

Error Correction Method in Classification by Using Multiple-Criteria and Multiple-Constraint Levels Linear Programming

B. Wang, Y. Shi

Bo Wang

School of Mathematical Sciences,
Graduate University of the Chinese Academy of Sciences,
Beijing 100190, China
E-mail: wangbo8014@126.com

Yong Shi (Corresponding author)

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:

In classification based on multiple-criteria linear programming (MCLP), we need to find the optimal solution of the MCLP problem as a classifier. According to dual theory, multiple criteria can be switched to multiple constraint levels, and vice versa. A MCLP problem can be logically extended into a multiple-criteria and multiple-constraint levels linear programming (MC2LP) problem. In many applications, such as credit card account classification, how to handle two types of error is a key issue. The errors can be caused by a fixed cutoff between a "Good" group and a "Bad" group. Two types of error can be systematically corrected by using the structure of MC2LP, which allows two alterable cutoffs. In order to do so, a penalty (or cost) is imposed to find the potential solution for all possible trade-offs in solving MC2LP problem. Some correction strategies can be investigated by the solution procedure. Furthermore, a framework of decision supporting system can be illustrated for various real-life applications of the proposed method.

Keywords: Classification, Two Types of Error, Multiple-Criteria Linear Programming, Multiple-Criteria and Multiple-Constraint Levels Linear Programming, Decision Supporting System.

1 Introduction

Taking advantage of many optimization based classification algorithms, the transactions data collected by the bank can be analyzed in many ways. As a result, the profit of the new accounts can be predicted, that is, which accounts are inclined to be bankrupt and which ones have good credit. However, for many real-life problems, the accounts cannot be separated by the hyperplane in spite of using some kernel techniques [1]. Thus, how to decrease the number of misclassified samples becomes a big issue. In some applications, misclassifications have unequal importance. For example, in credit card account classification, it is essential to classify credit card customers precisely in order to provide effective services while avoiding losses due to bankruptcy from users' debt. Actually, even 0.01 percent increase in early detection of bad accounts can save millions, while losing a good account does not influence much [2], [3].

Linear programming (LP) is a useful tool for discriminating analysis of a problem given appropriate groups (e.g., "Good" and "Bad") [4]. And multiple-criteria linear programming

(MCLP) has improved the result by minimizing the sum external deviations and maximizing the sum of the internal deviations simultaneously. Mostly, the cutoff of MCLP is fixed to be a given number (e.g., 1), while this will cause some other problems. For instance, it cannot involve those possible cases that can achieve the ideal cutoff score to be zero. Formally, this means that the solutions obtained by linear programming are not invariant under linear transformations of the data [2], [3]. Particularly, it is not invariant under vector addition. Furthermore, if the classes of the samples exchange, i.e. "Good" and "Bad" classes swap to each other, the solutions are different.

Noticing these problems, some researchers made many efforts on this topic. Consequently, a new model based on multiple-criteria and multiple-constraint levels linear programming (MC2LP) was posed [5], [6]. However, these methods were domain-driven, which meant they needed some domain knowledge to help finding the best cutoff. Nevertheless, our new models are all automatically solving. As a result, it is useful to solve the problems that no domain knowledge is prepared. In particular, it solves classification problem twice. The maximal external deviation is found for the first time, while MC2LP is exploited to search for the optimal hyperplane based on minimizing two types of error for the second time.

In addition to this, we also provide another MC2LP based model for the purpose of errors correction. In this model, we fix the cutoff to be 1 whereas we also add two more hyperplanes to detect the misclassified points carefully. Accordingly, a subtle discussion is involved regarding the relationship between two types of error and the deviations. In fact, in the statistics theory, two types of error influence on each other oppositely. In other words, reducing of Type I error will cause the increasing of Type II error, and vice versa. Hence, we focus on different types of error respectively. Moreover, some more elaborate introductions are demonstrated in the next sections.

2 Multiple-criteria and Multiple-constraint Levels Linear Programming for Classification

2.1 The MCLP model for classification

Given a set of n variables about the records $X^T = (x_1, x_2, \dots, x_l)$, and then let $x_i = (x_{i1}, x_{i2}, \dots, x_{in})^T$ be one sample of data, where $i = 1, 2, \dots, l$ and l is the sample size. In linear discriminant analysis, data separation can be achieved by two opposite objectives, that is, minimizing the sum of the deviations (MSD) and maximizing the minimum distances (MMD) of observations from the critical value [4]. That is to say, in order to solve classification problem, we need to minimize the overlapping of data, i.e. α , at the same time, to maximize the distances from the well classified points to the hyperplane, i.e. β .

However, it is difficult for traditional linear programming to optimize MMD and MSD simultaneously. According to the concept of Pareto optimality, we can check all the possible trade-offs between the objective functions by using multiple-criteria linear programming algorithm.

The MCLP model can be described by Figure 1:

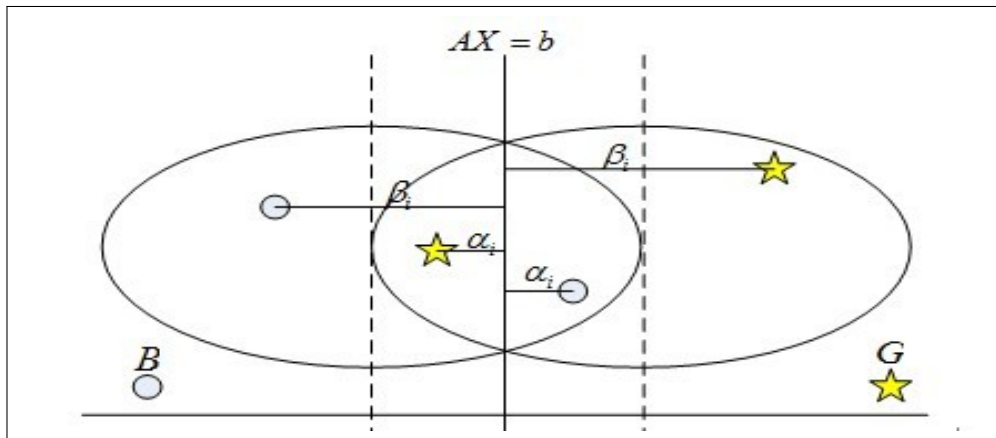


Figure 1: MCLP model

Moreover, the first Multiple Criteria Linear Programming (MCLP) model can be described as follows:

$$\begin{aligned}
 & \min \sum_i \alpha_i \\
 & \max \sum_i \beta_i \\
 \text{s.t. } & A_i X = b + \alpha_i - \beta_i, \quad A_i \in \text{Bad}, \\
 & A_i X = b - \alpha_i + \beta_i, \quad A_i \in \text{Good}, \\
 & \alpha_i, \beta_i \geq 0, i = 1, 2, \dots, l
 \end{aligned}$$

Here, α_i is the overlapping and β_i is the distance from the training sample x_i to the discriminator $(w \cdot x_i) = b$ (classification separating hyperplane).

2.2 Discussion of the cutoff b

A boundary value b (cutoff) is often used to separate two groups, where b is unrestricted. A problem caused by treating b as a variable will meet many cases of no solution. For some applications, the user can choose a fixed value of b ($b = 1$) to get a solution as the classifier. As a result, efforts to improve the accuracy rate of classification have been greatly confined to the unrestricted characteristics of b (that is, a given b is put into calculation to find coefficients w) according to the user's experience facing the real time data set [3].

In such procedure, the goal of finding the optimal solution for classification question is replaced by the task of testing boundary b . That is to say, if b is given, we find a classifier using an optimal solution by solving the model above.

However, now we will point out the drawbacks of handling model in this way. At first, fixing $b = 1$ will make the solutions different under vector addition (one kind of linear transformations). Besides, the solution will change if we swap the classes of "Bad" and "Good", which seems to be illogical. Thus, we will introduce the examples below to illustrate these two issues in the rest of this part.

Example 1. In figure 2, there are 20 points (10 points belong to class -1, 10 points belong to class 1). On the left, there are coordinates of the points. On the right, the distribution of the points is displayed. MCLP is applied to this data set and the solution is $X = (0.58, -0.77)$.

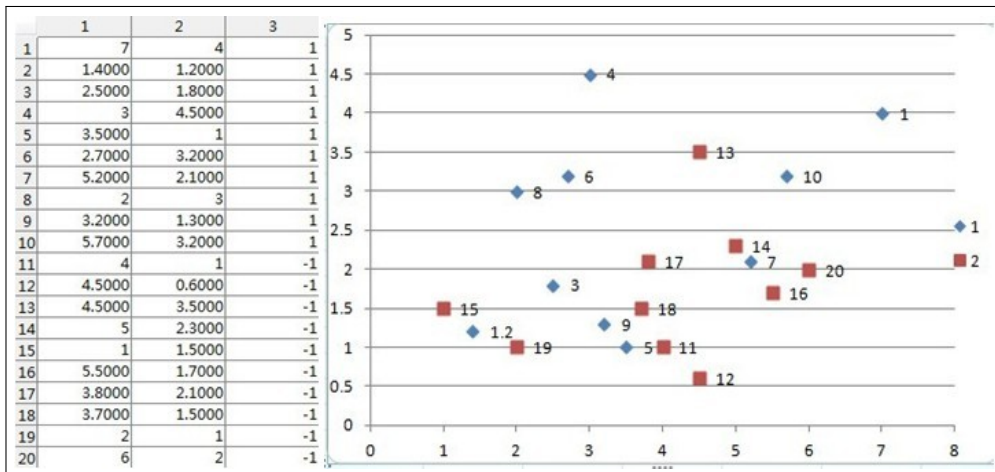


Figure 2: data1

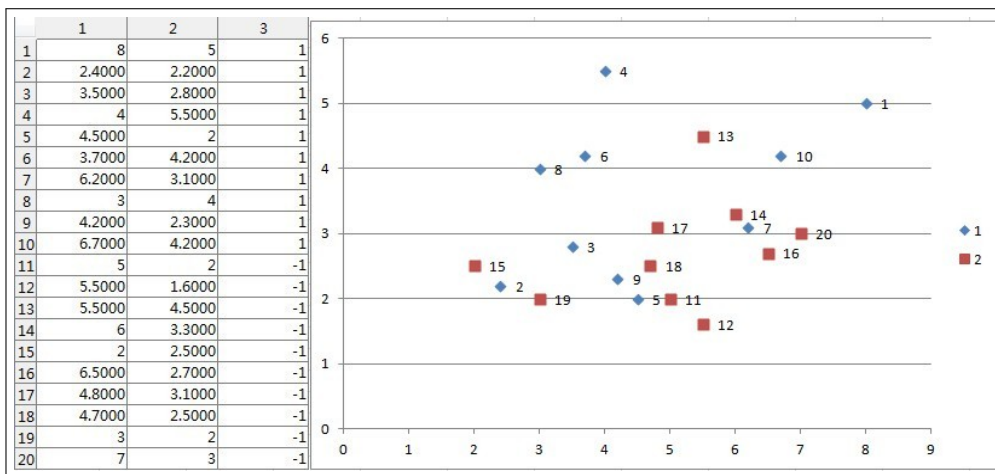


Figure 3: data2

And then, we transform the data by moving the points along the vector $(1, 1)$, and the new data coordinates and the distribution are shown in figure 3. It is obvious that the samples' distribution is the same as the former ones', which means the optimal hyperplane will parallel to the former one. Conversely, the result obtained by MCLP is $X = (0.4, -0.4)$, which is quite different from the former one.

In addition, if the classes are swapped, the result will change into $X = (-0.15, 0.77)$, which should be the same as the first one. For simplicity, in these experiments, the weight between two criteria is always fixed to be $(4, 1)$.

The discussion above shows that the result won't be invariant under linear transformation. This means when the data change, we cannot just keep b to be a fixed number, say 1, and then solve the MCLP problem for different data set. Moreover, we need to take advantage of MC2LP to trace a better b to decrease the number of errors.

2.3 The MC2LP model for classification

According to the discussion above, a non-fixed b is very important to our problem. At the same time, for the simplicity and existence of the solution, b should be fixed in some interval.

As a result, for different data, we fix b in different pairs of interval $[b_l, b_u]$, where b_l and b_u are two fixed numbers. Now our problem is to search the best cutoff between b_l and b_u at every level of their trade-offs, that is to say, to test every point in the interval $[b_l, b_u]$. We keep the multiple-criteria the same as MCLP, that is, MMD and MSD. And then, the following model is posed [3]:

$$\begin{aligned} & \min \sum_i \alpha_i \\ & \max \sum_i \beta_i \\ \text{s.t.} \quad & A_i X = [b_l, b_u] + \alpha_i - \beta_i, \quad A_i \in \text{Bad}, \\ & A_i X = [b_l, b_u] - \alpha_i + \beta_i, \quad A_i \in \text{Good}, \\ & \alpha_i, \beta_i \geq 0, i = 1, 2, \dots, l \end{aligned}$$

where A_i , b_l and b_u are given, and X is unrestricted.

In the model, $[b_l, b_u]$ represents a certain trade-off in the interval. By virtue of the technical of Multiple-criteria and multiple-constraint levels linear programming (MC2LP), we can test each trade-off between the multiple-criteria and multiple-constraint levels as follows:

$$\begin{aligned} & \min \lambda_1 \sum_i \alpha_i - \lambda_2 \sum_i \beta_i \\ \text{s.t.} \quad & A_i X = \gamma_1 b_l + \gamma_2 b_u + \alpha_i - \beta_i, \quad A_i \in \text{Bad}, \\ & A_i X = \gamma_1 b_l + \gamma_2 b_u - \alpha_i + \beta_i, \quad A_i \in \text{Good}, \\ & \alpha_i, \beta_i \geq 0, i = 1, 2, \dots, l \end{aligned}$$

Here, the parameters of $\lambda \times \gamma$ are fixed for each programming problem. Moreover, the advantage of MC2LP is that it can find the potential solutions for all possible trade-offs in the parameter space systematically [7] [8], where the parameter space is

$$\{(\lambda, \gamma) | \lambda_1 + \lambda_2 = 1, \gamma_1 + \gamma_2 = 1\}.$$

Of course, in this model, choosing a suitable pair for the goal problem is a key issue and needs domain knowledge. Consequently, a non-parameter choosing MC2LP method should be posed.

3 A New Two Alterable Cutoffs Model based on MC2LP

3.1 A framework of the new MC2LP model

For the original MCLP model, one cutoff is used to predict a new sample's class, that is to say, there is only one hyperplane. The former MC2LP model points out that we can define two cutoffs instead of the original single cutoff. And then a systematical method can be used to solve this problem. Consequently, all potential solutions at each constrain level trade-off can be acquired. However, one problem is how to find the cutoffs, that is, b_l and b_u .

On one hand, we utilize two cutoffs to discover the solution of higher accuracy; on the other hand, we hope the cutoffs can be obtained from the system directly. Inspired by the idea above, we address our first MC2LP model, which solves the classification problem twice.

For the first step, MCLP model is used to find the vector of external deviations α . It is a function of λ . For simplicity, we set $b = 1$. And then, we fix the parameter of λ to get

one potential solution. Now a non-parameter vector of external deviations α is acquired. The component ($\alpha_i > 0$) means the corresponding sample in the training set is misclassified. In other words, Type I and Type II errors occur. According to the idea of MC2LP, we can detect the result of every single MCLP by fixing the parameter of γ at each level in the interval $[b_l, b_u]$. Now, we find the maximal component of α :

$$\alpha_{max} = \max\{\alpha_i, 1 \leq i \leq l\}. \tag{1}$$

Indeed, the smaller the weight of external deviations is, the bigger α_{max} is.

The misclassified samples are all projected into the interval $[1 - \alpha_{max}, 1 + \alpha_{max}]$ according to the weight vector X obtained from the MCLP model. In this way, we define b_l and b_u as $1 - \alpha_{max}$ and $1 + \alpha_{max}$, respectively. It is easy to see, if we want to lessen the number of two types of error, in effect, we just need to inspect the cutoffs by altering the cutoff in the interval $[1 - \alpha_{max}, 1 + \alpha_{max}]$.

Moreover, for the second step, a new MC2LP classification model can be stated as follows:

$$\begin{aligned} \min \quad & \lambda_1 \sum_i \alpha_i - \lambda_2 \sum_i \beta_i \\ \text{s.t.} \quad & A_i X = [1 - \alpha_{max}, 1 + \alpha_{max}] + \alpha_i - \beta_i, \quad A_i \in \text{Bad}, \\ & A_i X = [1 - \alpha_{max}, 1 + \alpha_{max}] - \alpha_i + \beta_i, \quad A_i \in \text{Good}, \\ & \alpha_i, \beta_i \geq 0, i = 1, 2, \dots, l \end{aligned}$$

where A_i, α_{max} are given, and X is unrestricted, $[1 - \alpha_{max}, 1 + \alpha_{max}]$ means a certain trade-off in the interval. At the same time, $\lambda = (\lambda_1, \lambda_2)$ is the parameter chosen in the first step.

Claim 2. Furthermore, all the notations as $[a, b]$ in the models displayed in this paper mean a certain trade-off in the interval, where a and b are two real numbers.

3.2 Discussion of the new MC2LP model

The most direct modification of the new MC2LP model is to transfer the single objective function to be a multiple-criteria one. Because the vector of external deviations is a function of λ , it is easy to observe that if the weight between external deviations and internal deviations changes, α changes. Consequently, α_{max} alters. And the ideal α is the one that makes α_{max} not too huge. In other words, we do not hope to check the weight that satisfies λ_1 not too small. Actually, some papers have proved that only if $\lambda_1 > \lambda_2$, then $\alpha \cdot \beta = 0$, which makes the model meaningful [9]. As a result, we only need to check the parameters of objective functions that make α_{max} not too big, in short, not too far away from the original one.

On the other hand, we expect α_{max} not too small. That is to say, we hope the model has some generalization. Hence, two small positive numbers ϵ_1 and ϵ_2 are chosen manually. And then, the interval is builded as $[[1 - \alpha_{max} - \epsilon_1, 1 - \alpha_{max} + \epsilon_1], [1 + \alpha_{max} - \epsilon_2, 1 + \alpha_{max} + \epsilon_2]]$. This means that the lower and the upper bound of the interval should be trade-off of some intervals, i.e. the multiple-constrained levels are actually multiple-constrained intervals. Indeed, checking every trade-off of the intervals is the same as checking every trade-off of $1 - \alpha_{max} - \epsilon_1$ and $1 + \alpha_{max} + \epsilon_2$. In this case, we can consider the objective function as a multiple-criteria one. It

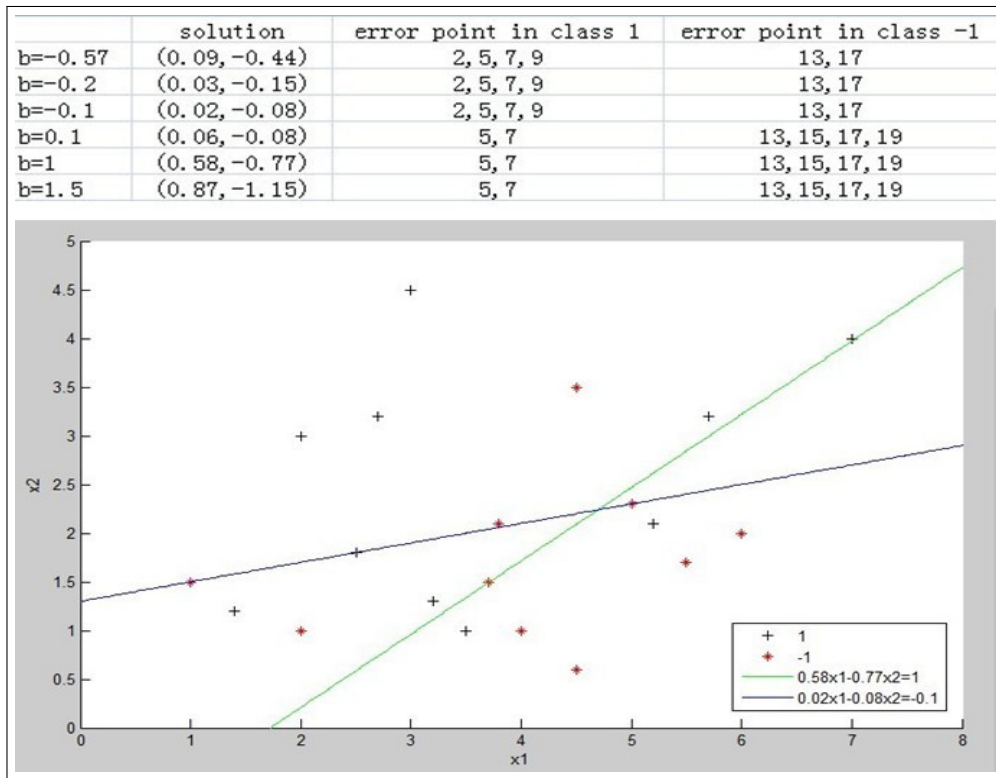


Figure 4: the solutions for MC2LP

can be stated as follows:

$$\begin{aligned}
 & \min \sum_i \alpha_i \\
 & \max \sum_i \beta_i \\
 & \text{s.t. } A_i X = [1 - \alpha_{max} - \epsilon_1, 1 + \alpha_{max} + \epsilon_2] + \alpha_i - \beta_i, \quad A_i \in \text{Bad}, \\
 & \quad A_i X = [1 - \alpha_{max} - \epsilon_1, 1 + \alpha_{max} + \epsilon_2] - \alpha_i + \beta_i, \quad A_i \in \text{Good}, \\
 & \quad \alpha_i, \beta_i \geq 0, i = 1, 2, \dots, l
 \end{aligned} \tag{2}$$

where A_i , α_{max} , ϵ_1 and ϵ_2 are given, and X is unrestricted. Here, ϵ_1 and ϵ_2 are two nonnegative number.

Example 3. A numeric experiment will be introduced to demonstrate that our new MC2LP model is effective. The data is the same as data1 that has been displayed in figure 2. As we have shown above, the solution for MCLP is $X = (0.58, -0.77)$. According to the result, we find there are 6 misclassified points in the data, that is, point 5, 7 for class 1 and point 13, 15, 17, 19 for class -1. Moreover, according to the solution, α_{max} is 1.57. Then, we keep $\epsilon_1 = \epsilon_2 = 0$. As a result, $[b_l, b_u] = [-0.57, 2.57]$ is obtained. Next, some trade-offs between multiple-constraint levels have been chosen and the solutions are listed below. For consistency, we keep the weight between the objective function to be (4,1). The result is laid out in Figure 4. The green hyperplane $0.02 * x_1 - 0.08 * x_2 = 1$ is obtained from negative cutoff. From the solutions, we can see that different trade-offs between the constraint levels yield different results, while the solutions will be the same when b is set to be the same sign. That is to say, for all the positive b , the solutions generate the same hyperplane, and then, for all the negative b , the solutions create the same hyperplane. Moreover, we want to prove this result as a lemma.

Lemma 4. *For certain trade-off between the objective functions, if b is maintained to be the same sign, then hyperplanes, which are obtained in the MCLP model, keep the same. Furthermore, different signs result in different hyperplanes.*

Proof: Let's assume that the trade-off between the objective functions is $\lambda = (\lambda_1, \lambda_2)$ and X_1 is the solution obtained by fixing b to be 1. And then, set b_1 as an arbitrary positive number. The MCLP model can be transformed as follows:

$$\begin{aligned} \min \quad & \lambda_1 \sum_i \alpha_i - \lambda_2 \sum_i \beta_i \\ \text{s.t.} \quad & A_i X = b_1 + \alpha_i - \beta_i, \quad A_i \in \text{Bad}, \\ & A_i X = b_1 - \alpha_i + \beta_i, \quad A_i \in \text{Good}, \\ & \alpha_i, \beta_i \geq 0, \quad i = 1, 2, \dots, l \end{aligned}$$

The problem above is the same as:

$$\begin{aligned} \min \quad & \lambda_1 \frac{\sum_i \alpha_i}{b_1} - \lambda_2 \frac{\sum_i \beta_i}{b_1} \\ \text{s.t.} \quad & A_i \frac{X}{b_1} = 1 + \frac{\alpha_i}{b_1} - \frac{\beta_i}{b_1}, \quad A_i \in \text{Bad}, \\ & A_i \frac{X}{b_1} = 1 - \frac{\alpha_i}{b_1} + \frac{\beta_i}{b_1}, \quad A_i \in \text{Good}, \\ & \alpha_i, \beta_i \geq 0, \quad i = 1, 2, \dots, l \end{aligned}$$

And then, we let $\alpha'_i = \frac{\alpha_i}{b_1}, \beta'_i = \frac{\beta_i}{b_1}, X' = \frac{X}{b_1}$. It is obvious that the solution is $X' = \frac{X_1}{b_1}$ and the hyperplane $AX' = b_1$ is the same as $AX_1 = 1$.

Similarly, we can prove that when b is a negative number, the solution is the same as the one that is obtained from $b = -1$.

As a result, we just need to compare the solutions (hyperplanes) resulted from $b = 1$ and $b = -1$. For this case, it is easy to see that the signs before α_i and β_i swap when we transform $b = -1$ into $b = 1$. If this happens, then the objective function changes into $-\lambda_1 \sum_i \alpha_i + \lambda_2 \sum_i \beta_i$. This means that the solutions will be different. \square

According to the lemma, we have the theorem below:

Theorem 5. *For our MC2LP model (2) above, according to the solutions (hyperplanes), space γ is divided into two non-intersect parts.*

Remark 6. When $[1 - \alpha_{max}, 1 + \alpha_{max}]$ is achieved, ϵ_1 and ϵ_2 are chosen to satisfy that 0 is contained by the interval $[1 - \alpha_{max} - \epsilon_1, 1 + \alpha_{max} + \epsilon_2]$. In this case, for any λ , the solutions belong to the trade-offs with same sign will result in the same hyperplane. In other words, there are only two different hyperplanes corresponding to model (2). In short, the flexibility of model (2) is limited.

4 A New Model based on Correcting of Two Types of Error

In many classification models, including original MCLP model, two types of error is a big issue. In credit card account classification, to correct two types of error can not only improve the accuracy of classification but also help to find some important accounts.

Accordingly, many researchers have focused on this topic. Based on this consideration, more attention should be paid to the samples that locate between two hyperplanes acquired by the original MCLP model, that is, the points in the grey zone [10]. Consequently, we define the external deviations and internal deviations related to two different hyperplanes, the left one and the right one, that is, $\alpha^l, \alpha^r, \beta^l$ and β^r .

Definition 7. The conditions the deviations should satisfy are stated as follows:

$$\alpha_i^l = \begin{cases} 0, & A_i X < 1 - \alpha_{max} \text{ and } A_i \in \text{Bad}; \\ A_i X - (1 - \alpha_{max}), & A_i X \geq 1 - \alpha_{max} \text{ and } A_i \in \text{Bad}; \\ 0, & A_i X \geq 1 - \alpha_{max} \text{ and } A_i \in \text{Good}; \\ (1 - \alpha_{max}) - A_i X, & A_i X < 1 - \alpha_{max} \text{ and } A_i \in \text{Good}. \end{cases}$$

$$\alpha_i^r = \begin{cases} 0, & A_i X < 1 + \alpha_{max} \text{ and } A_i \in \text{Bad}; \\ A_i X - (1 + \alpha_{max}), & A_i X \geq 1 + \alpha_{max} \text{ and } A_i \in \text{Bad}; \\ 0, & A_i X \geq 1 + \alpha_{max} \text{ and } A_i \in \text{Good}; \\ (1 + \alpha_{max}) - A_i X, & A_i X < 1 + \alpha_{max} \text{ and } A_i \in \text{Good}. \end{cases}$$

$$\beta_i^l = \begin{cases} (1 - \alpha_{max}) - A_i X, & A_i X < 1 - \alpha_{max} \text{ and } A_i \in \text{Bad}; \\ 0, & A_i X \geq 1 - \alpha_{max} \text{ and } A_i \in \text{Bad}; \\ A_i X - (1 - \alpha_{max}), & A_i X \geq 1 - \alpha_{max} \text{ and } A_i \in \text{Good}; \\ 0, & A_i X < 1 - \alpha_{max} \text{ and } A_i \in \text{Good}. \end{cases}$$

$$\beta_i^r = \begin{cases} (1 + \alpha_{max}) - A_i X, & A_i X < 1 + \alpha_{max} \text{ and } A_i \in \text{Bad}; \\ 0, & A_i X \geq 1 + \alpha_{max} \text{ and } A_i \in \text{Bad}; \\ A_i X - (1 + \alpha_{max}), & A_i X \geq 1 + \alpha_{max} \text{ and } A_i \in \text{Good}; \\ 0, & A_i X < 1 + \alpha_{max} \text{ and } A_i \in \text{Good}. \end{cases}$$

Figure 5 is a sketch for the model. In the graph, the green and the red lines are the left and right hyperplane, b^l and b^r respectively, which are some trade-offs in two intervals, i.e. $[1 - \alpha_{max} - \epsilon_2, 1]$ and $[1, 1 + \alpha_{max} + \epsilon_1]$. And all the deviations are measured according to them in different colors. For instance, if a sample in "Good" class is misclassified as "Bad" class, it means $\alpha_i^r > \beta_i^l \geq 0$ and $\alpha_i^l = \beta_i^r = 0$. And then, if a sample in "Bad" class is misclassified as "Good" class, it means $\alpha_i^l > \beta_i^r \geq 0$ and $\alpha_i^r = \beta_i^l = 0$. Thus, for the misclassified ones, $\alpha_i^r + \alpha_i^l - \beta_i^r - \beta_i^l$ should be minimized.

As a result, a more meticulous model could be stated as follows:

$$\begin{aligned} & \min \sum_i (\alpha_i^r + \alpha_i^l) \\ & \min \sum_i (\alpha_i^l - \beta_i^r) \\ & \min \sum_i (\alpha_i^r - \beta_i^l) \end{aligned}$$

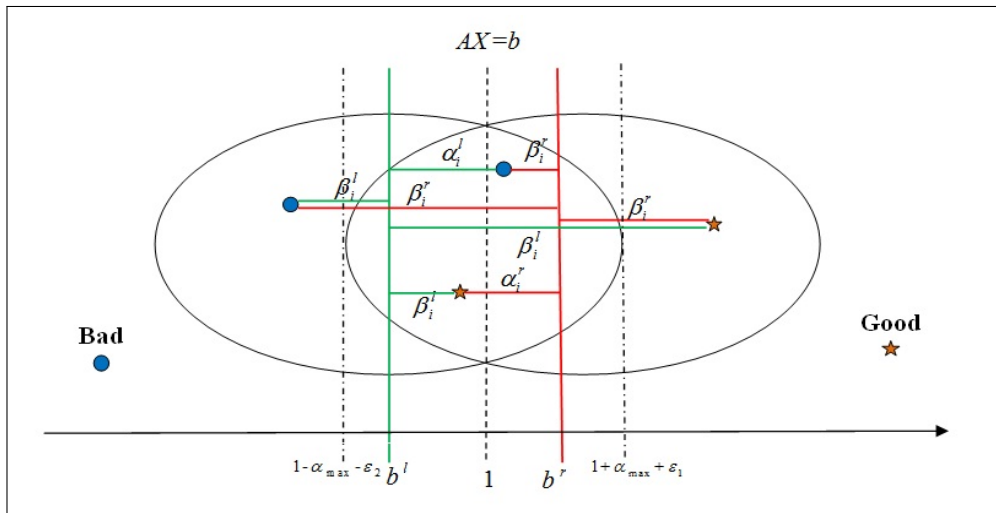


Figure 5: MC2LP model

$$\begin{aligned}
 & \max \sum_i (\beta_i^r + \beta_i^l) \\
 \text{s.t. } & A_i X = 1 + [0, \alpha_{max} + \epsilon_1] + \alpha_i^r - \beta_i^r, \quad A_i \in \text{Bad}, \\
 & A_i X = 1 - [0, \alpha_{max} + \epsilon_2] + \alpha_i^l - \beta_i^l, \quad A_i \in \text{Bad}, \\
 & A_i X = 1 + [0, \alpha_{max} + \epsilon_1] - \alpha_i^r + \beta_i^r, \quad A_i \in \text{Good}, \\
 & A_i X = 1 - [0, \alpha_{max} + \epsilon_2] - \alpha_i^l + \beta_i^l, \quad A_i \in \text{Good}, \\
 & \alpha_i^r, \alpha_i^l, \beta_i^r, \beta_i^l \geq 0, i = 1, 2, \dots, l.
 \end{aligned}$$

where $A_i, \alpha_{max}, \epsilon_1 > 0, \epsilon_2 > 0$ are given, and X is unrestricted.

In Figure 5, for each point, there are at most two kinds of deviations nonzero. The objective functions appear to deal with the deviations according to the position shown in Figure 5, respectively, whereas they have their own special meaning. That is to say, it measures two types of error in some degree by means of the second and third objective functions. As a result, in this new version of MC2LP, we not only consider the deviations respectively, but also take the relationship of the deviations based on two types of error into account in the objective functions. By virtue of MC2LP method, each trade-off between $1 - \alpha_{max} - \epsilon_2$ and 1 for the left hyperplane as well as each trade-off between 1 and $1 + \alpha_{max} + \epsilon_1$ for the right hyperplane can be checked.

After obtaining the weight vector X of the hyperplane, $AX = 1$ is still used to be the classification hyperplane. However, in our new model, we minimize the distance between the left hyperplane and the right one. In other words, we discover the hyperplane that genders the smallest grey area.

Example 8. We show how the new MC2LP model based on error correction works. The data is the same as data1 that has been displayed in figure 2. Two kinds of situations will be considered in this example, that is, the pairs of hyperplanes are $AX = 0.57, AX = 2.57$ and $AX = -6, AX = 10$. We choose ϵ_1, ϵ_2 bigger than 10, so that we can get the pair of hyperplanes $AX = -6, AX = 10$. From the new model, we obtain two hyperplanes, which are displayed in Figure 6. The blue hyperplane is obtained from $AX = 0.57, AX = 2.57$, while the green one is corresponding to $AX = -6, AX = 10$. In both experiments, we fix the weight between α_i^l, α_i^r and β_i^l, β_i^r to be (4,1). According to the result, we can see that the point marked is a misclassified sample corresponding to the former pair, while it is classified correctly based on the latter one.

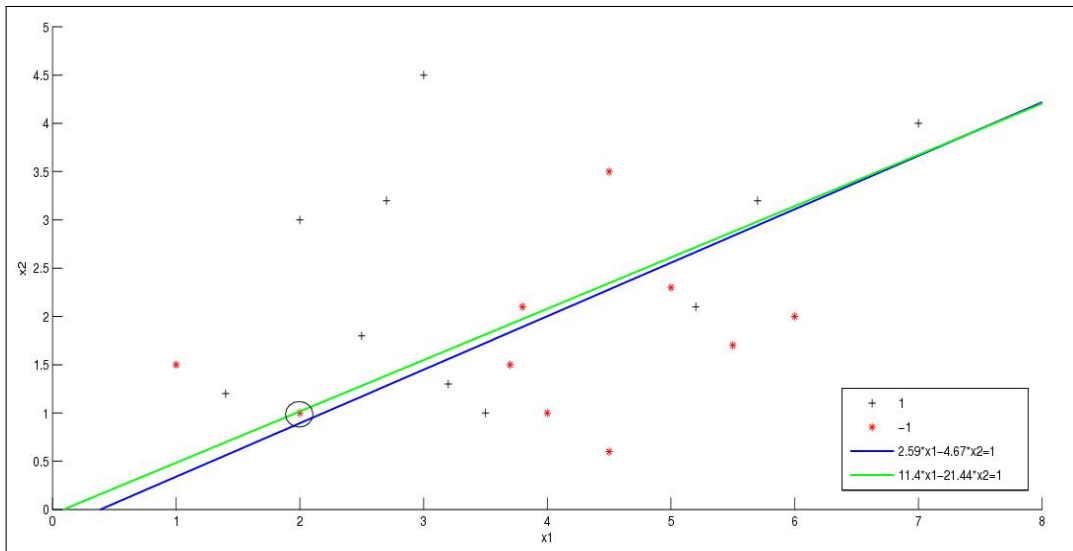


Figure 6: solutions for new MC2LP model

Actually, in statistics, Type I and Type II errors are two opposite objectives. That is to say, it is very hard to correct both of them at the same time. As a result, we modify the former model into two different models focusing on two types of error respectively as follows:

$$\begin{aligned}
 & \min \sum_i (\alpha_i^r + \alpha_i^l) \\
 & \min \sum_i (\alpha_i^l - \beta_i^r) \\
 & \max \sum_i (\beta_i^r + \beta_i^l) \\
 \text{s.t. } & A_i X = 1 + [0, \alpha_{max} + \epsilon] + \alpha_i^r - \beta_i^r, \quad A_i \in \text{Bad}, \\
 & A_i X = 1 + \alpha_i^l - \beta_i^l, \quad A_i \in \text{Bad}, \\
 & A_i X = 1 + [0, \alpha_{max} + \epsilon] - \alpha_i^r + \beta_i^r, \quad A_i \in \text{Good}, \\
 & A_i X = 1 - \alpha_i^l + \beta_i^l, \quad A_i \in \text{Good}, \\
 & \alpha_i^r, \alpha_i^l, \beta_i^r, \beta_i^l \geq 0, i = 1, 2, \dots, l.
 \end{aligned} \tag{3}$$

where A_i , α_{max} and $\epsilon > 0$ are given, and X is unrestricted. In this model, $\sum_i (\alpha_i^r - \beta_i^l)$ is not contained in the objective functions. This model can deal with Type II error, that is, classifying a "Good" point to be a "Bad" one. Now we provide an example to illustrate the effect of model (3).

Example 9. The data set is the same as data1 that has been displayed in figure 2. Here, the label "-1" is regarded as "Good" class, and the label "1" is regarded as "Bad" class. Two kinds of situations will be considered in this example, that is, the pairs of hyperplanes are $AX = 1, AX = 2.57$ and $AX = 1, AX = 8$. The left hyperplane is always kept to be $AX = 1$, which is also the hyperplane that is used to be the classification hyperplane at last. We set ϵ as 10. It is obvious that more remarkable result will be obtained from bigger ϵ . After fixing the weight of the objective functions as $(4, 0.1, 1)$, the outcome is shown in Figure 7. The blue hyperplane is obtained from $AX = 1, AX = 2.57$, while the green one is corresponding to $AX = 1, AX = 8$. Two marked points are the ones corrected by model (3) after the right hyperplane moving to the right.

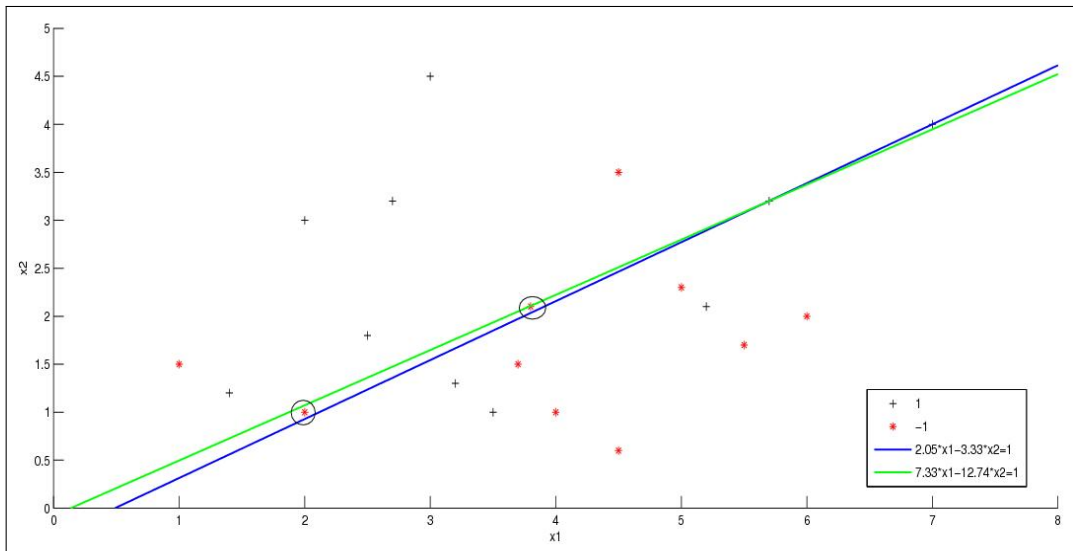


Figure 7: solutions for new MC2LP model (3)

As the result shown above, model (3) can correct Type II error in some degree. We conclude this in the proposition below.

Proposition 10. *Model (3) can correct Type II error by moving the right hyperplane to the right based on the concept of multiple-constraint levels.*

Remark 11. The second objective function in model (3) is nonzero for the samples in class "Bad" and getting negative when the right hyperplane moving to the right. That is to say, we tolerate some Type I errors. At the same time, the first objective function in model (3) renders Type II errors an increasing punishment with moving the right hyperplane to the right. As a result, it can correct Type II error in some degree.

Similarly to model (3), we pose model (4) which can deal with Type I error as follows:

$$\begin{aligned}
 & \min \sum_i (\alpha_i^r + \alpha_i^l) \\
 & \min \sum_i (\alpha_i^r - \beta_i^l) \\
 & \max \sum_i (\beta_i^r + \beta_i^l) \\
 \text{s.t. } & A_i X = 1 + \alpha_i^r - \beta_i^r, & A_i \in \text{Bad}, \\
 & A_i X = 1 - [0, \alpha_{max} + \epsilon] + \alpha_i^l - \beta_i^l, & A_i \in \text{Bad}, \\
 & A_i X = 1 - \alpha_i^r + \beta_i^r, & A_i \in \text{Good}, \\
 & A_i X = 1 - [0, \alpha_{max} + \epsilon] - \alpha_i^l + \beta_i^l, & A_i \in \text{Good}, \\
 & \alpha_i^r, \alpha_i^l, \beta_i^r, \beta_i^l \geq 0, i = 1, 2, \dots, l.
 \end{aligned} \tag{4}$$

where A_i , α_{max} and $\epsilon > 0$ are given, and X is unrestricted. In this model, $\sum_i (\alpha_i^l - \beta_i^r)$ is not contained in the objective functions. This model focuses on Type I error, that is, classifying a "Bad" point to be a "Good" one.

Example 12. *In order to compare model (3) and model (4), we draw two hyperplanes obtained from these two models in the same graph. Figure 8 is the result of using pair $AX = 1, AX = 8$ for model (3) and pair $AX = -0.57, AX = 1$ for model (4).*

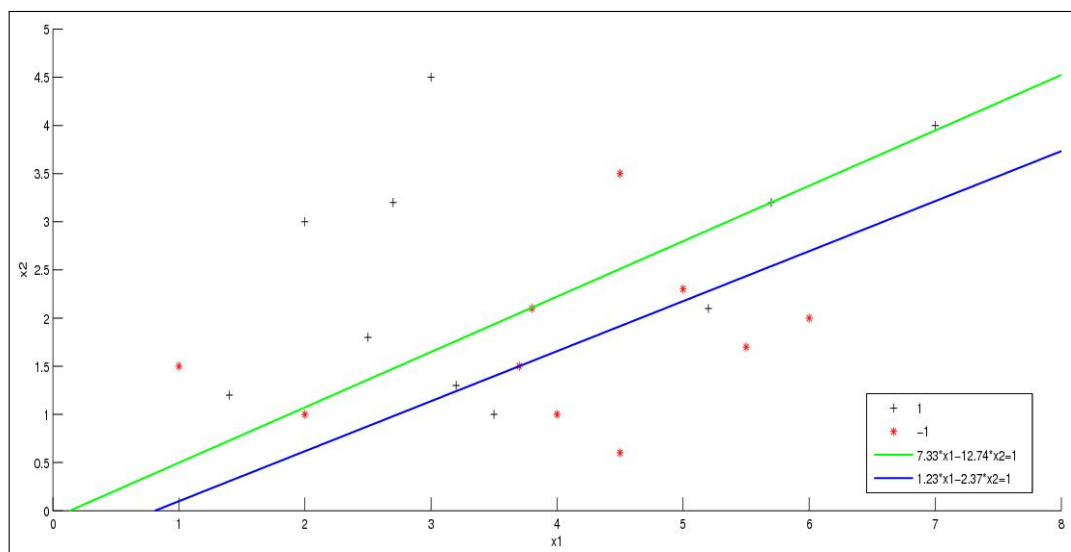


Figure 8: comparison of model (3) (4)

In figure 8, the green hyperplane is obtained from model (3) with less Type II errors, while the blue one is acquired from model (4) with less Type I errors.

In the last two models introduced above, the first one defines $AX = 1$ as the left hyperplane, at the same time, $AX = 1 + [0, \alpha_{max} + \epsilon]$ as the right one. Oppositely, the second model defines $AX = 1 - [0, \alpha_{max} + \epsilon]$ as the left hyperplane, and then, $AX = 1$ as the right one. For each trade-off, our model can deal with the misclassifications. In this way, two types of error can be corrected respectively.

5 Conclusion

This paper focuses on how to correct two types of error. At first, the disadvantage of MCLP by fixing the cutoff is addressed. And then, a MC2LP based model is introduced. According to the topic of error correction, we develop MC2LP by measuring the deviations of the misclassified points more sophisticatedly. In this way, we obtain a brand new hyperplane that is different from the one obtained by MCLP model. Moreover, focusing on correcting two types of error, a result with less errors is gained.

Acknowledgement

This work has been partially supported by grants from National Natural Science Foundation of China (No. 70921061, No. 71110107026) and the CAS/SAFEA International Partnership Program for Creative Research Teams.

Bibliography

- [1] Zhang Z., Zhang D., Tian Y., Shi Y., *Kernel-based Multiple Criteria Linear Programming Classifier*, *Procedia CS* 1(1): 2407-2415, 2010.
- [2] Thomas L.C., Edelman D.B., Crook J.N., *Credit Scoring and Its Applications*, SIAM, 2002.

- [3] He J., Zhang Y., Shi Y., Huang G., *Domain-Driven Classification Based on Multiple Criteria and Multiple Constraint-Level Programming for Intelligent Credit Scoring*, IEEE Transactions on Knowledge and Data Engineering, vol. 22, No. 6: 826-838, 2010.
- [4] Freed N., Glover F., *Simple but Powerful Goal Programming Models for Discriminant Problems*, European J. Operational Research, vol. 7: 44-60, 1981.
- [5] Shi Y., Tian Y., Kou G., Peng Y., Li J., *Optimization Based Data Mining: Theory and Applications, Advanced Information and Knowledge Processing*, Springer, 2011.
- [6] Shi Y., *Multiple Criteria Optimization-based Data Mining Methods and Applications: A Systematic Survey*, Knowl Inf Syst, 24: 369-391, 2010.
- [7] Shi Y., *Multiple Criteria and Multiple Constraint Levels Linear Programming: Concepts, Techniques and Applications*, World Scientific, 2001.
- [8] Shi Y., He J., Wang L., Fan W., *Computer-based Algorithms for Multiple Criteria and Multiple Constraint Level Integer Linear Programming*, Comput. Math. Appl. 49(5): 903-921, 2005.
- [9] Nakayama H., Yun Y., *Generating Support Vector Machines using Multiobjective Optimization and Goal Programming*, Studies in Computational Intelligence, vol. 16: 173-198, 2006.
- [10] Chen Y., Zhang L., Shi Y., *Post Mining of Multiple Criteria Linear Programming Classification Model for Actionable Knowledge in Credit Card Churning Management*, ICDMW, IEEE Computer Society: 204-211, 2011.