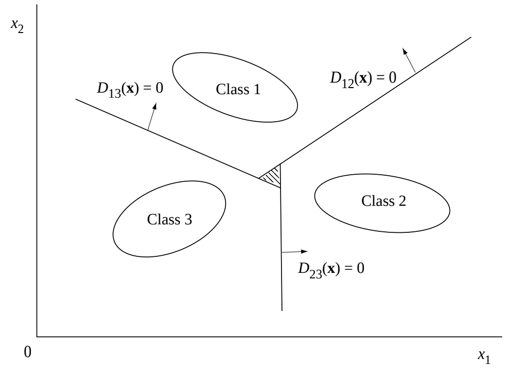

Figure 1: Unclassified regions by the pairwise formulation

3 Fuzzy Support Vector Machines

3.1 Conventional Pairwise Classification

Since the extension to nonlinear decision functions is straightforward, to simplify discussions, we consider linear decision functions. Let the decision function for class i against class j, with the maximum margin, be

$$D_{ij}(\mathbf{x}) = \mathbf{w}_{ij}^t\,\mathbf{x} + b_{ij}, \tag{6}$$

where $\mathbf{w}_{ij}$ is the m-dimensional vector, b_{ij} is a scalar, and $D_{ij}(\mathbf{x}) = -D_{ji}(\mathbf{x})$.

For the input vector $\mathbf{x}$ we calculate

$$D_i(\mathbf{x}) = \sum_{j \neq i, j=1}^{n} \mathrm{sign}(D_{ij}(\mathbf{x})), \tag{7}$$

where

$$\mathrm{sign}(x) = \begin{cases} 1 & x > 0, \\ 0 & x \leq 0 \end{cases} \tag{8}$$

and classify $\mathbf{x}$ into the class

$$\arg \max_{i=1,\ldots,n} D_i(\mathbf{x}). \tag{9}$$

If (9) is satisfied for plural i's, $\mathbf{x}$ is unclassifiable. In the shaded region in Fig. 1, $D_i(\mathbf{x}) = 1\,(i = 1, 2,$ and $3)$. Thus, the shaded region is unclassifiable.

3.2 Introduction of Membership Functions

Similar to the one-to-$(n-1)$ formulation [4], we introduce the membership functions to resolve unclassifiable regions while realizing the same classification results with that of the conventional pairwise classification. To do this, for the

ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 113-118

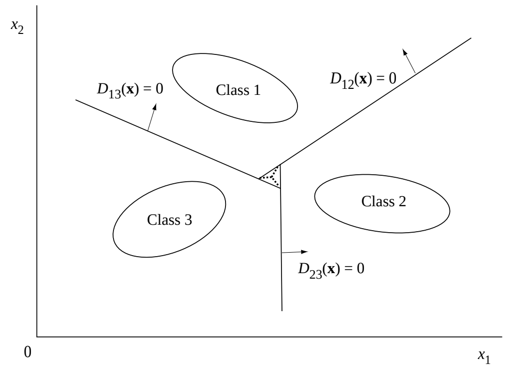

Figure 2: Extended generalization regions

optimal separating hyperplane $D_{ij}(\mathbf{x}) = 0\,(i \neq j)$ we define one-dimensional membership functions $m_{ij}(\mathbf{x})$ on the directions orthogonal to $D_{ij}(\mathbf{x}) = 0$ as follows:

$$m_{ij}(\mathbf{x}) = \begin{cases} 1 & \text{for} \quad D_{ij}(\mathbf{x}) \geq 1, \\ D_{ij}(\mathbf{x}) & \text{otherwise.} \end{cases} \tag{10}$$

Using $m_{ij}(\mathbf{x})\,(j \neq i, j = 1, \ldots, n)$, we define the class i membership function of $\mathbf{x}$ using the minimum operator:

$$m_i(\mathbf{x}) = \min_{j=1,\ldots,n} m_{ij}(\mathbf{x}). \tag{11}$$

Equation (11) is equivalent to

$$m_i(\mathbf{x}) = \min\left(1, \min_{j\neq,i,j=1,\ldots,n} D_{ij}(\mathbf{x})\right). \tag{12}$$

The shape of the membership function is shown to be a truncated polyhedral pyramid [1]. Since $m_i(\mathbf{x}) = 1$ holds for only one class, (12) reduces to

$$m_i(\mathbf{x}) = \min_{j\neq,i,j=1,\ldots,n} D_{ij}(\mathbf{x}). \tag{13}$$

Now an unknown datum $\mathbf{x}$ is classified into the class

$$\arg\max_{i=1,\ldots,n} m_i(\mathbf{x}). \tag{14}$$

Thus, the unclassified region shown in Fig. 1 is resolved as shown in Fig. 2.

4 Performance Evaluation

We evaluated our method using blood cell data [5], thyroid data[1], hiragana data with 50 inputs, and hiragana data with 13 inputs listed in Table 1 [1].

[1]ftp://ftp.ics.uci.edu/pub/machine-learning-databases/

ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 113-118

Table 1: Benchmark data specification

Data	Inputs	Classes	Training data	Test data
Blood cell	13	12	3097	3100
Thyroid	21	3	3772	3428
Hiragana-50	50	39	4610	4610
Hiragana-13	13	38	8375	8356

To compare our classification performance with other pairwise classification method, we used the software developed by Royal Holloway, University of London[2] [6]. The software resolved unclassifiable regions caused by the pairwise classification.

We used polynomial kernels: $(1 + \mathbf{x}\mathbf{x}')^d$ and RBF kernels: $\exp(-\gamma\|\mathbf{x} - \mathbf{x}'\|^2)$. To make comparison fair, we selected the values of d and γ so that the recognition rates of the training data became 100%. Table 2 lists the recognition rates of the test data for different kernels. In the table PW, PWM, and FPW mean pairwise classification, pairwise classification with some resolution by University of London, and our fuzzy pairwise classification, respectively. In most cases, the recognition rates by FPW are better than those by PW and PWM. FPW outperformed PWM for 12 cases out of 16 cases. The improvement of FPW over PW was especially evident for the blood cell data set, which is a very difficult classification problem.

5 Conclusions

In this paper, we proposed fuzzy support vector machines for pairwise classification that resolve unclassifiable regions caused by conventional support vector machines. In theory, the generalization ability of the fuzzy support vector machine is better than that of the conventional support vector machine. By computer simulations using four benchmark data sets, we demonstrated the superiority of our method over the support vector machines for pairwise classification.

References

[1] S. Abe. *Pattern Classification: Neuro-fuzzy Methods and Their Comparison*. Springer-Verlag, London, UK, 2001.

[2] http://svm.cs.rhbnc.ac.uk/

ESANN'2002 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 24-26 April 2002, d-side publi., ISBN 2-930307-02-1, pp. 113-118

Table 2: Performance for the benchmark data sets for different kernels

Data	Kernel	Parm	PW (%)	PWM (%)	FPW (%)
Blood cell	Poly	4	91.26	92.10	**92.35**
		5	91.03	91.90	**92.19**
		6	90.74	91.58	**91.74**
	RBF	10	91.52	91.58	**91.74**
Thyroid	Poly	4	96.27	96.56	**96.62**
	RBF	10	95.10	95.10	**95.16**
Hiragana-50	Poly	1	98.00	**98.29**	98.24
		2	98.89	98.94	98.94
		3	98.87	98.89	**98.94**
	RBF	0.1	99.02	99.02	99.02
		0.01	98.81	98.89	**98.96**
Hiragana-13	Poly	2	99.46	99.56	**99.63**
		3	99.47	99.53	**99.57**
		4	99.49	99.56	**99.57**
	RBF	1	99.76	**99.77**	99.76
		0.1	99.56	99.64	**99.70**

[2] U. H.-G. Kreßel. Pairwise classification and support vector machines. In B. Schölkopf, C. J. C. Burges, and A. J. Smola, editors, *Advances in Kernel Methods: Support Vector Learning*, pages 255–268. The MIT Press, Cambridge, MA, 1999.

[3] J. C. Platt, N. Cristianini, and J. Shawe-Taylor. Large margin DAGs for multiclass classification. In S. A. Solla, T. K. Leen, and K.-R. Müller, editors, *Advances in Neural Information Processing Systems 12*, pages 547–553. The MIT Press, 2000.

[4] T. Inoue and S. Abe. Fuzzy support vector machines for pattern classification. In *Proceedings of International Joint Conference on Neural Networks (IJCNN '01)*, volume 2, pages 1449–1454, July 2001.

[5] A. Hashizume, J. Motoike, and R. Yabe. Fully automated blood cell differential system and its application. In *Proceedings of the IUPAC Third International Congress on Automation and New Technology in the Clinical Laboratory*, pages 297–302, Kobe, Japan, September 1988.

[6] C. Saunders, M. O. Stitson, J. Weston, L. Bottou, B. Schölkopf, and A. Smola. Support vector machine reference manual. Technical Report CSD-TR-98-03, Royal Holloway, University of London, London, 1998.