The Ninth International Symposium on Operations Research and Its Applications (ISORA'10)
Chengdu-Jiuzhaigou, China, August 19–23, 2010

Leave-One-Out Bound for Crammer-Singer Multiclass Support Vector Machine*

Yan-Mei Zhao[2] Zhi-Xia Yang[1,†]

[1]College of Mathematics and System Science, Xinjiang University, Urumuchi China,830046
[2]College of Science, China Agricultural University, Beijing, China,100083

Abstract The selection of parameters in the support vector machine (SVM) is an important step for constructing a high performance learning machine. Minimizing the bound of leave-one-out (LOO) error is an efficient and time-economized approach for the SVM to select parameters. In fact, some famous bounds have been proposed. These researches focus on their issues for binary classification but not multiclass problem. In this paper we derive the leave-one-out error bound for the Crammer-Singer multiclass SVM. The numerical experiments on some benchmarks show that our method is reasonable.

Keywords Support vector machine; Parameters selection; Leave-one-out error bound; Crammer-Singer multiclass SVM

1 Introduction

The support vector machine (SVM)[11]has become one of the most popular methods in machine learning for both regression and classification problems during the last years. It performs structural risk minimization introduced to machine learning by Vapnik [12], and has yielded excellent generalization performance. However, SVM formulations require the user to set parameters which govern the training process, and those parameter settings can have a significant affect on the performance of engine. Clearly, the best performance is realized with an optimum choice of parameters. The most common approach to parameter selection is to realize an exhaustive grid search over the parameter space to find the best settings. Unfortunately, it is of little use in practical application due to its result of unacceptably long run time. Some other method such as cross-validation gives good effectiveness. It needs the machine learning engine to be trained multiple times in order to obtain a single performance for a single parameter setting.

The leave-one-out (LOO) error provides an almost unbiased estimate for the generation error and can be considered as a reliable criteria for the selection of parameters . But the computation of the LOO error is extremely time consuming. Thus the methods are sought to speed calculation of LOO error , or bound it with an easily computed quantity. Currently, there are some successfully bounds proposed for support vector classification

*This work is supported by the National Science Foundation of China (No.10801112) and the Ph.D Graduate Start Research Foundation of Xinjiang University Funded Project (No.BS080101).
†Corresponding author. E-mail: xjyangzhx@sina.com

machine [8, 13, 7], support vector regression machine[10, 4] and support vector ordinal regression machine [14].

In this paper, we concentrate on the bound of LOO error for the Crammer-Singer multiclass SVM [5] which solves the multiclass classification. It is based on "all together" idea by solving one optimization problem and applies the maximum margin principle used in binary SVM. This approach has been applied in many fields such as text classification [6],bioinformatics [1]. As we discussed above, selecting the optimal parameters for the Crammer-Singer multiclass SVM is also pivotal. Our approach is similar to the one described by Joachims[9] for the SVM classifier and can obtain the LOO error bound by computing all samples only once.

This paper is structured as follows. The Crammer-Singer multiclass SVM is reviewed in Section 2. In Section 3, we describe our derivation method for the bound of LOO error. Section 4 is some numerical results on benchmarks which are used to confirm the efficiency of our algorithm. Some conclusions are discussed in the last section.

2 Crammer-Singer multiclass SVM

In this section, we introduce Crammer-Singer multiclass SVM [5]. Consider the training set

$$T = \{(x_1, y_1), \cdots, (x_l, y_l)\} \in (\mathscr{X} \times \mathscr{Y})^l, \tag{1}$$

where $x_i \in \mathscr{X} \subseteq R^n$ is the input, and $y_i \in \mathscr{Y} = \{1, \cdots, k\}$ is the output or the class label. The input x is mapped into a Hilbert space $\mathscr{H}$ by a function $\mathrm{x} = \Phi(x) : x \in R^n \to \mathrm{x} \in \mathscr{H}$. We need to find k hyperplanes to construct the decision function

$$f(x) = \arg \max_{r=1,\cdots,k} \sum_{i=1}^{l} \alpha_i^r K(x_i, x), \tag{2}$$

where $K(x,x') = (\Phi(x) \cdot \Phi(x'))$ is the kernel function.

To get $\alpha_i^r, r = 1, \cdots, k, i = 1, \cdots, l$ in the decision function (2), we need solve the dual problem of Crammer-Singer multiclass SVM

$$\max_{\alpha} W(\alpha) = \quad -\frac{1}{2} \sum_{i=1}^{l} \sum_{j=1}^{l} K(x_i \cdot x_j) \alpha_i^{\mathrm{T}} \alpha_j - \sum_{i=1}^{l} \alpha_i^{\mathrm{T}} e_i, \tag{3}$$

$$\text{s.t.} \quad \sum_{r=1}^{k} \alpha_i^r = 0, i = 1, \cdots, l, \tag{4}$$

$$\alpha_i^r \leqslant 0, \ \ if \ \ y_i \neq r, i = 1, \cdots, l. r = 1, \cdots, k, \tag{5}$$

$$\alpha_i^r \leqslant C, \ \ if \ \ y_i = r, i = 1, \cdots, l. r = 1, \cdots, k, \tag{6}$$

where $\alpha = (\alpha_1, \cdots, \alpha_l), \alpha_i = (\alpha_i^1, \cdots, \alpha_i^k)^{\mathrm{T}}, i = 1, \cdots, l, e_i = (e_i^1, \cdots, e_i^k)^{\mathrm{T}}, e_i^r = 1 - \delta_{y_i, r}, i = 1, \cdots, l, r = 1, \cdots, k$, and $K(x,x') = (\Phi(x) \cdot \Phi(x'))$ is the kernel function.

3 Derivation of the LOO error bound

In order to derive the LOO error bound we give the definition of the LOO error and the support vector in advance.

Definition 1. (**LOO error**) Given the training set $T^t = T \setminus \{(x_t, y_t)\}, t = 1, \cdots, l$, where T is the training set given by (1). Suppose that $f_{T^t}(x)$ is the decision function obtained on the training set T^t, then the LOO error is defined as

$$R_{\mathrm{LOO}}(T) = \frac{1}{l}\sum_{t=1}^{l} c(x_t, y_t, f_{T^t}(x_t)), \tag{7}$$

where $c(x, y, f(x))$ is the 0-1 loss function.

Definition 2. (**Support Vector**) Suppose that $\alpha = (\alpha_1, \cdots, \alpha_l), \alpha_i = (\alpha_i^1, , \cdots, \alpha_i^k)^{\mathrm{T}}, i = 1, \cdots, l$, is the optimal solution of the dual problem (3) $\sim$(6) for the training set (1). Then

(1). The input x_i is called a margin support vector about $\alpha = (\alpha_1, \cdots, \alpha_l)$, $\alpha_i = (\alpha_i^1, , \cdots, \alpha_i^k)^{\mathrm{T}}$, $i = 1, \cdots, l$, if the corresponding multiplier vector $\alpha_i = (\alpha_i^1, \cdots, \alpha_i^k)^{\mathrm{T}} \neq \bar{0}$, where $\bar{0}$ represents the vector with 0 value for every element, and there exists at least a component $\alpha_i^r \in (0, C)$ and the others components are nor equal to C. Define the margin support vector set is the following index set

$$SV = \{i | \alpha_i \neq \bar{0}, \text{and } \exists r, \alpha_i^r \in (0, C), y_i = r; \alpha_i^r \neq C, y_i \neq r, i = 1, \cdots, l\}; \tag{8}$$

(2). The input x_i is called a non-margin support vector about $\alpha = (\alpha_1, \cdots, \alpha_l)$, $\alpha_i = (\alpha_i^1, , \cdots, \alpha_i^k)^{\mathrm{T}}$, $i = 1, \cdots, l$, if the corresponding multiplier vector $\alpha_i = (\alpha_i^1, \cdots, \alpha_i^k)^{\mathrm{T}} \neq \bar{0}$, where $\bar{0}$ represents the vector with 0 value for every element, and there exists at least a component $\alpha_i^r = C$. Define the non-margin support vector set is the following index set

$$SV^C = \{i | \alpha_i \neq \bar{0}, \text{and } \exists r, \alpha_i^r = C, i = 1, \cdots, l\}; \tag{9}$$

(3).The input x_i is called a non-support vector, if the corresponding multiplier vector $\alpha_i = (\alpha_i^1, \cdots, \alpha_i^k)^{\mathrm{T}} = \bar{0}$, where $\bar{0}$ represent the vector with 0 value for each element,

$$\alpha_i^r = 0, r = 1, \cdots, k. \tag{10}$$

Now, we address the derivation of LOO error bound for the Crammer-Singer multi-class SVM. After removing point $\{(x_t, y_t)\}$, the training set becomes $T^t = T \setminus \{(x_t, y_t)\}$, the corresponding dual problem is

$$\max_{\tilde{\alpha}} W_t(\tilde{\alpha}) = \quad -\frac{1}{2}\sum_{i \in I \setminus t}\sum_{j \in I \setminus t} K(x_i, x_j)\tilde{\alpha}_i^{\mathrm{T}}\tilde{\alpha}_j - \sum_{i \in I \setminus t}\tilde{\alpha}_i^{\mathrm{T}} e_i, \tag{11}$$

$$\text{s.t.} \quad \sum_{r=1}^{k}\tilde{\alpha}_i^r = 0, i \in I \setminus t, \tag{12}$$

$$\tilde{\alpha}_i^r \leq 0 \ \ if \ \ y_i \neq r, i \in I \setminus t, r = 1, \cdots, k, \tag{13}$$

$$\tilde{\alpha}_i^r \leq C \ y_i = r, i \in I \setminus t, r = 1, \cdots, k, \tag{14}$$

where $\tilde{\alpha} = (\tilde{\alpha}_1, \cdots, \tilde{\alpha}_l), \tilde{\alpha}_i = (\tilde{\alpha}_i^1, \cdots, \tilde{\alpha}_i^k)^{\mathrm{T}}, i \in I \setminus t$, $e_i = (e_i^1, \cdots, e_i^k)^{\mathrm{T}}, e_i^r = 1 - \delta_{y_i, r}, i \in I \setminus t, r = 1, \cdots, k$, and $I = \{1, 2, \cdots, l\}$.

In order to getting the LOO error bound, we give the following lemma firstly.

Lemma 1. *Given the training set T (1), suppose $\alpha = (\alpha_1, \cdots, \alpha_l), \alpha_i = (\alpha_i^1, \cdots, \alpha_i^k)^{\mathrm{T}}, i = 1, \cdots, l$ is the optimal solution of dual problem (3)∼(6). If the decision functions $f_T(x)$ and $f_{T^t}(x)$ are obtained by Crammer-Singer multiclass SVM for the training set T and $T^t = T \setminus \{(x_t, y_t)\}$ respectively, then we have*

(1) If $\alpha_t \neq \bar{0}$, $t \in SV$, and the removing point (x_t, y_t) belongs to the r-th class, when an error occurs on this point by the decision function $f_{T^t}(x)$, the following inequality holds

$$\frac{1}{\alpha_t^r}(\frac{1}{2}\alpha_t^{\mathrm{T}}\alpha_t K(x_t, x_t) + \sum_{i \in SV|t} \alpha_t^{\mathrm{T}}\alpha_i K(x_t, x_i)) - \frac{1}{2}\alpha_t^r s_m^{*\mathrm{T}} s_m^* < 0, \tag{15}$$

where

$$s_m^* = (0, \cdots, 0, 1, 0, \cdots, 0, -1, 0, \cdots, 0)^T, \tag{16}$$

and $s_m^* \in R^k$, the r-th element of vector s_m^* is 1, the m-th element of vector s_m^* is -1, the rest elements are 0, $m \in \{1, 2, \cdots, r-1, r+1, \cdots, k\}$ and SV is the index set of margin support vector given by (8);

(2) If $\alpha_t = \bar{0}$, where $\bar{0}$ represent the vector with 0 value for each element, it follows that $f_T(x) = f_{T^t}(x)$.

Proof: (1) Assume that $\alpha_t \neq \bar{0}$, $t \in SV = \{i | \alpha_i \neq \bar{0}, \text{and } \exists r, \alpha_i^r \in (0, C), y_i = r; \alpha_i^r \neq C, y_i \neq r, i = 1, \cdots, l\}$ and the removing point (x_t, y_t) belongs to the r-th class. We have $0 < \alpha_t^r < C$ according to the equations (4)∼(6), namely, the t-th points is a margin support vector. From the solution $\tilde{\alpha}$ of problem (11)∼(14), produce a feasible solution of problem (3)∼(6) by

$$\beta_i = \begin{cases} \tilde{\alpha}_i, & if \quad \tilde{\alpha}_i = \bar{0} \text{ and } \tilde{\alpha}_i \in SV^{Ct}; \\ \tilde{\alpha}_i, & if \quad i \in SV^t; \\ \alpha_t^r s_m^*, & if \quad i = t, \end{cases} \tag{17}$$

where $s_m^* = (0, \cdots, 0, 1, 0, \cdots, 0, -1, 0, \cdots, 0)^T$, $s_m^* \in R^k$, the r-th element of the vector s_m^* is 1, the m-th element of the vector s_m^* is -1, the rest elements are 0 and $m \in \{1, 2, \cdots, r-1, r+1, \cdots, k\}$, and SV^t and SV^{Ct} are the sets of index corresponding to the margin and non-margin support vectors in the solutions of problem (11)∼(14) respectively. It is easy to see that

$$\beta_i^r \leq 0, y_i \neq r, \tag{18}$$

$$\beta_i^r \leq C, y_i = r, \tag{19}$$

$$\sum_{i=1}^{k} \beta_i^r = 0, i = 1, \cdots, l. \tag{20}$$

So β is a feasible solution of the problem (3)∼(6). After the transformation, $W(\beta)$ can be written as

$$W(\beta) = W_t(\tilde{\alpha}) - \frac{1}{2}(\alpha_t^r)^2 s_m^{*\mathrm{T}} s_m^* K(x_t, x_t) + \alpha_t^r - \sum_{i \in SV^t} \alpha_t^r s_m^{*\mathrm{T}} \tilde{\alpha}_i K(x_t, x_i). \tag{21}$$

Similarly let us construct a feasible solution $\tilde{\gamma}$ of (11)∼(14) based on the solution α of (3)∼(6).

$$\tilde{\gamma}_i = \begin{cases} \alpha_i, & \text{if} \quad \alpha_i = \bar{0} \text{ and } \alpha_i \in SV^C; \\ \alpha_i, & \text{if} \quad i \in SV \setminus t, \end{cases} \tag{22}$$

where SV and SV^C are the index sets given by (8) and (9) respectively. $SV \setminus t$ excludes the index t corresponding the removing point. It is easy to see that

$$\tilde{\gamma}_i^r \leq 0, y_i \neq r, \tag{23}$$

$$\tilde{\gamma}_i^r \leq C, y_i = r, \tag{24}$$

$$\sum_{i=1}^{K} \tilde{\gamma}_i^r = 0, i \in I \setminus t, \tag{25}$$

where $I = \{1,2,\cdots,l\}$. So $\tilde{\gamma}$ is a feasible solution of the problem (11)$\sim$(14). After the transformation, $W_t(\tilde{\gamma})$ can be written as

$$W_t(\tilde{\gamma}) = W(\alpha) + \frac{1}{2}\alpha_t^{\mathrm{T}}\alpha_t K(x_t,x_t) - \alpha_t^r + \sum_{i \in SV \setminus t} \alpha_t^{\mathrm{T}}\alpha_i K(x_t,x_i). \tag{26}$$

Due to $W(\alpha) \geq W(\beta), W_t(\tilde{\alpha}) \geq W_t(\tilde{\gamma})$ and the equation (21) and (26), it results in the following inequality

$$\begin{aligned} &\sum_{i \in SV^t} \alpha_t^r s_m^{*\mathrm{T}} \tilde{\alpha}_i K(x_t,x_i) \geq \\ &\frac{1}{2}\alpha_t^{\mathrm{T}}\alpha_t K(x_t,x_t) + \sum_{i \in SV \setminus t} \alpha_t^{\mathrm{T}}\alpha_i K(x_t,x_i) - \frac{1}{2}(\alpha_t^r)^2 s_m^{*\mathrm{T}} s_m^* K(x_t,x_t). \end{aligned} \tag{27}$$

As the result $0 < \alpha_t^r < C$, it can be written as

$$\begin{aligned} &\sum_{i \in SV^t} s_m^{*\mathrm{T}} \tilde{\alpha}_i K(x_t,x_i) \geq \\ &\frac{1}{\alpha_t^r}(\frac{1}{2}\alpha_t^{\mathrm{T}}\alpha_t K(x_t,x_t) + \sum_{i \in SV \setminus t} \alpha_t^{\mathrm{T}}\alpha_i K(x_t,x_i)) - \frac{1}{2}\alpha_t^r s_m^{*\mathrm{T}} s_m^* K(x_t,x_t). \end{aligned} \tag{28}$$

According to the formulation of decision function (2), we know that an error occurs on the removing point (x_t,y_t) by $f_{T^t}(x)$ in the condition that, there exists s_m^* like the equation (16), $m \in \{1,\cdots,r-1,r+1,\cdots,k\}$, to make $\sum_{i \in SV^t} s_m^{*\mathrm{T}} \tilde{\alpha}_i K(x_t,x_i) < 0$. This means also that

$$\frac{1}{\alpha_t^r}(\frac{1}{2}\alpha_t^{\mathrm{T}}\alpha_t K(x_t,x_t) + \sum_{i \in SV \setminus t} \alpha_t^{\mathrm{T}}\alpha_i K(x_t,x_i)) - \frac{1}{2}\alpha_t^r s_m^{*\mathrm{T}} s_m^* K(x_t,x_t) < 0. \tag{29}$$

(2) Assume that $\alpha_t = \bar{0}$: We can obtain $\tilde{\mathrm{w}}_r = \mathrm{w}_r, r = 1,\cdots,k$ from the KKT condition. It follows that $f_{T^t}(x) = f_T(x)$. □

From Lemma 1 we can get the following theorem

Theorem 2. *The LOO error bound for Crammer-Singer multiclass SVM is given by*

$$\begin{aligned} R_{LOO}(T) \quad \leq \quad & \frac{1}{l}(|\{t|t \in SV, \exists s_m^*, s.t. \frac{1}{\alpha_t^r}(\frac{1}{2}\alpha_t^{\mathrm{T}}\alpha_t K(x_t,x_t) + \sum_{i \in SV|t} \alpha_t^{\mathrm{T}}\alpha_i K(x_t,x_i)) \\ & -\frac{1}{2}\alpha_t^r s_m^{*\mathrm{T}} s_m^* K(x_t,x_t) < 0\}| + |SV^C|, \end{aligned} \tag{30}$$

where $|\cdot|$ *represents the number of the elements in the set,* s_m^* *is given by* (16)*, and SV and* SV^C *are the index sets of margin support vector* (8) *and non-margin support vector* (9) *respectively.*

Proof: Assuming the removing point (x_t, y_t) belongs to the r-th class and $t \in SV$. According to Lemma 1, when the LOO error procedure commits an error at the removing point, there exists s_m^* by the equation (16), $m \in \{1, \cdots, r-1, r+1, \cdots, k\}$, to hold the following inequalities

$$\frac{1}{\alpha_t^r}\left(\frac{1}{2}\alpha_t^{\mathrm{T}}\alpha_t K(x_t, x_t) + \sum_{i \in SV|t} \alpha_t^{\mathrm{T}}\alpha_i K(x_t, x_i)\right) - \frac{1}{2}\alpha_t^r s_m^{*\mathrm{T}} s_m^* < 0. \tag{31}$$

If $t \in SV^C$, the removing point (x_t, y_t) is a margin support vector, then the number of error made by the LOO error procedure is $|SV^C|$, where SV^C is defined by (9) and $|\cdot|$ is the number of elements in the set.

So we get the LOO error bound (30) for the Crammer-Singer multiclass SVM. □

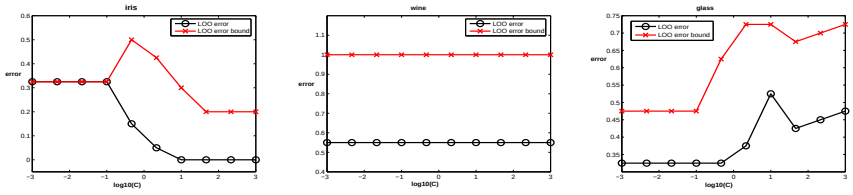
It can be seen easily that we have got the LOO error bound by training the whole training set only once which is valid for all kernel functions.

4 Numerical Experiments

In order to test the performance of the proposed LOO error bound for the Crammer-Singer multiclass SVM, we use three benchmark datasets from the UCI repository[2]: *iris, wine* and *glass*.

As a result of large scale of Crammer-Singer multiclass SVM, the computation of the LOO error is extremely time consumed. So we pick out 40 and 80 points using the stratified random method [3] to construct the new datasets in the first and the second group of experiments respectively. Each new dataset is normalized. And RBF kernel is selected $K(x, x') = \exp(-\gamma||x - x'||^2)$. So the parameters to be selected are C, γ, where they are selected respectively from the following two candidate sets $C = \text{logspace}(-3, 3, 10)$, $\gamma = \text{logspace}(-3, 3, 10)$, where logspace is a logarithmically spaced vector in MATLAB.

Each time we fix one parameter C or γ, and let the another parameter change according to one of the above two candidate sets. We compare the values of LOO error bounds and LOO errors in this manner to observe whether the trend of the LOO error bound consists with the trend of LOO error itself. Figure 1 and Figure 2 show the performance of the LOO error bound respectively for two groups of experiments. "LOO error" shows the actual LOO error and "LOO error bound" shows the bound obtained from the Theorem 2.



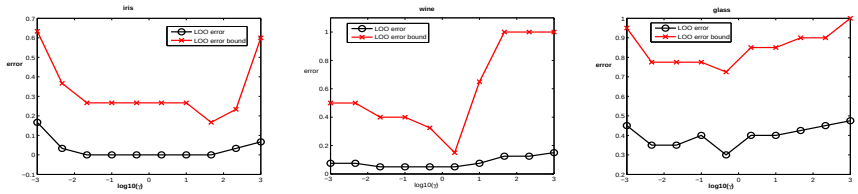


Figure 1. the comparison between LOO error and LOO error bound on the dataset consists of 40 points

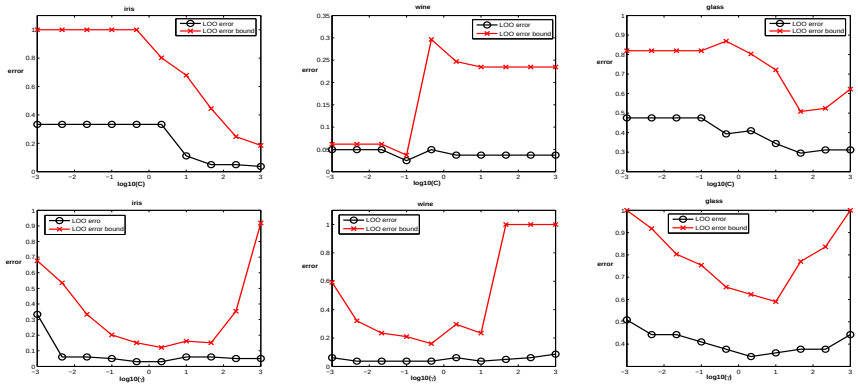


Figure 2. the comparison between LOO error and LOO error bound on the dataset consists of 80 points

As the above figures showed, the changing trend of the LOO error bounds is almost coincident with that of the LOO errors. As a whole, the point corresponding to the minimum of the LOO error bound is close to those of LOO error. So minimizing the LOO error bound to select parameters can obtain almost the same performance compared to minimizing the LOO error. However the time consumed by them has a great difference. Assume we have a k class sample set of l training points. It needs to solve a QP optimization problem with $(l-1)k$ variables l times for LOO error whereas a similar QP optimization problem with lk variables only once for LOO error bound. So the time computation cost of LOO error bound is so little relative to LOO error, approximatively its $\frac{1}{l}$. When the number of sample points increase, the execution time of LOO error increases fast while the execution time of LOO error bound increases slowly. So for the large sample data, the superior of our method is extreme prominent.

5 Conclusions

In this paper, we have proposed the LOO error bound for the Crammer-Singer multiclass SVM. The LOO error bound and the LOO error have almost the same changing trends when the parameters change, so minimizing the LOO error bound to select parameters can achieve almost the same performance compared to minimizing the LOO error, whereas the computation time required during the selection of parameters process

is drastically reduced. We have tested our bound on three benchmark datasets, which have produced promising results confirming our approach. Note that the computation time of LOO error grows fast along with the increase of the number of the sample set. Thus, for problems with large dataset, our strategy is very effective for performing good parameters selection for the Crammer-Singer multiclass SVM with limited time.

References

[1] M.T. Anwar Shamim, M. Anwaruddin, H.A. Nagarajaram, Support Vector Machine-based classification of protein folds using the structural properties of amino acid residues and amino acid residue pairs, Bioinformatics 23 (2007) 3320-3327.

[2] C.L. Blake, C.J. Merz, UCI repository of machine learning databases, University of California. [www.http://www.ics.uci.edu/ mlearn/MLRepository.html]

[3] L. Breiman, J. Friedman, R. Olshen, C. Stone, Classification and Regression Trees, Wadsworth and Brooks, Monterey, CA, 1984.

[4] M.W. Chang, C.J. Lin, Leave-one-out bounds for support vector regression model selection, Neural Computation, 17 (5) (2005) 1188-1222.

[5] K. Crammer, Y. Singer, On the learnability and design of output codes for multiclass problems, Machine Learning, 47 (2002) 201-233.

[6] D. Giorgetti, F.Sebastiani, Multiclass Text Categorization for Automated Survey Coding, Proceedings of SAC-03, 18th ACM Symposium on Applied Computing, Melbourne, US, 2003, pp. 798-802.

[7] A. Gretton ,R. Herbrich ,O. Chapelle, Estimating the leave-one-out error for classification learning with SVMs, http://www.kyb.tuebingen.mpg.de/publications/pss/ps1854.ps, May 15, 2003.

[8] T.S. Jaakkola, D. Haussler, Exploiting generative models in discriminative classifiers, In Advances in Neural Information Processing Systems 11,MIT Press, 1998, pp. 487-493.

[9] T. Joachims, Estimating the generalization performance of an SVM efficientily, In: Proceedings of the 17th International Conference on Machine Learning, San Franscisco, California: Morgan Kaufmann, 2000, pp. 431-438.

[10] Y.J. Tian, Support vector regession machine and its application, PH.D.Thesis, China Algricultural University, 2005.

[11] V. Vapnik, Statistical Learning Theory, Wiley, New York, 1998.

[12] V. Vapnik, Estimation of Dependences based on Empirical Data, Springer, Berlin, 1982.

[13] V. Vapnik, O. Chapelle, Bounds on error expectation for support vector machines, Neural Computation, 12 (9) (2000) 2013-2036.

[14] Z.X. Yang, Y.J. Tian, N.Y. Deng, Leave-one-out Bounds for Support Vector Ordinal Regression Machine, Neural Computing and Application,Vol. 18 Issue (2009), page 731–749.