A New Model of Multi-class Support Vector Machine with Parameter v

Xin Xue[1]
[1]Department of Mathematics & System Science,
Taishan University, Tai'an,271021,China
xuexin04@163.com

Taian Liu[2]
[2]Department of Information and Engineering,
Shandong University of Science
and Technology,Tai'an 271019,China

Abstract

A new model of multi-class support vector machine with parameter v (v-MC-SVM) is proposed firstly based on v-SVM. Existence of optimal solutions and dual problem of v-MC-SVM are also given. Because the constraints of v-MC-SVM are too complicated, one-class SVM problem is given by adding b_m to the objective function of v-MC-SVM and employing the Kesler's construction which simplify the original problem. The optimal solutions of one-class SVM problem are unchanged when its constraint $e^T\alpha \geq v$ is replaced with $e^T\alpha = v$. Numerical testing results show that the speed of v-MC-SVM algorithm is faster than that of QP-MC-SVM algorithm under the same accuracy rate.

Key words. Machine Learning, v-SVM, Multi-class SVM, Parameter v, One-class SVM

1. Introduction

Because SVMs can well resolve such practical problem as nonlinearity, high dimension and local minima, they have attracted more and more attention[1-3] and become a hot issue in the field of Machine Learning, such as handwritten numerals recognition, face recognition, texture classification, and so on[4-6]. Because the selection of value of C of the standard SVM(C-SVM) is difficult, Scholkopf proposed another support vector machine with parameter v(v-SVM) which is an improved algorithm. The value of v of v-SVM is related to the number of misclassified samples, support vectors and all train samples[7-8].

In this paper, the basic theory and specific meaning of v-SVM is introduced. A new model of multi-class support vector machine with parameter v(v-MC-SVM) is proposed based on v-SVM. The existence of optimal solutions and dual problem of v-MC-SVM are also given. Because the constraints of v-MC-SVM are too complicated, one-class SVM problem of v-MC-SVM is given by adding b_m to the objective function of v-MC-SVM and employing the Kesler's construction which simplifies the original problem. Then, we prove that the optimal solutions of one-class SVM of v-MC-SVM problem are unchanged when its constraint $e^T\alpha \geq v$ is replaced with $e^T\alpha = v$. Numerical testing results show lastly that the speed of v-MC-SVM algorithm is faster than that of QP-MC-SVM algorithm under the same accuracy rate.

2. v-SVM

Let us consider that we are given labelled training patterns $\{(x_i, y_i)|i \in I\}$, where a pattern x_i comes from an n-dimensional space X and its label attains a value from the set $\{-1, 1\}$. $I = \{1, \cdots, l\}$ denotes a set of indices.

2.1. The basic theory of v-SVM

The non-linear classification rules $f_m(x) = (w_m \cdot x) + b_m, m \in K$, (the dot product is denoted by$(\cdot)$) can be found directly by solving the v-SVM problem

$$\begin{aligned} &\min_{w,b,\xi,\rho} \frac{1}{2} \parallel w \parallel^2 - v\rho + \frac{1}{l}\sum_{i\in I}\xi_i \\ &s.t.\ \ y_i((w \cdot \varphi(x_i)) + b) \geq \rho - \xi_i, i \in I, \\ &\qquad \rho \geq 0, \xi_i \geq 0, i \in I. \end{aligned} \tag{1}$$

Where, $\frac{2\rho}{\|w\|}$ denotes the largest margin between classes. For a non-separable case, ξ_i is a nonnegative slack variable of a training data $x_i, i \in I$. $\varphi(x)$ denotes a mapping from input space X to feature space H. v is a parameter which will be selected in procedure.

The dual problem of problem (1) is as follows

$$\begin{aligned} &\min_{\alpha} \frac{1}{2}\sum_{i,j\in I} y_i y_j \alpha_i \alpha_j k(x_i, x_j) \\ &s.t. \sum_{i\in I} y_i\alpha_i = 0, \\ &\qquad \sum_{i\in I}\alpha_i \geq v, \\ &\qquad 0 \leq \alpha_i \leq \frac{1}{l}, i \in I. \end{aligned} \tag{2}$$

Where, kernel function $k(x_i, x_j) = (\varphi(x_i) \cdot \varphi(x_j))$. Crisp and Burges[100]improved that when the constraint of prob-

978-0-7695-3735-1/09 $25.00 © 2009 IEEE
DOI 10.1109/FSKD.2009.397

lem (2) $e^T\alpha \geq v$ is replaced with $e^T\alpha = v$, the optimal solutions of problem (2) are same as the following problem

$$\begin{array}{ll}\min\limits_{\alpha} \frac{1}{2}\sum\limits_{i,j\in I} y_i y_j \alpha_i \alpha_j k(x_i, x_j) \\ s.t. \ \sum\limits_{i\in I} y_i\alpha_i = 0, \\ \quad\ \sum\limits_{i\in I}\alpha_i = v, \\ \quad\ 0 \leq \alpha_i \leq \frac{1}{l}, i \in I.\end{array} \tag{3}$$

2.2. The specific meaning of v of v-SVM

$\tilde{\alpha}$ denotes the optimal solution of problem (3). If $\tilde{\alpha}_i \neq 0$, the training sample x_i is named as support vector. $(\tilde{w}, \tilde{b}, \tilde{\rho}, \tilde{\xi})$ denotes the optimal solution of (1). If $\tilde{\xi}_i > 0$, the training sample x_i is misclassified. If $\tilde{\rho}_i > 0$ [9], then

(1)If p denotes the number of input sample misclassified, we have $v \geq \frac{p}{l}$.

(2) If q denotes the number of support vector, we have $v \leq \frac{q}{l}$.

(3) When $l \to \infty$, v converges to $\frac{q}{l}$ with probability 1 under some condition.

So, the selection of value of v will be based on all the conclusions provided as above.

3. Models of multi-class support vector machine with parameter v

Let us consider that we are given labelled training patterns $\{(x_i, y_i)|i \in I\}$, where a pattern x_i comes from an n-dimensional space X and its label attains a value from a set $K = \{1, 2, \cdots, k\}$. $I = \{1, \cdots, l\}$ denotes a set of indices.

The original problem of multi-class support vector machine with parameter v(v-MC-SVM) is proposed based on v-SVM as follows

$$\begin{array}{l}\min\limits_{w,b,\xi,\rho} \frac{1}{2}\sum\limits_{m\in K} \parallel w_m \parallel^2 - v\rho + \frac{1}{l(k-1)}\sum\limits_{i\in I}\sum\limits_{m\in K\backslash y_i}\xi_i^m \\ s.t. \ (w_{y_i}\cdot x_i) + b_{y_i} - ((w_m \cdot x_i) + b_m) \geq \rho - \xi_i^m, \\ \qquad \rho \geq 0, \xi_i^m \geq 0, i \in I, m \in K\backslash y_i.\end{array} \tag{4}$$

Where the minimization of the sum of norms $\|w_m\|^2$ leads to maximization of the margin between classes. ξ_i denotes the nonnegative slack variant of training sample $x_i, i \in I$. $\varphi(x)$ denotes a mapping from input space X to feature space H. v is a parameter which will be selected in procedure.

Several theorems as to v-MC-SVM is given as follows.

Theorem 3.1 *Overall optimal solution of problem (4) is existent*

proof: Problem (4) is convex quadratic programming. Its feasible region is non-vacuous. Then, there are solutions of problem (4). By the minima theorem of convex function, we know that the set of solutions of problem (4) is convex set, and any solution is the overall optimal solution.

If $(\tilde{w}, \tilde{b}, \tilde{\xi}, \tilde{\rho})$ is a solution of problem (4), the classifying function is as follows

$$f(x) = \arg\max_j((\tilde{w}_j \cdot \varphi(x)) + \tilde{b}_j), j \in K. \tag{5}$$

Theorem 3.2 *Duel problem of problem (4) is as follows*

$$\begin{array}{l}\min\limits_{\alpha} \sum\limits_{i\in I}\sum\limits_{j\in I}(\frac{1}{2}c_j^{y_i}A_iA_j - \sum\limits_{m\in K}\alpha_i^m\alpha_j^{y_i} + \\ \quad \frac{1}{2}\sum\limits_{m\in K}\alpha_i^m\alpha_j^m)k(x_i, x_j) \\ s.t. \ \sum\limits_{i\in I}\alpha_i^m = \sum\limits_{i\in I}c_i^m A_i, m \in K, \\ \quad \sum\limits_{i\in I}\sum\limits_{m\in K\backslash y_i}\alpha_i^m \geq v, \\ \quad 0 \leq \alpha_i^m \leq C, \alpha_i^{y_i} = 0, i \in I, m \in K\backslash y_i.\end{array} \tag{6}$$

Where, $A_i = \sum\limits_{m\in K}\alpha_i^m$*, and*

$$c_j^{y_i} = \begin{cases} 1 & \text{if} y_i = y_j \\ 0 & \text{if} y_i \neq y_j \end{cases}, i \in I, j \in I. \tag{7}$$

Proof: The Lagrange function of problem (4) as follows[9]

$$\begin{array}{l}L(w, b, \xi, \rho, \alpha, \beta, \delta) = \frac{1}{2}\sum\limits_{m\in K} \parallel (w_m, b_m) \parallel^2 - v\rho \\ + \frac{1}{l(k-1)}\sum\limits_{i\in I}\sum\limits_{m\in K\backslash y_i}\xi_i^m - \sum\limits_{i\in I}\sum\limits_{m\in K\backslash y_i}\alpha_i^m((w_{y_i}\cdot x_i) \\ + b_{y_i} - ((w_m\cdot x_i) + b_m) - \rho + \xi_i^m) \\ - \sum\limits_{i\in I}\sum\limits_{m\in K\backslash y_i}\beta_i^m\xi_i^m - \delta\rho.\end{array} \tag{8}$$

Where α_i^m, β_i^m and δ are Lagrange multipliers. Let

$$\begin{array}{l}\frac{\partial}{\partial w_m}L(w, b, \xi, \rho, \alpha, \beta, \delta) = w_m - \sum\limits_{i\in I}(c_i^m A_i \\ \qquad\qquad - \alpha_i^m)\varphi(x_i) = 0, \\ \frac{\partial}{\partial b_m}L(w, b, \xi, \rho, \alpha, \beta, \delta) = \sum\limits_{i\in I}(c_i^m A_i - \alpha_i^m) = 0, \\ \frac{\partial}{\partial \xi_i^m}L(w, b, \xi, \rho, \alpha, \beta, \delta) = \frac{1}{l(k-1)} - \alpha_i^m - \beta_i^m = 0, \\ \frac{\partial}{\partial \rho}L(w, b, \xi, \rho, \alpha, \beta, \delta) = \sum\limits_{i\in I}\sum\limits_{m\in K\backslash y_i}\alpha_i^m - v - \delta = 0.\end{array} \tag{9}$$

Accordingly, we have that

$$\begin{array}{l}w_m = \sum\limits_{i\in I}(c_i^m A_i - \alpha_i^m)\varphi(x_i), \\ \sum\limits_{i\in I}c_i^m A_i = \sum\limits_{i\in I}\alpha_i^m, \\ \alpha_i^m - \beta_i^m = \frac{1}{l(k-1)}, \\ \sum\limits_{i\in I}\sum\limits_{m\in K\backslash y_i}\alpha_i^m - \delta = v.\end{array} \tag{10}$$

Then, we have that

$$
\begin{aligned}
&L(w,b,\xi,\rho,\alpha,\beta,\delta)=\tfrac{1}{2}(\sum_{i\in I}(c_i^m A_i-\alpha_i^m)\varphi(x_i))\\
&(\sum_{i\in I}(c_i^m A_i-\alpha_i^m)\varphi(x_i))-v\rho+\tfrac{1}{l(k-1)}\sum_{i\in I}\sum_{m\in K\backslash y_i}\xi_i^m\\
&-\sum_{i\in I}\sum_{m\in K\backslash y_i}\alpha_i^m((w_{y_i}\cdot x_i)+b_{y_i}-((w_m\cdot x_i)+b_m)\\
&-\rho+\xi_i^m)-\sum_{i\in I}\sum_{m\in K\backslash y_i}\beta_i^m\xi_i^m-\delta\rho\\
&=-\sum_{i\in I}\sum_{j\in I}(\tfrac{1}{2}c_j^{y_i}A_iA_j-\sum_{m\in K}\alpha_i^m\alpha_j^{y_i}+\\
&\quad\tfrac{1}{2}\sum_{m\in K}\alpha_i^m\alpha_j^m)k(x_i,x_j).
\end{aligned}
\tag{11}
$$

Finally, we get the duel problem of problem (4) by Wolfe Duel Theorem.

Theorem 3.3 *Let kernel matrix $\tilde{K}=(k(x_i,x_j))_{i,j\in I}$. If $\tilde{K}$ is positive definite kernel matrix, there are solutions of the duel problem (6).*

proof: If $\tilde{K}$ is positive definite kernel matrix, the duel problem (6) is convex quadratic programming. Its feasible region is non-vacuous. Because objective function of the duel problem (6) is nonnegative, and it has bound from below, there are overall optimal solutions of the duel problem (6).

4. The modified problems of v-MC-SVM

The constraints of v-MC-SVM are too complicated. One-class SVM problem of v-MC-SVM is given by adding b_m to the objective function and employing the Kesler's construction. The solution of one-class SVM problem of v-MC-SVM is unchanged when its constraint $e^T\alpha\geq v$ is replaced with $e^T\alpha=v$.

4.1. One-class SVM of v-MC-SVM

We modify v-MC-SVM by adding b_m to the objective function of problem (4) which leads to the BSVM model as follows

$$
\begin{aligned}
&\min_{w,b,\xi,\rho}\tfrac{1}{2}\sum_{m\in K}\parallel w_m,b_m\parallel^2-v\rho+\tfrac{1}{l(k-1)}\sum_{i\in I}\sum_{m\in K\backslash y_i}\xi_i^m\\
&s.t.\ \ (w_{y_i}\cdot x_i)+b_{y_i}-((w_m\cdot x_i)+b_m)\geq\rho-\xi_i^m,\\
&\qquad\rho\geq 0,\xi_i^m\geq 0,i\in I,m\in K\backslash y_i.
\end{aligned}
\tag{12}
$$

Similarly,we set

$$w=((w_1^T,b_1),\cdots,(w_k^T,b_k))^T\ . \tag{13}$$

and

$$z_i^m=((z_i^m(1),\cdots,(z_i^m(k))^T\ . \tag{14}$$

This Kesler's construction (13) and (14) maps the input n-dimensional space X to a new $(n+1)\cdot k$-dimensional space Y where the multi-class problem appears as the one-class problem[10-11]. Each training pattern x_i is mapped to new $(k-1)$ pattern $z_i^m, m\in K\backslash y_i$, defined as follows. We assume that coordinates of z_i^m are divided into k slots. Each slot

$$z_i^m(j)=\begin{cases}[x_i,1] & \text{for}\quad j=y_i\\ -[x_i,1] & \text{for}\quad j=m\\ 0 & \text{otherwise}\end{cases}\ ,\ j\in K. \tag{15}$$

By performing the transformation (13)-(15), the BSVM problem of v-MC-SVM can be equivalently expressed as the one-class SVM problem

$$
\begin{aligned}
&\min_{w,\xi,\rho}\tfrac{1}{2}\parallel w\parallel^2-v\rho+\tfrac{1}{l(k-1)}\sum_{i\in I}\sum_{m\in K\backslash y_i}\xi_i^m\\
&s.t.\ \ (w\cdot z_i^m)\geq\rho-\xi_i^m,i\in I,m\in K\backslash y_i,\\
&\qquad\rho\geq 0,\xi_i^m\geq 0,i\in I,m\in K\backslash y_i.
\end{aligned}
\tag{16}
$$

4.2. The duel problem of one-class SVM problem of v-MC-SVM

The Lagrange function of problem (16) is as follows

$$
\begin{aligned}
&L(w,\xi,\rho,\alpha,\beta,\delta)=\tfrac{1}{2}(w\cdot w)-v\rho\\
&+\tfrac{1}{l(k-1)}\sum_{i\in I}\sum_{m\in K\backslash y_i}\xi_i^m-\sum_{i\in I}\sum_{m\in K\backslash y_i}\alpha_i^m((w\cdot z_i^m)\\
&-\rho+\xi_i^m)-\sum_{i\in I}\sum_{m\in K\backslash y_i}\beta_i^m\xi_i^m-\delta\rho.
\end{aligned}
\tag{17}
$$

Where α_i^m ,β_i^m and δ are Lagrange multipliers. Let

$$
\begin{aligned}
&\tfrac{\partial}{\partial w_m}L(w,\xi,\rho,\alpha,\beta,\delta)=w-\sum_{i\in I}\alpha_i^m z_i=0,\\
&\tfrac{\partial}{\partial\xi_i^m}L(w,b,\xi,\rho,\alpha,\beta,\delta)=\tfrac{1}{l(k-1)}-\alpha_i^m-\beta_i^m=0,\\
&\tfrac{\partial}{\partial\rho}L(w,\xi,\rho,\alpha,\beta,\delta)=\sum_{i\in I}\sum_{m\in K\backslash y_i}\alpha_i^m-v-\delta=0.
\end{aligned}
\tag{18}
$$

Accordingly, we have the duel problem of problem (16) is as follows

$$
\begin{aligned}
&\min_{\alpha}\sum_{i,j\in I}\sum_{m\in K\backslash y_i}\alpha_i^m\alpha_j^n(z_i^m\cdot z_j^n)\\
&s.t.\quad 0\leq\alpha_i^m\leq\tfrac{1}{l(k-1)},i\in I,m\in K\backslash y_i,\\
&\qquad\sum_{m,i}\alpha_i^m\geq v.
\end{aligned}
\tag{19}
$$

Where, the dot product between z_i^m and z_j^n

$$
\begin{aligned}
(z_i^m\cdot z_j^n)&=(k(x_i,x_j)+1)\cdot(\delta(y_i,y_j)\\
&+\delta(m,n)-\delta(y_j,n)-\delta(y_i,m)),
\end{aligned}
\tag{20}
$$

and

$$\delta(i,j)=\begin{cases}1 & \text{for}\quad i=j,\\ 0 & \text{for}\quad i\neq j.\end{cases} \tag{21}$$

4.3. The equivalent problem of problem(19)

For convenience, the model of problem (19) is given as follows

$$\begin{aligned} &\min_{\alpha} \tfrac{1}{2}\alpha Q\alpha \\ &s.t. 0 \le \alpha \le \tfrac{1}{l(k-1)}, \\ &\quad e^T\alpha \ge v. \end{aligned} \tag{22}$$

Where kernal matrix $Q = ((z_i^m \cdot z_j^n))_{i,j,m,n}$, $\alpha = (\alpha_i^m)_{i,m}$ and $e = (1, \cdots, 1)^T$,$i, j \in I, m, n \in K \backslash y_i$.

Theorem 4.1 *If there are feasible solutions of problem (22), there is one optimal solution at least which is subject to* $e^T\alpha = v$. *If the optimal value of problem (22) is not zero, any optimal solution of problem (22) is subject to* $e^T\alpha = v$.

Proof: Because problem (22) is convex quadratic programming, and its feasible region is bound. There is one solution at least. Suppose that there is a optimal solution α^* which is subject to $e^T\alpha^*$ >v. We have $e^T\alpha^*$¿v¿0. Set

$$\bar{\alpha} \equiv \tfrac{v}{e^T\alpha^*}\alpha^*, \tag{23}$$

Obviously, $\bar{\alpha}$ is subject to $e^T\bar{\alpha}$ =v. Because α^* is a optimal solution of problem (22), and $e^T\alpha^* > v$, we have that

$$(\alpha^*)^T Q\alpha^* \le \bar{\alpha}^T Q\bar{\alpha} = (\tfrac{v}{e^T\alpha^*})^2(\alpha^*)^T Q\alpha^* \le (\alpha^*)^T Q\alpha^*. \tag{24}$$

So, $\bar{\alpha}$ is a optimal solution of problem (22), and we have $e^T\alpha^* = v$. Accordingly, there is one optimal solution of problem (22) at least which is subject to $e^T\alpha^* = v$.

From equation (24), We have that

$$\begin{aligned} &(\alpha^*)^T Q\alpha^* \le \bar{\alpha}^T Q\bar{\alpha} = (\tfrac{v}{e^T\alpha^*})^2(\alpha^*)^T Q\alpha^* \\ &\Longrightarrow (\alpha^*)^T Q\alpha^* \le (\tfrac{v}{e^T\alpha^*})^2 \\ &\Longrightarrow (1 - (\tfrac{v}{e^T\alpha^*})^2(\alpha^*)^T Q\alpha^* \le 0. \end{aligned} \tag{25}$$

Because $e^T\alpha^*$ >v,then

$$(1 - (\tfrac{v}{e^T\alpha^*})^2 > 0. \tag{26}$$

From equation (25) and (26), we know that $(\alpha^*)^T Q\alpha^* = 0$. Accordingly, we get the conclusion that if the optimal value of problem (22) is not zero, any optimal solution of problem (22) is subject to $e^T\alpha^* = v$.

Theorem 4.1 show that the constraint conditions $e^T\alpha^* \ge v$ of problem (22) is replaced with $e^T\alpha^*$ =v, the optimal solutions of problem (22) are unchanged.

So,we get the equivalent problem of problem (22) as follows

$$\begin{aligned} &\min_{\alpha} \sum_{i,j\in I} \sum_{m\in K\backslash y_i} \alpha_i^m \alpha_j^n (z_i^m \cdot z_j^n) \\ &s.t. \quad 0 \le \alpha_i^m \le \tfrac{1}{l(k-1)}, i \in I, m \in K\backslash y_i, \\ &\quad \sum_{m,i} \alpha_i^m = v. \end{aligned} \tag{27}$$

By solving problem (27), we get the classifying function

$$\begin{aligned} f(x) = \arg\max_j (&\sum_{i\in I} \sum_{m\in K\backslash y_i} \alpha_i^m (\delta(j, y_i) \\ &\delta(j, m))(k(x_j, x) + 1)). \end{aligned} \tag{28}$$

5. Numerical Testing

Our numerical testing is carried on three data sets of UCI-benchmark repository of artificial and real data sets. The first data set Glass includes 6 classes and 214 samples with 10 features. The second data set Hyroid includes 3 classes and 215 samples with 5 features. The second data set Wine includes 3 classes and 178 samples with 13 features. Software LIBSVM[12-14] is used and the kernel matrix is called directly which is calculated in advance. Suppose $k(x_i, x_j) = exp(-\|x_i - x_j\|^2/(2\sigma^2))$. The comparative results of numerical testing of v-MC-SVM algorithm with QP-MC-SVM algorithm are shown in Table 1.

Table 1. Results of numerical testing

Glass	v-MC-SVM	QP-MC-SVM
accuracy rate	61.3%	61.2%
training time(s)	3.87	4.06
number of SVs	166	167

Thyorid	v-MC-SVM	QP-MC-SVM
accuracy rate	97.2%	97.3%
training time(s)	0.11	0.13
number of SVs	44	43

Wine	v-MC-SVM	QP-MC-SVM
accuracy rate	97.7%	97.7%
training time(s)	0.58	0.67
number of SVs	37	37

6. Conclusion

Because the selection of value of C of C-SVM is difficult, Scholkopf proposed v-SVM. In this paper, the basic theory and specific meaning of v of v-SVM is introduced fistly. Then, we propose the model of v-MC-SVM which is based on v-SVM. Several theorems of v-MC-SVM are also given. The constraints of v-MC-SVM are too complicated. One-class v-SVM problem of v-MC-SVM is given by adding b_m to the objective function of v-MC-SVM and employing the Kesler's construction for simplifying the original problem. The optimal solutions of one-class SVM problem of v-MC-SVM are unchanged when its constraint $e^T\alpha \ge$v is replaced with $e^T\alpha = v$. Numerical testing results show that the speed of v-MC-SVM algorithm is faster than that of QP-MC-SVM algorithm under the same accuracy rate.

Acknowledgment

This work is supported by National Natural Science Foundation of China No.10571109 and Program of Shandong Tai'an Science and Technology No.20082025.

References

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

[2] Burges C J C. A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery*, 1998, 2(2): 121-167.

[3] Duda R O, Hart P E, Stork,D,G. *Pattern Recognition*.John Wiley & Sons, 2000.

[4] Weston J, Watkins C. Support vector machines for multi-class pattern recognition.*In: Proceedings of European Symposium on Artificial Neural Networks*, Brussels ,1999: 219-224.

[5] Weston J, Watkins C. *Multi-class Support vector machines*. Technical Report CSD-TR-98-04, Department of Computer Science, Royal Holoway, University of London, Egham, TW200EX, Uk, 1998.

[6] Hsu C, Lin C J. A comparison of methods for multi-class support vector machines. *IEEE Transactions on Neural Networks*, 2002, 13(2): 415-425.

[7] DENG Nai-Yang, TIAN Ying-jie. *A New Method of Data Mining-Support Vector Machine*. Beijing: Science Press, 2004.

[8] Scholkopf B, Platt J, Shawe-Taylor J, etal. Estimating the Supper of a High-Dimensional Distribution. *Neural Computition*, 2001, 13(7): 1443-1471.

[9] Yuan Yaxiang, Sun Wenyu. *Optimization theory and method*. Beijing :Science Press, 1997.

[10] Franc V, Hlavc V. *Multi-class Support Vector Machine*. In R Kasturi, D Laurendeau and Suen C, editors, ICPR 02: Proceedings 16th International Conference on Pattern Recognition, volume2, p236-239, CA 90720-1314, Los Alamitos, US, August 2002. IEEE Computer Society.

[11] Franc V, Hlavc V. *Kernel repression of the Kesler construction for Multi-class SVM classification*. In:H.Wildenauer and W. Kropatsch, editor, Proceedings of the CWWW'02, page 7, Wien, Austria, February 2002. PRIP. Available at ftp://cmp.felk.cvut.cz/pub/cmp/ articles/franc/ francmulitiKernel102.pdf.

[12] Platt J C. *Fast training of support vector machines using sequential minimal optimization*. In: Schokopf B, Burges C, Smola A, eds. Advances in Kernel Methods-Support Vector Learning. Cambridge, MA:MIT Press, 1999: 185-208.

[13] Keerthi S S, Shevade S K, Bhattacharyya C, et al. Improvement to Platt SMO algorithm for SVM classifier design. *Neural Computation*, 2001, 13: 637-649.

[14] Gunnar Ratsch, Member, IEEE, Sebastian Mika, Bernhard Scholkopf, and Klaus-Robert Muller. Constructing Boosting Algorithms form SVMs: An Application to One-Class Classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2002, 24(9): 1-16.