The Third International Symposium on Optimization and Systems Biology (OSB'09)
Zhangjiajie, China, September 20–22, 2009

Kernel Regularized Multiple Criteria Linear Programming

Wen-Jing Chen[1,*] Ying-Jie Tian[2,†]

[1]School of Mathematical Sciences of GUCAS, Beijing 100049
[2]Research Center on Fictitious Economy and Data Science, CAS, Beijing 100190

Abstract Based on our proposed regularized multiple criteria linear programming (RMCLP) for binary classification problems, this paper extends this method to treat with nonlinear case. By applying dual theory, we derived the dual problem of optimization problem constructed in RMCLP, and then proved the solution of RMCLP can be computed by the solution of its dual problem, finally, we constructed Algorithm Kernel RMCLP by introducing Kernel functions in RMCLP. A series of experimental tests are conducted to illustrate the performance of the proposed Kernel RMCLP with the outstanding support vector machine (SVM). The results of several publicly available datasets and a real-life credit dataset all show that our Kernel RMCLP is a competitive method in classification.

Keywords multiple criteria linear programming; Regularize; Kernel; support vector machine; classification

1 Introduction

For the last decade, the researchers have extensively applied optimization techniques to deal with problems in data mining or machine learning. Most data mining or machine learning problems reduce to optimization problems including: unconstrained, quadratic, linear, second-order cone, semi-definite, and semi-infinite convex programs. The research area of mathematical programming intersects with data mining or machine learning in two aspects: On one hand, mathematical programming theory supplies a definition of what constitutes an optimal solution — the optimality conditions. On the other hand, mathematical programming algorithms equip data mining or machine learning researchers with tools for training large families of models[1].

Among all the optimization techniques in data mining, 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[2, 3]. Then, since 1998 Shi and his colleagues extended such a research idea into classification via multiple criteria linear programming (MCLP) and multiple criteria quadratic programming (MCQP), which differ from statistics, decision tree induction, and neural networks[4] – [8]. These mathematical programming approaches to classification have been applied to handle many realworld data mining

*chenwenjing08@mails.gucas.ac.cn
†Corresponding author: tianyingjie1213@163.com

problems, such as credit card portfolio management[9, 10], bioinformatics [11, 12], fraud management[13], information intrusion and detection[14, 15], firm bankruptcy[16], etc.

In order to overcome some shortcoming of the MCLP models, we proposed regularized multiple criteria linear programming (RMCLP) with existence of solution for classification[17]. Numerical experiments have proved our Algorithm RMCLP's efficiency. However, RMCLP can only deal with linear classification problem which strongly restrict its application. Following the idea of SVM[18, 19], we will extend it to nonlinear case by introducing the kernel functions. That is the motivation of this paper.

Rest of the paper proceeds as follows. Section 2 introduces the basic notions and formulation of RMCLP. Then section 3 describes in detail our proposed Algorithm Kernel RMCLP (KRMCLP). Section 4 uses a series of experimental tests to illustrate the performance of the proposed KRMCLP with the existing methods: MCLP, MCQP, RMCLP and SVM. Section 5 gives the conclusions.

2 Primal problem

For a binary classification problem, given a training set

$$T = \{(x_1, y_1)\cdots,(x_l, y_l)\} \in (R^n \times \mathscr{Y})^l, \tag{2.1}$$

where $x_i \in R^n$, and $y_i \in \mathscr{Y} = \{-1, 1\}, i = 1, \cdots, l$. RMCLP[17] constructs the following problem

$$\begin{aligned}
\min_{z} \quad & \frac{1}{2}w^{\mathrm{T}}Hw + \frac{1}{2}u^{\mathrm{T}}Qu + de^{\mathrm{T}}u - ce^{\mathrm{T}}v, && (2.2)\\
\text{s.t.} \quad & (w \cdot x_i) + u_i - v_i = b, \ \ \text{for}\{i|y_i = 1\}, && (2.3)\\
& (w \cdot x_i) - u_i + v_i = b, \ \ \text{for}\{i|y_i = -1\}, && (2.4)\\
& u, v \geq 0, && (2.5)
\end{aligned}$$

where $z = (w^{\mathrm{T}}, u^{\mathrm{T}}, v^{\mathrm{T}}, b)^{\mathrm{T}} \in R^{n+l+l+1}$, $H \in R^{n\times n}$ and $Q \in R^{l\times l}$ are symmetric positive definite matrices, and $c, d \geq 0$, $e \in R^l$ be vectors whose all elements are 1, we call it the primal problem in this paper. If the solution of primal problem $z^* = (w^{*\mathrm{T}}, u^{*\mathrm{T}}, v^{*\mathrm{T}}, b^*)^{\mathrm{T}}$ is derived, we can construct the decision function as

$$f(x) = \mathrm{sgn}(g(x)) = \mathrm{sgn}((w^* \cdot x) - b^*), \tag{2.6}$$

therefore,c for any unknown input x, its label is deduced by (2.6).

Obviously, Problem (2.2)$\sim$(2.5) is a convex quadratic program and we have proved it has a bounded solution set if H, Q, d, c are chosen appropriately.

Without loss of generality, suppose inputs $x_1, \cdots, x_{l_1}$ belong to positive class, and inputs $x_{l_1+1}, \cdots x_{l_1+l_2}$ belong to negative class, $I_1 \in R^{l_1\times l_1}, I_2 \in R^{l_2\times l_2}$ be identity matrices,

$$A_1 = \begin{pmatrix} x_1^{\mathrm{T}} \\ \vdots \\ x_{l_1}^{\mathrm{T}} \end{pmatrix}_{l_1\times n}, \quad A_2 = \begin{pmatrix} x_{l_1+1}^{\mathrm{T}} \\ \vdots \\ x_{l_1+l_2}^{\mathrm{T}} \end{pmatrix}_{l_2\times n},$$

$$A = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix}_{l\times n}, \quad E = \begin{pmatrix} I_1 & 0 \\ 0 & -I_2 \end{pmatrix}_{l\times l}, \tag{2.7}$$

and $e \in R^l$ be vectors with all elements 1. In this paper we choose H to be identity matrix, and add a term $\frac{1}{2}b^2$ to the objective function, so the problem (2.2)∼(2.5) turns out to be

$$\begin{aligned}
\min_{z} \quad & \frac{1}{2}w^{\mathrm{T}}w + \frac{1}{2}b^2 + \frac{1}{2}u^{\mathrm{T}}Qu + de^{\mathrm{T}}u - ce^{\mathrm{T}}v, && (2.8)\\
\text{s.t.} \quad & Aw + Eu - Ev - be = 0, && (2.9)\\
& u \geq 0, v \geq 0. && (2.10)
\end{aligned}$$

Because the objective function (2.8) is strictly convex w.r.t (w,b,u), so we have the following theorem[20].

Theorem 1. *Solution of problem* (2.8)∼(2.10) *w.r.t* (w,b,u) *is unique.*

3 Kernel RMCLP

Naturally, we first derive the dual problem of problem (2.8)∼(2.10) by introducing its Lagrange Function

$$\begin{aligned}
L(w,u,v,b,\alpha,\beta,\eta) \quad = \quad & \frac{1}{2}w^{\mathrm{T}}w + \frac{1}{2}b^2 + \frac{1}{2}u^{\mathrm{T}}Qu + de^{\mathrm{T}}u - ce^{\mathrm{T}}v \\
& + \alpha^{\mathrm{T}}(Aw + Eu - Ev - be) - \beta^{\mathrm{T}}u - \eta^{\mathrm{T}}v, \qquad (3.1)
\end{aligned}$$

where $\alpha \in R^l$, and $\beta \geq 0, \eta \geq 0$ are the Lagrange multipliers. Therefore the dual problem of (2.8)∼(2.10) can be formulated as

$$\begin{aligned}
\max_{w,u,v,b,\alpha,\beta,\eta} \quad & L(w,u,v,b,\alpha,\beta,\eta), && (3.2)\\
\text{s.t.} \quad & \nabla_{w,u,v,b}L(w,u,v,b,\alpha,\beta,\eta) = 0, && (3.3)\\
& \beta, \eta \geq 0. && (3.4)
\end{aligned}$$

From equation (3.3) we get

$$\begin{aligned}
\nabla_w L &= w + A^{\mathrm{T}}\alpha = 0, && (3.5)\\
\nabla_v L &= -ce - E\alpha - \eta = 0, && (3.6)\\
\nabla_u L &= Qu + E\alpha + de - \beta = 0, && (3.7)\\
\nabla_b L &= b - e^{\mathrm{T}}\alpha = 0. && (3.8)
\end{aligned}$$

Substituting the above equations into problem (3.2)∼(3.4), we will get

$$\begin{aligned}
\max_{\alpha,u} \quad & -\frac{1}{2}\alpha^{\mathrm{T}}(AA^{\mathrm{T}} + ee^{\mathrm{T}})\alpha - \frac{1}{2}u^{\mathrm{T}}Qu, && (3.9)\\
\text{s.t.} \quad & -Qu - de \leq E\alpha \leq -ce. && (3.10)
\end{aligned}$$

We can find that the inputs $x_i, i = 1, \cdots, l$ only appear in the term $AA^{\mathrm{T}} = (x_i \cdot x_j)_{l \times l}$, which means we can introduce the kernel function $K(x,x') = (\Phi(x) \cdot \Phi(x'))$ to take place of $(x \cdot x')$, where $\Phi(\cdot)$ is a mapping from the input space R^n to some Hilbert space $\mathscr{H}$

$$\Phi: \quad \begin{array}{l} R^n \rightarrow \mathscr{H}\ , \\ x \rightarrow \Phi(x)\ , \end{array} \qquad (3.11)$$

Therefore, dual problem (3.9)∼(3.10) turns out to be

$$\min_{\alpha,u} \quad \frac{1}{2}\alpha^{\mathrm{T}}(K(A,A^{\mathrm{T}})+ee^{\mathrm{T}})\alpha+\frac{1}{2}u^{\mathrm{T}}Qu, \tag{3.12}$$

$$\text{s.t.} \quad -Qu-de \le E\alpha \le -ce, \tag{3.13}$$

where $K(A,A^{\mathrm{T}}) = \Phi(A)\Phi(A)^{\mathrm{T}} = (\Phi(x_i)\cdot\Phi(x_j))_{l\times l}$. Obviously problem (3.12)∼(3.13) is a convex quadratic problem and always has a solution if the parameters c,d chosen appropriately.

Theorem 2. *Suppose* (α^*,u^*) *is the solution of dual problem* (3.12)∼(3.13)*, the solution of corresponding primal problem* (2.8)∼(2.10) *in some Hilbert space* $\mathscr{H}$ *w.r.t* $(\mathrm{w},b) = (\Phi(w),b)$ *can be computed as follows:*

$$\mathrm{w}^* = -\Phi(A)^{\mathrm{T}}\alpha^* \quad \textit{and} \quad b^* = e^{\mathrm{T}}\alpha^*. \tag{3.14}$$

Proof The Lagrange Function of dual problem (3.12)∼(3.13) is

$$\begin{aligned} L(\alpha,u,\tilde{u},\tilde{v}) \;=\; & \frac{1}{2}\alpha^{\mathrm{T}}K(A,A^{\mathrm{T}})\alpha+\frac{1}{2}\alpha^{\mathrm{T}}ee^{\mathrm{T}}\alpha+\frac{1}{2}u^{\mathrm{T}}Qu \\ & -(Qu+de+E\alpha)^{\mathrm{T}}\tilde{u}+(E\alpha+ce)^{\mathrm{T}}\tilde{v}, \end{aligned} \tag{3.15}$$

where $\tilde{u}\ge 0,\tilde{v}\ge 0$ are the Lagrange multipliers. The KKT conditions of dual problem is

$$Qu^*+de+E\alpha^* \ge 0,\ E\alpha^*+c\le 0, \tag{3.16}$$

$$(Qu^*+de+E\alpha^*)^{\mathrm{T}}\tilde{u}=0, \tag{3.17}$$

$$(E\alpha^*+ce)^{\mathrm{T}}\tilde{v}=0, \tag{3.18}$$

$$K(A,A^{\mathrm{T}})\alpha^*+ee^{\mathrm{T}}\alpha-E\tilde{u}+E\tilde{v}=0, \tag{3.19}$$

$$Qu^*-Q\tilde{u}=0, \tag{3.20}$$

$$\tilde{u},\tilde{v}\ge 0. \tag{3.21}$$

It is easy to see that $\tilde{u}=u^*$ from condition (3.20).
Now let

$$\mathrm{w}^* = -\Phi(A)^{\mathrm{T}}\alpha^*, \quad b^* = e^{\mathrm{T}}\alpha^*, \tag{3.22}$$

then equation (3.19) turns out to be

$$\Phi(A)\mathrm{w}^*-b^*e+E\tilde{u}-E\tilde{v}=0, \tag{3.23}$$

together with (3.21) imply that $(\mathrm{w}^*,\tilde{u},\tilde{v},b^*)$ is the feasible point of primal problem.

Furthermore, from KKT conditions (3.16)∼(3.21) we know that

$$\begin{aligned} & \frac{1}{2}\mathrm{w}^{*\mathrm{T}}\mathrm{w}^*+\frac{1}{2}b^*+\frac{1}{2}\tilde{u}^{\mathrm{T}}Q\tilde{u}+de^{\mathrm{T}}\tilde{u}-ce^{\mathrm{T}}\tilde{v} \\ = \; & \frac{1}{2}\alpha^{*\mathrm{T}}(K(A,A^{\mathrm{T}})+ee^{\mathrm{T}})\alpha^*+\frac{1}{2}\tilde{u}^{\mathrm{T}}Q\tilde{u}+de^{\mathrm{T}}\tilde{u}-ce^{\mathrm{T}}\tilde{v} \\ = \; & \frac{1}{2}\alpha^{*\mathrm{T}}(K(A,A^{\mathrm{T}})+ee^{\mathrm{T}})\alpha^*+\frac{1}{2}\tilde{u}^{\mathrm{T}}Q\tilde{u}+de^{\mathrm{T}}\tilde{u}-ce^{\mathrm{T}}\tilde{v} \\ & -(K(A,A^{\mathrm{T}})\alpha^*+ee^{\mathrm{T}}\alpha^*-E\tilde{u}+E\tilde{v})^{\mathrm{T}}\alpha^* \\ = \; & -\frac{1}{2}\alpha^{*\mathrm{T}}(K(A,A^{\mathrm{T}})+ee^{\mathrm{T}})\alpha^*-\frac{1}{2}u^{*\mathrm{T}}Qu^*, \end{aligned} \tag{3.24}$$

i.e. the objective value of primal problem at $(\mathrm{w}^*, \tilde{u}, \tilde{v}, \tilde{b})$ equals to the objective value of dual problem at (α^*, u^*), so that $(\mathrm{w}^*, \tilde{u}, \tilde{v}, \tilde{b})$ is the solution of primal problem (2.8)∼(2.10) in some Hilbert space $\mathscr{H}$[20]. □

Now, based on Theorem 2 we construct the following Algorithm kernel RMCLP:

Algorithm 1.
Algorithm3.1 Algorithm of Kernel RMCLP(KRMCLP)

(1) Given a training set $T = \{(x_1, y_1), ..., (x_l, y_l)\} \in (R^n \times \{-1, 1\})^l$*;*
(2) Select an appropriate kernel $K(\cdot,\cdot)$*, symmetric positive definite matrices,* $Q \in R^{l\times l}$*,* $c, d \geq 0$*;*
(3) Solve problem (3.12)∼(3.13) *and get the solution* α^**;*
(4) Construct the decision function as

$$\begin{aligned} f(x) &= \mathrm{sgn}((\mathrm{w}^* \cdot \Phi(x)) - b^*) \\ &= \mathrm{sgn}(-K(A,x)^{\mathrm{T}}\alpha^* - e^{\mathrm{T}}\alpha^*), \end{aligned} \tag{3.25}$$

□

Obviously, we can see that when the kernel function $K(x, x')$ chosen to be linear kernel $(x \cdot x')$, algorithm Kernel RMCLP degenerates from algorithm RMCLP.

4 Numerical Experiments

Because RMCLP is the linear case of KRMCLP when the kernel function chosen to be linear kernel, so in this section, we will only compare the performance of KRMCLP with SVM on three publicly available datasets from UCI Machine Learning Repository[21] and credit card dataset, both algorithms will choose RBF kernel: $K(x, x') = \exp(-\frac{\|x-x'\|^2}{\sigma})$, where $\sigma > 0$.

4.1 UCI datasets

For every UCI dataset, we randomly separate it into two parts, one part is for training, and the other for testing, then apply the above KRMCLP and SVM to train and test. This process is performed ten times, every time the scores on training and testing are recorded, at last the average scores are computed and shown in Table 1—Table 3. Here, we apply three scores to evaluate two algorithms: sensitivity (Sn), specificity(Sp) and G-Mean(g)

$$Sn = \frac{TP}{TP + FN}, \tag{4.1}$$

$$Sp = \frac{TN}{TN + FP}, \tag{4.2}$$

$$g = \sqrt{Sn \times Sp}, \tag{4.3}$$

where TP is true positive, TN is true negative, FP is false positive and FN is false negative.

In every training, parameters in each algorithm are selected in some discrete set in order to get the best scores. For example, the parameters in KRMCLP needed to be chosen are $\sigma > 0$ in RBF kernel, matrix Q, penalty parameters $c > 0$ and $d > 0$, so we choose Q in a set of several given matrixes, σ, c, d in the sets of several given numbers. From Table 1 to Table 3 we can see that the performance of RMCLP is almost the same with SVM or even better than SVM in some scores.

Table 1: Experiments On Australian Dataset

Classification	Training (200 records)+Testing (490 records)							
Algorithms	*Sn*	*Sp*	*g*	*Acc*	*Sn*	*Sp*	*g*	*Acc*
RMCLP	85.0%	84.0%	84.5%	84.5%	90.3%	80.9%	85.5%	84.9%
SVM	—	—	—	87.0%	93.7%	78.8%	85.9%	85.1%

Table 2: Experiments On German Dataset

Classification	Training (200 records)+Testing (800 records)							
Algorithms	*Sn*	*Sp*	*Ac*	*g*	*Sn*	*Sp*	*Ac*	*g*
RMCLP	73.0%	57.0%	64.4%	65.0%	76.0%	63.5%	69.5%	66.6%
SVM	—	—	—	69.0%	79.0%	45.9%	60.2%	65.6%

Table 3: Experiments On Heart Dataset

Classification	Training (100 records)+Testing (170 records)							
Algorithms	*Sn*	*Sp*	*g*	*Acc*	*Sn*	*Sp*	*g*	*Acc*
RMCLP	80.0%	86.0%	82.9%	83.0%	77.1%	73.0%	80.0%	80.6%
SVM	—	—	—	87.0%	77.1%	88.0%	82.4%	83.5%

4.2 Credit Card Dataset

Now we test the performance of KRMCLP on credit card dataset. The 6000 credit card records used in this paper were selected from 25,000 real-life credit card records of a major US bank. Each record has 113 columns or variables to describe the cardholders' behaviors, including balance, purchases, payment cash advance and so on. With the accumulated experience functions, we eventually get 65 variables from the original 113 variables to describe the cardholders' behaviors.

In this paper we chose the holdout method on credit card dataset to separate data into training set and testing set: first, the bankruptcy dataset (960 records) is divided into 10 intervals (each interval has approximately 100 records). Within each interval, 50 records are randomly selected. Thus the total of 500 bankruptcy records is obtained after repeating 10 times. Then, with the same way, we get 500 current records from the current dataset. Finally, the total of 500 bankruptcy records and 500 current records are combined to form a single training dataset, with the remaining 460 lost records and 4540 current records merging into a testing dataset. This process is performed for ten times, for each time we apply KRMCLP to training and testing. In each training, we apply 5-fold cross-validation to choose best parameters in KRMCLP for testing. At last we recorded the corresponding average scores for each time in Table 4.

Table 4: Experiments on Credit Card Dataset

	Training Set				Testing Set			
	Sn	*Sp*	*Ac*	*g*	*Sn*	*Sp*	*Ac*	*g*
DS 1	87.50%	60.50%	74.00%	72.75%	90.13%	62.87%	66.57%	75.28%
DS 2	91.50%	64.00%	77.50%	76.46%	90.13%	62.87%	66.57%	75.28%
DS 3	92.00%	60.00%	76.00%	74.22%	90.13%	62.87%	66.57%	75.28%
DS 4	94.00%	67.00%	80.50%	79.34%	90.13%	62.87%	66.57%	75.28%
DS 5	89.00%	62.00%	75.50%	74.16%	90.13%	62.87%	66.57%	75.28%
DS 6	89.00%	63.50%	76.25%	75.04%	90.13%	62.87%	66.57%	75.28%
DS 7	89.00%	57.50%	73.25%	71.38%	90.13%	62.87%	66.57%	75.28%
DS 8	93.50%	63.00%	78.25%	76.68%	90.13%	62.87%	66.57%	75.28%
DS 9	94.00%	64.50%	79.25%	77.66%	90.13%	62.87%	66.57%	75.28%
DS 10	87.00%	63.5%	75.25%	74.20%	90.13%	62.87%	66.57%	75.28%

5 Conclusion

In this paper, a kernel regularized multiple criteria linear program (KRMCLP) has been proposed for classification problems in data mining. Comparing with the known multiple criteria linear program (MCLP) model and regularized multiple criteria linear program RMCLP, this model not only guarantees the existence of solution and mathematically solvable, but also can deal with nonlinear case, which extend the real application of MCLP and RMCLP. In addition to describing our algorithm's effiency, this paper has also conducted a series of experimental tests on several datasets comparing with the support vector machine (SVM), all results have shown that KRMCLP is a competitive method in classification.

References

[1] Bennett K. P., ,Emilio, P.H. The Interplay of Optimization and Machine Learning. Research Journal of Machine Learning Research, 2006, 7: 1265-1281.

[2] Freed N, Glover F. Simple but powerful goal programming models for discriminant problems. Simple but Powerful Goal Programming Models for Discriminant Problems, 1981, 7: 44-60.

[3] Freed N, Glover F. Evaluating alternative linear programming models to solve the two-group discriminant problem. Decision Science, 1986, 17: 151-162.

[4] Olson D, Shi Y. Introduction to Business Data Mining. McGraw- Hill/Irwin, 2007.

[5] Shi Y. Multiple criteria and multiple constraint levels linear programming: concepts, techniques and applications. New Jersey: World Scientific Pub Co Inc, 2001.

[6] He J, Liu X, Shi Y, et al.. Classifications of credit card holder behavior by using fuzzy linear programming. International Journal of Information Technology and Decision Making, 2004, 3: 633-650.

[7] Kou G, Liu X, Peng Y, et al. Multiple criteria linear programming approach to data mining: models, algorithm designs and software Development. Optimization Methods and Software, 2003, 18: 453-473.

[8] Shi Y, Peng Y, Kou G, et al. Classifying credit card accounts for business intelligence and decision making: a multiple-criteria quadratic programming approach. International Journal of Information Technology and Decision Making, 2005, 4: 581-600.

[9] Shi Y, Wise W, Lou M, et al. Multiple Criteria Decision Making in Credit Card Portfolio Management. Multiple Criteria Decision Making in New Millennium. 427-436, 2001.

[10] Shi Y, Y Peng, Xu W, et al. Data mining via multiple criteria linear programming: applications in credit card portfolio management. International Journal of Information Technology and Decision Making, 2002, 1: 131-151.

[11] Zhang J, Zhuang W, Yan N, et al. Classification of HIV-1 Mediated Neuronal Dendritic and Synaptic Damage Using Multiple Criteria Linear Programming. Neuroinformatics, 2004, 2: 303-326.

[12] Shi Y, Zhang X, Wan J, et al. Prediction the distance range between antibody interface residues and antigen surface using Multiple Criteria Quadratic Programming. International Journal of Computer Mathematics, 2004, 84:690-707.

[13] Peng Y, Kou G, Sabatka A, et al. Application of classification methods to individual disability income insurance fraud detection. In: ICCS 2007, Lecture Notes in Computer Science, 2007, 852-858.

[14] Kou G, Peng Y, Chen Z, et al. A multiple-criteria quadratic programming approach to network intrusion detection. Chinese Academy of Sciences Symposium on Data Mining and Knowledge Management. 2004, 7: 12-14.

[15] Kou G, Peng Y, Yan N, et al. Network intrusion detection by using multiple-criteria linear programming. In: International Conference on Service Systems and Service Management, 2004, 7: 19-21.

[16] Kwak W, Shi Y, Eldridge S, et al. Bankruptcy prediction for Japanese firms: using multiple criteria linear programming data mining approach. International Journal of Data Mining and Business Intelligence, 2006.

[17] Shi Y., Tian, Y.J., Chen, X.J., Zhang, P. Regularized Multiple Criteria Linear Programs for Classification. Science in China Series F: Information Sciences. Vol.52, No.1, 2009. to appear.

[18] Scholkopf, B., Smola, A.J. Learning with Kernels–Support Vector Machines, Regularization, Optimization, and Beyond. The MIT Press, 2002.

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

[20] Bazaraa, M.S., Sherali, H.D., Shetty, C.M. Nonlinear Programming: Theory and Algorithms. New York: John Wiley and Sons, 1993.

[21] Murphy P M, Aha D W, UCI Repository of Machine Learning Databases. Available online at: www.ics.uci.edu mlearnMLRepository.html, 1992.