USING A NONLINEAR DISCRIMINANT FUNCTION FOR SOLVING DISCRIMINANT ANALYSIS PROBLEMS

John Maloney and Yi-Hsin Liu

University of Nebraska, Omaha

Abstract. To solve a discriminant analysis problem, linear programming has been useful for finding a linear discriminant function. In this paper we present a mathematical programming approach to find a nonlinear discriminant function. It is believed that a nonlinear discriminant function can be more useful than a linear one.

1. Introduction. A discriminant analysis problem seeks a discriminant function $f(x)$ which is used to separate the given groups of vector-valued data $G_1, G_2, \ldots, G_m$ and provides an allocation rule for placing future unclassified data into one of the groups.

In this paper we construct a nonlinear discriminant function using a linear program. In fact, we construct a general polynomial, of degree n, in the vector x to be used as a discriminant function. Clearly, a nonlinear discriminant function is more effective than a linear one, since a linear function can be treated as a special case ($n = 1$) in the mentioned polynomial.

In this paper, we assume that a subjective ranking (order relation) has been imposed on the groups $G_1, G_2, \ldots, G_m$. 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 and it is possible to use, in some cases, artificial intelligence techniques to determine the subjective rankings discussed above in [5]. This will not be discussed in this paper, rather we will assume that the rankings have been given.

The problem to be considered is stated as follows.

Problem 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) $G_i = \{x_j^i \in E^n : j = 1, 2, \ldots, l_i\}$ where $i = 1, 2, \ldots, m$,

find a discriminant function $f(x)$, and the appropriate intervals $I_i = (L_i, U_i]$, such that

i. $I_i \cap I_k = \emptyset$, $\forall\ 1 \le i, k \le n$, $i \ne k$.

ii. $f(x_j^i) \in I_i$, $\forall\ j = 1, 2, \ldots, l_i$ and $\forall\ i = 1, 2, \ldots, m$.

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

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

2. Model and Discussion. In the following $f(x)$, $x = (x_1, x_2, \ldots, x_n)$ is restricted to the following form

$$f(x) = \sum_{j=0}^{n} a_{i_1 i_2 \cdots i_n} x_1^{i_1} x_2^{i_2} \cdots x_n^{i_n} \quad \text{where} \quad j = i_1 + i_2 + \cdots i_n.$$

Determining the degree of the polynomial can be difficult, since there are no general rules for doing this. It depends on the number of groups considered in the problem and the number of data (observations) in each class. However, trial and error can provide good information in determining n, the order of the discriminant polynomial (function). The quantities $\overline{x}_i$ in the objective function below are defined by

$$\overline{x}_i = \frac{\sum_{j=1}^{l_i} x_j^i}{l_i}.$$

We may think of the $\overline{x}_i$'s as being the center of mass of l_i unit masses located at x_j^i. This then becomes the center of mass of the group G_i. For this reason we shall refer to the points $\overline{x}_i$ as centroids.

The model used is

Q. P. 3. Maximize

$$\sum_{i=1}^{m} \sum_{j=i+1}^{m} \left| f(\overline{x}_i) - f(\overline{x}_j) \right| + 2K - \sum_{i=1}^{m} (U_i - L_i)$$

subject to the constraints

$$i. \quad -K \leq L_i \leq f(x_j^i) \leq U_i \leq K \quad i = 1, 2, \ldots, m, \quad j = 1, 2, \ldots, l_i \ .$$

$$ii. \quad 2K - \sum_{i=1}^{m} (U_i - L_i) \geq 0 \ .$$

We are finding the coefficient of the discriminant function, $a_{i_1 i_2 \cdots i_n}$, that will give maximum separation to the values $f(\overline{x}_i)$, and thus, lead to the most widely separated set of disjoint intervals I_j. We also obtain the intervals $[L_i, U_i)$ to be of minimum length in the sense that $f(x_j^i) = L_i$ for some j and $f(x_j^i) = U_i$ for some (different) value of j.

The first term in the objective function is maximized by giving maximum separation to the centroids. The second term in the objective function is maximized by making the intervals $[L_i, U_i]$ as small as possible subject to the constraint (i). One can, later, properly adjust the interval lengths to handle the future classification of new data.

Thus maximizing our objective function gives us the two properties that we sought above. K is a properly selected constant that guarantees a bounded solution for the intervals. In the example that follows a value of 100 was used, a larger or smaller value might well be suitable for a different problem.

The advantage of this form of the problem is that it determines what might be called a natural order for the intervals $[L_i, U_i]$ and removes any dependence upon the order relation prescribed for the G_i in the original problem.

3. Example. In this example the discriminant function is a third order polynomial in the two coordinates x and y that has the form

$$f(x) = a_{30}x^3 + a_{21}x^2y + a_{12}xy^2 + a_{03}y^3 + a_{20}x^2 + a_{11}xy + a_{02}y^2 + a_{10}x + a_{01}y + a_{00}.$$

The data set used in this example is given below. The model was solved on Gino [5].

Data Set 4. The data is

i. $G_1 = \{(1,3),(2,3),(2,4),(3,3),(4,2)\}$ $\quad \overline{x}_1 = (12/5, 3)$.

ii. $G_2 = \{(3,4),(5,4),(4,5),(5,5)\}$ $\quad \overline{x}_2 = (17/4, 9/2)$.

iii. $G_3 = \{(2,1),(3,1),(2,2),(3,2)\}$ $\quad \overline{x}_3 = (5/2, 3/2)$.

The model results are presented in the table below.

interval	*item*
$[-11.012358, -9.749569]$	$[L_1, U_1]$
$[-100.0, -69.838026]$	$[L_2, U_2]$
$[61.543902, 100.0]$	$[L_3, U_3]$

Table 4

Separable Data

The constants for the solution polynomial are

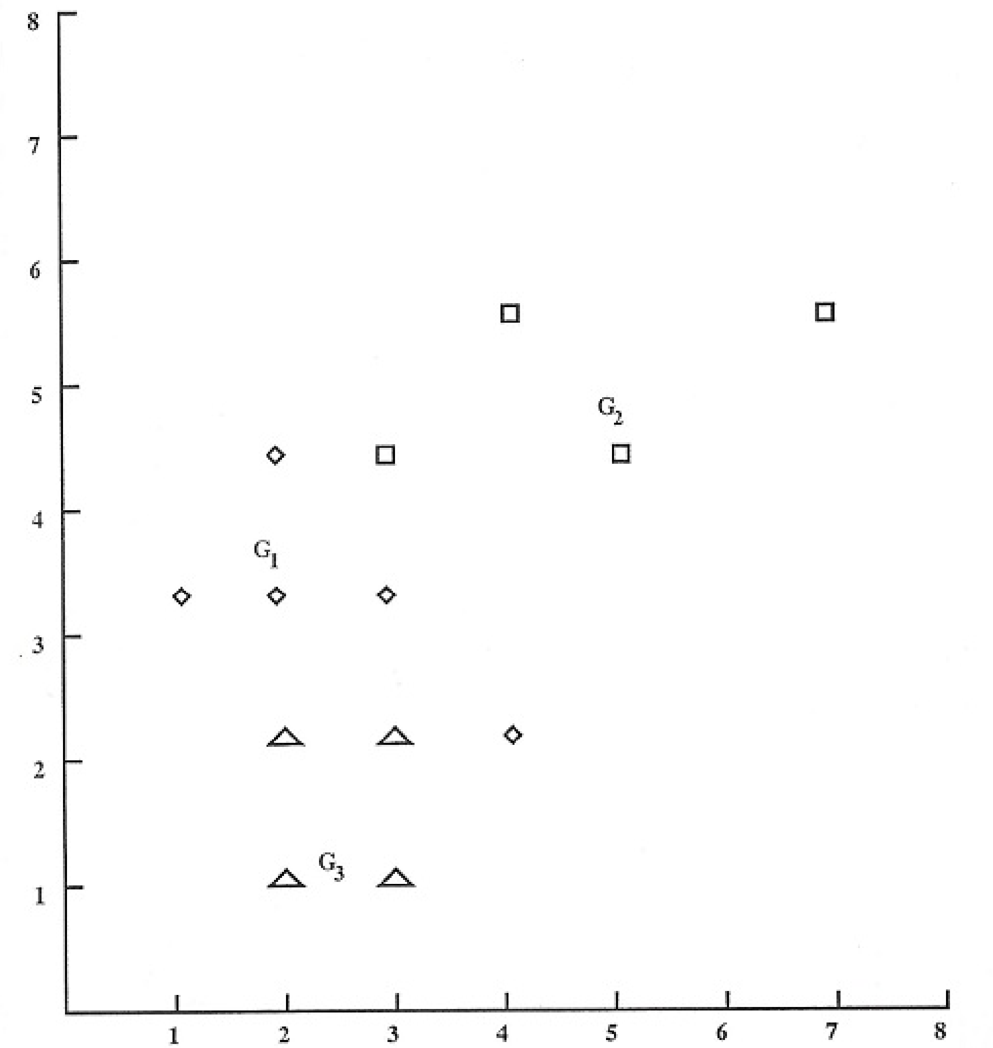
constant	*value*
a_{30}	-4.159692
a_{21}	-26.017895
a_{12}	30.588933
a_{03}	17.776081
a_{20}	-66.177253
a_{11}	-29.645071
a_{02}	-71.670812
a_{10}	274.182681
a_{01}	145.165613
a_{00}	-252.679563

Table 5

Solution Function Constants

In the following figure we show the data.

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. J. Liebman, L. Lasdon, L. Schrage, and A. Waren, *GINO*, The Scientific Press, San Francisco, California, 1986.
6. 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.
7. Y-H. Liu and J. Maloney, "Discriminant Analysis Using a Bicriteria Linear Program," *Missouri Journal of Mathematical Sciences*, 4 (1992), 61–69.
8. J. Maloney and Y-H. Liu, "A Quadratic Programming Approach for Solving Discriminant Analysis Problems," *Missouri Journal of Mathematical Sciences*, 4 (1992), 139-144.
9. P. A. Rubin, "Evaluating the Maximum Distance Formulation of the Discriminant Problem," *European Journal of Operational Research*, 41 (1989), 240–248.