Simple but powerful goal programming models for discriminant problems

Ned FREED and Fred GLOVER

Graduate School of Business Administration, University of Colorado, Boulder, CO 80309, U.S.A.

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. 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.

1. Overview

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 to a market segment, and the like (see [3,10,16 and 20]). 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.

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. Importantly, they enable 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

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 terminology among authors). Accordingly, the following descriptions are offered for clarification:

Cluster analysis encompasses those procedures which promote 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, or religious customs. The process begins with a standard data structure in which a number of cases (objects, individuals, items) 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. However, a number of the ideas advanced have application to the clustering problem as well.

2.1. Purpose

Our goal is to provide a simpler alternative to conventional discriminant 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-defined objective (goal) is to be maximized (minimized), subject to a set of constraining conditions. Given this perception, the task is to identify effective goals and appropriate constraints. While non-linear 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_1 norm and constrained regression procedures in which variants of the goal programming formulation first outlined by Charnes, Cooper and Ferguson [4] have been advanced [11,13,19] 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 impor-

Table 1

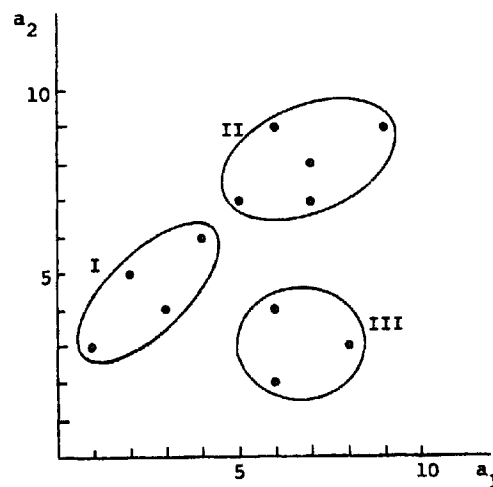|	Credit Customer	Responses	
		Quest 1 (a_1)	Quest 2 (a_2)
GROUP I (Poor Risk)	1	1	3
	2	2	5
	3	3	4
	4	4	6
GROUP II (Fair Risk)	5	5	7
	6	6	9
	7	7	8
	8	7	7
	9	9	9
GROUP III (Good Risk)	10	6	2
	11	6	4
	12	8	3

Fig. 1.

tance 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_j, find the linear transformation X, and the appropriate boundaries (interval subdivisions) b_j^L and b_j^u, to 'properly' categorize each A_i. (Bounds b_j^L and b_j^u represent respectively the lower and upper boundaries for points assigned to group j.) Thus the task is to determine a linear predictor or weighting scheme X and breakpoints b_j^L and b_j^u, such that

$$b_j^L \leqslant A_k X \leqslant b_j^u \Leftrightarrow A_k \in G_j , \tag{1}$$

and

$$b_1^L < b_1^u < b_2^L < b_2^u < \cdots < b_g^u . \tag{2}$$

The points A_i may of course be distributed in a way that makes complete group differentiation impossible (e.g., when the attributes of some credit applicants defy ready classification by risk category). Therefore, it becomes important to endow the weighting scheme with the power to establish the foregoing group differentiation with minimum exception. Two useful and direct LP formulations to accomplish such a goal are suggested below. The basic themes inherent in these forms will subsequently be extended to more complex applications.

Alternative 1. Determine a predictor X such that:

$$A_i X \geqslant b_j^L , \qquad A_i X \leqslant b_j^u \tag{3}$$

for all $A_i \in G_j$

and, to ensure that (2) is achieved *as nearly as possible*, impose as goal constraints [1]:

$$b_j^u < b_{j+1}^L + \alpha_j \quad \text{for } j = 1, ..., g - 1 , \tag{4}$$

where g = number of designated groups setting as the

[1] To maintain a strict inequality, the effective constraints here are: $b_j^u + \epsilon \leqslant b_{j+1}^L + \alpha_j$, where $\epsilon > 0$. For the examples that follow, $\epsilon = 1$.

objective

$$\text{Minimize } \Sigma \, c_j \alpha_j . \tag{5}$$

Importantly, the boundary constraints in (4) designate a specific ordering of discriminant scores, requiring, for example, that Group 1 scores be generally lower than Group 2 scores, Group 2 scores be generally lower than Group 3 scores, and so on. This particular sequencing may prove too restrictive to produce the most effective discriminant solution. Consequently, it may be necessary to examine the procedure for other sequencing possibilities. (The solution values of the α_j variables can suggest which alternative sequencings are worth exploring, or all possible sequencings may be implicitly explored by integer programming. Such embellishments, however, are obviated by a subsequent discussion.)

The objective function c_j's are weights which may reflect the relative importance of 'correct' assignment to a particular group, i.e., these weights should represent, proportionately, misclassification costs. Such an objective will serve to segregate the sets if possible, and otherwise narrow the range of overlap.

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_j^L and b_j^u, this formulation yields the basic constraint set:

Group 1

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

Group 2

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

Group 3

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

Table 2

	Credit Customer	Responses Quest 1 (a_1)	Quest 2 (a_2)	Transformed Scores Using $X = \left\lvert \begin{smallmatrix} 1 \\ -4 \end{smallmatrix} \right\rvert$	
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	$b_3^u = 0$
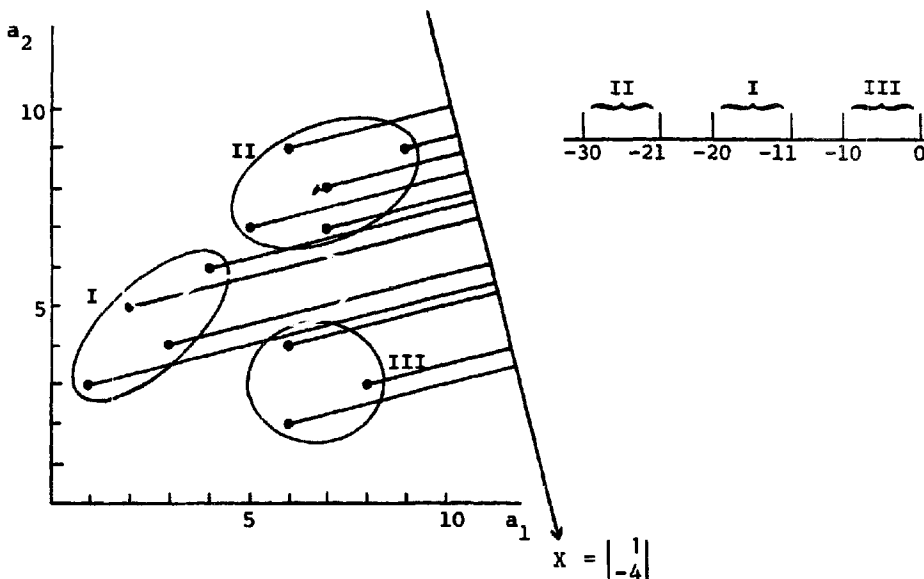
Fig. 2.

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, $X = 0$, the normalization $X_1 \geqslant 1$ completes the set.

Weighting equally the α_j's (which measure group overlap) produces a predictor, $X = \left[\begin{smallmatrix} 1 \\ -4 \end{smallmatrix} \right]$, Fig. 2 clearly demonstrates the ability of this transformation to segregate members of the three risk classes.

Each of the 12 sample customers has been classified correctly. Table 2 summarizes the results of our exercise.

Some additional observations appear in order. The formulation described here provides a significant degree of flexibility. It is possible, for example, to exploit the capacity of the procedure to specify the ordering of transformed group scores. Here it may be useful to ensure that 'poor' credit risks are generally assigned lower scores by the transformation vector

Table 3

	Credit Customer	Responses Quest 1 (a_1)	Quest 2 (a_2)	Transformed Scores Using $X = \left|{}^{\,2}_{-1}\right|$	
GROUP I (Poor Risk)	1	1	3	-1	
	2	2	5	-1	$b_1^L = -1$
	3	3	4	2	$b_1^u = 2$
	4	4	6	2	
GROUP II (Fair Risk)	5	5	7	3	
	6	6	9	3	
	7	7	8	6	$b_2^L = 3$
	8	7	7	7	$b_2^u = 9$
	9	9	9	9	
GROUP III (Good Risk)	10	6	2	10	$b_3^L = 8$
	11	6	4	8	
	12	8	3	13	$b_3^u = 13$
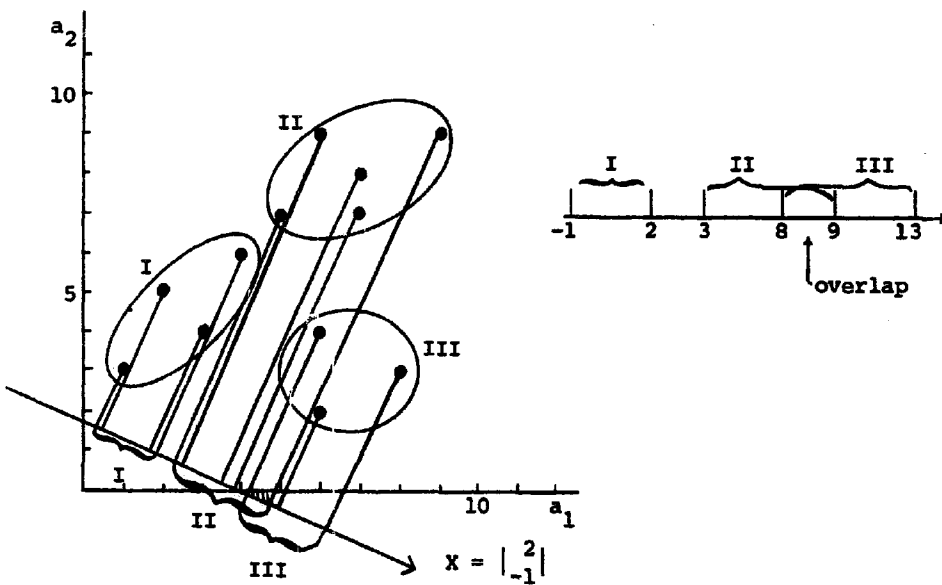
Fig. 3.

than scores computed for 'fair' or 'good' risks. Such a requirement in our example produces an 'optimal' transformation, $X = \left[{}^{\,2}_{-1}\right]$, resulting in a slight overlap between Group II and III (Fig. 3).

Further, the formulation can effectively accommodate the need to assign differential costs for misclassification. It may be determined, for instance, that the penalty for confusing 'fair' and 'good' credit risks (i.e., Group II and Group III members) is significantly greater than that associated with confusing 'poor'

and 'fair' risks. To impose such a condition on the problem, differing weights are assigned to the overlap variables (α's) in the objective function. Maintaining the ordering specification outlined above, and arbitrarily assigning a weight of 5 to Group II—Group III overlap and a weight of one to I—II overlap creates a transformation vector, $X = \left[{}^{\,3}_{-2}\right]$, which forces the overlap back to Groups I and II (Fig. 4). Added constraints or modified objectives may be readily introduced.

Table 4

		Responses		Transformed Scores
	Credit Customer	Quest 1 (a_1)	Quest 2 (a_2)	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 $b_3^u = 18$
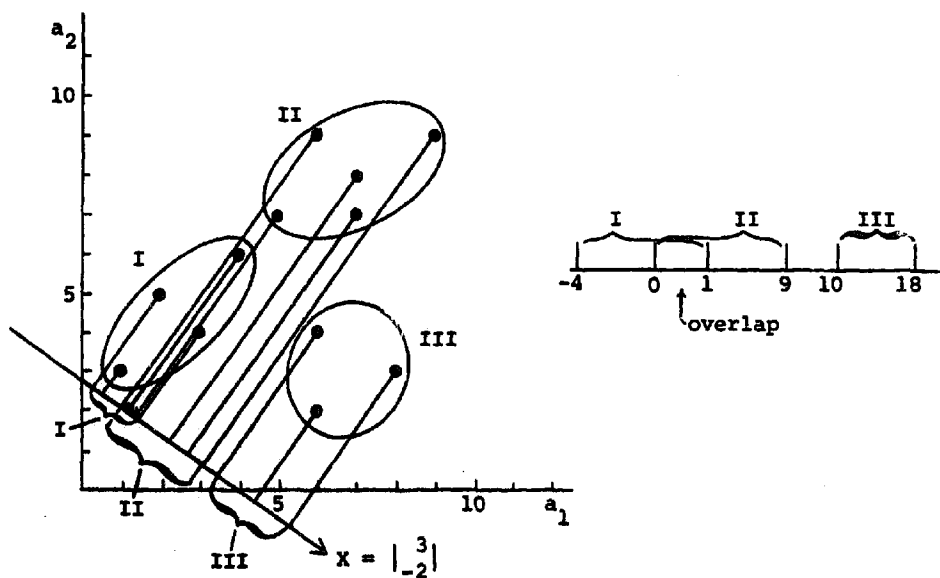
Fig. 4.

As indicated above, the capacity of the linear discriminator to effectively determine group membership can be readily evaluated. A simple count of misclassified points should provide an important empirical basis for judging the discriminant power of any candidate vector. (Computation of a chi-square statistic can be useful in providing added insight into the 'goodness of separation' produced by a given weighting scheme.) Such an indicator may be effec-tively supplemented by considering the magnitude of group overlap, α_j, as a proportion of overall interval width (i.e., $\alpha_j/(b_j^u - b_j^L)$) for each of the groups involved.

Finally, the proposed GP formulation can accom-modate fairly large problems. Setting n = number of points to be classified, g = number of designated groups and d = number of dimensions, the number of constraints required is simply $2n + g$; the number of

variables [2] is equal to $2d + 5g - 1$. Thus, for example, a problem involving 3 groups of 200 observations each, in 10 dimensions, would produce a GP with 1203 constraints and 34 variables — a problem of reasonably modest size. In an extension of this methodology, discussed fully in the next section, much larger discriminant problems can be cast as GPs of manageable dimension.

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,

$$b_j^u \leqslant b_{j+1}^L \qquad \text{for } j = 1, ..., g - 1 , \tag{6}$$

where g = number of designated groups.

$$A_i X \geqslant b_j^L - \alpha_j , \qquad A_i X \leqslant b_j^u + \alpha_j , \tag{7}$$

for all $A_i \in G_j$,

with the objective

Minimize $\sum c_j \alpha_j$.

Here again, weights c_j are assigned to reflect relative costs of misclassification [3].

2.4. The pair-wise discriminant problem

To see the power of these formulation ideas more clearly, consider now the following extension. Whereas 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.

Consider for example the data presented in Table 5 and graphically displayed in the succeeding Fig. 5. Here a single transformation would appear incapable of satisfactorily segregating the preassigned members of each of three distinct groups. (In fact, actual appli-

Table 5

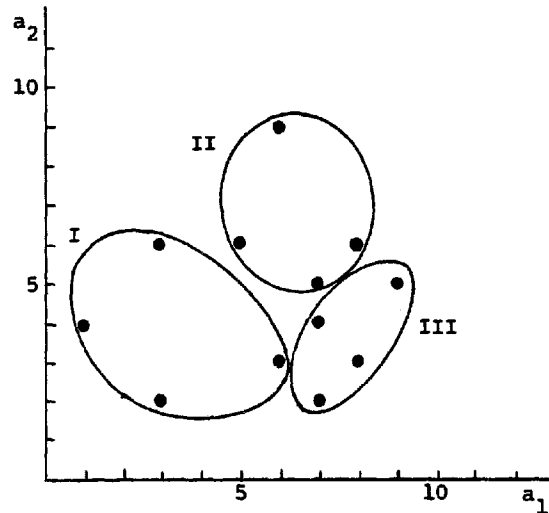| GROUP I		GROUP II		GROUP III	
A_1	3, 2	A_5	5, 6	A_9	7, 2
A_2	1, 4	A_6	6, 9	A_{10}	7, 4
A_3	3, 6	A_7	7, 5	A_{11}	9, 5
A_4	6, 3	A_8	8, 6	A_{12}	8, 3

Fig. 5.

cation of the formulations previously described shows this to be the case.)

To accommodate such a problem, the general GP discriminant procedure can be effectively modified. By extending the basic approach, it is possible to create a discriminant capacity far exceeding that of the previous method.

Here a pair-wise separating hyperplane formulation will identify the region designated to contain all points of the jth group as the intersection of half spaces determined by the 'best' hyperplanes separating group j points from each other group k [4].

Simply stated, the task is to generate a set of half spaces (here, three) that will serve to partition the n-space of the problem into appropriate regions (see

---

[2] Including the positive and negative components of unrestricted variables.

[3] Using (6—8) with the data for our credit application problem yields a transformation vector identical to that of the original exercise, producing perfect group separation with $X = [\frac{1}{-4}]$. (Here the c_j's were each set equal to 1.) Note: it would also be possible in this formulation to use two different α variables with each pair of inequalities of the form (7).

[4] If the convex hulls of the groups involved are disjoint, a collection of separating hyperplanes can always be found which will create regions providing 'complete discrimination'. In the non-convex case, sets can be partitioned into convex sets.
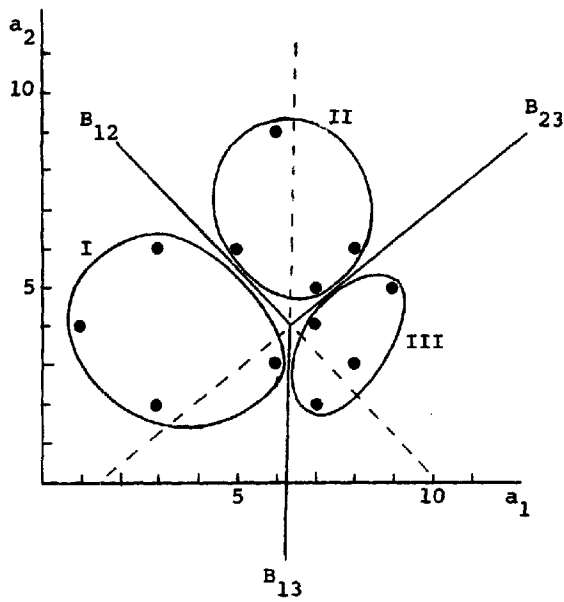
Fig. 6.

Fig. 6.) The boundary for each half-space will separate, insofar as possible, paired (adjacent) groups. Thus, to illustrate, in Fig. 6, any point falling below boundary B_{12} (which is assumed to extend beyond the portion shown) *could* be a member of Group I, but could not be a member of Group II. Conversely, any point *above* B_{12} could be a member of Group II, but could not be a Group I member. Similar statements apply to the roles of B_{13} and B_{23}. We can conclude, then, that any point bounded by B_{12} and B_{13} is a Group I member. Any point bounded by B_{12} and B_{23} is a Group II member, and any point falling within B_{13} and B_{23} is a Group III member. What remains is the determination of means by which appropriate boundaries may be drawn.

Importantly, the goal programming formulation proposed here decomposes into a collection of smaller two-group problems — one for each separating hyperplane — thereby facilitating computation. And, since in practical application the total number of group classifications one seeks to distinguish is typically small (as in credit rating categories), the number of separating hyperplanes is within reasonable limits — i.e., $g(g-1)/2$, where g = number of designated groups. Accordingly, the three-group example of Fig. 5 readily converts to an equivalent set of three distinct 2-group problems.

To exploit this capacity to decompose multi-group problems, a general two-group discriminant procedure is detailed below:

Given two groups, G_1 and G_2, determine an appropriate vector X and boundary value b such that, as nearly as possible,

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

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

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:

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

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

and the sum of boundary violations α_i (or some weighted sum of boundary violations $h_i\alpha_i$) is minimized.

Further, the separating hyperplane, $AX = b$, will be selected so that points which lie within the boundary are as *far* within the boundary as possible, thereby sharpening group differentiation. While it will generally not be possible to anticipate which points will lie within the 'true' boundary (i.e., those points which satisfy $A_iX \leqslant b$ for $A_i \in G_1$, or $A_iX \geqslant b$ for $A_i \in G_2$), it is clear that all points will lie within the 'adjusted' boundaries. That is, all points will satisfy $A_iX \leqslant b + \alpha_i$ for $A_i \in G_1$ or $A_iX \geqslant b - \alpha_i$ for $A_i \in G_2$. Letting d_i denote the distance of point A_i from its adjusted boundary, we can effectively combine the goal of minimizing boundary deviations with that of maximizing the (weighted) sum of these distances ($\Sigma k_i d_i$). Setting the problem in a minimization context, where maximizing $\Sigma k_i d_i$ corresponds to minimizing $-\Sigma k_i d_i$, the combined objective will have the form [5]:

Min $\sum h_i \alpha_i - \sum k_i d_i ,$

(A)

s.t. $A_iX + d_i = b + \alpha_i , \qquad A_i \in G_1 ,$

$\qquad -A_iX + d_i = -b + \alpha_i , \quad A_i \in G_2 .$

Note that the distances d_i are precisely the slack variables which change the inequalities of (2) into equalities.

Importantly, the procedure will yield a solution in which $d_i = 0$ whenever the weight for minimizing the boundary violation exceeds the weight for maximizing the distance of A_i from the adjusted boundary *and* A_i violates the true bound; that is, if $\alpha_i > 0$, then

---

[5] Fig. 7 is provided as a useful graphical reference.

I Perfect Separation

(i.e. α = 0)

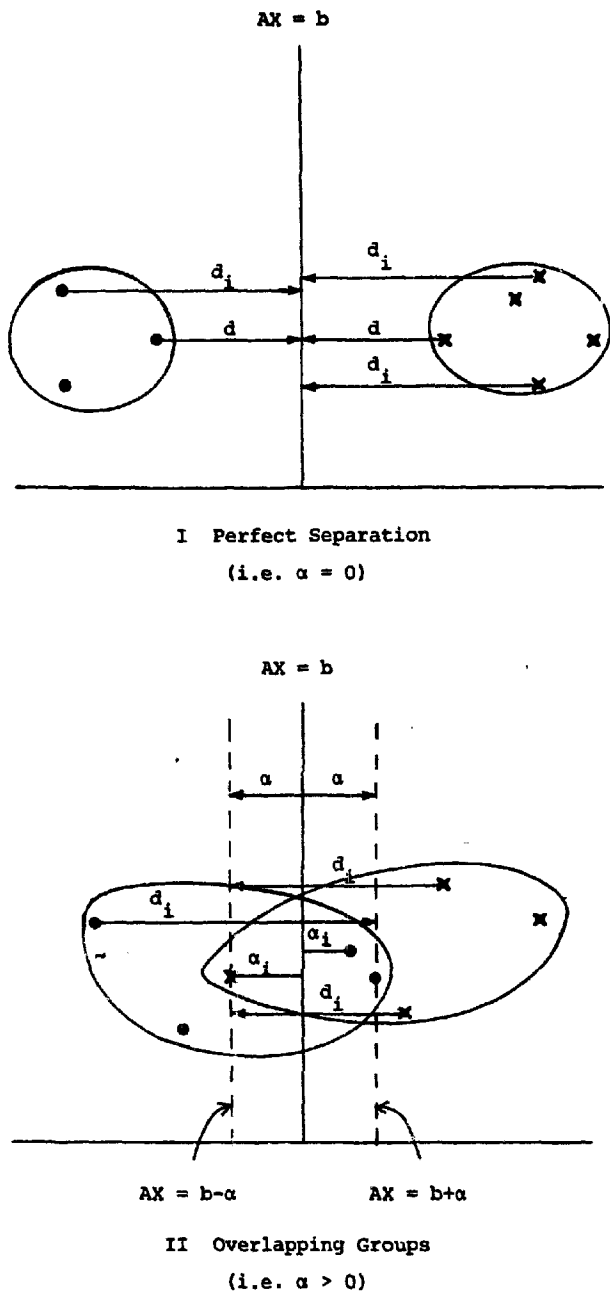II Overlapping Groups

(i.e. α > 0)

Fig. 7.

$d_i = 0$ for all $h_i > k_i$. Further, α_i will never be forced larger than necessary in order to let the slack d_i be positive. (The reverse is also true; for $h_i < k_i$, α_i will be pushed to its limit in order to increase the value of the slack d_i.) When $\alpha_i = 0$, i.e., when A_i is within the 'true' boundary, d_i will take on its largest value for any positive k_i.

Such a formulation proves extremely flexible. Usefully, the identity of individual data points is maintained here, enabling the user to assign distinct

misclassification penalties to each observation. For example, should it be judged that some subset of points, or even a single observation, need play a greater role in differentiating groups (i.e., be judged perhaps as most representative of 'typical' group members) increased weight can be readily assigned to the corresponding overlap variables, α_i. Similarly, less important points may be given appropriately reduced weighting. (Post-optimality procedures for monitoring the impact of alternative objective function weighting schemes and for establishing relative variable importance are discussed in a subsequent paper [8].)

Variants of the basic GP form which address related objectives can now be considered.

A significant reduction in the number of variables required in (A) can be achieved by aggregating terms. For example, by replacing the set of α_i variables with a single α term [6], the model becomes:

Min $H\alpha - \sum_i k_i d_i$,

(B)

s.t. $A_i X + d_i = b + \alpha$, $A_i \in G_1$,

    $-A_i X + d_i = -b + \alpha$, $A_i \in G_2$.

Here, α measures the *maximum* boundary violation associated with a candidate discriminant solution (i.e., $\alpha_i < \alpha$ for all i). The connection between α and d_i is similar to that between α_i and d_i described in (A) above. If $\alpha > 0, H > \Sigma\ k_i$ assures that the smallest d_i will always be 0. (d_i is smallest for those points which violate the true boundary by the greatest amount.) On the other hand, $H > \Sigma\ k_i$ assures that α will be forced to its upper limit.

Aggregating d_i's rather than α_i's produces the LP form:

Min $\sum_i h_i \alpha_i - Kd$,

(C)

s.t. $A_i X + d = b + \alpha_i$, $A_i \in G_1$,

    $-A_i X + d = -b + \alpha_i$, $A_i \in G_2$,

in which d measures the *minimum* distance of any group member from the boundary hyperplane $AX = b$, when all $\alpha_i = 0$. If $\alpha_i > 0$ for some i and $\Sigma\ h_i > K$, then $d = 0$. $\Sigma\ h_i < K$ assures that the α_i's will be pushed toward their upper bound.

Complete aggregation, an extreme variant of the formulation in which the set of α_i variables is replaced

[6] It should be noted that partial aggregation is possible, i.e., we may wish only to replace a subset of the α_i variables with α.
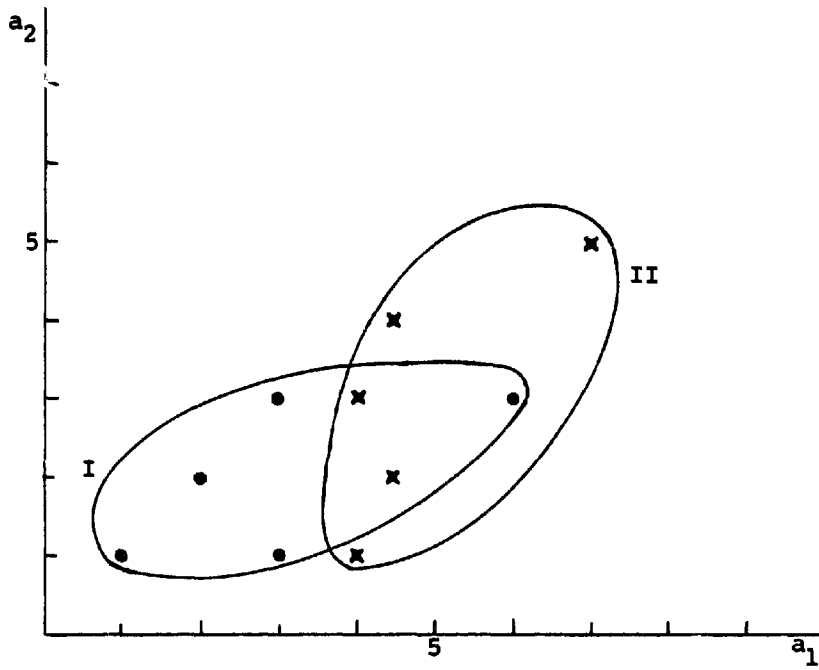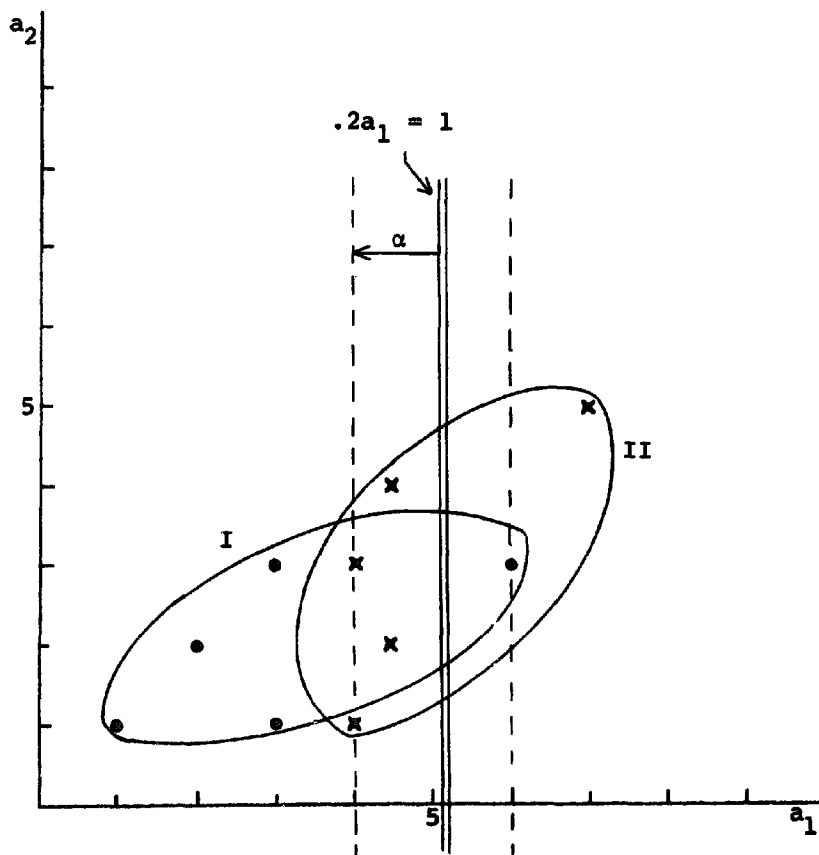
Fig. 8. Overlapping groups problem.
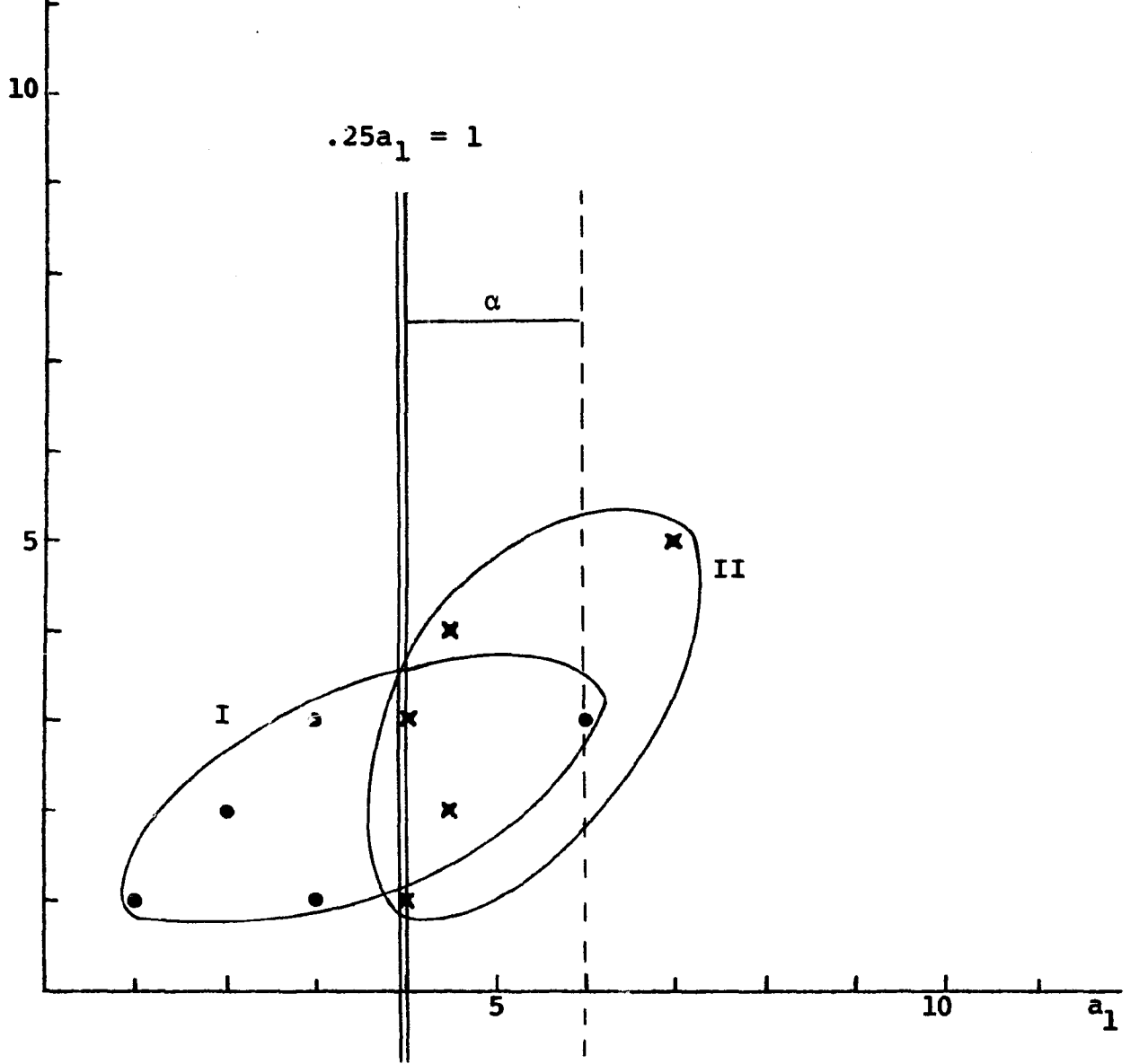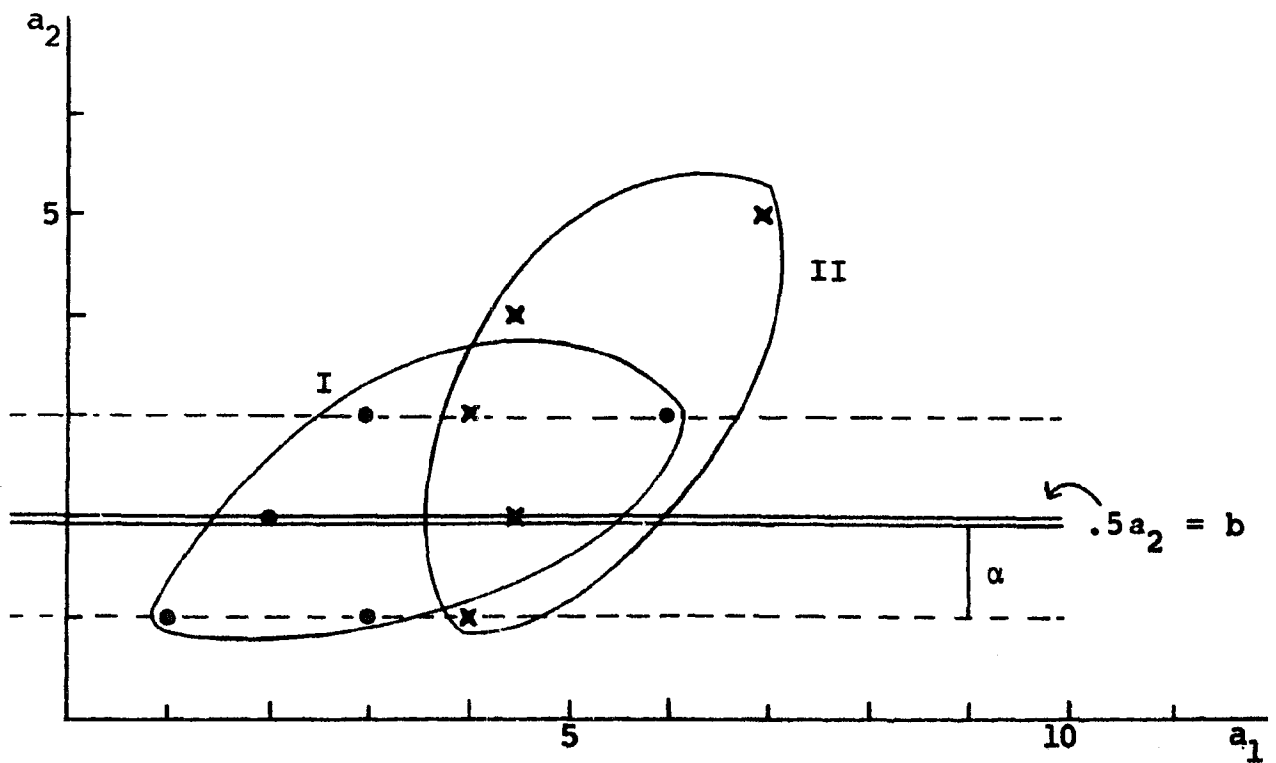

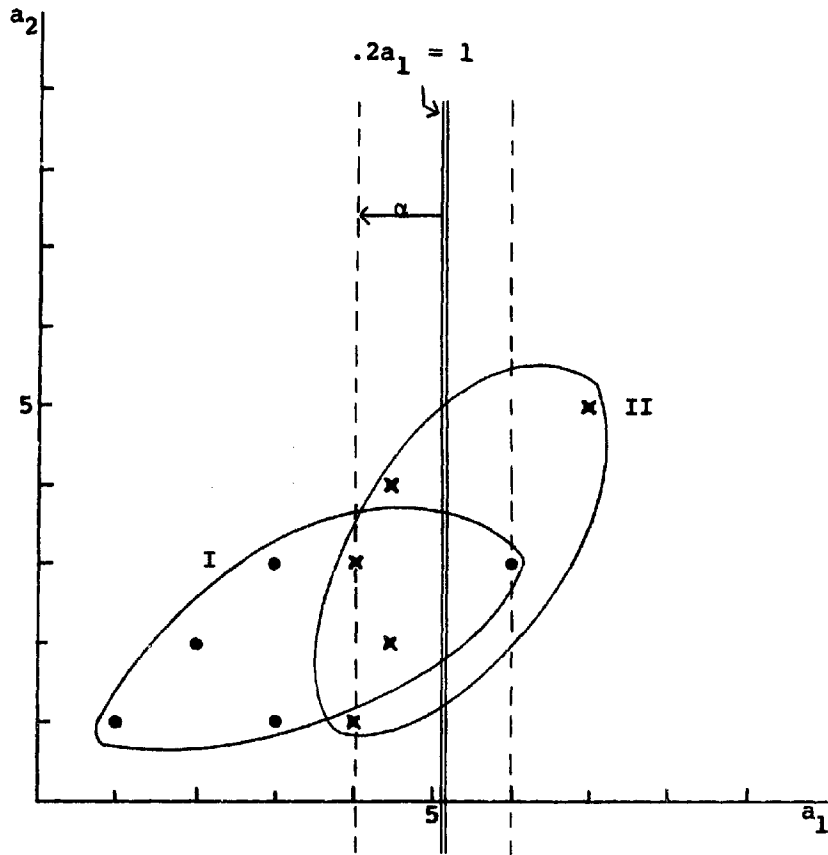Fig. 9. Objective: Min $2\alpha - d$.

Fig. 10. Objective: Min $\Sigma 2\alpha_i - d$.

Fig. 12. Objective: Min $50\alpha - \Sigma d_i$.
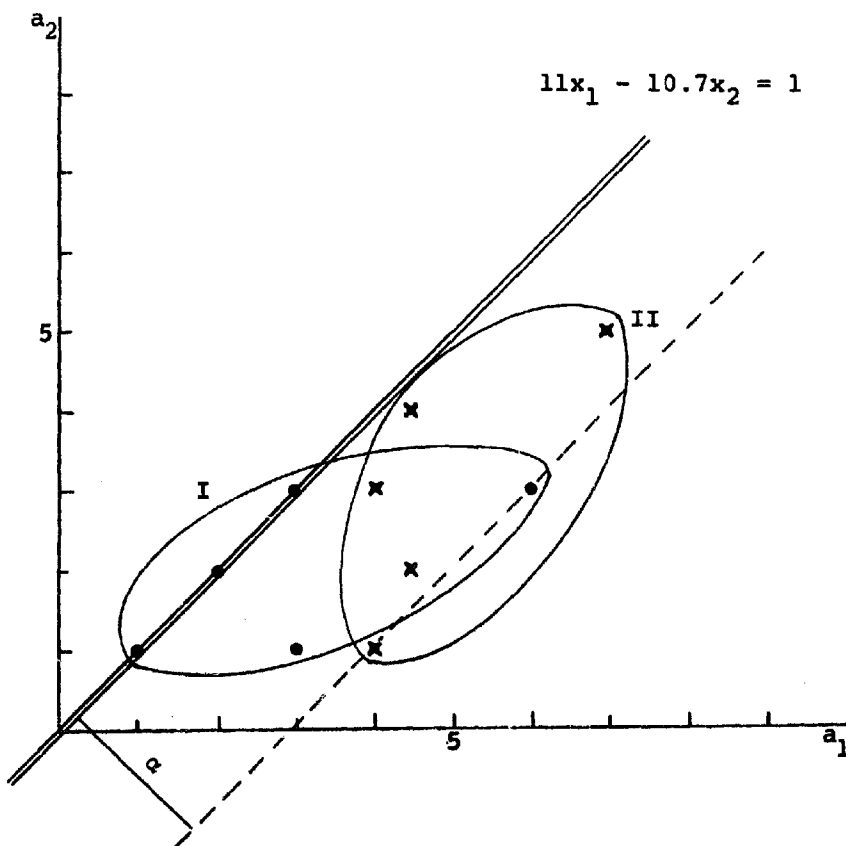

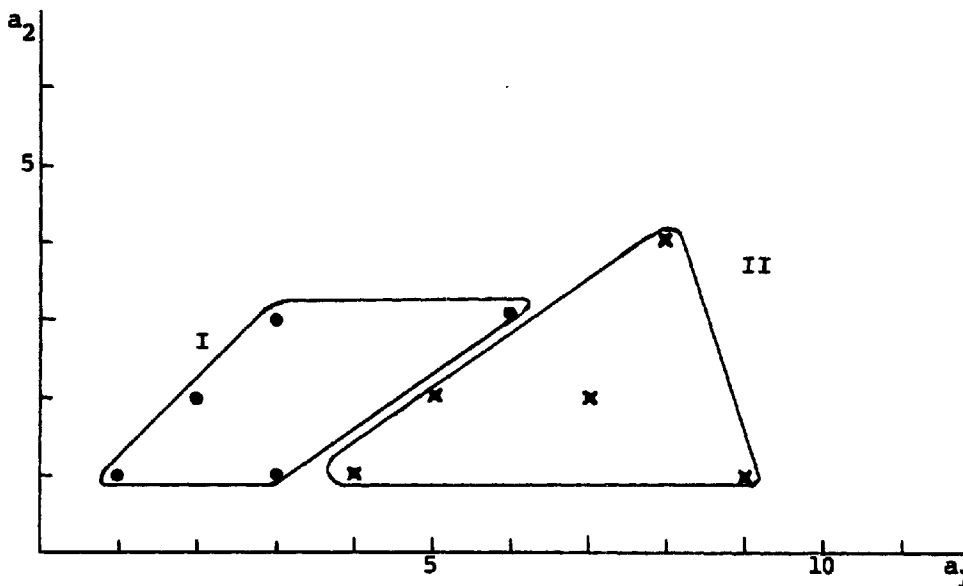Fig. 13. Objective: Min $\Sigma \alpha_i - \Sigma d_i$.

Fig. 14. Objective: Min $\Sigma 2\alpha_j - \Sigma d_j$.

with α and the set of d_j variables with d, provides a simple GP form in which the task is to:

$$\text{Min} \quad H\alpha - Kd ,$$

$$\text{s.t.} \quad A_i X + d = b + \alpha , \qquad A_i \in G_1 ,$$

$$-A_i X + d = -b + \alpha , \qquad A_i \in G_2 .$$

(D)

$H > K$ insures that the goal of minimizing maximum overlap (α) dominates the objective. For $\alpha > 0, d = 0$. For $\alpha = 0$, d is a measure of the minimum distance of any point from the separating hyperplane $AX = b$. $H < K$ assures that α is pushed to its upper bound.

The overlapping groups problem of Fig. 8 illustrates the connections between these GP forms and the essential nature of the discriminant capabilities which each posseses.

Note: The Group I and Group II data used in Figs. 8—14 is given in Table 6. In Figs. 8—19 ● denotes Group I members, X Group II members.

Fig. 9 shows the simple model solution produced by the aggregated objective (D) above. Notable here is the dominant role of the maximum group overlap variable α. Although the procedure has effectively produced an α-minimizing solution, the ability of this solution to appropriately classify group members is not wholly satisfactory (e.g., five of ten sample points are improperly assigned). Importantly, minimizing *maximum* overlap directs full attention to only those points which most seriously violate the between-groups boundary — i.e., those points from

each group which reach most deeply into the 'territory' of the other. All other points are ignored. Thus, for example, Group II points A_7 and A_9, although misclassified by the solution hyperplane $0.2a_1 = 1$, play no role in the procedure.

Disaggregation of variables encourages the fuller participation of additional group members. Fig. 10 shows the solution produced with the partially disaggregated objective in (C). Here *total* group overlap is minimized by the separating hyperplane $0.2a_1 = 1$. Only the extreme (and possibly 'atypical') Group I member A_5 is misclassified.

The disaggregation of (B) allows every point a role in establishing an optimal bound. While only the most serious boundary violators directly influence the goal of α-minimization, each d_i serves to register a counterbalancing influence on the overall objective. As previously described, the goal of maximizing some total d_i represents an attempt to force the bulk (mass) of each group back from the boundary hyperplane. Fig. 11 shows the result. By assigning a sufficiently large value to H (i.e., by weighting quite heavily the goal of minimizing overlap), this procedure can be expected to produce a solution equivalent to that of (D) above (see Fig. 12).

The fully disaggregated objective of (A) provides for maximum flexibility. Figs. 13 and 14 present some of the solution possibilities.

The capacity of these alternative forms to address the problem of differentiating members of *disjoint* groups bears mention. Using the two-group problem

Table 6
Data used in Figs. 8–14

Group I		Group II	
A_1	1,1	A_6	4,1
A_2	2,2	A_7	4.5,2
A_3	3,1	A_8	4,3
A_4	3,3	A_9	4.5,4
A_5	6,3	A_{10}	7,5

Table 7
Data used in Figs. 15–19

Group I		Group II	
A_1	1,1	A_6	4,1
A_2	2,2	A_7	5,2
A_3	3,1	A_8	7,2
A_4	3,3	A_9	8,4
A_5	6,3	A_{10}	9,1
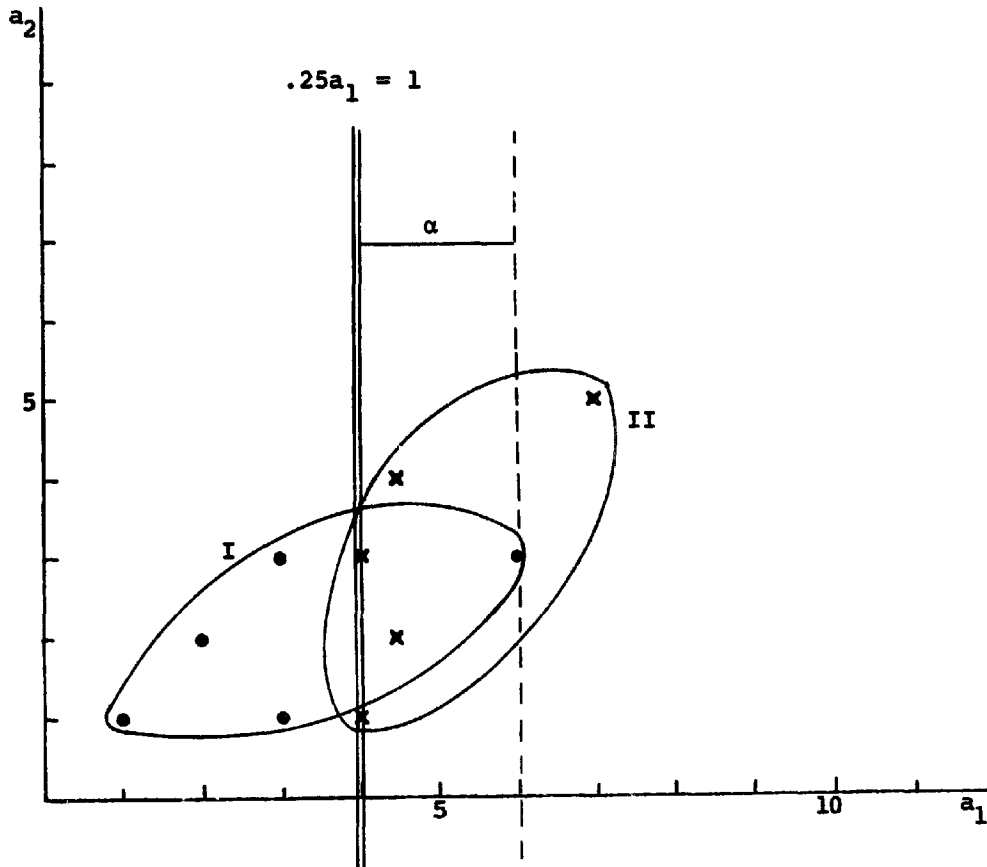
Fig. 15. Disjoint groups problem.
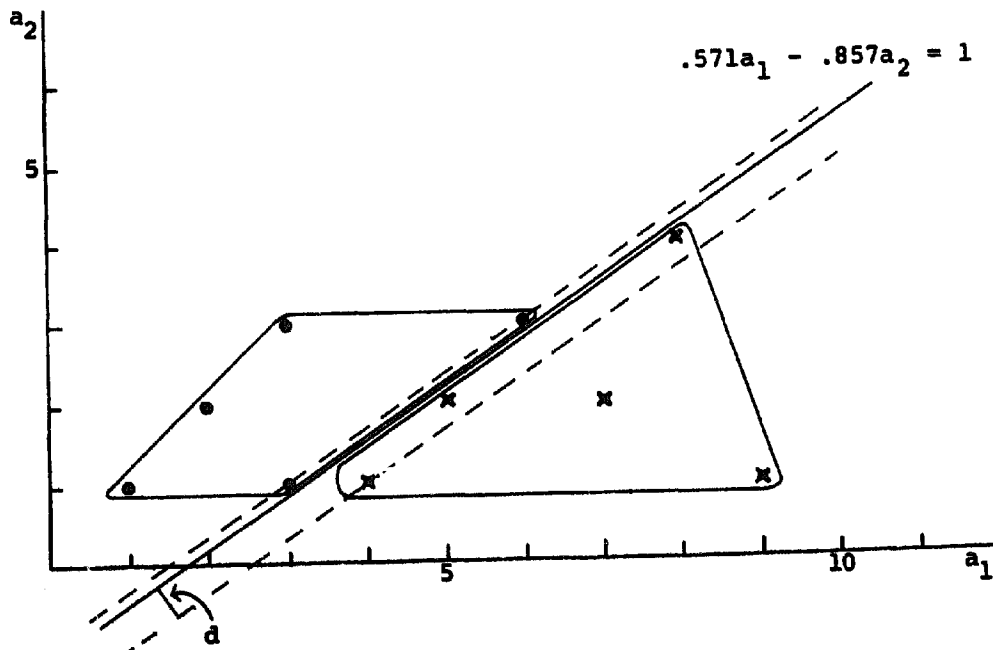
Fig. 16. Objective: Min $2\alpha - d$.
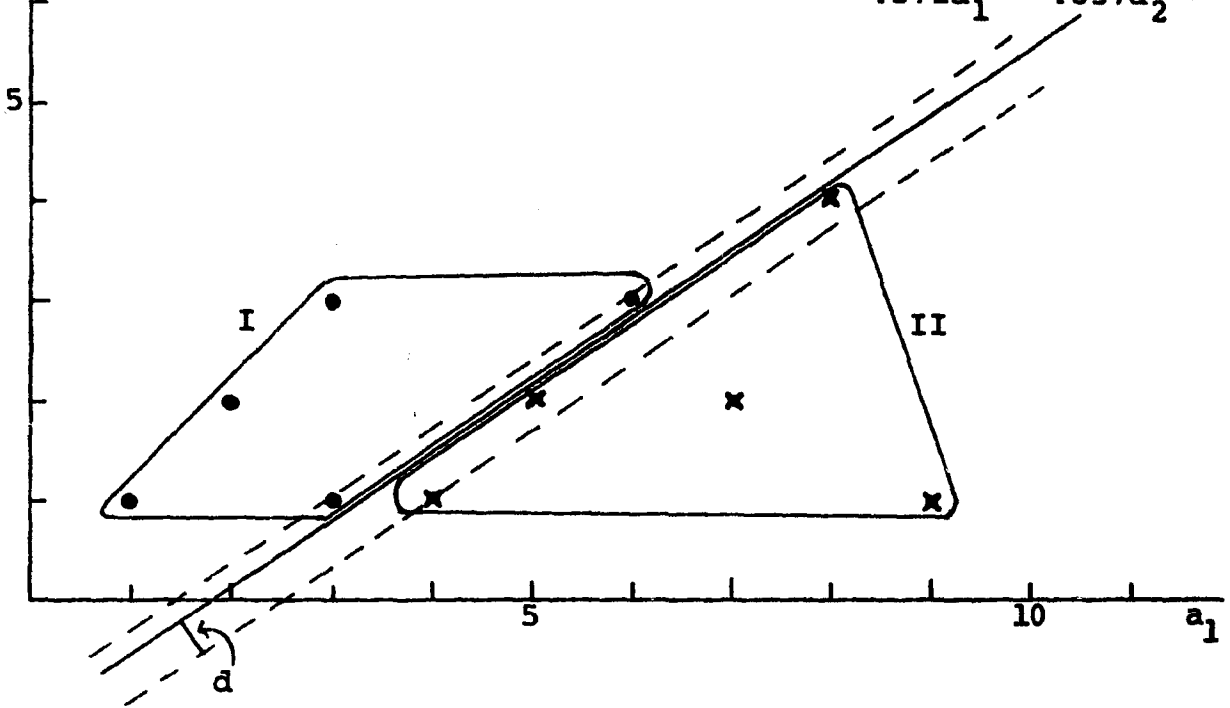
Fig. 17 Objective: Min $\Sigma 2\alpha_i - d$.

$.75a_1 - 1.25a_2 = 1$
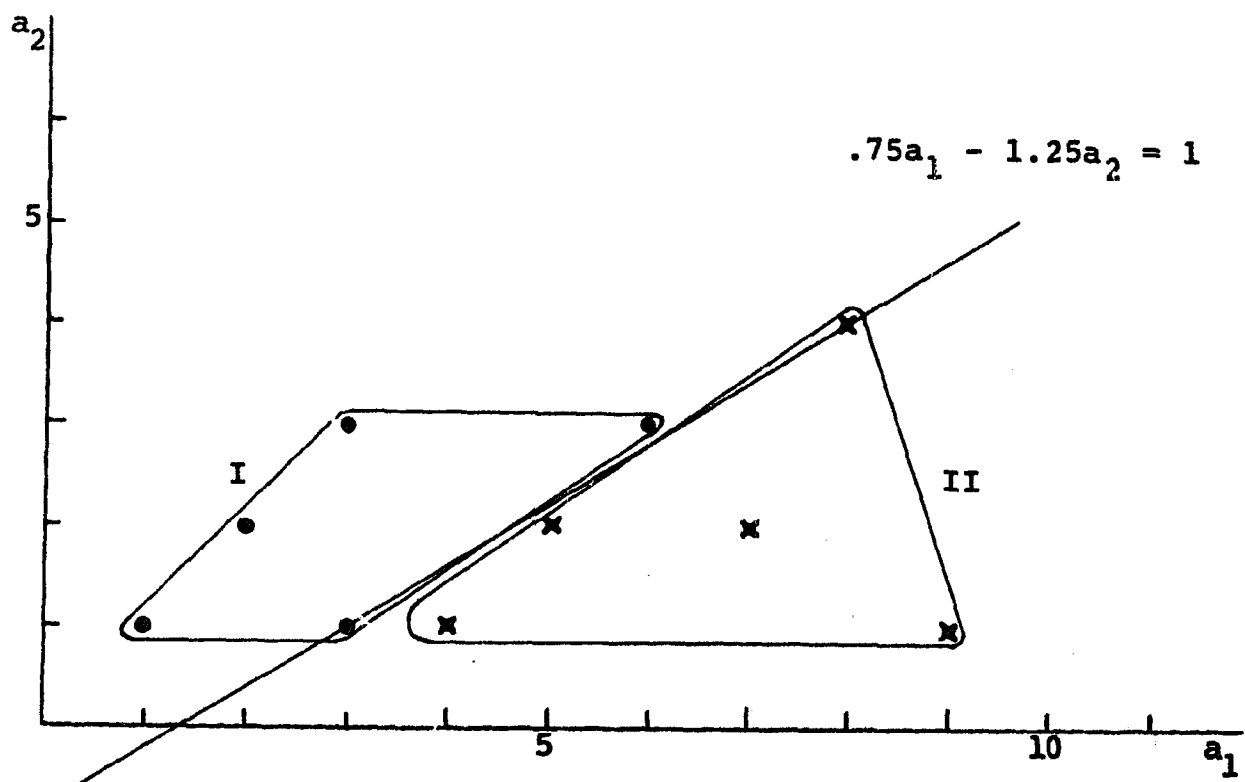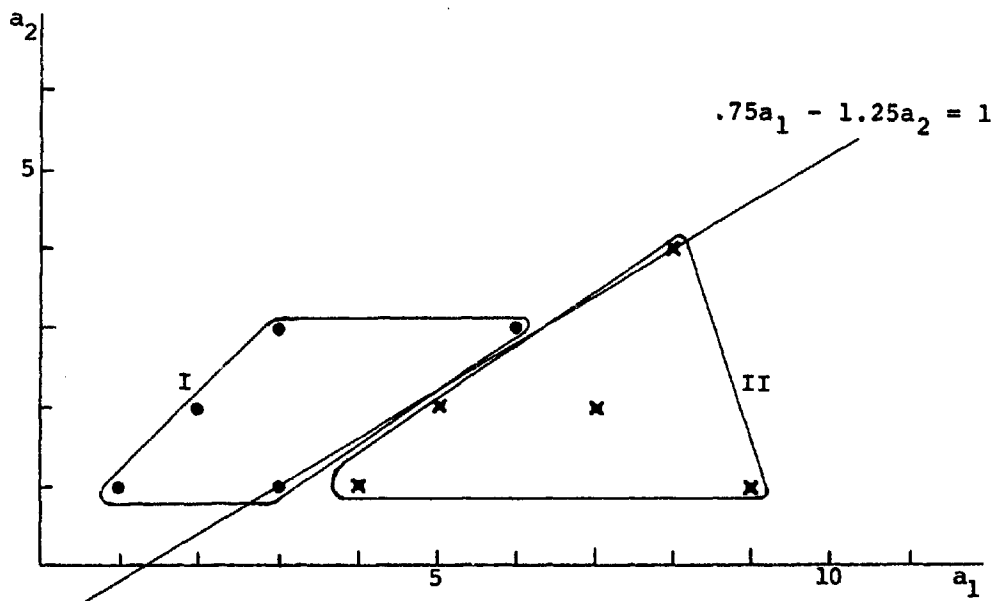
Fig. 18 Objective: Min $\Sigma 2\alpha_i$.

Fig. 19. Objective: Min $\Sigma 20\alpha_j - \Sigma d_j$.

of Fig. 15 (the data used in Figs. 15–19 is given in Table 7) to illustrate, it is clear that each form is capable of producing a discriminant solution which assures perfect separation. However, parameters of the 'optimal' hyperplane vary with the formulation (and weights) applied. For example, the boundary produced by those simpler forms in which the d_j's have been aggregated (i.e., (C) and (D)), is effectively 'centered' in the gap between groups (see Figs. 16 and 17). By contrast, the solution hyperplane for those forms in which the d_j's are wholly disaggregated (i.e., (A) and (B)) touches the hull of each group (see Figs. 18 and 19). (Here an additional adjustment in the weights assigned to the d_j, particularly in those weights corresponding to points A_3 and A_9, is necessary to produce an equivalent 'centered' solution.) Thus, a case has been identified in which the simplest model — requiring far fewer variables — compares favorably with a more complex, albeit more flexible, formulation.

3. Summary

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.

References

[1] V.G. Ashar and T.D. Wallace, A sampling study of minimum absolute deviation estimators, Operations Res. 11 (1963) 747–758.

[2] M. Blum, Failing discriminant analysis, J. Accounting Res. 12 (1974) 1–25.

[3] A. Charnes and W.W. Cooper, Management Models and Industrial Application 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. 1 (1955) 138–151.

[5] A. Charnes and W.W. Cooper, Goal programming and multiple objective optimizations, European J. 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.

[7] F.A. Eisenbeis and R.B. Avery, Two aspects of investigating group differences in linear discriminant analysis, Decision Sci. 4 (1973) 487–493.

[8] N. Freed and F. Glover, Discriminant analysis by the separating/goal hyperplane method, Management and Information Science Report Series, University of Colorado, to appear.

[9] F.R. Glahe, The small sample properties of simultaneous equations least absolute estimators vis-a-vis least squares estimators, Econometrics 38 (1970) 742–753.

[10] J.G. Hunt et al., L_1 Estimation in small samples with Laplace error distributions, Decision Sci. 5 (1974) 22–29.

[11] O.J. Karst, Linear curve fitting using least deviations, J. Amer. Statist. Assoc. 53 (1958) 118–132.

[12] M.G. Kendall, Discrimination and classification, in: ᵖ.R. Krishnaiah (Ed.), Multivariate Analysis (Academic Press, New York, 1966).

[13] E.A. Kiountouzis, Linear programming techniques in regression analysis, Appl. Statist. 1 (1973) 799–821.

[14] M.R. Rao, Cluster analysis and mathematical programming, J. Amer. Statist. Assoc. 66 (1971) 622–626.

[15] T.R. Rao, Is Brand loyalty a criterion for market segmentation: Discriminant analysis, Decision Sci. 4 (1973) 395–404.

[16] J.F. Sinkey, Multivariate statistical analysis of the characteristics of problem banks, J. Finance 30 (1975) 21–26.

[17] K.H. Usow, On L_1 approximation: Computation for discrete functions and discretization effects, SIAM J. Numer. Anal. 4 (1967) 233–244.

[18] J.M. Utterback, Successful industrial innovations: A multivariate analysis, Decision Sci. 6 (1975) 65–77.

[19] H.M. Wagner, Linear programming techniques for regression analysis, J. Amer. Statist. Assoc. 54 (1959) 206–212.

[20] R.B. Welker, Discriminant and classification analysis as an aid to employee selection, Accounting Rev. 49 (1974) 514–523.

[21] H.G. Wilson, Least squares versus minimum absolute deviations estimation in linear models, Decision Sci. 9 (1978) 322–335.