Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Letter to the Editor—A Note on Pattern Separation

Richard C. Grinold,

pected waiting time if the FIFO discipline were used (*see*, for example, reference 1) so,

$$E(L)=E(W_F)=[\sum_{i=1}^{i=k}\lambda_i E(P_i^2)]/2(1-\sum_{i=1}^{i=k}\rho_i)$$

By applying the theorem and simplifying, the corollary follows

REFERENCES

1 R W CONWAY, W L MAXWELL, AND L W MILLER, *The Theory of Scheduling*, Addison-Wesley, Reading, Mass, 1967
2 L KLEINROCK, "A Conservation Law for a Wide Class of Queueing Disciplines," *Naval Res Log Quart* **12**, 181–192 (1965)
3 J D C LITTLE, "A Proof of the Formula $L=\lambda W$," *Opns Res* **9**, 383–387 (1961)

A NOTE ON PATTERN SEPARATION

Richard C Grinold

University of California Berkeley California

(Received May 6 1969)

This note describes a linear programming method of separating two not necessarily disjoint convex polyhedral sets in E^n If the sets are disjoint, a strictly separating hyperplane is generated that maximizes and equalizes the distance between the sets and the hyperplane If the sets intersect a hyperplane is generated that minimizes the maximum error

THE SETS $\{A_i|i=1,2,\dots,m\}$ and $\{B_j|j=1,2,\dots,k\}$ are collections of points in E^n, their convex hulls being denoted C_a and C_b This note examines linear programming methods of separating C_a and C_b with a hyperplane If some hyperplane separates C_a and C_b strictly, the sets are separable If the sets are not separable, any hyperplane will leave points of C_a or C_b on the wrong side The distance of incorrect points from the hyperplane is called the error distance For example, if $\{x|xc=\gamma\}$ is the hyperplane, points in C_a should satisfy $xc>\gamma$, and $A_ic<\gamma$, then the error for point A_i is $\gamma-A_ic$

MANGASARIAN[1] has designed a linear program with $n+2$ constraints and $k+m$ nonnegative variables for separating C_a and C_b If the sets are disjoint, a strictly separating hyperplane is supplied that fends the sets off equally and maximizes the distance between set and hyperplane The technique detects a nonvoid intersection, but no hyperplane is supplied SMITH[3] has devised a computationally more difficult method for the nonseparable case that minimizes the total error distance If the sets are separable, any separating hyperplane has zero error and is thus optimal, and the program has $n+1$ constraints with $k+m$ upper and lower bounded variables (*see* reference 4) Mangasarian's multi-surface method[2] is designed for the nonseparable case, it constructs a piecewise linear function to separate the point sets In some situations, like the example at the end of this note, this may not help in deciding about new points

Before a pattern-separation problem is solved, the separability of C_a and C_b is unknown The method presented here is directed to this situation If the sets are separable, the results of reference 1 are duplicated, when the sets intersect, a hyperplane is generated that minimizes the largest error The linear program (3) has $n+2$ constraints, $k+m$ nonnegative variables, and one free variable A biomedical example and computational results using this algorithm and Smith's are available in reference 5

BACKGROUND, DEFINITIONS

THE NOTATION, motivation, and terminology used here are described in reference 1, pp 444–447, we will review these results by considering problem (19) on page 447 of this reference

$$\text{maximize } \alpha-\beta$$
$$\text{subject to } \quad Ac-e\alpha\geqq 0 \qquad -Bc+l\beta\geqq 0, \qquad f\geqq c\geqq -f \tag{1}$$

This problem always has an optimal solution, $(c, \bar{\alpha}, \bar{\beta})$

If the sets are separable, $\bar{\alpha}-\bar{\beta}$ is positive and $(\bar{c}, [\bar{\alpha}+\bar{\beta}]/2)$ is a separator, i e , $A\bar{c}-e(\bar{\alpha}+\bar{\beta})/2>0$ and $B\bar{c}-l(\bar{\alpha}+\bar{\beta})/2<0$ It can be shown in this case that $(\bar{c}, [\bar{\alpha}+\bar{\beta}]/2)$ solves

$$\max\{\min[A_i c-\gamma, B_j c+\gamma \mid i=1, 2, \ldots, m, j=1, 2, \ldots, k]\}$$

subject to $f\geqq c\geqq -f$ If the sets intersect, $\bar{\alpha}-\bar{\beta}$ is zero, and (0, 0, 0) solves (1)

Definitions (i) D is a $(k+m)\times(n+1)$ matrix

$$D=\begin{bmatrix} A & -e \\ -B & l \end{bmatrix}$$

(ii) D_i is the ith row of D

(iii) $d=(\sum_{i=1}^{k=m} D_i)/(k+m)$ is the row average

(iv) w is the $(n+1)$st column vector $\begin{pmatrix} c \\ \alpha \end{pmatrix}$

(v) For any w, the error on the ith pattern is $-\min[D_i w, 0]$

(vi) h is a $(k+m)$-dimensional column vector of ones

It is easy to see that w is a separator if and only if $Dw>0$ Also, for any w, the maximum error over all patterns is $-\min[0, D_i w \mid i=1, 2, \ldots, k+m]$

THE PROBLEM, RESULTS

WE PROPOSE to solve the following problem max min $[D_i w \mid i=1, 2, \ldots, k+m]$ subject to $dw=1$ The constraint is a normalization that bounds the solution set This problem can easily be transformed into a linear program

$$\text{maximize } \rho$$
$$\text{subject to } \quad Dw-h\rho\geqq 0 \quad \text{and} \quad dw=1, \tag{2}$$

where ρ is a scalar

Our main result is stated below, it is an extension of Theorem 1 of reference 1

THEOREM (i) *Problem* (2) *has an optimal solution iff* $d\neq 0$

(ii) *If* $(\bar{w}, \bar{\rho})$ *solves* (2) *and* $\bar{\rho}>0$, *then* $\bar{w}$ *defines a separation that maximizes the minimum of the* $D_i w$, *subject to* $dw=1$.

(iii) *If* $(\bar{w}, \bar{\rho})$ *solves* (2) *and* $\bar{\rho}<0$, *then no separation exists and* $\bar{w}$ *defines a hyperplane that minimizes the maximum error, subject to* $dw=1$.

Proof. First, note that some w solves $dw=1$ iff $d\neq 0$. If $d\neq 0$, then (w, ρ) satisfying $dw=1$ and $\rho=\min[D_i w \mid i=1, 2, \ldots, k+m]$ is feasible for (2).

Problem (3) is the dual of (2):

$$\begin{aligned} &\text{minimize } \beta \\ &\text{subject to } \quad D'u-d'\beta=0, \quad h'u=1, \quad u\geq 0 \end{aligned} \tag{3}$$

$(u, \bar{\beta})=(h/[k+m], 1)$ is always feasible for (3). When $d\neq 0$, both primal and dual are feasible and an optimal solution exists, establishing (i).

Optimality of $(\bar{w}, \bar{\rho})$ implies

$$\bar{\rho}=\min[D_i\bar{w} \mid i=1, 2, \ldots, k+m]\geq\min[D_i w \mid i=1, 2, \ldots, k+m] \tag{4}$$

for any w satisfying $dw=1$.

To demonstrate (iii), suppose $\bar{\rho}\leq 0$ and w satisfies $Dw>0$. For some positive scalar λ, $D\lambda w>0$ and $d\lambda w=1$, contradicting (4). Thus, if $\bar{\rho}\leq 0$, no separator exists. Note that

$$0\geq\bar{\rho}=\min[D_i\bar{w} \mid i=1, 2, \ldots, k+m]=\min[0, D_i\bar{w} \mid i=1, 2, \ldots, k+m] \tag{4}$$

This and (4) indicate that $\bar{w}$ minimizes the maximum error.

COMMENTS, AN EXAMPLE

THE HULLS C_a and C_b can be considered as figures in E^n with each point, A_i or B_j, having unit mass. Note that

$$(k+m)d=\left(\sum_{i=1}^{i=m} A_i-\sum_{j=1}^{j=k} B_j,\; m-k\right)$$

Thus, $d=0$ if and only if $m=k$ and $\sum_{i=1}^{i=m} A_i=\sum_{j=1}^{j=k} B_j$; this is equivalent to the two figures having equal mass and identical centers of gravity.

In most problems, $k+m\gg n$. Thus, solving (3) is more efficient.

As an example, suppose, unknown to the decision maker, that the points A_i and B_j arise with equal probability from two circular uniform distributions in the plane. The A_i distribution is centered at (1, 1) and the B_j at the origin. Both distributions have radius 1. The question is whether a new n-vector arises from the A distribution or the B. The linear decision function defined by $w=(1, 1, 1)$ minimizes the probability of a classification error. It is apparent that the solution of (2) will approximate the same decision surface for large sample size. Problem (1) would only indicate that the two sets overlap. This example, of course, presents (2) at its best.

ACKNOWLEDGMENT

THIS WORK was supported in part by the Office of Naval Research and the National Aeronautics and Space Administration at the Division of Engineering and Applied Physics, Harvard University.

REFERENCES

1 O L MANGASARIAN, "Linear and Nonlinear Separation of Patterns by Linear Programming," *Opns Res* **13,** 444–52 (1965)

2 ———, "Multi Surface Method of Pattern Separation," *IEEE Trans on Information Theory*, IT-4, **6,** 801-806 (1968)

3 F W SMITH, "Pattern Classifier Design by Linear Programming," *IEEE Trans on Computers*, C-17, **4,** 367–72 (1968)

4 R C GRINOLD, "A Comment on 'Pattern Classifier Design by Linear Programming,' " *IEEE Trans on Computers*, C-18, **4,** 378–9 (1969)

5 ———, "Mathematical Programming Methods of Pattern Classification," Div of Eng and App Physics Tech Rpt #591, Harvard (1969)