CENTER FOR PARALLEL OPTIMIZATION

MULTICATEGORY DISCRIMINATION VIA LINEAR PROGRAMMING

by

Kristin P. Bennett and O. L. Mangasarian

Computer Sciences Technical Report #1127

November 1992

Multicategory Discrimination via Linear Programming

Kristin P. Bennett & O. L. Mangasarian*

November 30, 1992

Abstract

A single linear program is proposed for discriminating between the elements of k disjoint point sets in the n-dimensional real space R^n. When the conical hulls of the k sets are $(k-1)$-point disjoint in R^{n+1}, a k-piece piecewise-linear surface generated by the linear program completely separates the k sets. This improves on a previous linear programming approach which required that each set be linearly separable from the remaining $k-1$ sets. When the conical hulls of the k sets are not $(k-1)$-point disjoint, the proposed linear program generates an error-minimizing piecewise-linear separator for the k sets. For this case it is shown that the null solution is never a unique solver of the linear program and occurs only under the rather rare condition when the mean of each point set equals the mean of the means of the other $k-1$ sets. This makes the proposed linear computational programming formulation useful for approximately discriminating between k sets that are not piecewise-linear separable. Computational results are reported for three previously available databases.

1 Introduction

We consider the k disjoint sets $\mathcal{A}^i$, $i = 1, \ldots, k$, in the n-dimensional real space R^n represented by the $m^i \times n$ matrices, A^i, $i = 1, \ldots, k$. Our objective here is to discriminate between these sets by a piecewise-linear convex function which is the maximum of k linear (affine) functions. The linear pieces of one such typical piecewise-linear surface projected on R^2 are depicted in Figure 1 together with the four sets in R^2 that are separated from each other. Many authors have considered this problem. Nilsson [16], Duda-Fossum [5], Duda-Hart [6], and Fukunaga [8], considered iterative methods which are extensions of the perceptron algorithm or the Motzkin-Schoenberg algorithm [14] for determining a piecewise-linear separator when it exists. Convergence of these methods is not known if such a piecewise-linear surface does not exist [8, page 374]. Smith [17] on the other hand considered solving k systems of linear inequalities by solving k linear programs to obtain a piecewise-linear separator. Unfortunately, this may not be possible for many simple piecewise-linear separable problems as we shall demonstrate below. By contrast our linear programming approach works for all piecewise-linear separable sets, and for those that are not some approximate separation will be achieved.

We note that piecewise-linear separation of k disjoint sets in R^n (see Definition 2.1) is a natural extension of the classical separation of two disjoint point sets $\mathcal{A}^1$ and $\mathcal{A}^2$ in R^n with nonintersecting convex hulls by using the piecewise-linear maximum of two linear functions. This is equivalent to separation by a single plane [9, 11, 4]. See Figure 2.

*Computer Sciences Department, University of Wisconsin, 1210 West Dayton Street, Madison, WI 53706, email: *bennett@cs.wisc.edu, olvi@cs.wisc.edu*. This material is based on research supported by Air Force Office of Scientific Research Grant AFOSR-89-0410 and National Science Foundation Grant CCR-9101801 and Air Force Laboratory Graduate Fellowship SSAN 531-56-2969.

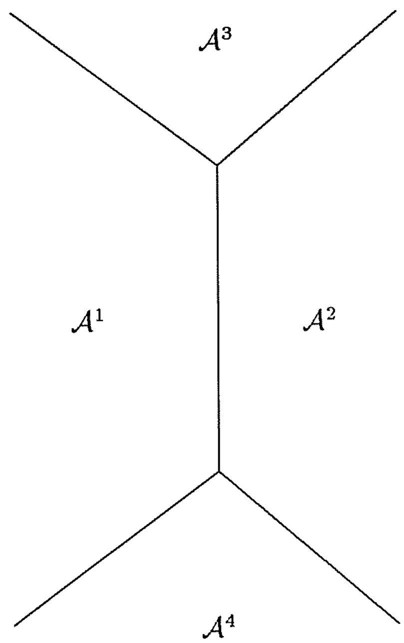

Figure 1: Projection of the linear pieces of a piecewise-linear surface on R^2 and the sets $\mathcal{A}^i$, $i = 1, \ldots, 4$, that it separates.

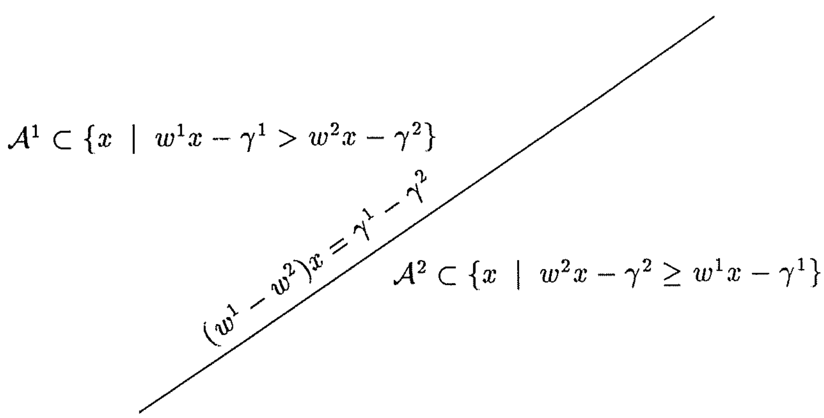

Figure 2: Separation of two sets $\mathcal{A}^1$ and $\mathcal{A}^2$ by a plane, or equivalently by the maximum of two linear functions.

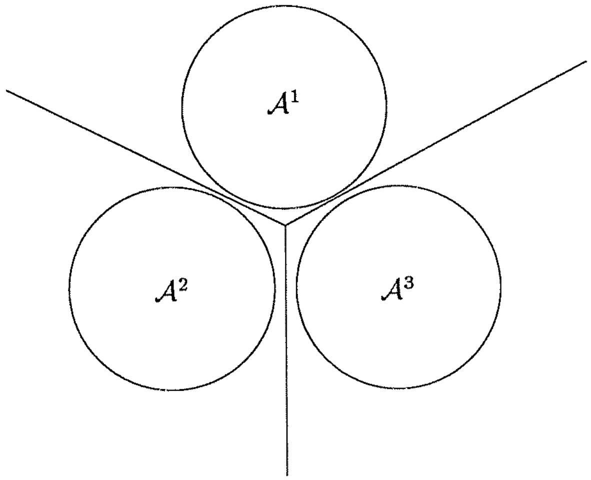

Figure 3: Three sets separable by a piecewise-linear function, but no one of which is linearly separable from the other two.

One of the first questions to resolve is: When is it indeed possible to discriminate between k sets by a piecewise-linear separator which is the maximum of k linear functions? Condition (6) of Theorem 2.1 gives a necessary and sufficient condition for such piecewise-linear separability of k sets. Geometrically, this condition can be interpreted as follows. For each $i = 1, \ldots, k$, choose $k-1$ points $\tilde{A}^i_j, j = 1, \ldots, k,\ j \neq i$, in the conical hull of $\bar{A}^i \subset R^{n+1}$, where $\bar{A}^i$ is the set of m^i points in R^{n+1} made up of the rows of $\bar{A}^i := [A^i \ \ e]$, where e is an $m^i \times 1$ column of ones. Condition (6) says that there are no points other than $0 \in R^{n+1}$ in these conical hulls of $\bar{A}^i,\ i = 1, \ldots, k$, satisfying

$$\sum_{\substack{j=1 \\ j\neq i}}^{k} \tilde{A}^i_j = \sum_{\substack{j=1 \\ j\neq i}}^{k} \tilde{A}^j_i, \quad i = 1, \ldots, k \tag{1}$$

We refer to this condition as the conical hulls in R^{n+1} of the k sets $\mathcal{A}^i,\ i = 1, \ldots, k$, being $(k-1)$-**point disjoint**. This is a considerably more relaxed requirement than that of Smith [17]. Smith proposed solving k systems of linear inequalities that are equivalent to the linear separation of each of the k sets from the remaining $k-1$ sets. This is far too restrictive an assumption that does not in general hold for simple piecewise-linear separable sets. Figure 3 depicts three sets separable by a 3-piece piecewise-linear function, but for which no one set is linearly separable from the remaining two. An even simpler example in R^1 is $\mathcal{A}^1 = \{-1\}$, $\mathcal{A}^2 = \{0\}$ and $\mathcal{A}^3 = \{1\}$. These three sets are piecewise-linear separable by $\max\left\{-x - \frac{1}{2}, \frac{1}{4}, x - \frac{1}{2}\right\}$ but $\mathcal{A}^2$ is not linearly separable from $\mathcal{A}^1 \cup \mathcal{A}^3$.

In Section 2 of the paper we begin with Definition 2.1 of piecewise-linear separability of k sets. Then, as indicated above, we characterize this separability in Theorem 2.1. In Remark 2.2 we note that piecewise-linear separation implies pairwise-linear separation, but not conversely, as depicted in Figure 5. A computationally constructive characterization of piecewise-linear separability of k sets is given in Theorem 2.2 by obtaining a zero minimum for the linear program (8) and a corresponding piecewise-linear separation (3). Since such piecewise-linear separation (3) for the k

sets $\mathcal{A}^i$, $i = 1, \ldots, k$, is determined by the quantities $(w^i - w^j, \gamma^i - \gamma^j)$, $i, j = 1, \ldots, k$, $j \neq i$, it is important to ensure the nonzeroness of $(w^i - w^j)$, $i, j = 1, \ldots, k$, $j \neq i$. This is done precisely in Theorem 2.3 where it is shown that the null solution $w^i - w^j = 0, i, j = 1, \ldots, k$, $j \neq i$, occurs under the rather rare condition that the mean of each point set equals the mean of the means of the other $k - 1$ sets. Theorem 2.4, however, shows that the null solution, even in this case, is never unique. Section 3 contains some computational results employing the proposed linear programming formulation (8).

A word about our notation now. For a vector x in the n-dimensional real space R^n, x_+ will denote the vector in R^n with components $(x_+)_i := \max\ \{x_i,\ 0\}$, $i = 1, \ldots, n$. The notation $A \in R^{m \times n}$ will signify a real $m \times n$ matrix. For such a matrix, A' will denote the transpose while A_i will denote the ith row. The 1-norm of x, $\sum_{i=1}^{n} |x_i|$, will be denoted by $\|x\|_1$, while the ∞-norm of x, $\max_{1 \leq i \leq n} |x_i|$, will be denoted by $\|x\|_\infty$. A vector of ones in a real space of arbitrary dimension will be denoted by e. The conical hull of the set $\mathcal{A}^1$, the set of all nonnegative linear combinations of points in $\mathcal{A}^1$, $\{z | z = uA^1,\ u \geq 0\}$, will be denoted by $K(\mathcal{A}^1)$.

2 Multicategory Separation by a Piecewise-Linear Surface

We begin by defining the concept of piecewise-linear separation of k sets in R^n [16].

Definition 2.1 (Piecewise-linear Separability) *The k sets $\mathcal{A}^i$, $i = 1, \ldots, k$, each consisting of m^i, $i = 1, \ldots, k$, points in R^n and represented by the $m^i \times n$ matrices, A^i, $i = 1, \ldots, k$, are piecewise-linear separable if there exist $w^i \in R^n$, $\gamma^i \in R$, $i = 1, \ldots, k$ such that*

$$A^i w^i - e\gamma^i > A^i w^j - e\gamma^j,\ i,\ j = 1, \ldots, k,\ i \neq j \tag{2}$$

or equivalently

$$A^i w^i - e\gamma^i \geq A^i w^j - e\gamma^j + e,\ i,\ j = 1, \ldots, k,\ i \neq j \tag{3}$$

Remark 2.1 *The piecewise-linear separability can be interpreted geometrically by the existence of a piecewise-linear convex function determined from $(w^i,\ \gamma^i)$, $i = 1, \ldots, k$, by*

$$p(x) = \max_{1 \leq \ell \leq k} xw^\ell - \gamma^\ell \tag{4}$$

and such that

$$\left\langle \begin{array}{l} p(x) = xw^i - \gamma^i \\ p(x) > xw^j - \gamma^j \end{array} \right\rangle \begin{array}{l} \textit{for}\ \ x \in \mathcal{A}^i,\ i = 1, \ldots, k \\ \textit{for}\ \ j \neq i \end{array} \tag{5}$$

Figure 4 depicts a simple piecewise-linear $p(x)$ on R that separates three sets.

Our first objective is to characterize piecewise-linear separability. This is done in the following theorem.

Theorem 2.1 (Characterization of Piecewise-Linear Separability) *Let $\mathcal{A}^i$, $i = 1, \ldots, k$ be nonempty point sets in R^n. The following are equivalent:*

(a) $\mathcal{A}^i$, $i = 1, \ldots, k$ are piecewise-linear separable; that is there exist $w^i \in R^n$, $\gamma^i \in R$, $i = 1, \ldots k$, satisfying (2) or (3).

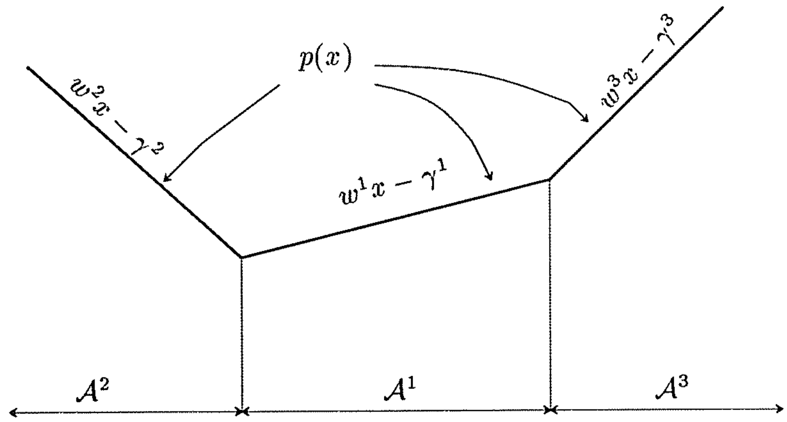

Figure 4: Piecewise-linear separation of sets $\mathcal{A}^1$, $\mathcal{A}^2$ and $\mathcal{A}^3$ by the convex piecewise-linear function $p(x) = \max_{i=1,2,3}\ w^i x - \gamma^i$.

(b) The conical hulls $K(\bar{\mathcal{A}}^i)$ of the k sets $\bar{\mathcal{A}}^i$, where $\bar{A}^i := [A^i \quad e]$, $i = 1,\ldots,k$, are $(k-1)$-point disjoint in R^{n+1}, that is

$$\sum_{\substack{j=1\\ j\neq i}}^{k} u^{ij}[A^i \quad e] = \sum_{\substack{j=1\\ j\neq i}}^{k} u^{ji}[A^j \quad e],\ u^{ij} \geq 0,\ i,\ j = 1,\ldots,k,\ j \neq i \qquad (6)$$
$$\implies u^{ij} = 0,\ i, j = 1,\ldots,k,\ j \neq i$$

Proof. Throughout the following arguments, $i,\ j = 1,\ldots,k$ and $j \neq i$.

(a) $\Leftrightarrow$ $A^i(w^i - w^j) - e(\gamma^i - \gamma^j) > 0$, have solution $(w^i,\ w^j,\ \gamma^i\ \gamma^j) \in R^n \times R^n \times R \times R$

$\Leftrightarrow$ $A^i(w^i - w^j) - e(\gamma^i - \gamma^j) - e\zeta \geq 0,\ \zeta > 0$, have solution $(w^i,\ w^j,\ \gamma^i, \gamma^j,\ \zeta) \in R^n \times R^n \times R \times R \times R$

$\Leftrightarrow$ $$-\sum_{\substack{j=1\\ j\neq i}}^{k} u^{ji}A^j + \sum_{\substack{j=1\\ j\neq i}}^{k} u^{ij}A^i = 0,\ \sum_{\substack{j=1\\ j\neq i}}^{k} u^{ji}e - \sum_{\substack{j=1\\ j\neq i}}^{k} u^{ij}e = 0$$
$$\sum_{i=1}^{k}\sum_{\substack{j=1\\ j\neq i}}^{k} u^{ij}e > 0, \text{ have no solution } u^{ij} \geq 0,\ u^{ij} \in R^{m^i}$$

(By Motzkin's Theorem of the Alternative[13])

$\Leftrightarrow$ (b).

□

Remark 2.2 *It is evident from Definition 2.1 that piecewise-linear separability of the sets $\mathcal{A}^i$, $i = 1,\ldots,k$ implies pairwise linear separability of the same sets. However the converse is not true as evidenced by the "whirlwind" counterexample depicted in Figure 5 for which three sets are pairwise linearly separable, but are not piecewise-linear separable. The latter fact, which may not be immediately evident from the figure, can be computationally verified by showing that the implication of (6) does not hold by solving the dual linear program (11) and showing that it has a positive maximum.*

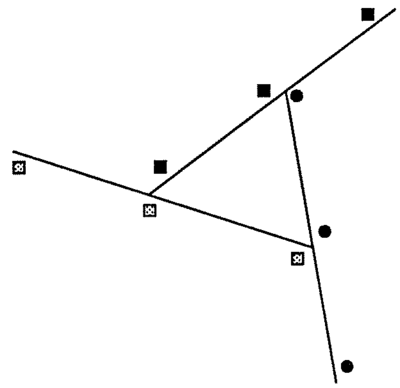

Figure 5: The whirlwind counterexample: Three sets that are pairwise linearly separable, but not piecewise-linear separable.

We can now specify a linear program that will generate a piecewise-linear separation between the sets $\mathcal{A}^i$, $i = 1, \ldots, k$, if one exists, otherwise it will generate an error-minimizing separation. The linear program will generate $(w^i,\ \gamma^i) \in R^n \times R,\ i = 1, \ldots, k$, that will satisfy (3) by minimizing the 1-norm of the averaged violations of (3), that is

$$\min_{w^i,\gamma^i} \sum_{i=1}^{k} \sum_{\substack{j=1 \\ j\neq i}}^{k} \frac{e}{m^i} \left(-A^i(w^i - w^j) + e(\gamma^i - \gamma^j) + e\right)_+ \tag{7}$$

This minimization problem can be written as the following linear program (LP):

$$\min_{w^i,\gamma^i,y^{ij}} \left\{ \sum_{i=1}^{k} \sum_{\substack{j=1 \\ j\neq i}}^{k} \frac{ey^{ij}}{m^i} \;\middle|\; \begin{array}{c} y^{ij} \geq -A^i(w^i - w^j) + e(\gamma^i - \gamma^j) + e,\ y^{ij} \geq 0, \\ i \neq j,\ i,\ j = 1, \ldots, k \end{array} \right\} \tag{8}$$

The purpose of the weights $\frac{1}{m^i}$, which are analogous to those of [4] for the two-category case, is to avoid the null $w^i - w^j = 0$ solution. See Theorem 2.3 below. Note that for any solution $(w^i,\ \gamma^j,\ y^{ij})$, of the LP (8), we have that

$$y^{ij} = (-A^i(w^i - w^j) + e(\gamma^i - \gamma^j) + e)_+,\ i \neq j,\ i,\ j = 1, \ldots, k \tag{9}$$

Since the inequalities (3) of the piecewise-linear separator are satisfied if and only if the minimum of (7) or equivalently of (8) is zero, we have the following result.

Theorem 2.2 (Multicategory Piecewise-linear Separation via Linear Programming) *The sets $\mathcal{A}^i$, $i = 1, \ldots, k$, represented by the $m^i \times n$ matrices A^i, $i = 1, \ldots, k$, are piecewise-linear separable if and only if the solvable linear program (8) has a zero minimum, in which case any solution $(w^i,\ \gamma^i,\ y^{ij})$, $i,\ j = 1, \ldots, k$, $j \neq k$, provides a piecewise-linear separation as defined by (3).*

As was the case for linear separation of two sets by linear programming [3, 4], it is important here also to rule out the null solution in case the sets $\mathcal{A}^i$, $i = 1, \ldots, k$ are not piecewise-linear separable. Since the piecewise-linear separation (3) is in effect achieved by a special pairwise linear separation between the sets $\mathcal{A}^i$, $i = 1, \ldots, k$, which is determined by $(w^i - w^j,\ \gamma^i - \gamma^i) \in R^n \times R^1$, $i \neq j$, $i, j = 1, \ldots, k$, it is therefore the nonzeroness of $w^i - w^j$, $i \neq j$, $i, j = 1, \ldots, k$ that matters. In [4] it was shown for the two-category case that $w^1 - w^2$ can be zero if and only if the $\mathcal{A}^1$ and $\mathcal{A}^2$ have equal means, in which case the null solution $w^1 - w^2 = 0$ is not a unique solution of the linear program. We shall derive generalizations of these two results to the multicategory case. Nonzeroness of $w^i - w^j$, $i \neq j$, $i, j = 1, \ldots, k$, is an important issue when one is trying to generate an approximate piecewise-linear separation (that is allow some errors in the separation) for sets which are not piecewise-linear separable. Zero $w^i - w^j$, $i \neq j$, $i, j = 1, \ldots, k$ will yield no information and no approximate separation for this case.

We give now a result that provides a necessary and sufficient condition for the occurrence of null $w^i - w^j$, $i \neq j$, $i, j = 1, \ldots, k$.

Theorem 2.3 (Null Solution Occurrence) *The linear program (8) has the null solution, $w^i - w^j = 0$, $i \neq j$, $i, j = 1, \ldots, k$ if and only if*

$$\frac{eA^i}{m^i} = \frac{1}{k-1} \sum_{\substack{j=1 \\ j \neq i}}^{k} \frac{eA^j}{m^j}, \quad i = 1, \ldots, k \tag{10}$$

Proof. The dual of the linear program (8) is

$$\max_{u^{ij}} \left\{ \sum_{i=1}^{k} \sum_{\substack{j=1 \\ j \neq i}}^{k} eu^{ij} \ \middle|\ \begin{array}{l} \sum_{\substack{j=1 \\ j \neq i}}^{k} (u^{ij} A^i - u^{ji} A^j) = 0, \quad i = 1, \ldots, k \\ \sum_{\substack{j=1 \\ j \neq i}}^{k} (-eu^{ij} + eu^{ji}) = 0, \quad i = 1, \ldots, k \\ 0 \leq u^{ij} \leq \frac{e}{m^i} \quad i \neq j,\ i = 1, \ldots, k \end{array} \right\} \tag{11}$$

The vectors $w^i - w^j = 0$, $i \neq j$, $i, j = 1, \ldots, k$ constitute an optimal solution for the primal LP (8) if and only if the equivalent minimization problem (7) is minimized by setting $w^i - w^j = 0$ and $\gamma^i - \gamma^j = 0$, $i \neq j$, $i, j = 1, \ldots, k$, which gives $y^{ij} = e$, $i \neq j$, $i, j = 1, \ldots, k$ and a primal minimum value of $k(k-1)$. Since $0 \leq u^{ij} \leq \frac{e}{m^i}$, $i \neq j$, $i = 1, \ldots, k$, and the dual optimal objective must equal the primal optimal objective of $k(k-1)$ we have

$$\sum_{i=1}^{k} \sum_{\substack{j=1 \\ i \neq j}}^{k} eu^{ij} = k(k-1)$$

It follows that

$$u^{ij} = \frac{e}{m^i}, \quad i \neq j,\ i = 1, \ldots, k$$

and

$$\frac{(k-1)}{m^i}\, eA^i = \sum_{\substack{j=1 \\ i \neq i}}^{k} \frac{eA^j}{m^j} \qquad i = 1, \ldots, k,$$

which is (10). $\square$

We now show even in the unlikely event that (10) is satisfied, the null solution $w^i - w^j = 0,\ i \neq j,\ i,\ j = 1, \ldots, k$, is not unique, and hence some nonzero solution will also be optimal.

Theorem 2.4 (Nonuniqueness of the Null Solution) *If condition (10) is satisfied, the null solution $w^i - w^j = 0,\ i \neq j,\ i,\ j = 1, \ldots, k$, to the linear program (8) is not unique.*

Proof. Let the primal solution to (8) be such that

$$\bar{w}^i - \bar{w}^j = 0,\ \bar{\gamma}^i - \bar{\gamma}^j = 0,\ \bar{y}^{ij} = e,\ i \neq j,\ i = 1, \ldots, k$$

Hence only the constraints $y^{ij} \geq 0$ of the linear program are inactive. It follows that this solution is unique in $\bar{w}^i - \bar{w}^j$ if and only if the following **has no solution for all** $h^{ij} \in R^n$, in the variables $(y^{ij} - \bar{y}^{ij},\ w^i - \bar{w}^i,\ \gamma^i - \bar{\gamma}^i),\ i \neq j,\ i,\ j = 1, \ldots, k$:

$$\begin{array}{l} \sum_{i=1}^{k} \sum_{\substack{j=1 \\ j \neq i}}^{k} - \frac{e}{m^i} (y^{ij} - \bar{y}^{ij}) \geq 0 \\ (y^{ij} - \bar{y}^{ij}) + A^i((w^i - w^j) - (\bar{w}^i - \bar{w}^j)) - e((\gamma^i - \gamma^j) - (\bar{\gamma}^i - \bar{\gamma}^j)) \geq 0 \\ \sum_{\substack{i,j=1 \\ i \neq j}}^{k} - h^{ij}((w^i - w^j) - (\bar{w}^i - \bar{w}^j)) > 0 \end{array}$$

By Motzkin's Theorem [13] this is equivalent to the following system **having a solution for all** $h^{ij} \in R^n,\ i \neq j,\ i,\ j = 1, \ldots, k$:

$$\begin{array}{l} \sum_{\substack{j=1 \\ j \neq i}}^{k} (u^{ij} A^i - u^{ji} A^j) = h^{ij},\ i = 1, \ldots, k \\ \sum_{\substack{j=1 \\ j \neq i}}^{k} (-eu^{ij} + eu^{ji}) = 0,\ i = 1, \ldots, k \\ -\frac{1}{m^i} e\zeta + u^{ij} = 0,\ i \neq j,\ i = 1, \ldots, k \\ (\zeta,\ u^{ij}) \geq 0,\ i \neq j,\ i = 1, \ldots, k \end{array}$$

This is obviously not true because there are $h^{ij},\ i \neq j,\ i,\ j = 1, ..., k$ in R^n that cannot be written as:

$$h^{ij} = \zeta(k-1) \left(\frac{eA^i}{m^i} - \frac{1}{k-1} \sum_{\substack{j=1 \\ j \neq i}}^{k} \frac{eA^j}{m^j} \right),\ \zeta \geq 0$$

Hence $\bar{w}^i - \bar{w}^j = 0$ is not unique. $\square$

3 Computational Results

We present now computational results that utilize the linear programming formulation (8) for discriminating between k sets for both the piecewise-linear separable and inseparable cases. Three different problems were considered: wine cultivar discrimination [1], iris classification [7], and breast

Problem	Classes k	Dimension n	Size	Percent Correct Training	Percent Correct Testing
Wine Cultivars	3	13	178	100.0	91.0
Iris Plants	3	4	150	98.7	96.7
Breast Cancer	3	11	122	66.3	56.6

Table 1: Performance of multicategory linear program on three problems. Correctness estimated using "leave-one-out" cross validation.

cancer prognosis [18]. The first two databases are available via anonymous file transfer protocol (ftp) from the University of California Irvine UCI Repository Of Machine Learning Databases [15]. The breast cancer prognosis data is available by request from the authors. Table 1 gives the number of classes, dimension, and size of each database as well as the results of using linear program (8) to discriminate between the classes. Cross validation testing was done by using the leave-one-out method [10, 8] to estimate the training set error and the testing set error (error on unseen examples). A brief discussion of the numerical results follows.

The Italian wine cultivar database is piecewise-linear separable. A single linear program was able to correctly separate the training set. The testing set performance was comparable to previously published results [1]. Fisher's classical iris discrimination problem is almost piecewise-linear separable and once again the multicategory linear programming approach performed quite well on both the training and testing sets. These problems illustrate that a single multicategory linear program can effectively discriminate sets that are totally (or almost totally) piecewise-linear separable.

The breast cancer prognosis problem is inherently more difficult. The breast cancer prognosis problem was created by dividing the Wisconsin breast cancer database into three classes: cancer which recurred (developed distant metastates) in less than 3 months, cancer which recurred in between 3 and 24 months, and cancers which did not recur in 24 months. The fact that the sets are not piecewise-linear separable is evidenced by the relatively poor training set accuracy shown in Table 1. A single linear program is insufficient for solving such problems. However, the results of several multicategory linear programs can be combined by using multisurface methods [12, 3, 2]. To demonstrate this, we used the multicategory linear program to create the multivariate splits in the multisurface method tree algorithm (MSMT) [2, 18].

MSMT-multicategory works by applying the linear program (8) to a k-class classification problem. The resulting piecewise-linear surface divides the space into k regions. If each of these k regions contains mostly points of one class, then we are done. If any region contains an unacceptable mixture of points then the linear program (8) is used again to divide that region into k or less regions. The resulting discriminant function can be thought of as a decision tree, thus the name **multisurface method tree - multicategory** (MSMT-MC).

The results using MSMT-MC are given in Table 2. These results show that using a multisurface approach enables the linear program (8) to be used for solving problems that are not piecewise-linear separable. By applying this approach to the breast cancer prognosis problem the training set accuracy improved over 40 percent and the testing set accuracy improved over 17 percent.

	Single LP	MSMT-MC	Change
Training Correctness (%)	66.3	93.0	+40.3%
Testing Correctness (%)	56.6	66.4	+17.3%
Average Number of LP's	1	5.1	+4.1

Table 2: Comparison of MSMT-MC and single multicategory linear program on breast cancer prognosis problem. Correctness estimated using "leave-one-out" cross validation.

References

[1] S. Aeberhard, D. Coomans, and O. de Vel. Comparison of classifiers in high dimensional settings. Technical Report 92-02, Departments of Computer Science and of Mathematics and Statistics, James Cook University of North Queensland, 1992.

[2] K. P. Bennett. Decision tree construction via linear programming. In M. Evans, editor, *Proceedings of the 4th Midwest Artificial Intelligence and Cognitive Science Society Conference*, pages 97–101, 1992.

[3] K. P. Bennett and O. L. Mangasarian. Neural network training via linear programming. In P. M. Pardalos, editor, *Advances in Optimization and Parallel Computing*, pages 56–67, Amsterdam, 1992. North Holland.

[4] K. P. Bennett and O. L. Mangasarian. Robust linear programming discrimination of two linearly inseparable sets. *Optimization Methods and Software*, 1:23–34, 1992.

[5] R. O. Duda and H. Fossum. Pattern classification by iteratively determined linear and piecewise linear discriminant functions. *IEEE Transactions on Electronic Computers Computers*, EC-15:220–232, 1966.

[6] R. O. Duda and P. E. Hart. *Pattern Classification and Scene Analysis.* John Wiley & Sons, New York, 1973.

[7] R.A. Fisher. The use of multiple measurements in taxonomic problems. *Annual Eugenics*, 7(Part II):179–188, 1936.

[8] K. Fukunaga. *Statistical Pattern Recognition.* Academic Press, New York, 1990.

[9] W. H. Highleyman. A note on linear separation. *IRE Transactions on Electronic Computers*, 10:777–778, 1961.

[10] P.A. Lahenbruch and R.M. Mickey. Estimation of error rates in discriminant analysis. *Technometrics*, 10:1–11, 1968.

[11] O. L. Mangasarian. Linear and nonlinear separation of patterns by linear programming. *Operations Research*, 13:444–452, 1965.

[12] O. L. Mangasarian. Multi-surface method of pattern separation. *IEEE Transactions on Information Theory*, IT-14:801–807, 1968.

[13] O. L. Mangasarian. *Nonlinear Programming.* McGraw–Hill, New York, 1969.

[14] T. S. Motzkin and I. J. Schoenberg. The relaxation method for linear inequalities. *Canadian Journal of Mathematics*, 6:393–404, 1954.

[15] P.M. Murphy and D.W. Aha. UCI repository of machine learning databases. Department of Information and Computer Science, University of California, Irvine, California, 1992.

[16] N. J. Nilsson. *Learning Machines.* MIT Press, Cambridge, Massachusetts, 1966.

[17] F. W. Smith. Design of multicategory pattern classifiers with two-category classifier design procedures. *IEEE Transactions on Computers*, 18:367–372, 1969.

[18] W.H. Wolberg, K.P. Bennett, and O.L. Mangasarian. Breast cancer diagnosis and prognostic determination from cell analysis. Manuscript, Departments of Surgery and Human Oncology and Computer Sciences, University of Wisconsin, Madison, WI 53706, 1992.