Efficient Global Programming Model for Discriminant Analysis

M.ANGULAKSHMI,Assitant Professor
Department of Information Technology and Engineering
VIT UNIVERSITY
Vellore,India

Abstract— **Conventional statistical analysis includes the capacity to systematically assign individuals to groups. We suggest alternative assignment procedures, utilizing a set of interrelated goal programming formulations. This paper represents an effort to suggest ways by which the discriminant problem might reasonably be addressed via straightforward linear goal programming formulations. Simple and direct, such formulations may ultimately compete with conventional approaches - free of the classical assumptions and possessing a stronger intuitive appeal. We further demonstrate via simple illustration the potential of these procedures to play a significant part in addressing the discriminant problem, and indicate fundamental ideas that lay the foundation for other more sophisticated approaches.**

Keywords- Credit scoring, mathematical programming, multiple criteria and multiple constraint-level programming, Discriminant Analysis.

1. Introduction

. Included in the role of conventional statistical procedures is the capacity to systematically assign individuals to groups. Such a capability is assured widespread application: assigning patients to a disease,loan applicants to a risk category, potential product purchasers . Yet while considerable effort has been made to generate appropriate classification techniques grounded firmly in the principles of classical and Bayesian statistics, little has been done to explore the potential of alternative management science approaches to the problems of group discrimination.

. We suggest alternative assignment procedures, utilizing a set of interrelated goal programming formulations. Importantly, Discriminant analysis enables the user to play an active part in the analysis, encouraging user participation in the selection of appropriate discriminant criteria and allowing flexibility in setting relative penalties for misclassification.

2 Cluster vs Discriminant Analysis

Cluster vs. discriminan**t** analysis While the task and importance of assigning individuals to groups is easily understood, it should be noted that two sets of assignment-related procedures **-** those classed as clustering techniques and those identified with standard statistical discrimination are often confused (a situation further complicated by varying tetminology among authors) .Accordingly, the following descriptions are offered for clarification:

Cluster analysis encompasses those procedures Which romote the formation of readily identifiable groupings of 'similar' objects. Thus, for example, a clustering procedure might be used to group human diseases, product lines, archeological artifacts. The process begins with a standard data structure in which a number of cases (objects, individuals, tems) have been measured on a number of dimensions (properties, characteristics, traits). Cases, initially ungrouped, are ultimately clustered (grouped) according to some criterion of proximity (and hence, similarity).

Discriminant analysis also addresses the need to distinguish groups of cases, but here appropriate groupings are defined *prior* to application of the technique. That is, a sample of members (cases) from each of a number of known groups is given. For each case, measurements are taken on a set of dimensions (variables). A discriminant procedure is used to mathematically combine variables into a single dimension that will 'best' differentiate the groups. That combination of variables can then be used to (1) establish the relative importance of the original dimensions in separating group members, and (2) assign new cases with unknown group membership to an appropriate group. Issues generally associated with the *discriminant* task, as described above, serve as the principal focus of this paper.

2.1. Purpose

Our goal is to provide a simpler alternative to conventional dicriminant procedures where by 'simpler' we mean easier to understand and manipulate (due to increased flexibility). It should be stressed that we are not undertaking a thoroughgoing critique of classical methods, nor suggesting that they are not useful. Rather, emphasis is placed on disclosing the positive aspects of proposed options.

Efforts to cast discriminant-type problems in linear goal programming form derive from a recognition that such problems are inherently problems in constrained optimization: that is, problems in which some well-def'ined objective (goal) is to be maximized (minimized), subject to a set of constraining conditions. Given this perception, the task is to identify effective goats and appropriate constraints. While nonlinear formulations are clearly possible, linearity serves to promote conceptual simplicity and ensures a fair degree of computational efficiency. More complex extensions of the essential theme are left to another place.

2.2. Related research

To date, efforts to promote the application of LP-based techniques to typically statistical problems have been largely restricted to L, norm and constrained regression procedures in which variants of the goal programming formulation first outlined by Charnes, Cooper and Ferguson [4] have been advanced as attractive alternatives to the conventional least squares approach. In such procedures, the standard goal of producing a set of *squared deviations* is replaced by the task of producing coefficients which minimize a sum of *absolute deviations.* Beyond these regression-related applications, the extension of basic LP techniques to common problems has been quite modest. Kendall [12], for example, suggests a convex hull method for discriminating group membership, which he ultimately rejects as too cumbersome, insufficiently general and lacking the capability to measure the relative importance of discriminant variables. Rao [14] offers an interesting set of linear and non-linear integer programming formulations for a class of clustering problems, but observes that such formulations appear extremely difficult to solve with existing computational procedures. While not wholly successful, such efforts do suggest the potential of alternative perspectives on problem types generally conceded to conventional statistics.

2.3. Producing a single linear discriminator for the multi-group discriminant problem

The basic problem initially to be addressed may be briefly described as follows. Group membership for a set of p-dimensional points is known. A simple weighting scheme is sought to 'score' each p-dimensional point by weighting its components. The scores will be divided into intervals designed to insure, insofar as possible, proper group assignment. By extension, the scoring (weighting) scheme may then be applied to additional points in the space in order to determine likely group membership and, significantly, should provide insight into the relative importance of dimensions in segregating groups.

Let the task of assigning credit applicants to risk classifications serve as a simple example. An applicant is to be classified as a 'poor', 'fair', or 'good' credit risk based on his/her responses to two questions appearing on a standard credit application. Previous experience with 12 customers produced the data shown in Table 1 and displayed graphically in Fig. 1.A simple weighting scheme (linear transformation) will be produced to score the 12 customer-points so that they can be appropriately classified upon subdividing the scores into intervals.

The problem can now be recast more formally: Given points A_i and sets G_i find the linear transformation X, and the appropriate boundaries (interval subdivisions) b^L_j and b^U_j, to 'properly' categorize each A_i. Thus the task is to determine a linear predictor or weighting scheme X and breakpoints b^L_j and b^U_j, and such that

$$b_j^L \leqslant A_k X \leqslant b_j^u \Leftrightarrow A_k \in G_j \,, \qquad (1)$$

$$b_1^L < b_1^u < b_2^L < b_2^u < \cdots < b_g^u \,. \qquad (2)$$

Alternative 1. Determine a predictor X such that:

$$b_j^u < b_{j+1}^L + \alpha_j \quad \text{for } j = 1, \ldots, g-1 \,,$$

For all A_i € G_i. and, to ensure that (2) is achieved *as nearly as possible,*impose as goal constrants :

b^U_{j}, <. $b^L_{j+1} + \alpha_i$ = 1 .... , g -- 1 ,

where g = number of designated groups setting as the objective
Minimize $\sum c_j\alpha_j$.

Accordingly, the task of assigning credit applicants to risk classifications *is* here cast as a linear goal programming (GP) problem. Removing non-negativity constraints from the interval bounds, b^L_j and b^U_j, this formulation yields the basic constraint set:

Group 1

$$1X_1 + 3X_2 \geqslant b_1^L, \qquad 1X_1 + 3X_2 \leqslant b_1^u,$$
$$2X_1 + 5X_2 \geqslant b_1^L, \qquad 2X_1 + 5X_2 \leqslant b_1^u,$$
$$3X_1 + 4X_2 \geqslant b_1^L, \qquad 3X_1 + 4X_2 \leqslant b_1^u,$$
$$4X_1 + 6X_2 \geqslant b_1^L, \qquad 4X_1 + 6X_2 \leqslant b_1^u,$$

Group2

$$5X_1 + 7X_2 \geqslant b_2^L, \qquad 5X_1 + 7X_2 \leqslant b_2^u,$$
$$6X_1 + 9X_2 \geqslant b_2^L, \qquad 6X_1 + 9X_2 \leqslant b_2^u,$$
$$7X_1 + 8X_2 \geqslant b_2^L, \qquad 7X_1 + 8X_2 \leqslant b_2^u,$$
$$7X_1 + 7X_2 \geqslant b_2^L, \qquad 7X_1 + 7X_2 \leqslant b_2^u,$$
$$9X_1 + 9X_2 \geqslant b_2^L \qquad 9X_1 + 9X_2 \leqslant b_2^u,$$

Group3

$$6X_1 + 2X_2 \geqslant b_3^L, \qquad 6X_1 + 2X_2 \leqslant b_3^u,$$
$$6X_1 + 4X_2 \geqslant b_3^L, \qquad 6X_1 + 4X_2 \leqslant b_3^u,$$
$$8X_1 + 3X_2 \geqslant b_3^L, \qquad 8X_1 + 3X_2 \leqslant b_3^u.$$

Adding the boundary sequencing constraints

$$b_1^u + 1 \leqslant b_2^L + \alpha_1$$
$$b_2^u + 1 \leqslant b_3^L + \alpha_2$$

and, to preclude the trivial null solution, X = O, the normalization XI >1 completes the set.

Table 1

	Responses				
	Credit Customer	Quest 1 (a_1)	Quest 2 (a_2)	Transformed Score Using $X = \begin{vmatrix} 1 \\ -4 \end{vmatrix}$	
GROUP I (Poor Risk)	1	1	3	-11	
	2	2	5	-18	$b_1^L = -20$
	3	3	4	-13	$b_1^u = -11$
	4	4	6	-20	
GROUP II (Fair Risk)	5	5	7	-23	
	6	6	9	-30	$b_2^L = -21$
	7	7	8	-25	
	8	7	7	-21	$b_2^u = -30$
	9	9	9	-27	
GROUP III (Good Risk)	10	6	2	-2	
	11	6	4	-10	$b_3^L = -10$
	12	8	3	-4	$b_3^u = 0$

Weighting equally (which measure group verlap) produces a predictor, $X =[^{1}{}_{4}]$ Fig. 1 clearly demonstrates the ability of this transformation to segregate members of the three risk classes. Here it may be useful to ensure that 'poor' credit risks are generally assigned lower scores by the transformation vector than scores computed for 'fair' or 'good' risks**.** Further, the formulation can effectively accommodate the need to assign differential costs for misclassification. To impose such a condition on the problem, differing weights are assigned to the overlap variables in the objective function. Maintaining the ordering specification outlined above, and arbitrarily assigning a weight of 5 to Group II Group III overlap and weight of one to I-II overlap creates a transformation vector, $X = [^{3}{}_{2}]$, which forces the overlap back to Groups I and II.(Fig 2)

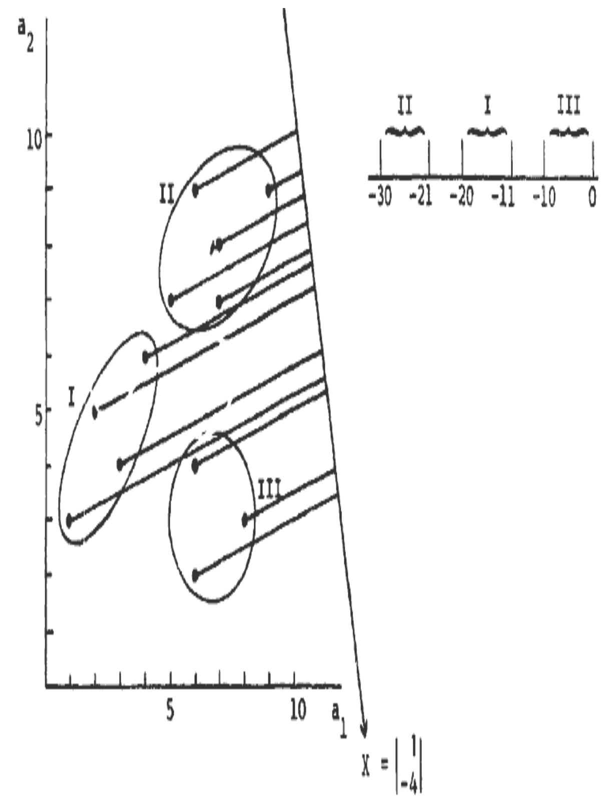

FIG 1 : Graphical representation of three risk classes

An alternative formulation may now be considered.

***Alternative* 2**. This formulation would impose boundary separation as a common constraint, setting as a goal the inclusion of points within appropriate bounds. Thus, $b_j^u \leqslant b_{j+1}^L$ for $j = 1, \ldots, g - 1$, where g = number of designated groups.

$A_iX \geqslant b_j^L - \alpha_j$, $A_iX \leqslant b_j^u + \alpha_j$, For all A_i€ G_i as the objective Minimize

TABLE2

	Credit Customer	Responses: Quest 1 (a_1)	Responses: Quest 2 (a_2)	Transformed Scores Using $X = \begin{vmatrix} 3 \\ -2 \end{vmatrix}$	
GROUP I (Poor Risk)	1	1	3	-3	
	2	2	5	-4	$b_1^L = -4$
	3	3	4	1	$b_1^u = 1$
	4	4	6	0	
GROUP II (Fair Risk)	5	5	7	1	
	6	6	9	0	$b_2^L = 0$
	7	7	8	5	
	8	7	7	7	$b_2^u = 9$
	9	9	9	9	
GROUP III (Good Risk)	10	6	2	14	
	11	6	4	10	$b_3^L = 10$
	12	8	3	18	$b_3^u = 18$

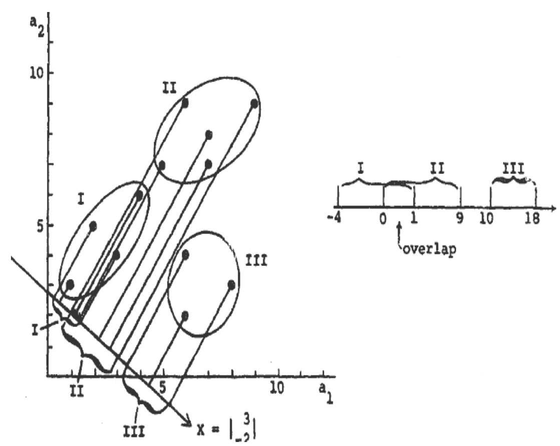

FIG 2 Graphical representation of three risk classes

2.4 The pair-wise discriminant problem

To see the power of these formulation ideas more
cearly, consider now the 'following extension. Where.as the development thus far has undertaken to produce a single suitable weighting scheme by which data points can be transformed and aggregated, it is apparent that in many cases such a 'one-dimensional' approach may prove too restrictive to provide adequate group discrimination.

Given two groups, G and B, determine an appropriate vector X and boundary value b such that, as nearly as possible,

$$(1) \quad A_iX \leqslant b, \quad A_i \in G_1,$$

$$A_iX \geqslant b, \quad A_i \in G_2.$$

Introducing α_i to measure the degree to which group members A_i violate the two-group boundary, we thus seek to insure a solution in which:

$$A_iX \leqslant b + \alpha_i, \quad A_i \in G_1$$

$$A_iX \geqslant b - \alpha_i, \quad A_i \in G_2$$

and the sum of boundary violations α_i. While it will generally not be possible to anticipate which points will lie within the 'true' boundary,it is clear that all points will lie within the 'adjusted' boundaries.

. To accurately measure the separation of G and B, we defined four parameters for the criteria and constraints as follows:

- .α_i: the overlapping of two-class boundary for case A_i (external measurement);
- α: the maximum overlapping of two-class boundary for all cases Ai ($\alpha i < \alpha$);
- . βi: the distance of case Ai from its adjusted boundary (internal measurement); and
- β: the minimum distance for all cases Ai from its adjusted boundary ($\beta i > \beta$).To achieve the separation, the Sum of the Deviations (MSD) of the observations is minimized. The second separates the observations by Maximizing the Minimal Distances of observations from the critical value. This deviation is also called "overlapping". A simple version of Freed and Glover's [19] model which seeks MSD can be written as:

$$\text{(M1)} \quad \text{Min} \quad \Sigma_i \alpha_i, \quad s.t. \quad A_i X \le b + \alpha_i, A_i \in B \quad (1) \quad A_i X > b - \alpha_i, A_i \in G,$$

where Ai, b are given, X is unrestricted and $\alpha i >= 0$.

The alternative of the above model is to find MMD:

$$\text{(M2)} \quad \text{Max} \quad \Sigma_i \beta_i, \quad s.t. \quad A_i X \ge b - \beta_i, A_i \in B \quad (2) \quad A_i X < b + \beta_i, A_i \in G,$$

where Ai, b are given, X is unrestricted and $\beta i >= 0$.A graphical representation of these models in terms of α and β is shown in Fig. 3 We note that the key point of the two-class linear classification model is to use a linear combination of the minimization of the sum of αi or maximization of the sum of βi. The advantage of this conversion is that it allows easy utilization of all techniques of LP for separation, while the disadvantage is that it may miss the scenario of trade-offs between these two separation criteria

A hybrid model (M3) in the format of multiple criteria linear programming (MC) that combines models of (M1) and (M2)

$$\text{(M3)} \quad \begin{array}{l} \text{Minimize} \quad \Sigma_i \alpha_i, \\ \text{Maximize} \quad \Sigma_i \beta_i, \\ A_i X = b + \alpha_i - \beta_i, \quad A_i \in B, \\ A_i X = b - \alpha_i + \beta_i, \quad A_i \in G, \end{array} \quad (3)$$

Where A_i ,b are given, X is unrestricted, $\alpha_i >=0$ and $\beta_i >=0$.

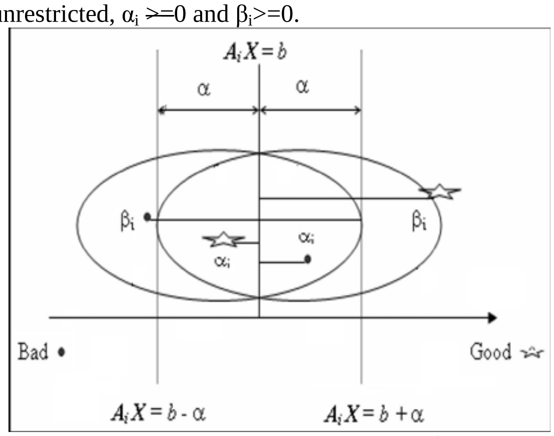

Fig. 3 Overlapping case in two-class separation of LP model

3. Conclusion

The assumption-free GP procedure offers a simple and direct approach to the discriminant problem. Although a full evaluation of the proposed goal programming formulations must await detailed testing,the technique holds significant promise. The flexibility of these forms and their ability to handle side conditions make them a potentially desirable alternative to standard statistical methods.

4. References

[1] V.G. Ashar and T.D. Wallace, A sampling study of minimum absolute deviation estimators , operations Res. It (1963) 747-758.

[2] M. Blum, Fairing disctiminant analysis, I. Accounting Res. 12 (1974) 1-25.
[3] A. Charnes and W.W. Cooper, Management Models and Industrial Appfication of Linear Programming, Vol. 1 (Wiley, New York, 1961).
[4] A. Charnes, W.W. Cooper and R.O. Ferguson, Optimal estimation of executive compensation by linear programming, Manag. Sci. I (1955) 138-151.
[5] A. Charnes and W.W. Cooper, Goal programming and multiple objective optimizations, European I. Operational Res. 1 (1977) 39-54. [6| M.R. Crasic and W.D. Perreault, Validation of discriminant analysis in marketing research, J. Marketing Res. 14 (1977) 60-68.
[6] F.A. Eisenbeis and R.B. Avery, Two aspects of investigating group differences in linear discriminant analysis, Decision Sci. 4 (1973) 487-493.
[7] N. Freed and F. Glover, Discriminant analysis by the separating/goal hyperplane method, Management and Information Science Report Series, University of Colorado, to appear.
[8] F.R. Ghhe, The small sample properties of simultaneous equations least absolute estimators vis-a-vis least squares estimators, Econometrics 38 (1970) 742-753.
[9] G. Hunt et al., L I Estimation in small samples with Laplace error distributions, Decision Sci. 5 (1974) 22-29.
[10] O.J. Karst, Linear curve fitting using least deviations, J. Amer. Statistic. Assoc. 53 (1958) 118-132.
[11] M.G. Kendall, Discrimination and classification, in: R. Krishna (Ed.), Multivariate Analysis (Academic Press, New York, 1966).
[12] E.A. Kiountouzis, Linear programming techniques in regression analysis, Appl. Statist. 1 (1973)799--821
[13] M.R. Rao, Cluster analysis and mathematical programming, J. Amer. Statistic. Assoc. 6(1971) 622-626.
[14] T.R. Rao, Is Brand loyalty a criterion for market segmentation: Discriminant analysis, Decision Sci. 4 (1973) 398-404.
[15] J.F. Sinkey, Multivariate statistical analysis of the characteristics of problem banks, J. Finance 30 (1975) 21-26.
[16] K.H. Usow, On L 1 approximation: computation for discrete functions and discretization effects, SIAM J. Numer. Anal. 4 (1967) 233-244.