ESANN'2004 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 28-30 April 2004, d-side publi., ISBN 2-930307-04-8, pp. 429-434

Fuzzy LP-SVMs for Multiclass Problems

Shigeo Abe
Graduate School of Science and Technology, Kobe University

Rokkodai, Nada, Kobe, Japan

Abstract. In this paper, we propose fuzzy linear programming support vector machines (LP-SVMs) that resolve unclassifiable regions for multiclass problems. Namely, in the directions orthogonal to the decision functions obtained by training the LP-SVM, we define membership functions. Then by the minimum or average operation for these membership functions we define a membership function for each class. We evaluate one-against-all and pairwise fuzzy LP-SVMs for some benchmark data sets and demonstrate the superiority of our fuzzy LP-SVMs over conventional LP-SVMs.

1 Introduction

Since support vector machines are formulated for two-class classification problems [1], there are several ways to extend to multiclass problems. Vapnik [1] proposed one-against-all classification, in which one class is separated from the remaining classes. By this classification, however, unclassifiable regions exist. Instead of discrete decision functions, Vapnik [1, p. 438] proposed to use continuous decision functions. Inoue and Abe [2] proposed fuzzy support vector machines, in which membership functions are defined using the decision functions. Abe [3] showed that support vector machines with continuous decision functions and fuzzy support vector machines are equivalent.

Kreßel [4] used pairwise classification, in which an n-class problem is converted into $n(n-1)/2$ two-class problems and decision is made by voting. But by this method also unclassifiable regions remain. To resolve unclassified regions for the pairwise classification, Platt et al. [5] 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 Abe and Inoue [6] proposed pairwise fuzzy support vector machines. This method is extended to least-squares SVMs [7].

In this paper, we extend our method to pairwise linear programming support vector machines (LP-SVMs). Namely, using the decision functions obtained by training the LP-SVM we define membership functions in the directions orthogonal to the decision functions. Then, by the minimum or average operation for these functions, for each class we define a membership function.

ESANN'2004 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 28-30 April 2004, d-side publi., ISBN 2-930307-04-8, pp. 429-434

We evaluate one-against-all and pairwise fuzzy LP-SVMs with minimum and average operators using two benchmark data sets.

In Section 2, we explain two-class LP-SVMs, and in Section 3 we discuss pairwise fuzzy LP-SVMs. In Section 4 we compare performance of the fuzzy LP-SVMs with minimum and average operators with that of the conventional LP-SVMs.

2 Two-class LP-SVMs

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.

In the LP-SVM, we define the decision function as follows [8]:

$$D(\mathbf{x}) = \sum_{i=1}^{M} \alpha_i H(\mathbf{x}, \mathbf{x}_i) + b, \tag{1}$$

where α_i take on real values, $H(\mathbf{x}, \mathbf{x}')$ is a kernel, and b is a bias. Unlike the conventional SVMs, the kernel needs not be positive semi-definite.

We consider minimizing

$$Q(\boldsymbol{\alpha}, \boldsymbol{\xi}) = \sum_{i=1}^{M} (|\alpha_i| + C\xi_i) \tag{2}$$

subject to

$$y_j \left(\sum_{i=1}^{M} \alpha_i H(\mathbf{x}_j, \mathbf{x}_i) + b \right) \geq 1 - \xi_j \quad \text{for} \quad j = 1, \ldots, M, \tag{3}$$

where ξ_i are positive slack variables and C is a margin parameter.

Letting $\alpha_i = \alpha_i^+ - \alpha_i^-$ and $b = b^+ - b^-$, where $\alpha_i^+ \geq 0$, $\alpha_i^- \geq 0$, $b^+ \geq 0$, and $b^- \geq 0$, we can solve (2) and (3) for α_i, ξ_i, and b by linear programming.

Similar to conventional SVMs, LP-SVMs have degenerate solutions [9]. Namely, α_i are all zero. The difference is that LP-SVMs have degenerate solutions when C is small as the following theorem shows.

Theorem For the LP-SVM, there exists a positive C_0 such that for $0 \leq C \leq C_0$, the solution is degenerate.

Proof Because of the slack variables ξ_i,(3) has a feasible solution. Thus, for large C, (2) and (3) have the optimal solution with some α_i being non-zero.

For $\alpha_i = 0$ $(i = 1, \ldots, M)$, (3) reduces to

$$y_i\, b \geq 1 - \xi_i. \tag{4}$$

For $b = 0$, (4) is satisfied for $\xi_i = 1$. Then (2) is

$$Q(\boldsymbol{\alpha}, \boldsymbol{\xi}) = MC. \tag{5}$$

Thus, by decreasing the value of C from a large value, we can find a maximum value of C, C_0, in which (2) is minimized for $\alpha_i = 0$. For $0 < C \leq C_0$, it is evident that $\alpha_i = 0$ are the optimal solution for (2) and (3).

ESANN'2004 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 28-30 April 2004, d-side publi., ISBN 2-930307-04-8, pp. 429-434

3 Fuzzy LP-SVMs

Since there is no much difference in pairwise and one-against-all fuzzy LP-SVMs, in the following we discuss pairwise fuzzy LP-SVMs.

3.1 Conventional Pairwise Classification

Let the decision function for class i against class j, with the maximum margin, be

$$D_{ij}(\mathbf{x}) = \sum_{i=1}^{M} \alpha_{ij} H(\mathbf{x}, \mathbf{x}_i) + b_{ij}, \tag{6}$$

where α_{ij} take on real values, 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 n is the number of classes and

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

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.

3.2 Introduction of Membership Functions

To resolve unclassifiable regions, for the optimal separating hyperplanes $D_{ij}(\mathbf{x}) = 0\,(i \neq j)$ we define one-dimensional membership functions $m_{ij}(\mathbf{x})$ in 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:

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

or using the average operator

$$m_i(\mathbf{x}) = \frac{1}{n-1} \sum_{i \neq j, i=1}^{M} m_{ij}(\mathbf{x}). \tag{12}$$

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

$$\arg \max_{i=1,\ldots,n} m_i(\mathbf{x}). \tag{13}$$

Table 1: Benchmark data specification

Data	Inputs	Classes	Training data	Test data
Iris	4	3	75	75
Numeral	12	10	810	820
Blood cell	13	12	3097	3100
Hiragana	50	39	4610	4610

4 Performance Evaluation

We evaluated our method using the data sets [10] listed in Table 1.

We used polynomial kernels: $(1+\mathbf{x}^t\mathbf{x}')^d$ and RBF kernels: $\exp(-\gamma\|\mathbf{x}-\mathbf{x}'\|^2)$ where γ is a positive constant. We determined the value of C by 5-fold cross validation for the training data.

Here we show the performance of one-against-all and pairwise fuzzy LP-SVMs. Table 2 shows the parameters, recognition rates (in %) of the one-against-all LP-SVM and fuzzy LP-SVM, and the number of support vectors (SVs). The recognition rates of the training data are shown in the brackets when they are not 100%. Because of the large memory consumption, we could only train LP-SVMs for iris and numeral data sets.

For one-against-all classification, continuous SVMs are equivalent to fuzzy SVMs with minimum operators [3]. We can also prove that fuzzy SVMs with minimum and average operators are equivalent. Thus in the table we show the results of discrete SVMs and fuzzy SVMs (FSVMs). By the introduction of membership functions, recognition rates were improved.

Table 3 shows the results for the fuzzy L1-SVMs. Comparing Tables 2 and 3, the recognition rates of the test data for the fuzzy L1-SVMs and fuzzy LP-SVMs are comparable but the numbers of support vectors of LP-SVMs are smaller. This is because in LP-SVMs, the sum of $|\alpha_i|$ is minimized. Thus it leads to a smaller number of support vectors.

Table 2: Recognition rates of one-against-all fuzzy LP-SVMs (%)

Data	Parm	SVM	FSVM	SVs
Iris	$d=2$	**97.33** (97.33)	**97.33** (97.33)	3
	$\gamma=1$	92.00 (96.00)	**93.33** (97.33)	2
Numeral	$d=2$	99.15	**99.63**	8
	$\gamma=1$	99.27 (99.63)	**99.63** (99.88)	7

ESANN'2004 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 28-30 April 2004, d-side publi., ISBN 2-930307-04-8, pp. 429-434

Table 3: Recognition rates of one-against-all fuzzy SVMs (%)

Data	Parm	SVM	FSVM	SVs
Iris	$d = 2$	93.33 (96.00)	**96.00** (96.00)	23
	$\gamma = 1$	92.00 (96.00)	**94.67** (97.33)	21
Numeral	$d = 2$	99.15	**99.39**	15
	$\gamma = 1$	99.02 (99.63)	**99.27** (99.88)	17

Table 4 shows the parameters, recognition rates of the pairwise LP-SVM and fuzzy LP-SVMs with minimum and average operators, and the number of support vectors. The recognition rates of the training data are shown in the brackets when they are not 100%.

The recognition rates of the fuzzy LP-SVMs with minimum and average operators are almost the same. Thus, we can choose either of the operators.

As seen from Tables 2 and 4, there is no much difference between pairwise classification and one-against-all classification.

Table 4: Recognition rates of pairwise fuzzy LP-SVMs (%)

Data	Parm	SVM	FSVM (Min)	FSVM (Ave.)	SVs
Iris	$d = 2$	**97.33** (97.33)	**97.33** (97.33)	**97.33** (97.33)	2
	$\gamma = 1$	**94.67** (98.67)	**94.67** (98.67)	**94.67** (98.67)	2
Numeral	$d = 2$	**99.76**	**99.76**	**99.76**	3
	$\gamma = 1$	**99.39** (99.63)	**99.39** (99.63)	**99.39** (99.63)	3
Blood cell	$d = 4$	91.61 (98.58)	92.55 (98.61)	**92.58** (92.39)	7
	$\gamma = 1$	91.48 (96.74)	92.32 (96.84)	**92.39** (96.87)	6
Hiragana	$d = 2$	97.68	**98.05**	97.94	7
	$\gamma = 1$	97.31	**97.87**	97.61	6

5 Conclusions

In this paper, we proposed fuzzy LP-SVMs that resolve unclassifiable regions caused by conventional support vector machines. Namely, in the directions orthogonal to the decision functions obtained by training the LP-SVM, we define membership functions. Then by the minimum or average operation for these membership functions we define a membership function for each class. We evaluated one-against-all and pairwise fuzzy LP-SVMs for some benchmark data sets and showed that the generalization ability was improved by the introduction of membership functions.

References

[1] V. N. Vapnik. *Statistical Learning Theory.* John Wiley & Sons, New York, NY, 1998.

[2] 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.

[3] S. Abe. Analysis of multiclass support vector machines. In *Proceedings of International Conference on Computational Intelligence for Modelling Control and Automation (CIMCA'2003)*, pages 385–396, Vienna, Austria, February 2003.

[4] 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.

[5] 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, Cambridge, MA, 2000.

[6] S. Abe and T. Inoue. Fuzzy support vector machines for multiclass problems. In *Proceedings of the Tenth European Symposium on Artificial Neural Networks (ESANN"2002)*, pages 113–118, Bruges, Belgium, April 2002.

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

[8] B. Schölkopf and A. J. Smola. *Learning with Kernels.* The MIT Press, Cambridge, MA, 2002.

[9] S. Abe. Analysis of support vector machines. In *Neural Networks for Signal Processing XII—Proceedings of the 2002 IEEE Signal Processing Society Workshop*, pages 89–98, 2002.

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