Ellipsoidal Separation: A Medical Diagnostic Systems Technique.

by: ROBERT W. NEWCOMB, Ph.D. (Stanford), Mem. IEEE

ABSTRACT

A technique for separating clustered date through the use of n-dimensional ellipsoids surrounding the clusters is discussed. The ellipsoids are minimal in the sense that a maximum number of data points lie on their boundaries. The theory is set up for the computer aided classification of patients according to the presence or absence of a disease, though it is applicable to other similar situations.

LIST OF PRINCIPAL SYMBOLS

α test data n-vector

α_i i th initially classified data n-vector

c_α center of gravity n-vector of α_i's

d_i diagonal entry D = ellipsoid axis intercept (squared)

D diagonal matrix determining the ellipsoid

I_j jth index set = numbers from 1 to N with j–1 numbers deleted

n_i ith norm

$n_i^{(j)}$ ith norm in jth iteration

σ orthogonal n x n matrix =rotation to principal axes

$\phi_{ik}^{(j)}$ k th entry of ϕ_i in the jth iteration

$\phi, \phi^{(j)}$ shifted data vectors

$\Phi, \Phi^{(j)}$ nonsingular matrices, columns consisting of vector ϕ's

$R, R^{(j)}$ rotation n x n matrices

$R_{i_j}^{(i)}$ rotation (n–j+1) x (n–j+1) matrices

x, y column n-vectors of entries x^i, y^i

$\langle x,y \rangle$ scalor product of vector x and y

$\|x\|$ norm of vector x

+Professor Robert Newcomb is a visiting Professor in Electrical Engineering at Universiti Teknologi Malaysia, Gurney Road, Kuala Lumpur, Malaysia.

$'$ matrix transpose

I_n n x n identity matrix

$\dot{+}$ direct sum of matrics $A \dot{+} B = \begin{bmatrix} A & O \\ O & B \end{bmatrix}$

INTRODUCTION

In the area of medical diagnostics one desires methods of accurately specifying the presence or absence in a patient of a given disease. For, it is clear that in certain cases the decision to apply curative techniques, including operation, involves a question of life or death upon which one desires no mistakes. Toward computerization for these decisions, present techniques use linear separating (hyperplane) surfaces for classifying data as either normal or diseased. But linear surfaces may not give an accurate picture, especially for borderline or otherwise unclassifiable casses. Consequently, we begin here a discussion of the next step, that is, the use of quadratic separating surfaces.

In the medical diagnostic case data tends to cluster. Consequently the idea to be developed is that of passing an n-dimensional ellipsoid around the center of gravity of a cluster of n-dimensional data points. Medically, initial clusters are obtained from a large sample of patients known to either possess or not to possess the disease under question. This gives two clusters each within an ellipsoid with classification of a new patient under test occuring if the patient's data points lie inside a region of one of the ellipsoids nonoverlapping with the other; other data points outside or inside both ellipsoids are unclassifiable, necessitating further testing and observation.

The following then gives a discussion of possible mathematics for the testing of data points to see if they lie within a given ellipsoid. For this the primary problem is that of characterising the ellipsoid, so the setting up of an encircling ellipsoid is the main concern of our treatment.

MATHEMATICAL NOTATION AND FORMULATION OF THE PROBLEM

Although the ideas are relatively simple, notation is rather a problem and worth getting well fixed. Our data is assumed to be real and consists of many characteristics (such as body temperature, blood pressure, etc.); hence we will work with n-vectors $x = \begin{bmatrix} x^1 \\ \vdots \\ x^n \end{bmatrix}$ in the real n-dimensional vector space R^n. Here superscript indices are used for the vector's components in order to reserve subscripts for indexing different vectors; one can consider that the superscripts indicate components of contravariant vectors. The transpose of a vector, or a matrix, is denoted by a superscript prime $'$ and the scalar product of any two n-vectors x and y by $< x,y >$. Thus,

$$< x,y > = x'y = \sum_{i=1}^{n} x^i y^i$$

If K is any positive definite (real symmetric) matrix, we take as the definition of an n-dimensional (solid) ellipsoid the set of vectors $x \in R^n$ satisfying

$$< x, Kx > \leqslant 1 \tag{1}$$

The situation is somewhat simplified by observing that K can be diagonalized by an orthogonal matrix σ

$$\sigma K \sigma' = D^{-1}, D = \text{diagonal } [d_1 \ldots, d_n], d_i > 0 \tag{2}$$

This diagonalization represents a rotation of axes through

$$y = \sigma x \tag{3}$$

which transforms (1) into

$$< y, D^{-1} y > = \sum_{i=1}^{n} \frac{(y i)^2}{d_i} \leqslant 1 \tag{4}$$

Our problem is to determine the orthogonal matrix σ and the diagonal weighting factor matrix D given data n-vectors α_i for N patents, i = 1, . . ., N, of known classification. For these determinations of D and σ we will deal with the norm $\| x \| = \sqrt{< x,x >}$, working initially with the maximum norm to insure the proper major axis of an ellipsoid.

A PROCEDURE

Although it is easy to find a number of solid spheres which encompass finite data sets, by simply choosing large enough radii, we desire minimal ellipsoids. Thus, we develop ellipsoids which contain a maximum number of data points on their boundary surfaces.

By perhaps the exclusion of redundant data, consider that N linearly independent data n-vectors α_i, i = 1, 2, . . ., N with $N \leqslant n$, of similar and known classification are given and with an ordering to be specified shortly. We first form their "center of gravity" vector.

$$c_\alpha = \frac{1}{N} \sum_{i=1}^{N} \alpha_i \tag{5}$$

A new origin which will be the center for our ellipsoid is choosen as c_α, that is, we work with shifted vectors

$$\phi_i = \alpha_i - c_\alpha \qquad i \in [1, 2, \ldots, N] \tag{6}$$

The ordering assumed, which is always obtianable by a permutation of subscripts, is that the norms

$$n_i = \| \phi_i \| \tag{7}$$

satisfy

$$n_1 \geqslant n_i \qquad i \epsilon [2, \ldots, N] \tag{8}$$

A rotation of coordinates, around the new origin at c_α, is now undertaken through the use of an orthogonal rotation transformation matrix R. The desired effect of R is to bring ϕ_1 in line with the x^1 coordinate axis. For this we form the nonsingular n x n matrix.

$$\Phi = [\Phi_1 \mid \Phi_2 \mid \ldots \mid \Phi_n] \tag{9}$$

The columns of Φ are orthogonalized by using the Gram-Schmidt procedure, implemented through a matrix T, by beginning with the first column of Φ and working toward the nth. Using a partition to emphasize the first column we can write.

$$R' = \Phi T = [\Phi_1 \mid \ldots \mid \Phi_n] \begin{bmatrix} 1 & \\ n_1 & \\ 0 & \\ \vdots & \\ 0 & \end{bmatrix} = \frac{[\Phi_1 \mid \ldots]}{n_1} \tag{10}$$

To this point, using 1_n as the n x n identity matrix, we have

$$RR' = 1_n \tag{11}$$

which shows that

$$R \phi_1 = \begin{bmatrix} n_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{12}$$

Thus R is an orthogonal matrix which rotates the vector ϕ_1 of maximum norm to the x^1 axis. In line with equations (3) and (4) we set

$$y_1 = R \phi_1 \tag{13a}$$

$$< y_1, D^{-1} y_1 > = \frac{(n_1)^2}{d_1} \leqslant 1 \tag{13b}$$

Choosing ϕ_1 on the boundary of the ellipsoid being generated gives, from (13b),

$$d_1 = (n_1)^2 \tag{14}$$

We proceed in somewhat the same manner with the other coordinates, though the presence of a prior coordinate introduces further steps.
As ϕ has already been used we wish to remove it by deleting $i = 1_1$ from our index set, while we also wish to concentrate on the (n–1) remaining dimensions. Thus, using super-script (2) to indicate the second (interative) step, we write

$$R \phi_i = \begin{bmatrix} \phi_{i1}{}^{(2)} \\ \phi_i{}^{(2)} \end{bmatrix} i \epsilon [2, \ldots, N] = [1, 2, \ldots, N] - [1] \tag{15a}$$

where the scaler $\phi_{i1}{}^{(2)}$ is the first component and

$$\phi_i{}^{(2)} = \begin{bmatrix} \phi_{i2}{}^{(2)} \\ \vdots \\ \phi_{in}{}^{(2)} \end{bmatrix} \tag{15b}$$

is an (n–1) - vector. Again form the norms of these

$$n_i{}^{(2)} = \| \phi_i{}^{(2)} \| \qquad i \epsilon\ 2, \ldots, N \tag{16}$$

Now, however, due to the different contributions of the first components $\phi_{i1}{}^{(2)}$ to the ellipsoid, we desire to rotate each of these vectors $\phi_i{}^{(2)}$ individually in its turn to the x^2 axis. Thus, by placing the vector $\phi_i{}^{(2)}$ first in $\phi_i{}^{(2)}$ constructed as at (9), we generate, using the method described above, N–1 rotation (n–1) x (n–1) matrices $R_i{}^{(2)}$ such that

$$R_i{}^{(2)} \phi_i{}^{(2)} = \begin{bmatrix} n_i{}^{(2)} \\ 0 \\ \vdots \\ 0 \end{bmatrix} i \epsilon \quad 2, \ldots, N \tag{17}$$

Having done this, and using + to denote the direct sum of two matrices, we set

$$y_i{}^{(2)} = [1 + r R_i{}^{(2)}]\ R \phi_i \qquad i \epsilon\ 2, \ldots, N \tag{18a}$$

$$< y_i{}^{(2)}, D^{-1}\ y_i{}^{(2)} > = \frac{(\phi_{i1}(2))^2}{d_1} + \frac{(n_i(2))^2}{d_2} \leqslant 1 \tag{18b}$$

where d is known from (14).

We wish to choose d_2 such that this inequality holds for all i and with equality for at least one i. If we define the index set

$$I_2 = [2, \ldots, N] = [1, \ldots, N] - [i_1], i_1 = 1 \tag{19}$$

then we have

$$d_2 = \max_{i \in I_2} \frac{(n_i^{(2)})^2}{1-[\phi_{i1}^{(2)})^2 / d_1]} \tag{20}$$

Choosing one of the indices, i_2, giving this maximum defines

$$R^{(2)} = 1 \dot{+} R_{i_2}^{(2)} \tag{21}$$

We next repear using the index set $1_3 = [1, \ldots, N] - [i_1, i_2]$ concentrating on the (n-2) - vectors $\phi_i^{(3)}$ in

$$R^{(2)} \; R\phi_i = \begin{bmatrix} \phi_{i1}^{(3)} \\ \phi_{i2}^{(3)} \\ \phi_i^{(3)} \end{bmatrix} \quad i \in I_3 \tag{22}$$

Continuing yields, on defining the index set $1_j = [1, \ldots, N] - [i_1, i_2, \ldots, i_j - 1]$

$$d_j = \max_{i \in I_j} \frac{[(n_i^{(j)})^2]}{1 - \sum_{k=1}^{j-1} [(\phi_{ik}^{(j)})^2 / d_k]} \tag{23a}$$

$$\sigma = R^{(n)} \; R^{(n-1)} \; \ldots . R^{(2)} \; R \tag{23b}$$

$$R^{(j)} = 1_{j-1} \dot{+} R_{i_j}^{(j)} \tag{23c}$$

Our ellipsoid is defined by (4)

$$\langle y, D^{-1} y \rangle \leq 1 \quad , \; D = d_1 \dot{+} d_2 \dot{+} \ldots \dot{+} d_n \tag{24a}$$

where for any test data n-vector α we form

$$\phi = \alpha - c_\alpha \tag{24b}$$

$$y = \sigma \alpha$$

This ellipsoid has at least n of the N known classification vectors α_i on the boundary and all others inside; in this sense it is minimal and any vector α which satisfies $\langle y, D^{-1} y \rangle \leq 1$ is to be classified with the classification of the original α_i.

CONCLUSIONS – DISCUSSION

A method has been described for setting up ellipsoids to surround clustered data, as occurs in the field of medical diagnostics. For each cluster of data a separate ellipse is set up and new data (patients) are tested to see if classification can be obtained within any of the clusters.

As yet the method remains to be programmed, though this should be straightforward. For this, more efficiency may be obtained by using the maximum $n_i^{(j)}$ at each, jth, step, to avoid the maximization of (20), followed by a test, (24a), of all remaining ϕ_i; if, however, one of the ϕ_i fail this test it must replace the one used and the process repeated. Some numerical examples need to also be carried out for n>2, especially to determine the nature of overlaps and to obtain a feeling for the minimality property.

ACKNOWLEDGEMENTS

The author wishes to acknowledge Professor N. Declaris as the source for the problem and O. Ijaola for discussion on various aspects. Further, the hospitality of his Malay hosts is gratefully acknowledged as well as their enthusiasm which lends a solid dimension to life. For YAMPSHbSHJ.

"Orang yang bertanam pokok nyior, kadang-kadang tiada makan buahnya"

[1, p.22]

REFERENCES

1. Winstedt, Sir Richard, *Malay Proverbs* (John Murray, London, W., 1950).