homework solutions for:

Homework #3: Classification using Linear Discriminant Analysis

submitted to:

Dr. Joseph Picone
ECE 8993 Fundamentals of Speech Recognition

April 30, 1998

submitted by:

Jonathan Hamaker

Institute for Signal and Information Processing
Department of Electrical and Computer Engineering
Mississippi State University
Box 9571, 216 Simrall, Hardy Rd.
Mississippi State, Mississippi 39762
Tel: 601-325-8335, Fax: 601-325-3149
Email: hamaker@isip.msstate.edu

Introduction

In this exercise, we use Linear Discriminant Analysis (LDA) to explore a problem similar to that solved with Principle Components Analysis (PCA) in a previous homework assignment. In particular, we study the classification characteristics of class-independent and class-specific LDA using synthetic data distributions. LDA is, in essence, a transform-based method which attempts to minimize the within-class distance from the mean and attempts to maximize the out-of-class distance from the mean. This provides maximum discriminability for classification problems. The theory for these processes is well-founded and clearly explained in [1, 2]. Our main concern in this assignment is to compare the decision regions for each of the LDA techniques and to determine why and how LDA chooses these regions as the discriminating points.

The Famous Ellipses

In the first portion of this problem we examine two identical elliptical sets, one offset from the other, as shown in Figure 1. We see that these sets have variances in the same directions. This is the type of situation which should be well-suited for either class-independent or class-specific
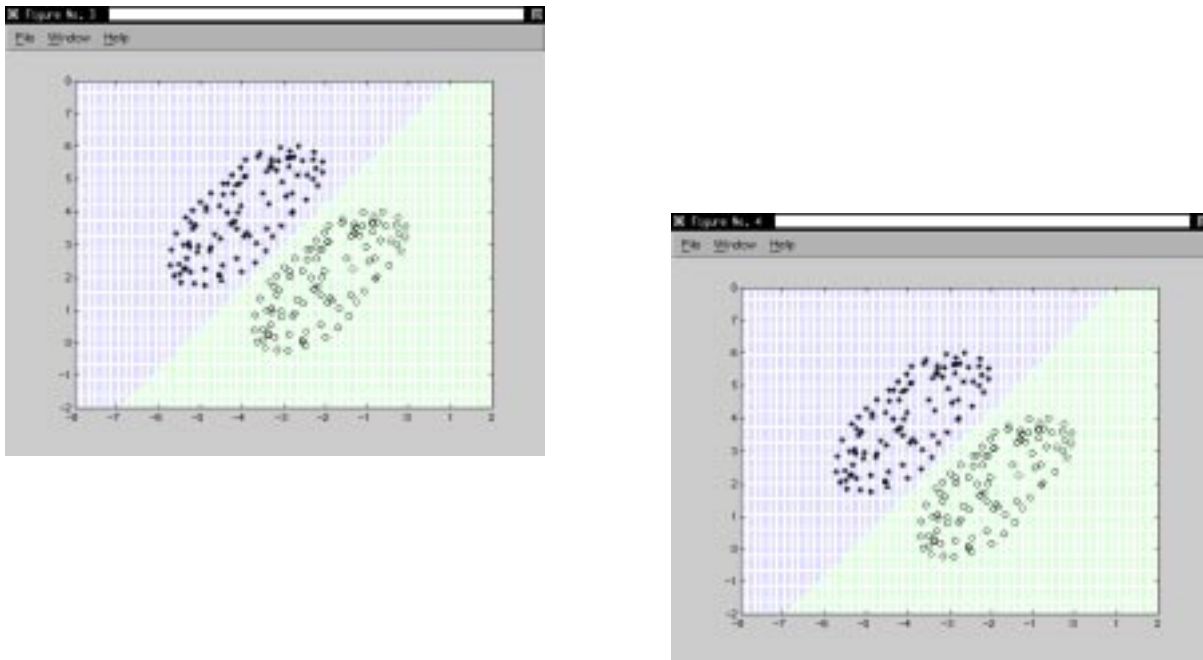

Figure 1. Decision regions shown with LDA. Notice that the region defined without a transform is coincident with the LDA decision region and that both trace out an equidistant line from the untransformed means. This equidistant line is orthogonal to a line connecting the means of the distributions. In this case LDA is an optimal classifier since its decision region traces the optimal decision region. The line perpendicular to the line drawn between means is optimal because the sets are identical and parallel.

LDA as it only requires that a linear decision region be found. We see from Figure 1 that the decision region for both class-independent and class-specific LDA are equivalent. Another note is that the LDA decision line falls on top of the line representing the decision region in the untransformed space (i.e. the decision boundary found if you just measure the distance from the means of each class and classify based on the smaller distance). This, again, is due to the equivalent shapes and parallel positioning of the classes. This fact would not hold true if we were to rotate either of the classes by even a small amount.

Overlapping Sets

Our second example deals with another classical case in which it is known that both PCA and -LDA fail. Figure 4 gives the distributions and decision regions for both class-specific and class-independent LDA. The figure shows that this is a particularly difficult case because the distribution means are identical and the classes overlap. There is some available discriminability since the classes are rotated about the mean from each other, but this discriminability is slight at best. This is a case where we would want to add another feature to separate the classes along a different dimension. We see from Figure 4 that class-specific LDA does a pretty good job of dividing the two classes with minimal errors while the class-independent LDA model does a very poor job. This is due to the class-specific model's ability to simultaneously model the within-class data and reject the out-of-class data individually for each class.

Ellipses and Pears

The last example is the same problem tackled with PCA in a previous homework assignment.

1)   Two sets were defined with the distributions shown in Figure 2. Set 1 was constructed with a mean of approximately μ_1=(-2, 2) and Set 2 with a mean of approximately μ_2=(2, -2). Each set contains 100 points. The sets were designed such that the primary directions of variance for each were not equivalent.

2)   Next we defined a sample set of four points — three of which lie on the line y=x and the other shifted slightly toward set 2. Since the means among the sets are symmetric about the line y = x, we expect that points on this line will not be easily classified.

$$x_1 = (-1, -1);$$
$$x_2 = (0, 0);$$
$$x_3 = (1/2, 1/2);$$
$$x_4 = (1/2, -1/2)$$

3)   The euclidean distances were calculated using equation 1. These distances are shown in Table 1.

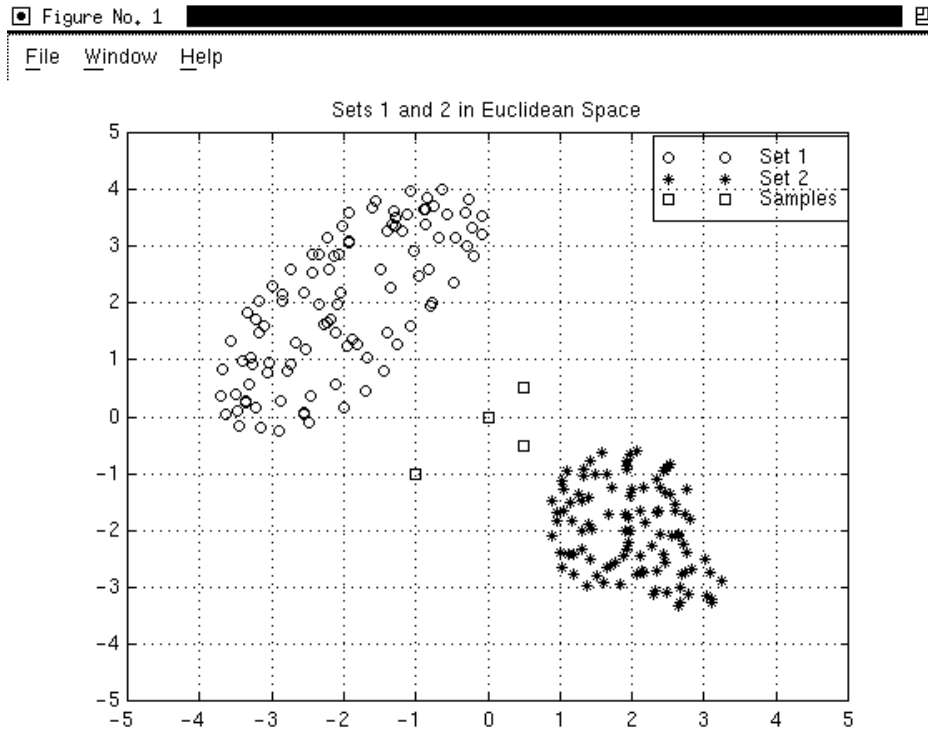$$d(z, \mu) \;=\; \sqrt{(x - \mu_x)^2 + (y - \mu_y)^2} \text{ where z = [x, y].} \qquad (1)$$

Figure 2. Distributions of sets 1 and 2 in the original space and the sample points used in this experiment

	x_1	x_2	x_3	x_4
μ_1	**3.1623**	**2.8285**	**2.9155**	3.5356
μ_2	3.1623	2.8285	2.9155	**2.1213**

Table 1: Distance between the indicated vectors in the untransformed space. The bold type indicates the set to which each sample point was classified. Note that, though x_1, x_2, and x_3 are equidistant from both sets — each is arbitrarily classified into set 1.

4) From euclidean distances we are also able to determine the curve which represents equidistance from both sets. The functional form of this curve is y = x and is plotted in Figure 3.

5) Next we use class-independent LDA transforms to find the classification of the test points. The distances in the transform space are shown in Table 1. With class-independent transforms, the decision region in a two dimensional space is simply a
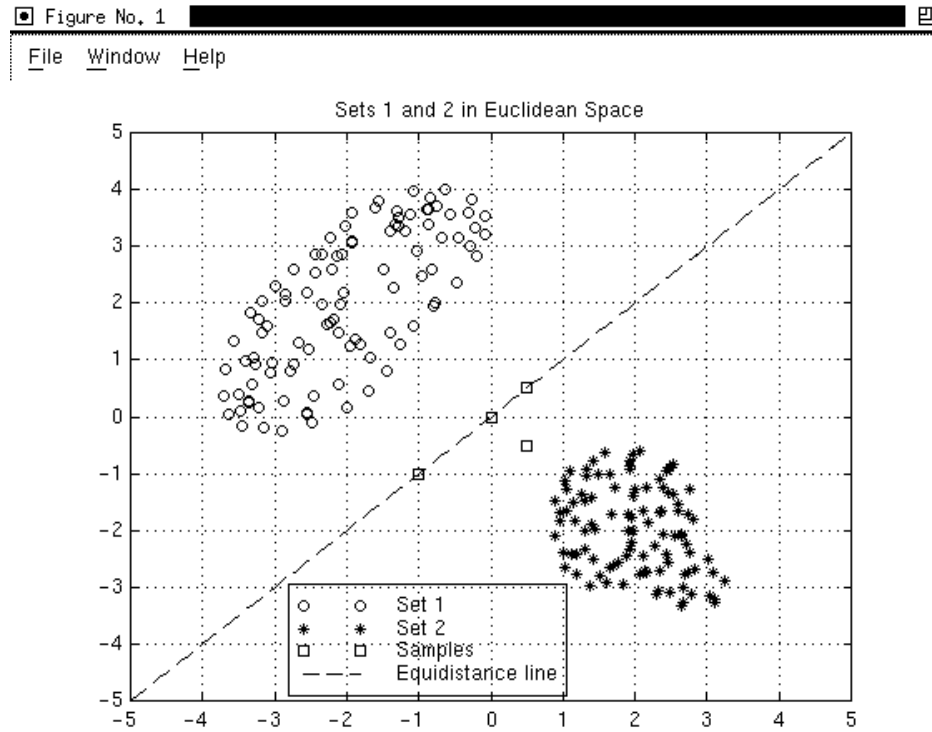
Figure 3. Equidistant curve plotted in the untransformed space

line. This differs from the class-specific transform in that a single linear transform is used instead of two, separate transforms. The transform produces a linear decision region which gives more importance to the pear-shaped class in its primary direction of variance and likewise to the elliptical set. The decision region for this system is shown in Figure 5.

	x_1	x_2	x_3	x_4
μ_1	**2.6093**	2.8021	2.8984	3.5026
μ_2	2.9948	**2.8020**	**2.7057**	**2.1015**

Table 2: Distance between the indicated vectors in the transformed space using class-independent LDA. The bold type indicates the set to which each sample point was classified. Note that the classification has changed quite a bit from the Euclidean classification. Those points hovering close to the origin have now been grouped into the pear-shaped class since its primary direction of variance passes approximately through the origin.

6)   Lastly, we apply a class-specific LDA transform procedure to the data sets. We note that the primary direction of variance of each set is approximately orthogonal to that of the other set. This orthogonality leads to a slight parabolic shape of the decision

regions given by the class-specific transforms as shown in Figure 5. The elliptical set has a variance at approximately a 45 degree angle, while the pear-shaped distribution has a primary variance in the direction of 135 degrees from zero. Thus, the model for set 1 gives more importance to outlying points in the direction of 45 degrees than does the model for set 2. The classification of the data points is shown in Table 3.

	x_1	x_2	x_3	x_4
μ_1	**2.6416**	2.8087	2.8922	3.5108
μ_2	3.0937	**2.7369**	**2.5585**	**2.0527**

Table 3: Distance between the indicated vectors in the transformed space using class-specific LDA. The bold type indicates the set to which each sample point was classified. Notice that the class-specific transforms created only a slight change in the decision region and no change in the classification of the data points.

Software

All software for this assignment was written using the Matlab scripting language and can be found at: *http://www.isip.msstate.edu/resources/courses/ece_8993_speech/homework/1998/ problem_03/hamaker.* To use the software: download all files in the directory given above; switch to the directory on your local system containing all of the "m-files" and start Matlab. At the command prompt, type "homework_03". The script will display a number of useful data items and will produce plots used to analyze the performance of LDA.

References

[1]   K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, San Diego, California, 1990, pp 24-30.

[2]   S. Balakrishnama, A. Ganapathiraju, J. Picone "Linear Discriminant Analysis - A Brief Tutorial," *Institute for Signal and Information Processing*, March 2, 1998.
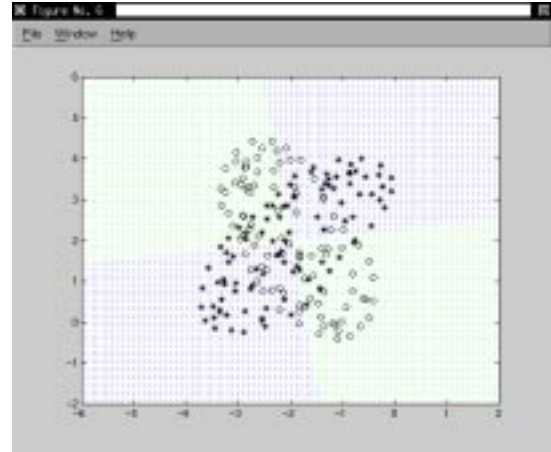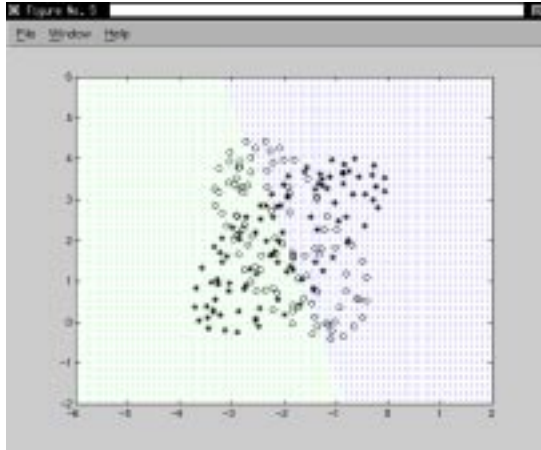
Figure 4. Classes transformed using LDA with a class-independent transforms and class-specific transforms. Notice that the class-specific transform does a much better job of classifying the two region though it retains some unavoidable error due to the overlap in the classes.
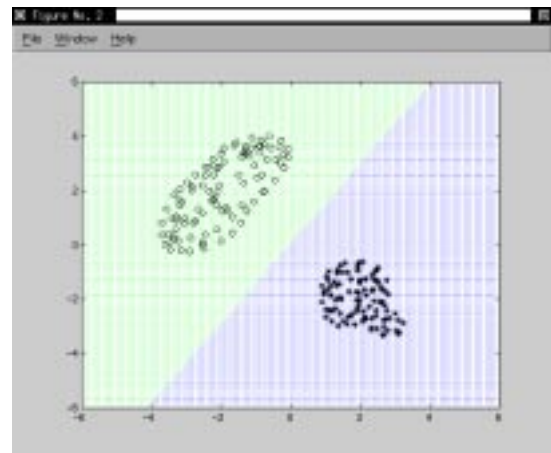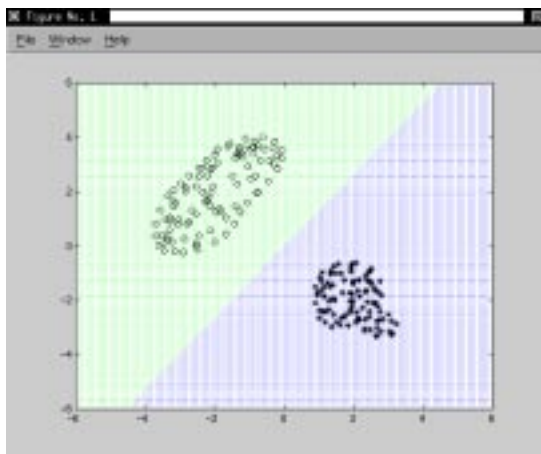

Figure 5. Classes transformed using LDA with class-independent and class-specific transforms. Note the slight curvature in the class-specific case. The decision regions in these cases are nearly equivalent since the classes are well separated.