homework for

Linear Discriminant Analysis

EE 8993: Fundamentals of Speech Recognition

May 15, 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:

Janna Shaffer

I.   Original Data Set Calculations

This homework assignment applies linear discriminant analysis(LDA) to a classification problem. The data sets in Figure 1 were hand drawn using a MATLAB graphical user interface. The means for the two sets are $\begin{bmatrix} 2 & -2 \end{bmatrix}$ and $\begin{bmatrix} 2 & -2 \end{bmatrix}$. The test set consists of the points $\begin{bmatrix} -1 & 1 \end{bmatrix}$, $\begin{bmatrix} 0 & 0 \end{bmatrix}$, $\begin{bmatrix} 0.5 & 0.5 \end{bmatrix}$, and $\begin{bmatrix} 0.5 & -0.5 \end{bmatrix}$.
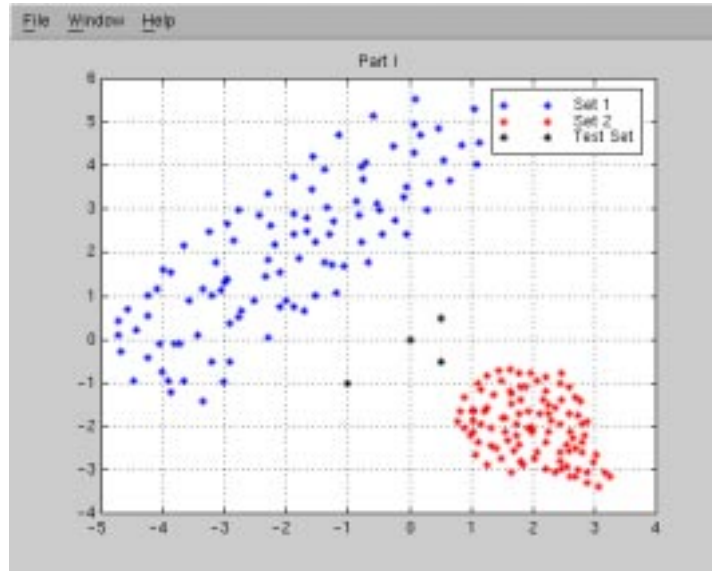

Figure 1.   Data sets and test set used for LDA

After the data and test sets were defined, the Euclidean distances between each point in the test set and the data set means were calculated. Set membership was based upon this distance. Three out of the four test points were equidistant from the two sets. In these cases, set membership could be with either set. The fourth test point clearly belongs with data set 2. These distances are shown in Table 1.

	Distance from mean 1	**Distance from mean 2**
x1	3.1623	3.1623
x2	2.8284	2.8284
x3	2.9155	2.9155
x4	3.5355	**2.1213**

Table 1: Euclidean Distances Between Test Sets and Data Set Means

II.   LDA

In order to find a better classification of the test data, a linear discriminant analysis was applied to the data sets and test sets [1]. Table 2 displays the new distances calculated from the data in the transform region. Figure 2 shows the two data sets in the new transform region.

	Distance from mean 1	Distance from mean 2
x1	3.1176	**2.9287**
x2	2.7614	**2.4705**
x3	2.8546	**2.5926**
x4	3.4518	**1.8529**

Table 2: Euclidean Distances Between Test Sets and Data Set Means
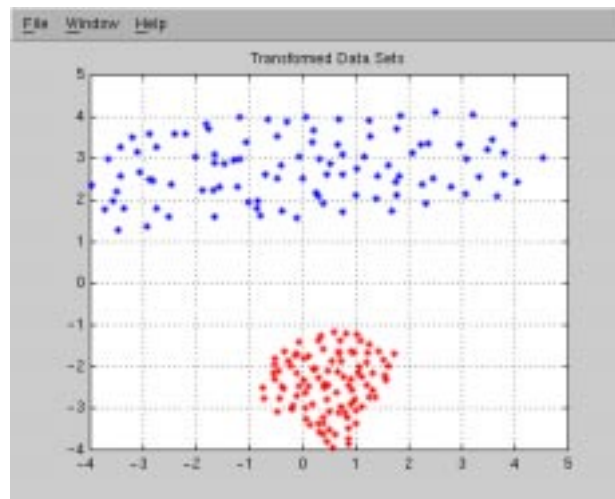

Figure 2. Data Sets in the Transform Region

The non-transformed data sets and test set are plotted in Figure 3 with the lines representing the classification of data in the original space and in the transform space. From this plot the set membership of each test point is clear. The division line in the original space is linear while that of the transform space is slightly parabolic. The reason that the transformed space is parabolic is because two transforms are used to divide the original space for classification [2].
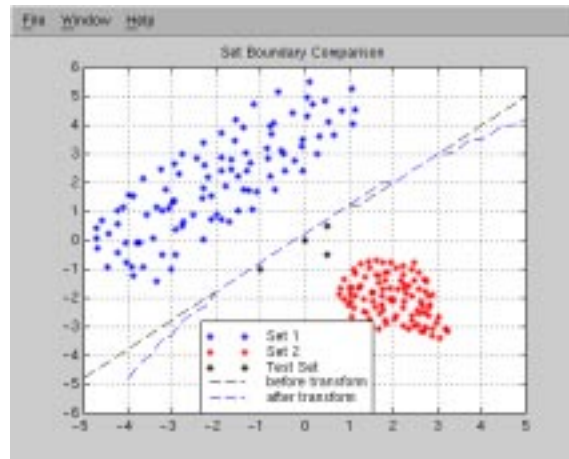
Figure 3.   Classification Lines in the Original and Transform Regions

III.   REFERENCES

[1]   Brown, S. Balakrishnama, J. Picone, "Scenic Beauty Estimation using Linear Discriminant Analysis," MS State ECE 4773 Semester Project December 11, 1997.

[2]   Nagendra Kumar, Ph. D. Thesis. John Hopkins University, 1996.