The First International Symposium on Optimization and Systems Biology (OSB'07)
Beijing, China, August 8–10, 2007

A Framework of Optimization Method for Classification*

Juliang Zhang[1,†] Yong Shi[2]

[1] Research Institute of Material Flow, School of Economics and Management
Beijing Jiaotong University, Beijing 100044, China
[2] Graduate University of Chinese Academy of Sciences, Beijing 100080, China

Abstract In this paper, we propose a framework of optimization method for classification problem and show that Support Vector Machine (SVM), Glover's method and Shi's Method (MCLP) are special cases of the model.

Keywords Optimization; data mining; classification.

1 Introduction

Within last decades, data mining has been drawing increasing attention and research on data mining has mushroomed. Numerous innovative methods have been proposed and the methodology of data mining has been applied successfully in many practical and scientific areas, for example, bioinformatics, information retrieval, adaptive hypermedia, electronic commerce, etc. [2].

Classification is an important area of data mining because it straddles the areas of management science and artificial intelligence as well as statistics [7]. Management science applications include decision to make or buy, lend or invest, hire or reject and artificial intelligence includes pattern recognition, signal differentiation, diagnostic classifications, code signatures etc..

Classification utilizes techniques from many areas, such as statistics, decision tree and optimization. The optimization based methods have enhanced both theoretical foundation and practical applications of data mining [4]. Now, there are two classes of optimization based classification methods: Support Vector Machine (SVM), which was proposed and developed by [3], [1] and [8] respectively, and linear programming methods, which was firstly proposed by [6] and then developed by [9] (called MCLP). In this paper, we give a general framework of optimization method for classification. Then we prove that SVM, Glover's method and MCLP are special cases of the framework. This demonstrates that SVM, Glover's method and MCLP have the similar modeling philosophy.

*This research has been partially supported by a grant from National Natural Science Foundation of China (70671100, 70302003) and Beijing Jiaotong University Science and Technology Foundation(2007RC014)
†Corresponding author: Email: zhangjl4@sem.tsinghua.edu.cn

This paper is organized as follows. Section 2 briefly reviews the existed methods. Section 3 gives the general framework of the model and shows that the existed methods are special cases of our model. Section 4 gives several new models based on our model. Section 5 gives some conclusions.

2 Problem Formulations and Existed Method

This section gives classification problem formulation and reviews some existed methods for this problem.

Given an r-dimensional attribute vector $a = (a_1, ..., a_r)$, let $A_i = (A_{i1}, ..., A_{ir}) \in R^r$ be one of the sample records, where $i = 1, ..., n$. Suppose two groups, G_1 and G_2, are predefined. A boundary scalar b can be set to separate these two groups. A vector $\boldsymbol{X}$ = $(x_1, ..., x_r)^T \in R^r$ can be identified to establish the following linear inequality [5]:

$\boldsymbol{A}_i\boldsymbol{X} \geq b$, $\forall\, \boldsymbol{A}_i \in G_1$;

$\boldsymbol{A}_i\boldsymbol{X} < b, \forall\, \boldsymbol{A}_i \in \mathrm{G}_2$;

There are many methods to find $\boldsymbol{X}$ and b. Here we only introduce the methods based on optimization.

2.1 Support Vector Machine (SVM)

The philosophy of SVM is to find a hyperplane which maximizes the margin between bounding planes. Consider separable case. For hyperplane l, the corresponding bounding plane is l_1 and l_2, which parallel to l and pass the point in G_1 and G_2 respectively and are closest to l. The margin is the distance between l_1 and l_2. SVM finds a hyperplane l such that the margin corresponding to hyperplane l is maximized. We can formulate this problem as follows. Let the equation of the hyperplane l be $\boldsymbol{AX}= b$.Then the equations of l_1 and l_2 (after transformation) are

$l_1 : AX{=}b{+}1$

$l_2 : AX{=}b{-}1$

The distance between l_1 and l_2 is $\frac{2}{\|X\|}$. Therefore, we can obtain the desired X and b by solving the following quadratic programming

$$\begin{array}{rll} \min & \|X\|^2 & \\ \text{Subject to} & A_iX \geq b+1, & A_i \in G_1 \\ & A_iX \leq b-1, & A_i \in G_2 \end{array} \tag{1}$$

We can formulate this problem in another way. Let β_i be the distance of record $\boldsymbol{A}_i$ from the hyperplane l. Then this problem can be formulated as follows

$$\begin{array}{rll} \max & (\min\{\beta_i\}) & \\ \text{Subject to} & A_iX - b - \beta_i = 0, & A_i \in G_1 \\ & A_iX - b + \beta_i = 0, & A_i \in G_2 \\ & \|X\| = 1, \beta_i \geq 0 & \end{array} \tag{2}$$

For the nonseparable case, we introduce a new variable α_i, which represents the error of misclassification of sample i. α_i= 0 means that sample i is classified correctly and $\alpha_i > 0$ means that sample i is classified incorrectly. Then the problem can be formulated as follows

$$\begin{array}{ll} \min & \|X\|^2 + C\sum_{i=1}^{\infty}\alpha_i \\ \text{Subject to} & A_iX - b - 1 + \alpha_i > 0, \quad A_i \in G_1 \\ & A_iX - b + 1 - \alpha_i \leq 0, \quad A_i \in G_2 \\ & \alpha_i \geq 0 \end{array} \tag{3}$$

where C is a penalty parameter.

By the similar argument as above, (3) is equivalent to the following problem

$$\begin{array}{ll} \max & (\min\{\beta_i\}) - C\sum_{i=1}^{\infty}\alpha_i \\ \text{Subject to} & A_iX - b - \beta_i + \alpha_i = 0, \quad A_i \in G_1 \\ & A_iX - b + \beta_i - \alpha_i = 0, \quad A_i \in G_2 \\ & \|X\| = 1, \alpha_i, \beta_i \geq 0 \end{array} \tag{4}$$

where C is a penalty parameter.

Note that for SVM, we first construct the basic model for separable case and then generalize it to nonsepaprable case.

2.2 Glover's Method

In 1981, Glover gave another linear programming model for classification problem. The basic idea of the model is as follows. Let α_i be the overlapping of two-group boundary for record $\boldsymbol{A}_i$ and β_i be the distance of record $\boldsymbol{A}_i$ from its adjusted boundary. The overlapping α_i means the distance of record $\boldsymbol{A}_i$ to the boundary b if $\boldsymbol{A}_i$ is misclassified into another group. Then the problem can be formulated as follows

$$\begin{array}{ll} \min & w_\alpha \sum_{i=1}^{\infty}\alpha_i - w_\beta \sum_{i=1}^{\infty}\beta_i \\ \text{Subject to} & A_iX - b - \beta_i + \alpha_i = 0, \quad A_i \in G_1 \\ & A_iX - b + \beta_i - \alpha_i = 0, \quad A_i \in G_2 \\ & \alpha_i, \beta_i \geq 0 \end{array} \tag{5}$$

where $\boldsymbol{A}_i$ is given, $\boldsymbol{X}$ and b are unrestricted, and $w_\alpha$and w_β are weights determined previously by decision maker.

2.3 Shi's Method (MCLP)

Recently, Shi et al. generalized the Glover's method by considering the following multi-criterion linear programming

$$\begin{array}{ll} \min & \sum_{i=1}^{\infty}\alpha_i \text{ and } \max\sum_{i=1}^{\infty}\beta_i \\ \text{Subject to} & A_iX - b - \beta_i + \alpha_i = 0, \quad A_i \in G_1 \\ & A_iX - b + \beta_i - \alpha_i = 0, \quad A_i \in G_2 \\ & \alpha_i, \beta_i \geq 0 \end{array} \tag{6}$$

where $\boldsymbol{A}_i$ is given, $\boldsymbol{X}$ and b are unrestricted.

Note that Glover's method and MCLP directly construct model for both separable and nonseparable cases.

3 The General Framework

This section introduces a general model for classification. Then we show that SVM, Glover's method and MCLP are special cases of our model. Let α, $\boldsymbol{X}$ and b be defined as in 2.2 and α^* represent the maximum of overlapping. Then, $f(\alpha)$ and $g(\beta)$ can be used to describe the relation of all overlapping α_i and all distances β_i respectively. The final classification accuracies depend on simultaneously minimizing $f(\alpha)$ and maximizing $g(\beta)$. Thus, a general method for classification can be formulated as:

$$\begin{array}{rll} \text{(general model)} & \min f(\alpha) \text{ and } \max g(\beta) & \\ \text{Subject to} & A_iX - b - \beta_i + \alpha_i = 0, \quad A_i \in G_1 & \\ & A_iX - b + \beta_i - \alpha_i = 0, \quad A_i \in G_2 & \\ & \|X\| = 1, \alpha_i, \beta_i \geq 0 & \end{array} \tag{7}$$

where $\boldsymbol{A}_i$ is given, b is unrestricted, and $\alpha = (\alpha_1, \ldots, \alpha_n)$, $\beta = (\beta_1, \ldots, \beta_n)$.

From (5), (6) and (7), we know that the feasible set in Glover's method and MCLP is the same as that in the general model in addition to $\|X\| = 1$. Note that $\|X\| = 1$ does not affect the separate hyperplane. The objective function in Glover's method and MCLP are linear functions while the objective functions in the general model are general functions. Hence, it is obvious that Glover's method and MCLP are special cases of the general model.

Now we show that SVM is a special case of our general model. First we consider the separable case. If we set $g(\beta)$=min$\{\beta_i\}$ and $f(\alpha)$ is a component-wise nondecreasing function, then we can show that the optimal solution of (7) satisfies α_i=0, i=1,...,n. This implies that (7) is equivalent to (2). For nonsepaprable case, let $f(\alpha)=\sum_{i=1}^{\infty}\alpha_i$, $g(\beta)$=min$\{\beta_i\}$, then (7) turns to

$$\begin{array}{rl} \min & \sum_{i=1}^{\infty}\alpha_i \text{ and } \max(\min\{\beta_i\}) \\ \text{Subject to} & A_iX - b - \beta_i + \alpha_i = 0, \quad A_i \in G_1 \\ & A_iX - b + \beta_i - \alpha_i = 0, \quad A_i \in G_2 \\ & \|X\| = 1, \alpha_i, \beta_i \geq 0 \end{array} \tag{8}$$

(2) is a special case of (8). This implies that SVM is a special case of (7).

4 Some New Methods

Based on this general model, we can propose some new models for classification. Let $f(\alpha) = \|\alpha\|_p^p$ and $g(\beta) = \|\beta\|_q^q$. We use weights $w_\alpha > 0$ and $w_\beta > 0$ for $\|\alpha\|_p^p$ and $\|\beta\|_q^q$, respectively. Thus, the general model turns to:

$$\begin{array}{rl} (Model1) & \min w_\alpha\|\alpha\|_p^p - w_\beta\|\beta\|_q^q \\ \text{Subject to} & A_iX - b - \beta_i + \alpha_i = 0, \quad A_i \in G_1 \\ & A_iX - b + \beta_i - \alpha_i = 0, \quad A_i \in G_2 \\ & \|X\| = 1, \alpha_i, \beta_i \geq 0 \end{array} \tag{9}$$

where $\boldsymbol{A}_i$ is given, $\boldsymbol{X}$ and b are unrestricted.

Based on Model 1, mathematical programming models with any norm can be theoretically defined. Now we are only interested in formulating a linear or a quadratic programming model.

Case 1: $p = q = 1$

In this case, Model 1 turns to be

$$\begin{array}{rl} (linear) & \min w_\alpha \sum_{i=1}^{\infty} \alpha_i - w_\beta \sum_{i=1}^{\infty} \beta_i \\ \text{Subject to} & A_iX - b - \beta_i + \alpha_i = 0, \quad A_i \in G_1 \\ & A_iX - b + \beta_i - \alpha_i = 0, \quad A_i \in G_2 \\ & \|X\| = 1, \alpha_i, \beta_i \geq 0 \end{array} \tag{10}$$

This is Glover's and Shi's methods.

Case 2: p =2, q =1

In this case, Model 1 turns to be a convex programming

$$\begin{array}{rl} (convex) & \min w_\alpha \sum_{i=1}^{\infty} \alpha_i^2 - w_\beta \sum_{i=1}^{\infty} \beta_i \\ \text{Subject to} & A_iX - b - \beta_i + \alpha_i = 0, \quad A_i \in G_1 \\ & A_iX - b + \beta_i - \alpha_i = 0, \quad A_i \in G_2 \\ & \|X\| = 1, \alpha_i, \beta_i \geq 0 \end{array} \tag{11}$$

Case 3: p=1, q=2

In this case, Model 1 turns to be a concave programming

$$\begin{array}{rl} (concave) & \min w_\alpha \sum_{i=1}^{\infty} \alpha_i - w_\beta \sum_{i=1}^{\infty} \beta_i^2 \\ \text{Subject to} & A_iX - b - \beta_i + \alpha_i = 0, \quad A_i \in G_1 \\ & A_iX - b + \beta_i - \alpha_i = 0, \quad A_i \in G_2 \\ & \|X\| = 1, \alpha_i, \beta_i \geq 0 \end{array} \tag{12}$$

Case 4: $p = q = 2$

In this case, Model 1 turns to be an indefinite quadratic programming

$$\begin{array}{rl} (indefinite) & \min w_\alpha \sum_{i=1}^{\infty} \alpha_i^2 - w_\beta \sum_{i=1}^{\infty} \beta_i^2 \\ \text{Subject to} & A_iX - b - \beta_i + \alpha_i = 0, \quad A_i \in G_1 \\ & A_iX - b + \beta_i - \alpha_i = 0, \quad A_i \in G_2 \\ & \|X\| = 1, \alpha_i, \beta_i \geq 0 \end{array} \tag{13}$$

These four models have been studied extensively in our recent paper [10]. Numerical experiments show the efficiency and effectiveness of these methods.

5 Conclusion

In this paper, we give a general framework of optimization method for classification and show that SVM, Glover's method and MCLP are special cases of our general model. This demonstrates that the famous SVM, Glover's method and MCLP share the same philosophy. This is interesting because this study gives deeper insights to SVM and Glover's method and Shi's method. Moreover, based on the general model, we can propose some new methods. Because there are many choices of $f(\alpha)$ and $g(\beta)$, maybe we can obtain some more efficient methods for classification based on our general model.

References

[1] Bradley, P.S., Fayyad, U.M., and Mangasarian, O.L. (1999) Mathematical Programming for Data Mining: Formulations and Challenges. *INFORMS Journal on Computing*, 11:217–238.

[2] Chen, S.Y. and Liu X. (2005) Data mining from 1994 to 2004: an application-orientated review, *Int. J. Business Intelligence and Data Mining*, 1:4–21.

[3] Cortes, C. and Vapnik, V. (1995) Support-vector Network. *Machine Learning*, 20:273–297.

[4] Felici, G., Lampariello, F., Rinaldi, G. and Sciandrone M. (2002) Description of OLDAM, Optimization Laboratory for Data Mining, IASI, Italian National Research Council (CNR), (Downloadable from website http://www.iasi.rm.cnr.it/).

[5] Fisher, R.A. (1936) The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, 7:179–188.

[6] Freed, N. and Glover F. (1981) A Linear Programming Approach to the Discriminant Problem. *Decision Sciences* 12:68–79.

[7] Glover, F. (1993) Improved Linear and Integer Programming Models for Discriminate Analysis, in Yuji Ijiri (Ed.) *Creative and Innovative Approaches to the Science of Management*, 365–395.

[8] Mangasarian, O.L. (2001) Data Mining via Support Vector Machines. IFIP Conference on System Modelling and Optimization, Trier, Germany, July 23-27.

[9] Shi, Y. (2001) *Multiple Criteria and Multiple Constraint Levels Linear Programming: Concepts, Techniques and Applications*. World Scientific Pub Co Inc, New Jersey, USA.

[10] Zhang, J., Shi, Y. and Zhang, P. (2005) Several Multi-criteria Programming Methods for Classification, Working Paper.