DISCRIMINANT ANALYSIS USING A BICRITERIA LINEAR PROGRAM

Yi–Hsin Liu and John Maloney

University of Nebraska at Omaha

Abstract. In the past decade, there has been some interest shown in the use of linear programs to solve discriminant analysis problems. In this paper we present and clarify an alternative to the statistical approach for discriminating between two (or more) groups of vector–valued data in which an order relation is given.

By the nature of this problem, a bicriteria linear program can be used instead of a statistical approach and hence no statistical assumptions need to be imposed on the problem.

1. Introduction. Classical discriminant analysis is a multivariate statistical technique concerned with

(1) separating distinct sets of objectives and

(2) allocating new objectives to previously defined groups [4].

In the past decade, this problem has been studied using linear programs [1, 2, 3]. However, in many cases, a "complete separation" using a linear discriminant function is impossible. Therefore, by the nature of the problem a bicriteria linear programming approach is preferred.

The bicriteria linear program contains the following two objective criteria:

(1) maximize the total separation and

(2) minimize the total misclassification

Of course, (2) above can be omitted if it is known that the groups are separable. This will be discussed further in the later sections.

A linear discriminant analysis problem seeks a vector c which is used to construct a linear discriminant function $f(x) = cx$. This linear discriminant function is then used to separate the given groups of vector–valued data G_1, G_2, …, G_m and provides an allocation rule for placing future unclassified data into one of the groups.

In this paper we assume that a subjective ranking (order relation) has been imposed on the groups G_1, G_2, ..., G_m. That is, for any two distinct groups of data G_i and G_j either G_i is preferred to G_j or G_j is preferred to G_i. Without loss of generality, we may assume that G_i is preferred to G_j whenever $i > j$. This order relation is denoted by writing $G_j \prec G_i$ if $i > j$. Thus, we are given

$$G_1 \prec G_2 \prec \cdots \prec G_m$$

This assumption arises in many problems [5] and it is possible to use, in some cases, artificial intelligence techniques to determine the subjective rankings discussed above. This will not be discussed in this paper, rather we will assume that the rankings have been given.

We now formally state the problem.

Problem 1.1. Given m groups of vector–valued data (that is values in E^n) such that
(1) $G_1 \prec G_2 \prec \cdots \prec G_m$
(2) $G_i = \{x^i_j \in E^n : j = 1, 2, \ldots, l_i\}$ where $i = 1, 2, \ldots, m$.
find a vector c (and hence a linear discriminant function $f(x) = cx$), and the appropriate intervals $I_i = (L_i, U_i]$, such that

i. $I_i \cap I_k = \emptyset$, $1 \le i, k \le n$, $i \ne k$.
ii. $f(x^i_j) \in I_i$, $j = 1, 2, \ldots, l_i$ $i = 1, 2, \ldots, m$.
iii. $L_1 < U_1 \le L_2 < U_2 \le \cdots \le L_m < U_m$

Definition 1.2. The groups G_1, G_2, ..., G_m are said to be separable if there exists a linear function $f(x) = cx$ such that $f(x^i_j) \in I_i$ for all $j = 1, 2, \ldots, l_i$ and for all $i = 1, 2, \ldots, m$ provided that (iii) above holds. Otherwise the groups are said to be nonseparable.

2. Models and Discussion. First, we solve a simple basic mode, L. P. 2.1, to follow, which assumes that the groups are separable. We will then generalize the basic model to attack the case where the groups are not separable.

L. P. 2.1. Maximize

$$\sum_{c=1}^{m} \alpha_{ii} + \sum_{i=1}^{m-1} \alpha_{i+1,i} \tag{1}$$

subject to the constraints

$$(2) \qquad L_i \le cx_j^i \le U_i \qquad i = 1, 2, \ldots, m, \quad j = 1, 2, \ldots, l_i$$

$$(3) \qquad U_i - L_i = \alpha_{ii} \qquad i = 1, 2, \ldots, m$$

$$(4) \qquad L_{i+1} - U_i = \alpha_{i+1,i} \qquad i = 1, 2, \ldots, m-1$$

$$(5) \qquad 0 \le \alpha_{ii}, \alpha_{i+1,i} \le K$$

L. P. 2.1 finds the vector c and the boundaries of the intervals U_i, L_i while the x_j^i's are given data and K is chosen so as to bound the unknown quantites to be found.

The assumption of separability guarantees that $f(x_j^i) \in (L_i, U_i]$ and the remaining constraints guarantee that

$$L_1 < U_1 \le L_2 < U_2 \le \cdots \le L_m < U_m \ .$$

The last constraint in problem 2.1 guarantees that the solution will be bounded. Without it the model is unbounded as can be seen from (2) or from the dual linear program of L. P. 2.1. In fact, the upper bound K can be chosen to be any real number, since by (2) one can see that in the discriminant function $f(x) = cx$, that c is a vector of relative weights. Or, if one considers the geometry, it will be the normal of the hyperplane defined by $f(x) = cx = z^*$, for some real number z^*.

Let c^* and $(L_i^*, U_i^*]$ be an optimal solution found from L. P. 2.1. Then the allocation rule for a new vector $\overline{x}$ states that if $f(\overline{x}) = c^*\overline{x} \in J_i$ then $\overline{x} \in G_i$, where

$$J_i = \left(\frac{L_i^* + U_{i-1}^*}{2}, \frac{U_i^* + L_{i+1}^*}{2} \right] \ .$$

Then $J_i \cap J_j = \emptyset$ if $i \ne j \quad \forall \quad 1 \le i, j \le m$.

Next, we consider the nonseparable case. This case is a generalization of the separable case. In the nonseparable case, we assume that

$$G_1 \prec G_2 \prec \cdots \prec G_m$$

and

$$L_1 < U_1 \le L_2 < U_2 \le \cdots \le L_m < U_m \ .$$

To find $f(x) = cx$ and L_i,U_i for all i, however, we note that the condition

$$f(x^i_j) = cx^i_j \in I_i$$

may very well be false for some of the $x^i_j \in G_i$. Thus for each $x^i_j \in G_i$, one of the following three cases must hold

a. $L_i \leq cx^i_j \leq U_i$
b. $cx^i_j \leq L_i$
c. $cx^i_j \geq U_i$

The above three cases can be combined into

$$L_i - \beta^i_j \leq cx^i_j \leq U_i + \gamma^i_j \tag{6}$$

where β^i_j and γ^i_j both are nonnegative quantities and can be considered as the "distance" away from the interval, or the "error" due to misclassification. When the groups are separable we have that $\beta^i_j = \gamma^i_j = 0 \quad \forall \quad i, j$. Based on this, we develop the following bicriteria linear program

B. C. L. P. 2.2. Maximize

$$\sum_{i=1}^{m} \alpha_{ii} + \sum_{i=1}^{m-1} \alpha_{i+1,i} \tag{7}$$

and minimize

$$\sum_{i=1}^{m} \sum_{j=1}^{l_i} \beta^i_j + \sum_{i=1}^{m} \sum_{j=1}^{l_i} \gamma^i_j \tag{8}$$

subject to the constraints (2), (3), (4), (5), of L. P. 2.1 and

$$0 \leq \beta^i_j \leq K, \qquad 0 \leq \gamma^i_j \leq K \tag{9}$$

In the above (7) maximizes the total interval length as in (1), while (8) minimizes the total misclassification. This last problem simultaneously optimizes both goals of the discriminant analysis problem.

To solve the bicriteria linear program one can use the weighted linear program to obtain an optimal solution for B. C. L. P. 2.2. By the scalarization theorem of the multicritieria linear program [6] and the nature of the problem, one can then obtain the following weighted linear program with equal weights for both objectives as in the problem.

L. P. 2.3. Maximize

$$\sum_{i=1}^{m} \alpha_{ii} + \sum_{i=1}^{m-1} \alpha_{i+1,i} - \left(\sum_{i=1}^{m} \sum_{j=1}^{l_i} \beta_j^i + \sum_{i=1}^{m} \sum_{j=1}^{l_i} \gamma_j^i \right) \tag{10}$$

subject to the constraints (2), (3), (4), (5), (9).

An optimal solution of L. P. 2.3 c^*, $\beta_j^{i^*}$, $\gamma_j^{i^*}$, U_i^*, L_i^* then can be used to define a linear discriminant function $f(x) = c^* x$ and the allocation rule then would be the same as the separable case. When the optimal solution has $\beta_j^i = \gamma_j^i = 0$ for all i, j, such that $1 \leq j \leq l_i$ and $1 \leq i \leq m$, then we have the separable case.

3. Conclusions and Examples. L. P. 2.3, in fact, is a general formulation of L. P. 2.1. In general, among groups G_i, $i = 1, 2, \ldots, m$ with $G_1 \prec G_2 \prec \cdots \prec G_m$, it is very difficult to detect whether the groups are separable or nonseparable. So for any given problem one should always apply problem 2.3 to solve the discriminant analysis problem.

An optimal solution of L. P. 2.3 then shows whether the groups are separable or nonseparable and determines a linear discriminant function. Also one can then obtain an allocation rule for the future data allocation.

Example 3.1. Given the data shown in table 1, find a discriminant function.

G_1	G_2	G_3
$x_1^1 = (2, 1)$	$x_1^2 = (2, 3)$	$x_1^3 = (4, 8)$
$x_2^1 = (3, 1)$	$x_2^2 = (3, 4)$	$x_2^3 = (4, 9)$
$x_3^1 = (3, 2)$	$x_3^2 = (3, 5)$	$x_3^3 = (5, 9)$
$x_4^1 = (5, 3)$	$x_4^2 = (6, 6)$	$x_4^3 = (5, 10)$
$x_5^1 = (6, 2)$	$x_5^2 = (7, 8)$	$x_5^3 = (7, 10)$
	$x_6^2 = (8, 7)$	
	$x_7^2 = (1, 4)$	

Table 1

Let the order relation be given by $G_1 \prec G_2 \prec G_3$. Applying L. P. 2.3 and (6) we obtain the discriminant function $f(x) = -33.3x_1 + 53.3x_2$ and the intervals $J_1 = (-\infty, 50]$, $J_2 = (50, 243.33]$ and $J_3 = (243.33, \infty]$. The allocation rule for new data $x = (x_1, x_2)$ will be

i. $x \in G_1$ if $-33.3x_1 + 53.3x_2 \in J_1$.

ii. $x \in G_2$ if $-33.3x_1 + 53.3x_2 \in J_2$.

iii. $x \in G_3$ if $-33.3x_1 + 53.3x_2 \in J_3$.

Figure 1

References

1. T. M. Cavalier, J. P. Ignizio, and A. L. Soyster, "Discriminant Analysis via Mathematical Programming: Certain Problems and Their Causes," *Computer and Operations Research*, 16 (1989), 353–362.

2. N. Freed and F. Glover, "A Linear Programming Approach to the Discriminant Problem," *Decision Sciences*, 12 (1981), 67–84.

3. N. Freed and F. Glover, "Simple But Powerful Goal Programming Models for Discriminant Problems," *European Journal of Operational Research*, (1989), 353–362.

4. R. A. Johnson and D. W. Wichen, *Applied Multivariate Statistical Analysis*, Prentice Hall, Englewood Cliffs, N.J., (1988).

5. Y–H. Liu and Z. Chen, "Rule Base Assisted Consistence Checking for Decision Making," *Proceedings of the Twentieth Annual Model and Simulation Conference*, May 1989.

6. J. Philip, "Algorithm for the Vector Maximization Problem," *Mathematical Programming*, 2 (1972), 207–229.

7. P. A. Rubin, "Evaluating the Maximum Distance Formulation of the Discriminant Problem," *European Journal of Operational Research*, 41 (1989), 240–248.