

## Structural Regular Multiple Criteria Linear Programming for Classification Problem

Z. Qi, Y. Shi

### Zhiquan Qi

Research Center on Fictitious Economy & Data Science,  
Chinese Academy of Sciences, Beijing 100190, China  
E-mail: qizhiquan@gucas.ac.cn

### Yong Shi

1. Research Center on Fictitious Economy & Data Science,  
Chinese Academy of Sciences, Beijing 100190, China and  
2. College of Information Science & Technology,  
University of Nebraska at Omaha Omaha, NE 68182, USA  
E-mail: yshi@gucas.ac.cn

#### Abstract:

Classification problem has attracted an increasing amount of interest. Various classifiers have been proposed in the last decade, such as ANNs, LDA, and SVM. Regular Multiple Criteria Linear Programming (RMCLP) is an effective classification method, which was proposed by Shi and his colleagues and have been applied to handle different real-life data mining problems. In this paper, inspired by the application potential of RMCLP, we propose a novel Structural RMCLP (called SRMCLP) method for classification problem. Unlike RMCLP, SRMCLP is sensitive to the structure of the data distribution and can construct more reasonable classifiers by exploiting these prior data distribution information within classes. The corresponding optimization problem of SRMCLP can be solved by a standard quadratic programming. The effectiveness of the proposed method is demonstrated via experiments on synthetic and available benchmark datasets.

**Keywords:** classification, RMCLP, structural information of data, SVM

## 1 Introduction

For the last decade, the researchers have extensively developed various optimization techniques to deal with the classification problem in data mining or machine learning. Support Vector Machine (SVM) ([1,2]) is one of the most popular methods. However, Applying optimization techniques to solve classification has seventy years history. Linear Discriminant Analysis(LDA) ([3]) was first proposed in 1936. Mangasarian ([4]) has proposed a large margin classifier based on linear programming in 1960's. From 1980's to 1990's, Glover proposed a number of linear programming models to solve discriminant problems with a small sample size of data ([5,6]). Recently, Shi and his colleagues([7]) extend Glover's method into classification via multiple criteria linear programming (MCLP), and then various improved algorithms were proposed one after the other ([8,9]). These mathematical programming approaches to classification have been applied to handle many real world data mining problems, such as credit card portfolio management ([11,12]), bioinformatics ([13]), firm bankruptcy ([14]), and etc.

Recently, how to apply the structural information of data to build a good classifier is a new research focus. Many new large margin classifiers based on structural information have been proposed. Exploiting clustering algorithms to extract the structural information embedded with classes is one popular strategy [15–17]. The structured large margin machine (SLMM) [15] is a representative work based on the strategy. Firstly, SLMM explores the structural information within classes by Ward's agglomerative hierarchical clustering method on input

data [18], and then introduces the related structure information into the constraints. Finally, SLMM can be solved by a sequential second order cone programming (SOCP). Experimentally, SLMM is superior to support vector machine minimax probability machine (MPM) [19] and maximum margin machine (M4) [20]. However, as we all know, solving the involved SOCP problem is more difficult than the Quadratic Programming Problem (QPP) as in SVM, so SLMM has more higher computational complexity than traditional SVM. Consequently, a novel structural support vector machine (SRSVM) was proposed by Xue et. al [17]. Unlike SLMM, SRSVM exploits the classical framework of SVM rather than as constraints in SLMM and the corresponding optimization problem can still be solved by the QPP. SRSVM has been shown to be theoretically and empirically better in generalization than SVM and SLMM.

In this paper, inspired by the success of SRSVM and the application potential of RMCLP, we propose a novel Structural RMCLP (called SRMCLP) method for classification problem. Unlike RMCLP, SRMCLP is sensitive to the structure of the data distribution and can construct more reasonable classifiers by exploiting these prior data distribution information within classes.

The remaining parts of the paper are organized as follows. Section 2 introduces the basic notions and formulation of MCLP; Section 3 describes in detail our proposed Algorithms; All experimental results are shown in section 4; Conclusions are given in the last section.

## 2 Background

We give a brief introduction of MCLP in the following. For classification about the training data

$$T = \{(x_1, y_1), \dots, (x_l, y_l)\} \in (\mathfrak{R}^n \times \mathcal{Y})^l, \quad (1)$$

where  $x_i \in \mathfrak{R}^n, y_i \in \mathcal{Y} = \{1, -1\}, i = 1, \dots, l$ , data separation can be achieved by two opposite objectives. The first objective separates the observations by minimizing the sum of the deviations (MSD) among the observations. The second maximizes the minimum distances (MMD) of observations from the critical value [6]. The overlapping of data  $u$  should be minimized while the distance  $v$  has to be maximized. However, it is difficult for traditional linear programming to optimize MMD and MSD simultaneously. According to the concept of Pareto optimality, we can seek the best trade-off of the two measurements [11, 12]. So MCLP model can be described as follows:

$$\min_u \quad e^T u \quad \& \quad \max_v \quad e^T v, \quad (2)$$

$$\text{s.t.} \quad (w \cdot x_i) + (u_i - v_i) = b, \quad \text{for } \{i|y_i = 1\}, \quad (3)$$

$$(w \cdot x_i) - (u_i - v_i) = b, \quad \text{for } \{i|y_i = -1\}, \quad (4)$$

$$u, v \geq 0, \quad (5)$$

where  $e \in \mathfrak{R}^l$  is a vector whose all elements are 1,  $w$  and  $b$  are unrestricted,  $u_i$  is the overlapping and  $v_i$  the distance from the training sample  $x_i$  to the discriminator  $(w \cdot x_i) = b$  (classification separating hyperplane). By introducing penalty parameter  $c, d > 0$ , MCLP has the following version

$$\min_{u,v} \quad ce^T u - de^T v, \quad (6)$$

$$\text{s.t.} \quad (w \cdot x_i) + (u_i - v_i) = b, \quad \text{for } \{i|y_i = 1\}, \quad (7)$$

$$(w \cdot x_i) - (u_i - v_i) = b, \quad \text{for } \{i|y_i = -1\}, \quad (8)$$

$$u, v \geq 0, \quad (9)$$

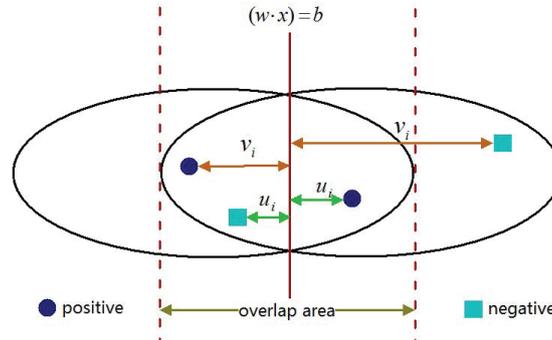


Figure 1: Geometric meaning of MCLP.

The geometric meaning of the model is shown in Figure 1.

A lot of empirical studies have shown that MCLP is a powerful tool for classification. However, we cannot ensure this model always has a solution under different kinds of training samples. To ensure the existence of solution, recently, Shi et al proposed a RMCLP model by adding two regularized items  $\frac{1}{2}w^T H w$  and  $\frac{1}{2}u^T Q u$  on MCLP as follows (more theoretical explanation of this model can be found in [8]):

$$\min_z \quad \frac{1}{2}w^T H w + \frac{1}{2}u^T Q u + de^T u - ce^T v, \quad (10)$$

$$\text{s.t.} \quad (w \cdot x_i) + (u_i - v_i) = b, \quad \text{for } \{i|y_i = 1\}, \quad (11)$$

$$(w \cdot x_i) - (u_i - v_i) = b, \quad \text{for } \{i|y_i = -1\}, \quad (12)$$

$$u, v \geq 0, \quad (13)$$

where  $z = (w^T, u^T, v^T, b)^T \in R^{n+l+l+1}$ ,  $H \in R^{n \times n}$ ,  $Q \in R^{l \times l}$  are symmetric positive definite matrices. Obviously, the regularized MCLP is a convex quadratic programming.

Compared with traditional SVM, we can find that the RMCLP model is similar to the Support Vector Machine model in terms of the formation by considering the minimization of overlapping of the data. However, RMCLP tries to measure all possible distances  $v$  from the training samples  $x_i$  to separating hyperplane, while SVM fixes the distance as 1 (through bounding planes  $(w \cdot x) = b \pm 1$ ) from the support vectors. Although the interpretation can vary, RMCLP addresses more control parameters than the SVM, which may provide more flexibility for better separation of data under the framework of the mathematical programming. In addition, different with SVM, RMCLP considers all the samples to solve classification problem. These make RMCLP have stronger insensitivity to outliers.

### 3 Structural Regular Multiple Criteria Linear Programming for Classification Problem

#### 3.1 Extracting Structural Information within Classes

Following the strategy of the SLMM and SRSVM, SRMCLP also has two steps. The first step is to extract structural information within classes by some clustering method; the second step is the model learning. In order to compare the main different of the second step between SRMCLP and the other two methods, here we adopt the same clustering method: Ward's linkage clustering(WIL) [15–18], which is one of the hierarchical clustering analysis. A main advantage of WIL is that clusters derived from this method are compact and spherical, which provides a

meaningful basis for the computation of covariance matrices [15]. Concretely, if  $S$  and  $T$  are two clusters with means  $\mu_S$  and  $\mu_T$ , the Ward's linkage  $W(S, T)$  between clusters  $S$  and  $T$  can be computed as [15]

$$W(S, T) = \frac{|S| \cdot |T| \cdot \|\mu_S - \mu_T\|}{|S| + |T|}. \quad (14)$$

Initially, each sample is considered as a cluster. The Ward's linkage of two samples  $x_i$  and  $x_j$  is  $W(x_i, x_j) = \|x_i - x_j\|^2/2$ . When two clusters are being merged to a new cluster  $A'$ , the linkage  $W(A', C)$  can be conveniently derived from  $W(A, C)$ ,  $W(B, C)$  and  $W(A, B)$  by [15]

$$W(A', C) = \frac{(|A| + |C|)W(A, C) + (|B| + |C|)W(B, C) - |C|W(A, B)}{|A| + |B| + |C|}. \quad (15)$$

During the hierarchical clustering, the Ward's linkage between clusters to be merged increases as the number of clusters decreases [15]. A relation curve between the merge distance and the number of clusters can be drawn to represent this process. The optimal number of clusters can be determined by finding the knee point. Furthermore, the WIL can also be extended to the kernel space. More details of WIL can be found in [15].

### 3.2 Model Learning

We obtained two groups of  $P$  and  $N$  clusters in class  $C_P$  and  $C_N$  by the first step, i.e.,  $P = P_1 \cup \dots \cup P_i \cup \dots \cup P_{C_P}$ ,  $N = N_1 \cup \dots \cup N_j \cup \dots \cup N_{C_N}$ . Consider the optimization(10), choosing  $H, Q$  to be identity matrix and introducing  $\frac{1}{2}b^2$  and  $\frac{1}{2}w^\top \Sigma w$  into the object function, SRMCLP can be formulated as:

$$\begin{aligned} \min_z \quad & \frac{1}{2}\|w\|^2 + \frac{1}{2}c_1 w^\top \Sigma w + \frac{1}{2}\|u\|^2 + \frac{1}{2}b^2 + c_2 e^T u - c_3 e^T v, \\ \text{s.t.} \quad & (w \cdot x_i) + (u_i - v_i) = b, \quad \text{for } \{i|y_i = 1\}, \\ & (w \cdot x_i) - (u_i - v_i) = b, \quad \text{for } \{i|y_i = -1\}, \\ & u, v \geq 0, \end{aligned} \quad (16)$$

where  $z = (w^T, u^T, v^T, b)^T \in R^{n+l+l+1}$ ,  $c_1, c_2, c_3 \geq 0$  are the pre-specified penalty factors,  $\Sigma = \Sigma_+ + \Sigma_-$ , where  $\Sigma_+ = \Sigma_{P_1} + \dots + \Sigma_{P_{C_P}}$ ,  $\Sigma_- = \Sigma_{N_1} \dots + \Sigma_{N_{C_N}}$ ,  $\Sigma_{P_i}$  and  $\Sigma_{N_j}$  are respectively the covariance matrices corresponding to the  $i$  th and  $j$  th clusters in the two classes,  $i = 1, \dots, C_P$ ,  $j = 1, \dots, C_N$ . Obviously, the regularized SRMCLP is a convex quadratic programming. By introducing its Lagrange function

$$L(w, u, v, b, \alpha, \beta, \eta) = \frac{1}{2}\|w\|^2 + \frac{1}{2}c_1 w^\top \Sigma w + \frac{1}{2}\|u\|^2 + \frac{1}{2}b^2 + c_2 e^T u - c_3 e^T v + \quad (17)$$

$$\sum_{i=1}^l \alpha_i (y_i ((w \cdot x_i) - b) + u_i - v_i) - \sum_{i=1}^l \beta_i u_i - \sum_{i=1}^l \eta_i v_i, \quad (18)$$

where  $\alpha_i, \beta_i, \eta_i \in R$  are the Lagrange multipliers, Therefore the dual problem of (39) can be formulated as

$$\begin{aligned} \max_{w, u, v, b, \alpha, \beta, \eta} \quad & L(u, v, w, b, \alpha, \beta, \eta), \\ \text{s.t.} \quad & \nabla_{w, u, v, b} L(u, v, w, b, \alpha, \beta, \eta) = 0, \\ & \beta_i, \eta_i \geq 0, i = 1, \dots, l. \end{aligned} \quad (19)$$

From (19) we get

$$\nabla_w L = (I + c_1 \Sigma)w + \sum_{i=1}^l y_i \alpha_i x_i = 0, \quad (20)$$

$$\nabla_{u_i} L = u_i + c_2 + \alpha_i - \beta_i = 0, \quad i = 1, \dots, l, \quad (21)$$

$$\nabla_{v_i} L = -c_3 - \alpha_i - \eta_i = 0, \quad i = 1, \dots, l, \quad (22)$$

$$\nabla_b L = b - \sum_{i=1}^l y_i \alpha_i = 0, \quad (23)$$

where  $I$  is an identity matrix. Substituting the above equations into problem (19), the dual problem can be expressed as

$$\max_{\alpha, u, b} \quad -\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j \alpha_i \alpha_j [x_i^\top (2I + c_1 \Sigma) x_j] - \frac{1}{2} \sum_{i=1}^l u_i^2, \quad (24)$$

$$\text{s.t.} \quad \sum_{i=1}^l y_i \alpha_i = 0, \quad i = 1, \dots, l, \quad (25)$$

$$-c_2 - u_i \leq \alpha_i \leq -c_3, \quad i = 1, \dots, l. \quad (26)$$

Solving the convex quadratic programming problem, we can obtain its solution.

$$w = -(I + c_1 \Sigma)^{-1} \sum_{i=1}^l y_i \alpha_i x_i, \quad (27)$$

$$b = \sum_{i=1}^l y_i \alpha_i. \quad (28)$$

$$(29)$$

So the decision function can be formulated as follows

$$f(x) = - \sum_{i=1}^l y_i \alpha_i x_i^\top (I + c_1 \Sigma)^{-1} x + b. \quad (30)$$

Applying the kernel trick, we also extend the linear SRMCLP to the nonlinear case. Introduce the kernel function  $K(x, x') = (\Phi(x) \cdot \Phi(x'))$ , where  $\Phi(\cdot)$  is a mapping from the input space  $R^n$  to Hilbert space  $\mathcal{H}$

$$\Phi : \begin{array}{l} R^n \rightarrow \mathcal{H}, \\ x \rightarrow \Phi(x). \end{array} \quad (31)$$

Then the optimization problem of SRMCLP in the kernel space can be described as:

$$\max_{\alpha, u, b} \quad -\frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l y_i y_j \alpha_i \alpha_j [\Phi(x_i)^\top (2I + c_1 \Sigma^\Phi) \Phi(x_j)] - \frac{1}{2} \sum_{i=1}^l u_i^2, \quad (32)$$

$$\text{s.t.} \quad \sum_{i=1}^l y_i \alpha_i = 0, \quad i = 1, \dots, l, \quad (33)$$

$$-c_2 - u_i \leq \alpha_i \leq -c_3, \quad i = 1, \dots, l. \quad (34)$$

$2\Phi(x_i)^\top I\Phi(x_j) = K(x_i, x_j)$ , so we only need to consider how to compute the kernel matrix  $c_1\Phi(x_i)^\top \Sigma^\Phi \Phi(x_j)$ . Suppose  $T_{P_i}$  is a matrix corresponding to the cluster  $P_i$ ,  $T_{P_i} \in \mathfrak{R}^{P_i \times n}$ , in which the  $k$ -th row is  $x_k^\top$ .  $O_{P_i}$  is a mean matrix of cluster  $P_i$ ,  $O_{P_i} \in \mathfrak{R}^{P_i \times n}$ . Each row of  $O_{P_i}$  is the same, i.e.

$$\mu_{P_i} = \frac{1}{P_i} \sum_{x_k \in P_i} x_k. \quad (35)$$

The related covariance matrix for cluster  $P_i$  can be expressed as

$$\Sigma_{P_i}^\Phi = \frac{1}{P_i} (\Phi(T_{P_i}) - \Phi(O_{P_i}))^\top (\Phi(T_{P_i}) - \Phi(O_{P_i})). \quad (36)$$

So we obtain

$$\begin{aligned} \Phi(x_i)^\top \Sigma_+^\Phi \Phi(x_j) &= \left( \frac{1}{\sqrt{P_i}} (\Phi(T_{P_i}) - \Phi(O_{P_i})) \Phi(x_i) \right)^\top \\ &\quad \left( \frac{1}{\sqrt{P_i}} (\Phi(T_{P_i}) - \Phi(O_{P_i})) \Phi(x_j) \right) \\ &= \left( \frac{1}{\sqrt{P_i}} (K(T_{P_i}, x_i) - K(O_{P_i}, x_i)) \right)^\top \\ &\quad \left( \frac{1}{\sqrt{P_i}} (K(T_{P_i}, x_j) - K(O_{P_i}, x_j)) \right). \end{aligned} \quad (37)$$

Similarly,  $\Phi(M)^\top \Sigma_-^\Phi \Phi(M)$  of  $F_\Phi$  can be computed

$$\begin{aligned} \Phi(x_i)^\top \Sigma_-^\Phi \Phi(x_j) &= \left( \frac{1}{\sqrt{P_i}} (K(T_{N_i}, x_i) - K(O_{N_i}, x_i)) \right)^\top \\ &\quad \left( \frac{1}{\sqrt{P_i}} (K(T_{N_i}, x_j) - K(O_{N_i}, x_j)) \right), \end{aligned} \quad (38)$$

where  $T_{N_i}$  is a matrix of cluster  $N_i$ ,  $O_{N_i}$  is a mean matrix of cluster  $N_i$ .

SRMCLP has a similar structure of RMCLP, We can easily proof that RMCLP are the special case of SRMCLP. Suppose the variance-covariance matrix of each cluster is  $\Sigma_{P_i} = \sigma_{N_j} = I$ ,  $i = 1, \dots, C_P$ ,  $j = 1, \dots, C_N$ . For an example of linear SRMCLP, the primal optimization problem (39) of SRMCLP becomes

$$\begin{aligned} \min_z \quad & \frac{1}{2} \|w\|^2 + c_1 \frac{C_P + C_N}{2} \|w\|^2 + \frac{1}{2} \|u\|^2 + \frac{1}{2} b^2 + c_2 e^T u - c_3 e^T v, \\ \text{s.t.} \quad & (w \cdot x_i) + (u_i - v_i) = b, \quad \text{for } \{i | y_i = 1\}, \\ & (w \cdot x_i) - (u_i - v_i) = b, \quad \text{for } \{i | y_i = -1\}, \\ & u, v \geq 0, \end{aligned} \quad (39)$$

It is not difficult to see that the optimization problem (10) is equivalent to one of the primal problem of RMCLP.

## 4 Experiments

We compare the SRMCLP against RMCLP and SRSVM [16, 17] on various data sets in this section.

The testing accuracies of all experiments are computed using standard 10-fold cross validation.  $c_1, c_2, c_3$  and RBF kernel parameter  $\sigma$  are all selected from the set  $\{2^i | i = -7, \dots, 7\}$

by 10-fold cross validation on the tuning set comprising of random 10% of the training data. Once the parameters are selected, the tuning set is returned to the training set to learn the final decision function. The "quadprog" function is used to solve the QP problems in SRMCLP, RMCLP and SRSVM. The "1 vs r" method [2] is used to solve the multi-class classification. All algorithms are implemented by using MATLAB 2010. The experiment environment: Intel Core i7-2600 CPU, 4 GB memory.

#### 4.1 Toy data

In the subsection, we use a 2-D toy data to show the intuitive performance of SRMCLP. The 2-D toy data is the synthetic XOR dataset [17], which is a typical linearly nonseparable problem in classification and randomly generated under two Gaussian distributions in each class. In practise, samples in each class are designed to two clusters  $P_1, P_2$  and  $N_1, N_2$  (the number of samples in each cluster is equal), and each gaussian distribution contains 100 samples. We respectively use 10%, 20%, 30%, 50% of data in each cluster as the training set, and others for testing. The comparative results of SRMCLP and SRSVM are shown in Figure 2.

In the XOR dataset, the positive class and negative class have both the horizontal distribution and the vertical distribution. How to fully exploit these prior knowledge will be a very difficult task. From Figure 2, we can find that SRMCLP's discriminant boundaries basically enclose those of RMCLP, which means that SRMCLP has better generalization performance than RMCLP. Figure 3, we can also find that the accuracy's different of these methods decreases with the increase of training samples. Those show that SRMCLP can fully exploit these prior structural information to design a more reasonable classifier.

#### 4.2 UCI datasets

In this subsection, we perform these methods on the UCI datasets [21]. For each dataset, we randomly select the same number of data from different classes to compose a dataset. 50% percent of each extracted dataset are for training, 50% for testing. The results are shown in the Table 1. From the Table 1, we can draw the conclusion as follows: 1) SRMCLP and SRSVM have the better predictive ability than RMCLP in all cases. This shows that these priori structural information embedded in classes has a great help to improve the classification performance of the classifier. 2) SRMCLP is superior to SRSVM in most cases. This shows SRMCLP is a strong competitive method.

## 5 Conclusion

In this paper, we proposed a novel Structural RMCLP (called SRMCLP) method for classification problem. Unlike RMCLP, SRMCLP is sensitive to the structure of the data distribution and can construct more reasonable classifiers by exploiting these prior data distribution information within classes. The corresponding optimization problem of SRMCLP can be solved by a standard quadratic programming. The effectiveness of the proposed method is demonstrated via experiments on synthetic and available benchmark datasets and applications on the decision supporting system. In the future work, we will apply SRMCLP into other actual classification problems such as stock forecast, credit card analysis to further test its effectiveness.

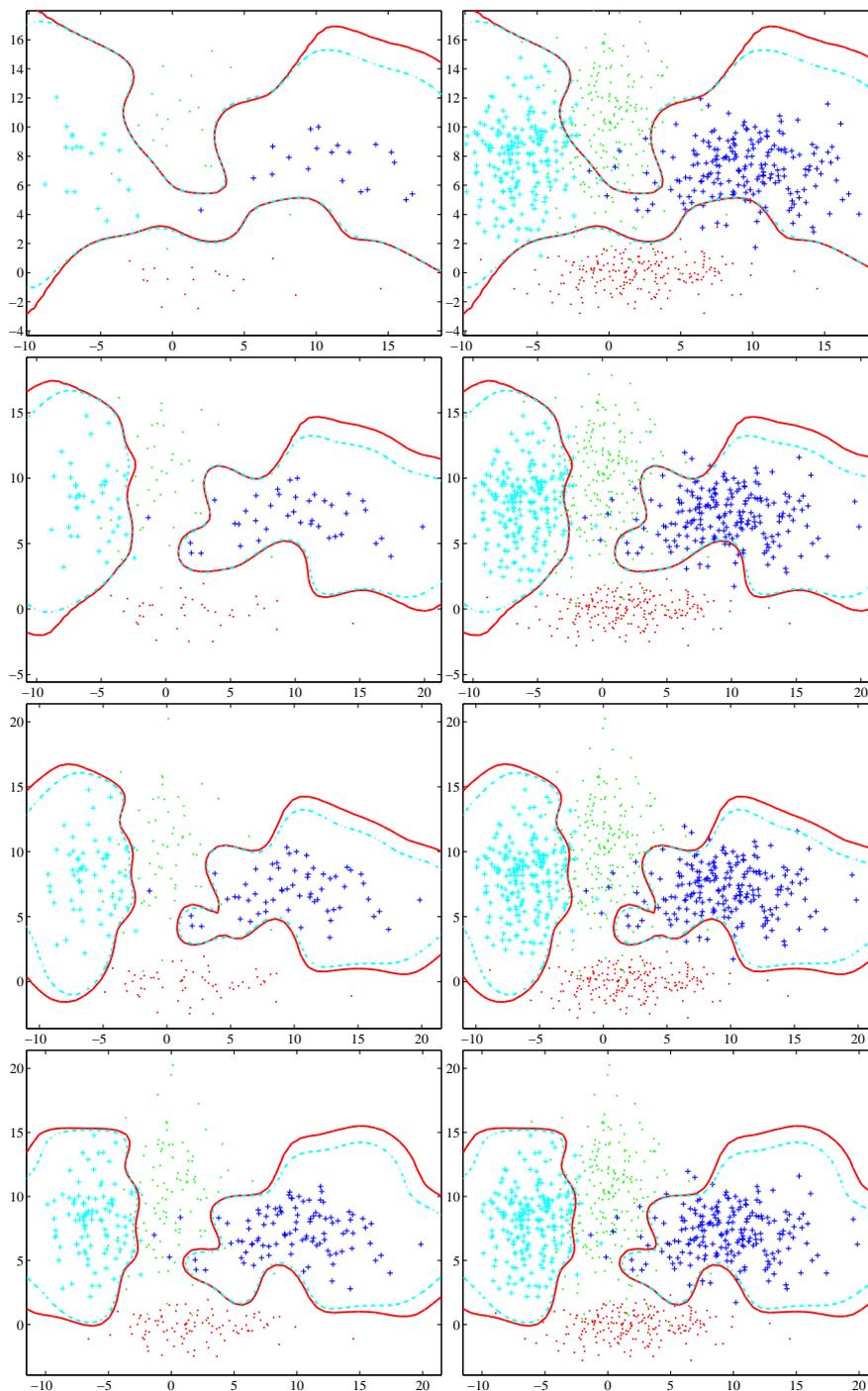


Figure 2: The performance of SRMCLP and RMCLP in the case of RBF case. The first column and second column are the results on the training set and testing set. Each row is the result on 10%, 20%, 30% and 50% training sets, respectively. The magenta dotted curve and red solid curve denote the hyperplanes of SRMCLP and RMCLP, respectively.

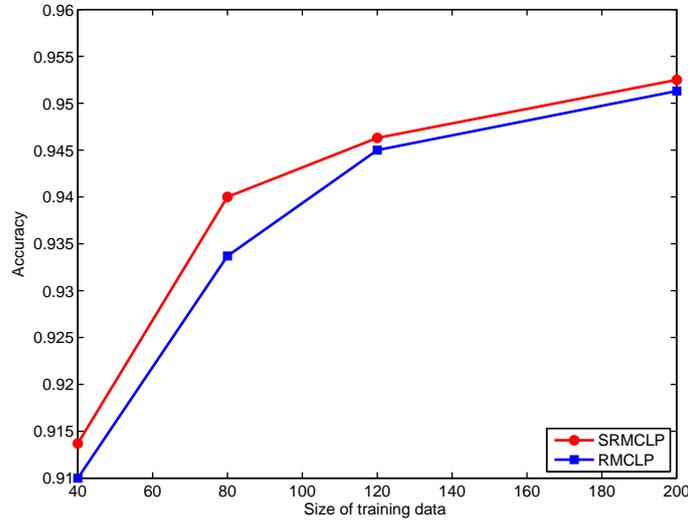


Figure 3: Accuracy of SRMCLP and RMCLP on the XOR dataset

Table 1: The testing accuracy and training times on UCI datasets

Datasets	SRMCLP Accuracy	SRSVM Accuracy	RMCLP Accuracy
Hepatitis (155×19)	<b>79.91±2.55</b>	79.83±1.27	77.82±4.22
Australian (690×14)	68.18±2.12	<b>69.32±2.31</b>	67.73±1.56
BUPA liver (345×6)	68.12±1.34	<b>68.96±1.09</b>	67.61±2.71
CMC (844×9)	<b>65.71±2.54</b>	65.27±2.35	64.53±3.62
Credit (690×19)	<b>76.36±2.18</b>	76.11±2.61	75.82±2.87
Diabetes (768×8)	63.98±1.90	<b>64.19±2.43</b>	62.44±3.47
Flare-Solar (1066×9)	<b>59.11 ±2.98</b>	58.43±2.77	57.96±2.51
German (1000×20)	<b>63.91±1.93</b>	63.84±1.88	62.55±2.86
Heart-Statlog (270×14)	<b>76.13 ±2.50</b>	76.04±2.47	76.21±2.83
Image (2310×18)	83.18 ±2.57	<b>83.44±1.40</b>	82.64±2.88
Ionosphere (351×34)	<b>76.93 ±2.61</b>	76.55±2.63	76.43± 3.26

## 6 Acknowledgment

This work has been partially supported by grants from National Natural Science Foundation of China( NO.70921061, NO.10601064), the CAS/SAFEA International Partnership Program for Creative Research Teams, Major International(Ragional) Joint Research Project(NO.71110107026), the President Fund of GUCAS.

## Bibliography

- [1] Vapnik V.N. The Nature of Statistical Learning Theory. 2nd ed. New York: Springer, 2000.
- [2] Deng N.Y., Tian Y.J. Support vector machines: Theory, Algorithms and Extensions. Science Press, Beijing, 2009.
- [3] Fisher R.A. The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics* 7(2): 179-188, 1936.
- [4] Mangasarian O.L. Generalized support vector machines. *Advances in Large Margin Classifiers*. Cambridge, MA: MIT Press, 2000.
- [5] Freed N., Glover F. Simple but powerful goal programming models for discriminant problems. *European Journal of Operational Research* 7: 44-60, 1981.
- [6] Freed N., Glover F. Evaluating alternative linear programming models to solve the two-group discriminant problem. *Decision Science* 17: 151-162, 1986.
- [7] Olson D., Shi Y. *Introduction to Business Data Mining*. McGraw-Hill/Irwin, 2007.
- [8] Shi Y., Tian Y.J., Chen X.J., Zhang P. A Regularized Multiple Criteria Linear Program for Classification. In: *ICDM Workshops*, pp 253-258, 2007.
- [9] Kou G., Shi Y., Wang S.Y. Multiple criteria decision making and decision support systems - Guest editor's introduction. *Decision Support Systems* 51(2): 247-249, 2011.
- [10] Zhang D., Tian Y.J., Shi Y. A regression method by multiple criteria linear programming. In: *19th International Conference on Multiple Criteria Decision Making (MCDM)*, pp 7-12, 2008.
- [11] Shi Y., Wise W., Lou M. Multiple Criteria Decision Making in Credit Card Portfolio Management. *Multiple Criteria Decision Making in New Millennium*, pp 427-436, 2001.
- [12] Shi Y., Peng Y., Xu W. Data mining via multiple criteria linear programming: applications in credit card portfolio management. *International Journal of Information Technology and Decision Making* 1: 131-151, 2002.
- [13] Zhang J., Zhuang W., Yan N. Classification of HIV-1 Mediated Neuronal Dendritic and Synaptic Damage Using Multiple Criteria Linear Programming. *Neuroinformatics* 2: 303-326, 2004
- [14] Kwak W., Shi Y., Eldridge S. Bankruptcy prediction for Japanese firms: using multiple criteria linear programming data mining approach. *International Journal of Data Mining and Business Intelligence*, 2006.

- [15] Yeung D., Wang D., Ng W., Tsang E., Wang X., Structured large margin machines: sensitive to data distributions, *Machine Learning* 68 (2): 171-200, 2007.
- [16] Xue H., Chen S., Yang Q., Structural support vector machine, in: *The 15th International Symposium on Neural Networks*, pp. 501-511, 2008.
- [17] Xue H., Chen S., Yang Q., Structural Regularized Support Vector Machine: A Framework for Structural Large Margin Classifier, *Neural Networks, IEEE Transactions on* 22 (4): 573-587, 2011.
- [18] Ward, J.R. Hierarchical Grouping to Optimize an Objective Function, *Journal of the American Statistical Association*, 58 (301): 236-244, 1963.
- [19] G. R. G. Lanckriet, L. E. Ghaoui, C. Bhattacharyya, M. I. Jordan, A robust minimax approach to classification, *Journal of Machine Learning Research*, 3: 555-582, 2002.
- [20] Kzhuang K. H., Yang H. , King I., Learning large margin classifiers locally and globally, in: *In The Twenty-First International Conference on Machine Learning*, pp. 401-408, 2004.
- [21] Murphy P.M. and Aha D.W. *UCI machine learning repository*, 1992.