Published in final edited form as:

Nonlinear Anal Theory Methods Appl. 2001 August ; 47(3): 1893–1904. doi:10.1016/

S0362-546X(01)00319-4

Robust Set Separation Via Exponentials

Yelena Dandurova, Lana Yeganova, and James E. Falk

School of Engineering and Applied Science, The George Washington University, 707 22nd str. NW, Washington DC 20052

Abstract

Given a pair of finite disjoint sets A and B in Euclidean n -space, a fundamental problem with numerous applications is to efficiently determine a hyperplane $H(\omega, \gamma)$ which separates these sets when they are separable, or ‘nearly’ separates them when they are not. We seek a hyperplane that separates them in the sense that a measure of the Euclidean distance between the separating hyperplane and *all* of the points is as large as possible. This is done by ‘weighting’ points relative to $A \cup B$ according to their distance to $H(\omega, \gamma)$, with the closer points getting a higher weight, but still taking into account the points distant from $H(\omega, \gamma)$. The negative exponential is chosen for that purpose.

In this paper we examine the optimization problem associated with this set separation problem and characterize it (convex or non-convex).

AMS 1991 Subject Classification

15A45; 49N15; 90C25; 65K10

1. Introduction

Let $A = \{a_1, \dots, a_s\}$ and $B = \{b_1, \dots, b_t\}$ be two disjoint finite sets of points in R^n . We wish to *efficiently* determine a hyperplane $H(\omega, \gamma) = \{x: x \cdot \omega = \gamma\}$ which separates these sets in the sense that

$$a \cdot \omega \geq \gamma \text{ for all } a \in A$$

and

$$b \cdot \omega \leq \gamma \text{ for all } b \in B$$

if possible, or which ‘nearly’ separates them when A and B are not separable. If the sets *are* separable, there are generally an infinite number of hyperplanes that separate. Falk and Lopez Cardona [1] define the notion of a ‘robust’ separating hyperplane. This is a hyperplane which separates, and of all hyperplanes that separate, is of maximal (Euclidean) distance from the points in $A \cup B$. In non-separable case, Falk and Lopez Cardona determine a hyperplane which has the property that the distance between those points which are misclassified is minimal.

In either case, one criticism of these separators is that they are determined by a relatively small number of points in the set $A \cup B$ - those points which are closest to it.

In this paper we introduce a new separator which takes into account the distance between it and the set $A \cup B$ but is determined by *all* of the points in $A \cup B$. This is done by ‘weighting’ points relative to $A \cup B$ according to their distance to $H(\omega, \gamma)$, with the closer points getting a higher weight, but still taking into account the points distant from $H(\omega, \gamma)$. The negative exponential is chosen for that purpose.

Such problems typically arise in, e.g., medical diagnosis. In one well-known study [6] the points in $A \cup B$ are 9-dimensional vectors whose components correspond to attributes related to breast cancer (clump thickness, uniformity of cell size, ...). Each vector corresponds to a tissue sample of a patient who discovers a lump on her breast and each component measures pathologist’s judgment as to the severity of that attribute. The set A is the set of all patients in $A \cup B$ who’s lumps were cancerous, and B is the complement of that set (See [4],[5]).

After a separating hyperplane H is identified, it can be used to predict the set A or B to which a new patient $c \in A \cup B$ belongs, according to whether $c\omega - \gamma > 0$ or $c\omega - \gamma < 0$. Ties are not addressed here.

2. The Basic Optimization Problem

Let A and B be two disjoint finite sets in R^n , and $H(\omega, \gamma) = \{x: x \cdot \omega - \gamma\}$ be a hyperplane which separates these sets in the sense that

$$\begin{aligned} a \cdot \omega &\geq \gamma \text{ for all } a \in A \\ \text{and} \\ b \cdot \omega &\leq \gamma \text{ for all } b \in B. \end{aligned}$$

The distance between a point x and a hyperplane $H(\omega, \gamma)$ is

$$d(x, H) = \frac{|x \cdot \omega - \gamma|}{\|\omega\|}$$

or simply

$$d(x, H) = |x \cdot \omega - \gamma|$$

if we restrict ω to have norm 1. Points $a \in A$ should, if possible, have a distance $d(a, H) = a \cdot \omega - \gamma \geq 0$ and points $b \in B$ should, if possible, have a distance $d(b, H) = \gamma - b \cdot \omega \geq 0$.

The function

$$e^{-(x \cdot \omega - \gamma)}$$

will be smaller for those properly classified points in A farthest from H than for those properly classified points in A closest to H . It also will be larger for any misclassified point in A than for any properly classified point in A .

Likewise the function

$$e^{-(\gamma - x \cdot \omega)}$$

will be largest for misclassified points in B , and smallest for properly classified points in B , with the farthest from H points in B having the smallest value.

In order to account for possible differences in the number of points in A and B (s and t), we

will weight the terms $\sum_{i=1}^s e^{-(a_i \cdot \omega - \gamma)}$ and $\sum_{j=1}^t e^{-(\gamma - b_j \cdot \omega)}$ with the coefficients $(1/s)$ and $(1/t)$ respectively.

Thus, the function

$$\begin{aligned} f(\omega, \gamma) &= \left(\frac{1}{s}\right) \sum_{i=1}^s e^{-(a_i \cdot \omega - \gamma)} + \left(\frac{1}{t}\right) \sum_{j=1}^t e^{-(\gamma - b_j \cdot \omega)} \\ &= e^{\gamma} \left(\frac{1}{s}\right) \sum_{i=1}^s e^{-a_i \cdot \omega} + e^{-\gamma} \left(\frac{1}{t}\right) \sum_{j=1}^t e^{b_j \cdot \omega} \end{aligned}$$

will tend to be minimized over the set $S = \{(\omega, \gamma): \|\omega\| = 1\}$ at a point which will separate A and B when they are separable, or give a ‘nearly’ separating hyperplane.

Thus, we are led to study the problem

$$\min_{\|\omega\|=1} f(\omega, \gamma) \quad (1)$$

First we can eliminate γ .

For fixed ω the problem has the form

$$\min S e^{\gamma} + T e^{-\gamma}$$

where $S = \left(\frac{1}{s}\right) \sum_{i=1}^s e^{-a_i \cdot \omega}$ and $T = \left(\frac{1}{t}\right) \sum_{j=1}^t e^{b_j \cdot \omega}$ are positive constants. This is a convex function, so the minimizer satisfies

$$S e^{\gamma} - T e^{-\gamma} = 0$$

so that

$$e^{\gamma} = \sqrt{\frac{T}{S}}.$$

Thus, with ω given, the threshold γ is

$$\gamma = \left(\frac{1}{2}\right) \ln \left(\frac{T}{S}\right)$$

and above problem becomes

$$\min 2 \sqrt{ST}$$

which is equivalent to the problem

$$\min ST$$

or

$$\left(\frac{1}{st}\right) \min_{\|\omega\|=1} \left(\sum_{i=1}^s e^{-a_i \cdot \omega} \right) \left(\sum_{j=1}^t e^{b_j \cdot \omega} \right).$$

This seems somewhat cumbersome, but since the terms in brackets are positive, one can take logs and address the problem

$$\min_{\|\omega\|=1} \left[\ln \left(\sum_{i=1}^s e^{-a_i \cdot \omega} \right) + \ln \left(\sum_{j=1}^t e^{b_j \cdot \omega} \right) \right]. \quad (2)$$

In this form, we can characterize the objective as convex.

Lemma 1

The function

$$g(s) = \ln \left(\sum_{i=1}^n e^{s_i} \right)$$

is convex in $s = (s_1, \dots, s_n)$.

Proof—We compute

$$g'(s) = \left(\frac{1}{\sum_{i=1}^n e^{s_i}} \right) (e^{s_1}, \dots, e^{s_n}).$$

Let $S = \left(\sum_{i=1}^n e^{s_i} \right)^{-1}$. The Hessian of g is

$$g''(s) = \text{diag}(e^{s_1}/S, \dots, e^{s_n}/S) - \begin{pmatrix} e^{s_1}/S \\ \vdots \\ e^{s_n}/S \end{pmatrix} (e^{s_1}/S, \dots, e^{s_n}/S)$$

and we need to show this is positive semi-definite.

Note that the entries $e^{s_1}/S, \dots, e^{s_n}/S$ are positive and sum to one.

Let $p_1, \dots, p_n$ be positive numbers which sum to one. We need to show that

$$(z_1, \dots, z_n) \left[\text{diag}(p_1, \dots, p_n) - \begin{pmatrix} p_1 \\ \vdots \\ p_n \end{pmatrix} (p_1, \dots, p_n) \right] \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix} \geq 0$$

for all z , i.e.,

$$\sum_{i=1}^n p_i z_i^2 \geq \left(\sum_{i=1}^n p_i z_i \right)^2 \quad (3)$$

Let Z be a random variable defined by

$$P(Z=z_i)=p_i, \quad i=1, \dots, n.$$

The variance of Z , denoted by $\text{Var}(Z)$, is defined as follows [8]

$$\text{Var}(Z)=E\left[(Z-E(Z))^2\right]=\sum_{i=1}^n p_i(z_i-\bar{z})^2.$$

Since $\text{Var}(Z)$ is the expected value of the nonnegative random variable $(Z-E(Z))^2$, it follows that $\text{Var}(Z) \geq 0$. On the other hand, the value of the variance of Z is

$$\text{Var}(Z)=E(Z^2)-(E(Z))^2=\sum_{i=1}^n p_i z_i^2-\left(\sum_{i=1}^n p_i z_i\right)^2 \geq 0.$$

Therefore, inequality (3) follows from the non-negativity of variance.

Nevertheless, problem (2) is non-convex due to the constraint $\|\omega\|=1$.

3. Isovalue Contours and Separability

The problem which we are solving is

$$\min_{\|\omega\|=1} \left[\ln \left(\sum_{i=1}^s e^{-a_i \cdot \omega} \right) + \ln \left(\sum_{j=1}^t e^{b_j \cdot \omega} \right) \right].$$

Let

$$G(\omega) = \ln \left(\sum_{i=1}^s e^{-a_i \cdot \omega} \right) + \ln \left(\sum_{j=1}^t e^{b_j \cdot \omega} \right).$$

Note that

$$G(0) = \ln s + \ln t$$

which is positive since $s, t \geq 1$. A ‘level’ set of G is any set of the form

$$G(\omega) \leq \sigma$$

where σ is some real number.

Theorem

The level sets of $G(\omega)$ are unbounded if and only if the sets A and B are separable.

Proof—Assume the sets A and B are separable. Let the hyperplane $H(\omega, \gamma)$ separate them, so that

$$a_i \cdot \omega \geq \gamma \text{ for } i=1, \dots, s$$

and

$$b_j \cdot \omega \leq \gamma \text{ for } j=1, \dots, t$$

Now look at $G(\lambda\omega)$ for $\lambda > 0$:

$$\begin{aligned} G(\lambda\omega) &= \ln \left(\sum_{i=1}^s e^{-a_i \cdot \lambda\omega} \right) + \ln \left(\sum_{j=1}^t e^{b_j \cdot \lambda\omega} \right) \\ &= \ln \left(\sum_{i=1}^s (e^{-a_i \cdot \omega})^\lambda \right) + \ln \left(\sum_{j=1}^t (e^{b_j \cdot \omega})^\lambda \right). \end{aligned}$$

As a function of λ , $G(\lambda\omega)$ is convex.

Now since A and B are separable,

$$e^{-a_i \cdot \omega} \leq e^{-\gamma} \text{ for all } i$$

and

$$e^{b_j \cdot \omega} \leq e^{\gamma} \text{ for all } j$$

which implies that

$$\begin{aligned} G(\lambda\omega) &\leq \ln \left(\sum_{i=1}^s (e^{-\gamma})^\lambda \right) + \ln \left(\sum_{j=1}^t (e^{\gamma})^\lambda \right) \\ &= \ln (s \cdot e^{-\gamma\lambda}) + \ln (t \cdot e^{\gamma\lambda}) \\ &= \ln s + \ln t. \end{aligned}$$

Thus we see the picture:

Hence $G(\lambda\omega)$ is bounded above by $\ln s + \ln t$ and below by zero. For $\sigma = \ln s + \ln t$, we know that above the ray $\lambda\omega$, the function $G(\lambda\omega)$ decreases.

Now suppose that

$$G(\lambda\omega) = \ln \left(\sum_{i=1}^s (e^{-a_i \cdot \omega})^\lambda \right) + \ln \left(\sum_{j=1}^t (e^{b_j \cdot \omega})^\lambda \right) > \ln s + \ln t$$

for some pair (ω, γ) with $\omega \neq 0$ and for some $\bar{\lambda} > 0$. Since $G(\lambda\omega)$ is convex in λ , $G(\lambda\omega)$ must remain larger than $\ln s + \ln t$ for all $\lambda \geq \bar{\lambda}$. it follows that

$$\left(\sum_{i=1}^s (e^{-a_i \cdot \omega})^\lambda \right) \left(\sum_{j=1}^t (e^{b_j \cdot \omega})^\lambda \right) > st$$

or

$$\sum_{i=1, j=1}^{s, t} (e^{-a_i \cdot \omega + b_j \cdot \omega})^\lambda > st$$

for all $\lambda \geq \bar{\lambda}$. This means that, for at least one i, j pair,

$$e^{-a_i \cdot \omega + b_j \cdot \omega} > 1$$

or, equivalently,

$$-a_i \cdot \omega + b_j \cdot \omega > 0$$

which means that there can be no γ such that

$$a_i \cdot \omega \geq \gamma \geq b_j \cdot \omega.$$

Thus the sets A and B are not separable.

Only one case remains – the case where

$$G(\lambda\omega) = \ln s + \ln t$$

For some $\omega \neq 0$ and for all $\lambda \geq 0$ (i.e. G is constant). This implies that

$$G(\lambda\omega) = \ln \left(\sum_{i=1}^s (e^{-a_i \cdot \omega})^\lambda \right) + \ln \left(\sum_{j=1}^t (e^{b_j \cdot \omega})^\lambda \right) = \ln s + \ln t$$

for all $\lambda \geq 0$. It follows that

$$\sum_{i=1, j=1}^{s, t} (e^{-a_i \cdot \omega + b_j \cdot \omega})^\lambda = st$$

for all $\lambda \geq 0$ which can only happen if

$$-a_i \cdot \omega + b_j \cdot \omega = 0$$

for all pairs i, j . But this means there is some common value $\gamma = a_i \cdot \omega = b_j \cdot \omega$ which simply means that all of the vectors a_i and b_j lie on the same hyperplane $H(\omega, \gamma)$. They are thus weakly separable.

The implication of this result is that when A and B are strongly separable, we can replace the problem

$$\min_{\|\omega\|=1} \left[\ln \left(\sum_{i=1}^s e^{-a_i \cdot \omega} \right) + \ln \left(\sum_{j=1}^t e^{b_j \cdot \omega} \right) \right]$$

by the *convex* problem

$$\min_{\|\omega\| \leq 1} \left[\ln \left(\sum_{i=1}^s e^{-a_i \cdot \omega} \right) + \ln \left(\sum_{j=1}^t e^{b_j \cdot \omega} \right) \right]$$

However, this is not so in the non-separable case and Branch and Bound [2] must be used.

Note that the use of the exponential function e^{-t} is arbitrary and, in fact, can be replaced by the more general function $e^{-\lambda t}$ where λ is a positive parameter. Then, more general problem is

$$\min_{\|\omega\|=1} \left[\ln \left(\sum_{i=1}^s e^{-\lambda \cdot a_i \cdot \omega} \right) + \ln \left(\sum_{j=1}^t e^{\lambda \cdot b_j \cdot \omega} \right) \right] \quad (4)$$

with the threshold

$$\gamma = \left(\frac{1}{2\lambda} \right) \ln \left(\frac{T}{S} \right)$$

$$\text{where } S = \left(\frac{1}{s} \right) \sum_{i=1}^s e^{-\lambda a_i \cdot \omega} \text{ and } T = \left(\frac{1}{t} \right) \sum_{j=1}^t e^{\lambda b_j \cdot \omega}.$$

In some cases a hyperplane produced by our model with $\lambda=1$ does not separate the sets. However, setting λ large enough we can always find a separating hyperplane (See Fig. 3.1).

It can be shown that the hyperplane produced with large λ approaches the hyperplane produced by the Falk-Lopez model ('Simple' Separator) [1] (Fig. 3.1).

Also the separating hyperplane produced is invariant under transformations that preserve congruence, i.e., if A and B undergo a transformation T which is a translation, reflection, rotation, scaling, or combination of these, then our model chooses the hyperplane $T(H)$ separating $T(A)$ and $T(B)$ when our model will choose H separating A and B .

4. Numerical Results

Example 1

Separable Sets

Let

$$A = \begin{bmatrix} -1.4 & 0 \\ 0.3 & -1.05 \\ -0.35 & -1.5 \\ -1.65 & -1.2 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} -0.7 & 1.2 \\ 0.6 & -0.1 \\ 0.7 & 0.2 \\ 0.85 & 0.7 \\ 1.25 & 1.15 \\ 1.65 & 0.7 \end{bmatrix}$$

where the rows of the matrices are the points of the sets A and B (with some abuse of notation).

The appropriate problem is

$$\min_{\|\omega\|=1} \left[\ln \left(e^{1.4\omega_1} + e^{-0.3\omega_1 + 1.05\omega_2} + e^{0.35\omega_1 + 1.5\omega_2} + e^{1.65\omega_1 + 1.2\omega_2} \right) + \ln \left(e^{-0.7\omega_1 + 1.2\omega_2} + e^{0.6\omega_1 - 0.1\omega_2} + e^{0.7\omega_1 + 0.2\omega_2} + e^{0.85\omega_1 + 0.7\omega_2} + e^{1.25\omega_1} \right) \right]$$

The plot below (Fig. 4.1) shows the isovalue contours of the objective function plotted against the circle $\|\omega\|=1$.

The optimal solution (By MATLAB) is $\omega = (-0.618, -0.7862)$ with $\lambda = 0.1372$, i.e., the separating hyperplane is $(-0.618 \ x_1 - 0.7862 \ x_2 - 0.1372 \ 0)$ (See Fig. 4.2 below).

Note that the level sets depicted in Fig. 4.1 above are unbounded. Note also that the sets A and B are separable.

Example 2

Non-Separable Sets

Let

$$A = \begin{bmatrix} 0.49 & -2.1 \\ 0.4 & -0.7 \\ 1.1 & 0 \\ 1.25 & -0.7 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} -1.2 & -0.9 \\ -0.1 & 0.9 \\ 0.49 & -1.7 \\ 0.6 & 0.2 \\ 0.8 & -1.1 \\ 1 & 0.6 \end{bmatrix}.$$

The appropriate problem is

$$\min_{\|\omega\|=1} \left[\ln \left(e^{-0.49\omega_1 + 2.1\omega_2} + e^{-0.4\omega_1 + 0.7\omega_2} + e^{-1.1\omega_1} + e^{-1.25\omega_1 + 0.7\omega_2} \right) + \ln \left(e^{-1.2\omega_1 - 0.9\omega_2} + e^{-0.1\omega_1 + 0.9\omega_2} + e^{0.49\omega_1 - 1.7\omega_2} + e^{0.6\omega_1 + 0.2\omega_2} + e^{0.8\omega_1 - 1.1\omega_2} + e^{1\omega_1 + 0.6\omega_2} \right) \right]$$

The isovalue contours of the objective function are shown in Fig. 4.3 below.

The optimal solution (By MATLAB) is $\omega = (0.8992, -0.4375)$ with $\lambda = 0.8423$. The sets A and B and their separator are shown below (Fig. 4.4).

In this example, the level sets of the objective function are bounded, and also the sets A and B are not separable.

References

1. Falk JE, Lopez-Cardona E. The Surgical Separation of Sets. *Journal of Global Optimization*. 1997; 11:1–31.

2. Falk JE, Soland RM. An Algorithm for Separable Nonconvex Programming Problems. *Management Science*. 1969; 15:550–569.
3. Mangasarian OL. Linear and Nonlinear Separation of Patterns. *Operations Research*. 1965; 13:444–452.
4. Mangasarian, OL., Setiono, WH., Wolberg, R. *Large-Scale Numerical Optimization*. SIAM; Philadelphia: 1990. *Pattern Recognition via Linear Programing: Theory and Applications to Medical Diagnosis*; p. 22-30.
5. Mangasarian O, Street N, Wolberg W. Breast Cancer Diagnosis and Prognosis via Linear Programming. *Operations Research*. 1995; 43:570–577.
6. Murphy, PM., Aha, DW. *UCI repository of machine learning databases*. Department of Information and Computer Science, University of California; Irvine, California: 1992.
7. Panik, MJ. *Fundamentals of Convex Analysis – Duality, Separation, Representation, and Resolution*. Kluwer; Amsterdam: 1993.
8. Ross, SM. *Introduction to Probability Models*. Academic Press; San Diego: 1997.

Figure 3.1.
'Simple' separator approximates the 'weighted' separator with large λ .

Figure 4.1.
Isovalue contours of the objective function of example 1

Fig. 4.2.
The sets A and B of example 1 and their separator

Figure 4.3.
Isovalue contours of the objective function of example 2.

Fig. 4.4.
The sets A and B of example 2 and their separator