

**EE 8993: Speech Recognition**

Homework Assignment #3  
Linear Discriminant Analysis

January 30, 1998

*submitted to:*

Dr. Joseph Picone

Department of Electrical and Computer Engineering  
413 Simrall, Hardy Rd.  
Mississippi State University  
Box 9571  
MS State, MS 39762

*submitted by:*

Julie Ngan

Department of Electrical and Computer Engineering  
Mississippi State University  
Box 9571  
Mississippi State, Mississippi 39762  
Tel: 601-325-8335  
Fax: 601-325-3149  
email: [ngan@isip.msstate.edu](mailto:ngan@isip.msstate.edu)



## I. Problem Definition

Define two sets such that the distributions approximate an ellipse and a pear shape. The elliptical distribution should stretch from lower-left to upper-right at about a  $45^\circ$  angle and should be longer in that direction than it is wide. Set 2 should look like a pear with the stem pointing at  $-45^\circ$ . Set 1 should have a mean of approximately  $m_1 = (-2, 2)$  and Set 2 should have a mean of approximately  $m_2 = (2, -2)$ . Each set contains 100 points.

Define a test set of four points:

$$x_1 = (-1, -1)$$

$$x_2 = (0, 0)$$

$$x_3 = (1/2, 1/2)$$

$$x_4 = (1/2, -1/2)$$

These two data sets will be analyzed to find the decision regions using Linear Discriminant Analysis (LDA). The sample points will be classified using the results of class independent and class specific LDAs.

## II. Data Set Generation

The two test sets are generated using the point operation function in *xmgr*. Points are randomly drawn to obtain the required shapes and then vertically or horizontally shifted to obtain the required means. A plot of the two data sets is shown in Figure 1. Two other test cases are also used to test the robustness of LDA. Plots of the two cases, case 2 and case 3, are shown in Figure 2 and Figure 3 respectively.

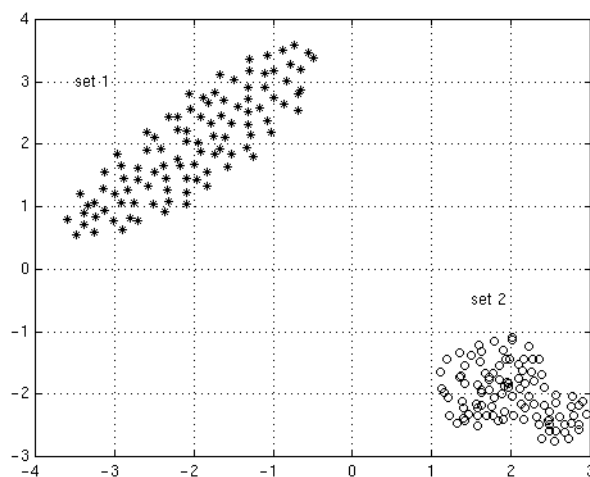


Figure 1 The two generated data sets.

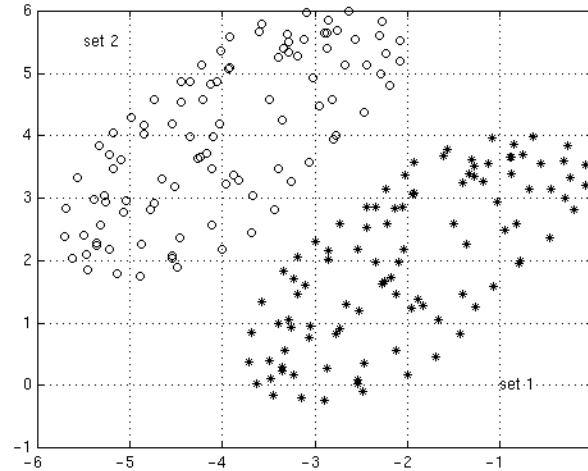


Figure 2 Plot of data sets for case 2.

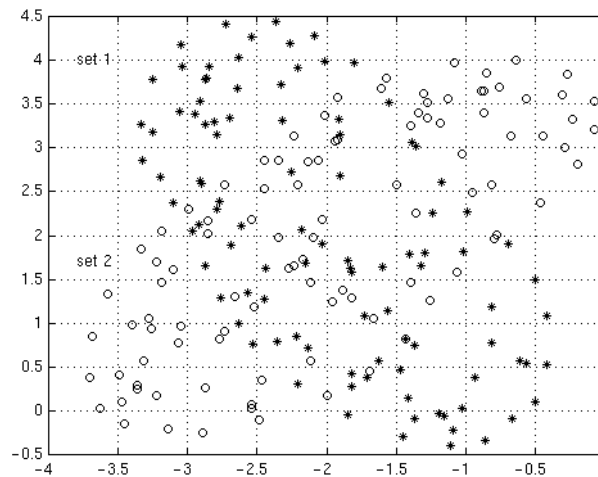


Figure 3 Plot of data sets for case 3.

It is observed that in the original space, the data sets in case 1 and 2 are linearly separable, whereas the sets in case 3 are not. A class-dependent multitransformation method is performed on the first test case using the steps describe below. Then a class-independent single transformation method is used, and the results are compared. Using the same methods, experiments on case 2 and 3 are performed.

### III. Scatter Matrices

In order to formulate criteria of class separability [1] for discriminant analysis, we need to find the within-class, between-class, and mixture scatter matrices. The within class scatter matrix is calculated by:

$$S_w = \sum_{i=1}^L P_i E\{(X - M_i)(X - M_i)^T | \omega_i\} = \sum_{i=1}^L P_i \Sigma_i \quad (1)$$

The mixture mean or the expected vector of the mixture distribution, is computed using:

$$M_0 = E[X] = \sum_{i=1}^L P_i M_i \quad (2)$$

Since there are 100 data points in both set 1 and set 2, the two classes are equally probable.  $M_0$  can then be calculated by:

$$\frac{1}{2} \times (M_1 + M_2) \quad (3)$$

where  $M_1$  and  $M_2$  are the means of set 1 and set 2 respectively. The between class scatter matrix is then calculated by:

$$S_b = \sum_{i=1}^L P_i (M_i - M_0)(M_i - M_0)^T. \quad (4)$$

Therefore, the mixture scatter matrix is the covariance matrix of all samples regardless of their class assignment:

$$S_m = E\{(X - M_0)(X - M_0)^T\} = S_b + S_w. \quad (5)$$

The calculated values for the first test case are:

$$S_b = \begin{bmatrix} 4 & -4 \\ -4 & 4 \end{bmatrix}$$

$$S_w = \begin{bmatrix} 0.4480 & 0.2441 \\ 0.2441 & 0.4173 \end{bmatrix}$$

$$S_m = \begin{bmatrix} 4.4480 & -3.7559 \\ -3.7559 & 4.4173 \end{bmatrix}$$

The criterion of class separability used in this exercise is:

$$J_1 = tr(S_2^{-1} S_1), \quad (6)$$

where  $S_2$  and  $S_1$  are one of the scatter matrices. Hence,  $J_1$  is large when the between class scatter is large or the within class scatter is small. The optimal solution for the problem would try to maximize the value of  $S_1$  while minimizing the value of  $S_2$ . In this exercise, we have used the between class scatter as  $S_1$  and the within class scatter as  $S_2$ . Using the new feature vectors, the eigen values and eigen vectors are found and arranged in order according to the values of the eigen vectors. Only the most significant eigen vector values are kept to avoid the problem of singularity matrix. Then the data sets are transformed into the new space. Since we are performing class specific LDA, the within class scatter is nothing but the covariances of the classes. Therefore, instead of using the  $S_w$  as calculated above, we have two separate  $S_w$  for the two data sets:

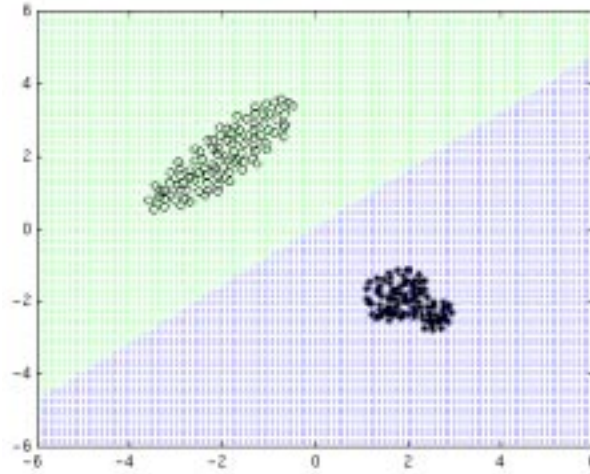


Figure 4 The new decision region using class specific LDA.

$$S_{w1} = \begin{bmatrix} 0.6617 & 0.5572 \\ 0.5572 & 0.6633 \end{bmatrix} \quad S_{w2} = \begin{bmatrix} 0.2343 & -0.0690 \\ -0.0690 & 0.1713 \end{bmatrix}.$$

The decision regions are found by arbitrarily scanning through the space for points that are equidistant from both sets. Region that is closer to the mean of set 1 is shaded with green while region that is closer to the mean of set 2 is shaded with blue. Figure 4 shows a plot of the two data sets in the original space along with the decision regions. The test vectors are transformed into the new space and the Euclidean distance between the test vectors and the means of the transformed sets are calculated. The results are shown in Table 1.

Sample point	Euclidean distance from data set 1	Euclidean distance from data set 2	Classification
x1	2.8275	2.4288	set 2
x2	2.8284	2.7531	set 2
x3	2.8289	2.9152	set 1
x4	3.5355	2.0648	set 2

Table 1 The Euclidean distance from the transformed sample points to the transformed means using class specific LDA.

#### IV. Class Independent LDA

In order to perform class independent LDA for data classification, the within class matrix is calculated by combining the two data sets together and only one feature space is found. The two data sets are transformed into the new space using the same transformation, rather than two different transformations. Figure 5 shows a plot of the two data sets in the original space along with the decision regions. The test vectors are transformed into the new space and the Euclidean distance between the test vectors and the means of the transformed sets are calculated. The results

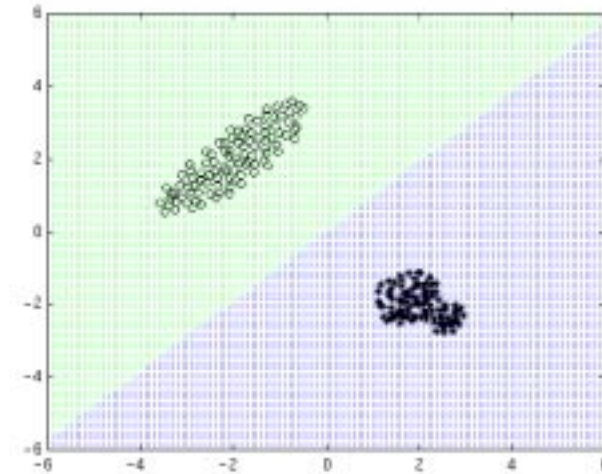


Figure 5 The new decision region using class independent LDA.

are shown in Table 2.

Sample point	Euclidean distance from data set 1	Euclidean distance from data set 2	Classification
x1	2.8598	2.7956	set 2
x2	2.8277	2.8277	set 2
x3	2.8117	2.8437	set 1
x4	3.5346	2.1208	set 2

Table 2 The Euclidean distance from the transformed sample points to the transformed means using class independent LDA.

It is noted that test vector x2 is equidistant from both data sets, even though it is assigned to set 2 by the algorithm default.

## V. Case 2—Parallel Ellipses

Both the class independent LDA and class specific LDA are applied to the second case. The second case consists of two elliptically shaped data sets that are identical. One set is shifted from the other both horizontally and vertically. However, their variances stay the same. Figure 6 and Figure 7 shows plots of the two data sets in the original space along with the decision regions using class specific LDA and class independent LDA respectively. The region that is closer to the mean of set 1 is shaded with green while the region that is closer to the mean of set 2 is shaded with blue. It is observed that the decision regions using the two different LDA methods are the same. This is because the two data sets are of the same shape, have the same variance, and are parallel to each other.

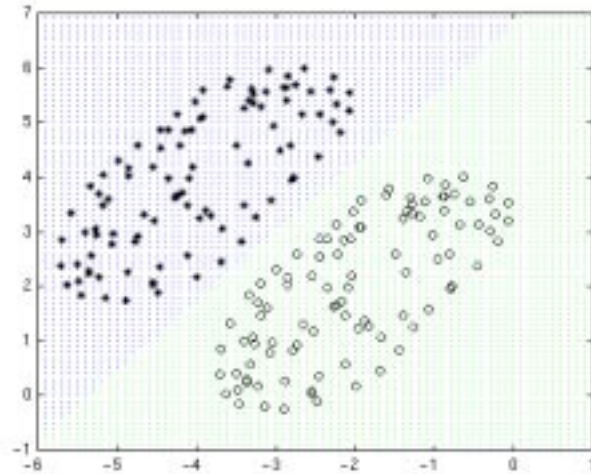


Figure 6 Decision regions for class specific LDA for case 2.

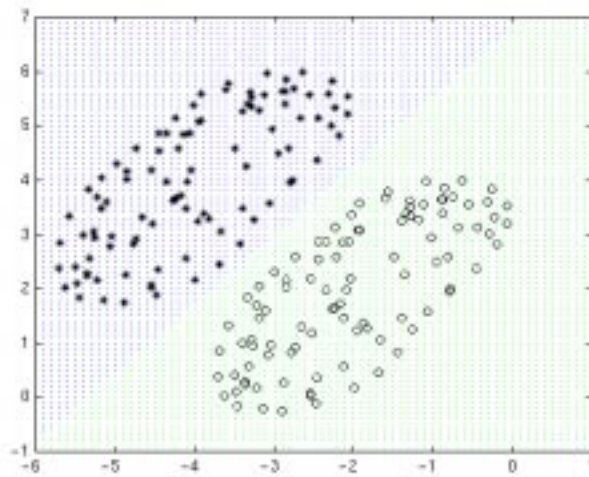


Figure 7 Decision regions for class independent LDA for case 2.

## VI. Case 3—Overlapping

In this case, the two data sets are of elliptical shapes with one stretching from lower-left to upper-right at about a  $45^\circ$  angle and the other one perpendicular to the first set. Both ellipses are longer in the stretching direction than their width. Both class specific LDA and class independent LDA are applied to both data sets to find the decision regions. This problem is of particular interest since the two data sets are not separable due to the overlap and that there will always exist some errors in classifying the two sets. Figure 8 and Figure 9 shows plots of the two data sets in the original space along with the decision regions using class specific LDA and class independent LDA respectively. The region that is closer to the mean of set 1 is shaded with green while the region that is closer to the mean of set 2 is shaded with blue. It is observed that with the class

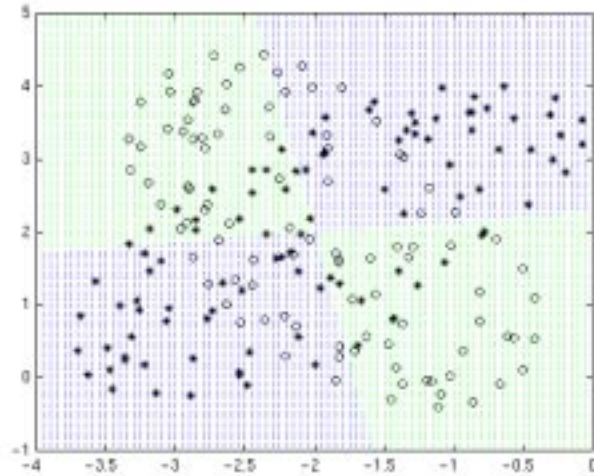


Figure 8 Decision regions for class specific LDA for case 3.

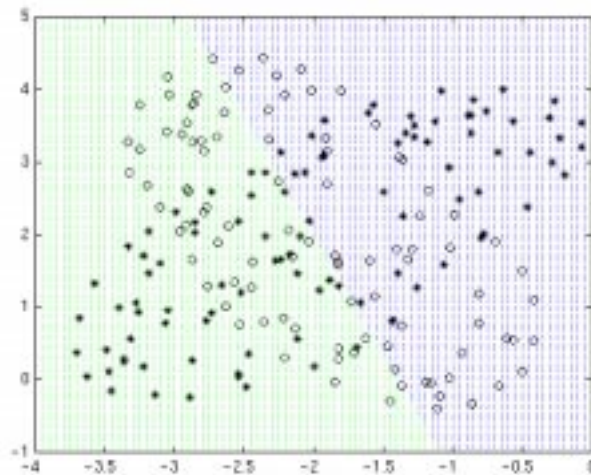


Figure 9 Decision regions for class independent LDA for case 3.

independent LDA, it does a poor job in classifying the data since there are only two decision regions. This is because in class independent LDA, the two data sets are transformed into the same space and in order to separate the data, a decision line is drawn where about half of each set of data is classified correctly. With the class specific LDA, it does a much better job in finding the decision regions with minimal errors. This is because the two data sets are transformed separately into their own new space and LDA is able to define more class-specific decision regions for each set.

## VII. REFERENCES

- [1] K. Fukunaga, "Introduction to Statistical Pattern Recognition," Academic Press, San Diego, California, 1990.