Avik Bag
Professor Matt Burlick
CS 383 – Assignment 1
25th January 2017

Theory Questions

<u>Question 1</u>

a) The goal is to perform Principal Component Analysis on the raw data that is provided.

$$Raw\ Data = \begin{bmatrix} -2 & 1 \\ -5 & -4 \\ -3 & 1 \\ 0 & 3 \\ -8 & 11 \\ -2 & 5 \\ 1 & 0 \\ 5 & -1 \\ -1 & -3 \\ 6 & 1 \end{bmatrix}$$

This raw dataset will need to be standardized. This is done by finding the average and standard deviation along the columns. Using these values, the data is centered by subtracting the average value and then dividing the standard deviation along the respective columns. Here are the values computed below.

$$\mu = [-0.9 \quad 1.4]$$
$$\sigma = [4.228 \quad 4.273]$$

After standardization:

$$Standardized\ Data = \begin{bmatrix} -0.2602 & -0.0936 \\ -0.9697 & -1.2635 \\ -0.4967 & -0.0936 \\ 0.2129 & 0.3744 \\ -1.6792 & 2.2462 \\ -0.2602 & 0.8423 \\ 0.4494 & -0.3276 \\ 1.3954 & -0.5615 \\ -0.0237 & -1.0295 \\ 1.6319 & -0.0936 \end{bmatrix}$$

Now that the data is centered, the covariance matrix can be computed using the following equation.

$$Covariance\ Matrix = \frac{X^T X}{N-1}\ where\ X\ is\ the\ standardized\ data$$

Therefore, the computed covariance matrix is the following

$$Covariance\ matrix = \begin{bmatrix} 1 & -0.4082 \\ -0.4082 & 1 \end{bmatrix}$$

The next step is to evaluate the eigenvalues and eigenvectors based on the covariance matrix. The computation process is as follows;

- $|A - \lambda I|\ and\ then\ solve\ for\ \lambda\ using\ quadratic\ solver$
- $use\ \lambda\ to\ solve\ for\ eigenvector\ x\ in\ (A - \lambda I)x = 0$

$$|A - \lambda I| = \begin{vmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{vmatrix} = \begin{vmatrix} 1 - \lambda & -0.408 \\ -0.408 & 1 - \lambda \end{vmatrix}$$

$$= (1 - \lambda)(1 - \lambda) - (-0.408 * -0.408) = 0$$

$$= 1 - 2\lambda + \lambda^2 - 0.166 = 0$$

$$= -2\lambda + \lambda^2 - 0.8335 = 0$$

$$= \lambda^2 - 2\lambda - 0.8335 = 0$$

Using Quadratic formula;

$$\lambda = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{2 \pm \sqrt{-2^2 - (4 * 1 * 0.8335)}}{2} = [1.408, 0.591]$$

Since we are interested in the best eigenvector to project the data on, the the eigenvector corresponding to the highest eigenvalue will be used. This is computed with the help of python's numpy library.

Let λ = 1.408

$$(A - \lambda I)x = 0$$

$$\left(\begin{bmatrix} 1 & -0.408 \\ -0.408 & 1 \end{bmatrix} - 1.408 * \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) X = 0$$

$$\begin{bmatrix} -0.408 & -0.408 \\ -0.408 & -0.408 \end{bmatrix} X = 0$$

$$Eigenvector = \begin{bmatrix} 0.7071 \\ -0.7071 \end{bmatrix}$$

Finally, the standardized data is projected onto the eigenvector to complete the PCA process.

$$PCA = \begin{bmatrix} -0.2602 & -0.0936 \\ -0.9697 & -1.2635 \\ -0.4967 & -0.0936 \\ 0.2129 & 0.3744 \\ -1.6792 & 2.2462 \\ -0.2602 & 0.8423 \\ 0.4494 & -0.3276 \\ 1.3954 & -0.5615 \\ -0.0237 & -1.0295 \\ 1.6319 & -0.0936 \end{bmatrix} * \begin{bmatrix} 0.7071 \\ -0.7071 \end{bmatrix} = \begin{bmatrix} -0.1178 \\ 0.2077 \\ -0.2850 \\ -0.1141 \\ -2.7756 \\ -0.7795 \\ 0.5494 \\ 1.3837 \\ 0.7112 \\ 1.2201 \end{bmatrix}$$

b) On Projecting the data on the principal component, this is the result.

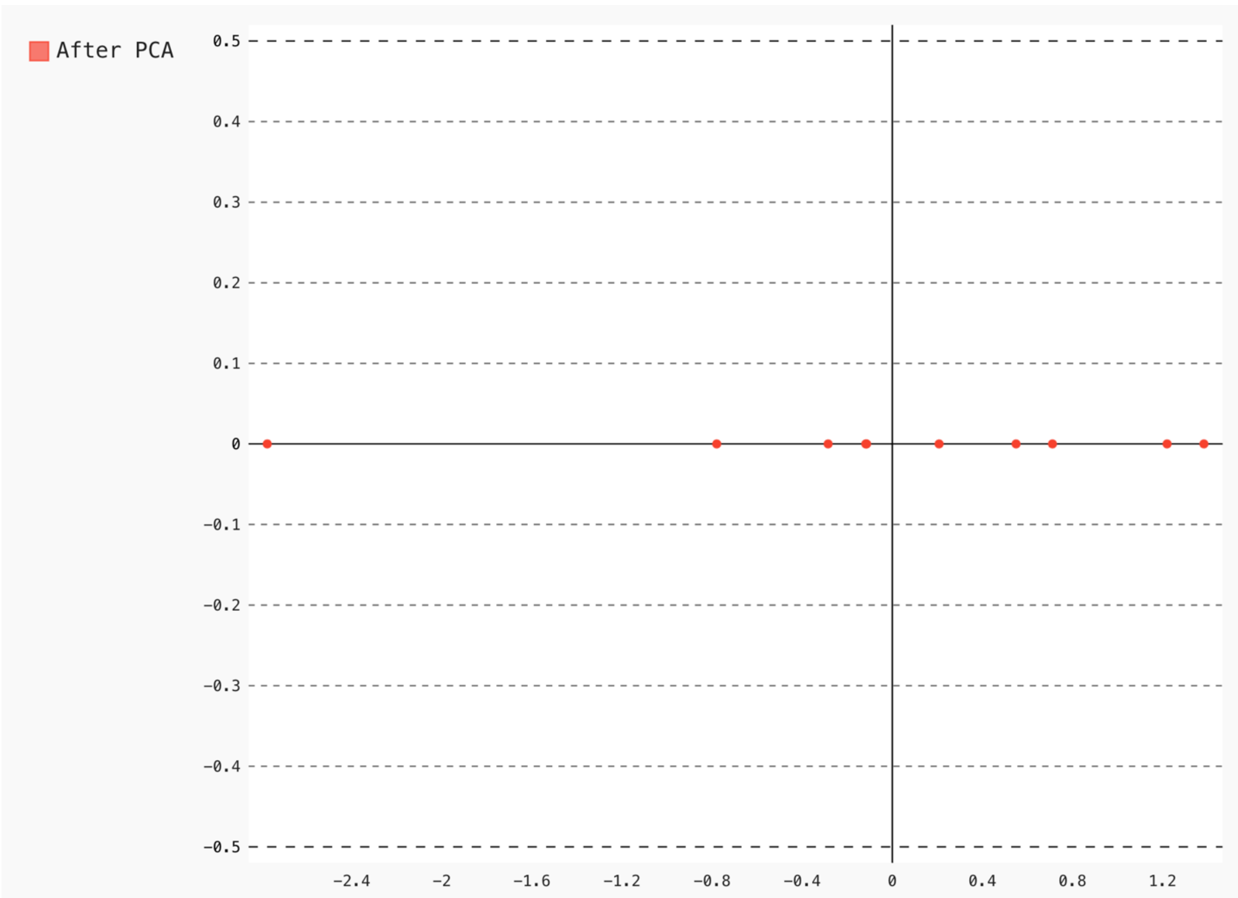

Figure 1: Plot of Data Projection onto principle component PCA

Question 2

$$Class\ 1 = \begin{bmatrix} -2 & 1 \\ -5 & -4 \\ -3 & 1 \\ 0 & 3 \\ -8 & 11 \end{bmatrix}, Class\ 2 = \begin{bmatrix} -2 & 5 \\ 1 & 0 \\ 5 & -1 \\ -1 & -3 \\ 6 & 1 \end{bmatrix}$$

a) The goal here is to find the information gain from the two classes. For this question, the assumption will be made that column 0 will correspond to p_i, and column 1 will correspond to n_i.

For Feature 1;

$$Dimensionality = 2$$
$$k = \{-8, -5, -4, -3, -2, 0, 1, 3, 11\}$$

K	p_i	n_i
-8	1	0
-5	1	0
-4	0	1
-3	1	0
-2	1	0
0	1	0
1	0	2
3	0	1
11	0	1

$$Remainder = \sum_{i=1}^{k} \frac{p_i + n_i}{p + n} H(\frac{p_i}{p_i + n_i}, \frac{n_i}{p_i + n_i})$$

$$Entropy = \sum_{i=1}^{n} (-P(v_i) \log_2 p_i(v_i))$$

$$\begin{aligned}\therefore Remainder = &\left(\frac{1}{10}\right)(-1\log_2 1 - 0\log_2 0) + \left(\frac{1}{10}\right)(-1\log_2 1 - 0\log_2 0) \\ &+ \left(\frac{1}{10}\right)(-1\log_2 1 - 0\log_2 0) + \left(\frac{1}{10}\right)(-1\log_2 1 - 0\log_2 0) \\ &+ \left(\frac{1}{10}\right)(-1\log_2 1 - 0\log_2 0) + \left(\frac{1}{10}\right)(-1\log_2 1 - 0\log_2 0) \\ &+ \left(\frac{1}{10}\right)(-1\log_2 1 - 0\log_2 0) + \left(\frac{1}{10}\right)(-1\log_2 1 - 0\log_2 0) \\ &+ \left(\frac{2}{10}\right)(-2\log_2 2 - 0\log_2 0) = -0.400\end{aligned}$$

$$Entropy = -\left(\frac{5}{10}\right)\log_2\frac{5}{10} - \frac{5}{10}\log_2\frac{5}{10} = \frac{5}{10} + \frac{5}{10} = 1$$

$$Informtion\ gain\ for\ feature\ 1 = 1 - (-0.400) = 1.400$$

For Feature 2;

$$Dimensionality = 2$$
$$k = \{-3, -2, -1, 0, 1, 5, 6\}$$

K	p_i	n_i
-3	0	1
-2	1	0
-1	1	1
0	0	1
1	1	1
5	1	1
6	1	0

$$Remainder = \sum_{i=1}^{k}\frac{p_i + n_i}{p + n}H(\frac{p_i}{p_i + n_i}, \frac{n_i}{p_i + n_i})$$

$$Entropy = \sum_{i=1}^{n}(-P(v_i)\log_2 p_i(v_i))$$

$$\begin{aligned}\therefore Remainder = &\left(\frac{1}{10}\right)(-1\log_2 1 - 0\log_2 0) + \left(\frac{1}{10}\right)(-1\log_2 1 - 0\log_2 0)\\ &+ \left(\frac{1}{10}\right)(-1\log_2 1 - 0\log_2 0) + \left(\frac{1}{10}\right)(-1\log_2 1 - 0\log_2 0)\\ &+ \left(\frac{2}{10}\right)\left(-\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2}\right) + \left(\frac{2}{10}\right)\left(-\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2}\right)\\ &+ \left(\frac{2}{10}\right)\left(-\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2}\right) = 0.6\end{aligned}$$

$$Entropy = -\left(\frac{5}{10}\right)\log_2\frac{5}{10} - \frac{5}{10}\log_2\frac{5}{10} = \frac{5}{10} + \frac{5}{10} = 1$$

$$Informtion\ gain\ for\ feature\ 2 = 1 - (0.600) = 0.400$$

b) Thus here we see that the information gain for feature 1 is 1.4, as opposed to a value of 0.4 for the information gain for feature 2. Based on this information, feature 1 is more discriminating.

c) The next step is to use the Linear Discriminant Analysis for feature projection.

$$Class\ 1 = \begin{bmatrix} -2 & 1 \\ -5 & -4 \\ -3 & 1 \\ 0 & 3 \\ -8 & 11 \end{bmatrix}, Class\ 2 = \begin{bmatrix} -2 & 5 \\ 1 & 0 \\ 5 & -1 \\ -1 & -3 \\ 6 & 1 \end{bmatrix}$$

The first step is to concatenate the entire dataset and then standardize. The standardized data is as follows.

$$Standardized\ Data = \begin{bmatrix} -0.2602 & -0.0936 \\ -0.9697 & -1.2635 \\ -0.4967 & -0.0936 \\ 0.2129 & 0.3744 \\ -1.6792 & 2.2462 \\ -0.2602 & 0.8423 \\ 0.4494 & -0.3276 \\ 1.3954 & -0.5615 \\ -0.0237 & -1.0295 \\ 1.6319 & -0.0936 \end{bmatrix}$$

We then split it to get the respective class data

Table 1: Standardized data with mean and std

	1	2
Class data	$\begin{bmatrix} -0.2602 & -0.0936 \\ -0.9697 & -1.2635 \\ -0.4967 & -0.0936 \\ 0.2129 & 0.3744 \\ -1.6792 & 2.2462 \end{bmatrix}$	$\begin{bmatrix} -0.2602 & 0.8423 \\ 0.4494 & -0.3276 \\ 1.3954 & -0.5615 \\ -0.0237 & -1.0295 \\ 1.6319 & -0.0936 \end{bmatrix}$
μ	$[-0.63858 \quad 0.23398]$	$[0.638 \quad -0.23398]$
σ	$[0.7212 \quad 1.277]$,	$[0.8428 \quad 0.694]$

Calculating Scatter Matrices is done by using the following equation

$$\sigma^2 = (N-1)cov(C) = \frac{(N-1)(X^T X)}{N-1} = X^T X$$

These are the computed scatter matrices.

Table 2: Scatter Matrix data

Class data	1	2
Scatter Matrix	$\begin{bmatrix} 4.119 & -2.396 \\ -2.396 & 6.799 \end{bmatrix}$	$\begin{bmatrix} 4.8804 & -1.278 \\ -1.278 & 2.200 \end{bmatrix}$

Within class scatter matrix

$$S_W = \sigma_1^2 + \sigma_2^2 = \begin{bmatrix} 9 & -3.674 \\ -3.674 & 9 \end{bmatrix}$$

$$S_W^{-1} = \begin{bmatrix} 0.1333 & 0.0544 \\ 0.0544 & 0.1333 \end{bmatrix}$$

Between class scatter matrix

$$S_B = (\mu_1 - \mu_2)^T (\mu_1 - \mu_2) = 1.850$$

Eigen decomposition is performed on $S_W^{-1} S_B$. This is done with the help of python.

Eigenvalues = [0.1459, 0.3473]

Eigenvectors = $\begin{bmatrix} -0.7071 & -0.7071 \\ 0.7071 & -0.7071 \end{bmatrix}$

Since the value at index 1 for eigenvalues is the greatest (at 0.3473), the corresponding eigenvector that will be used is column 1 of eigenvector, i.e, $[-0.7071, -0.7071]^T$. This will then be projected onto the dataset.

d) Data projection onto principal component.

<u>Projected data for class 1</u>

$$\begin{bmatrix} -2 & 1 \\ -5 & -4 \\ -3 & 1 \\ 0 & 3 \\ -8 & 11 \end{bmatrix} * \begin{bmatrix} -0.7071 \\ -0.7071 \end{bmatrix} = \begin{bmatrix} 0.250 \\ 1.579 \\ 0.417 \\ -0.415 \\ -0.400 \end{bmatrix}$$

Projected data for class 2

$$\begin{bmatrix} -2 & 5 \\ 1 & 0 \\ 5 & -1 \\ -1 & -3 \\ 6 & 1 \end{bmatrix} * \begin{bmatrix} -0.7071 \\ -0.7071 \end{bmatrix} = \begin{bmatrix} -0.411 \\ -0.086 \\ -0.589 \\ 0.744 \\ -1.08 \end{bmatrix}$$

The following is the plot for the projected data

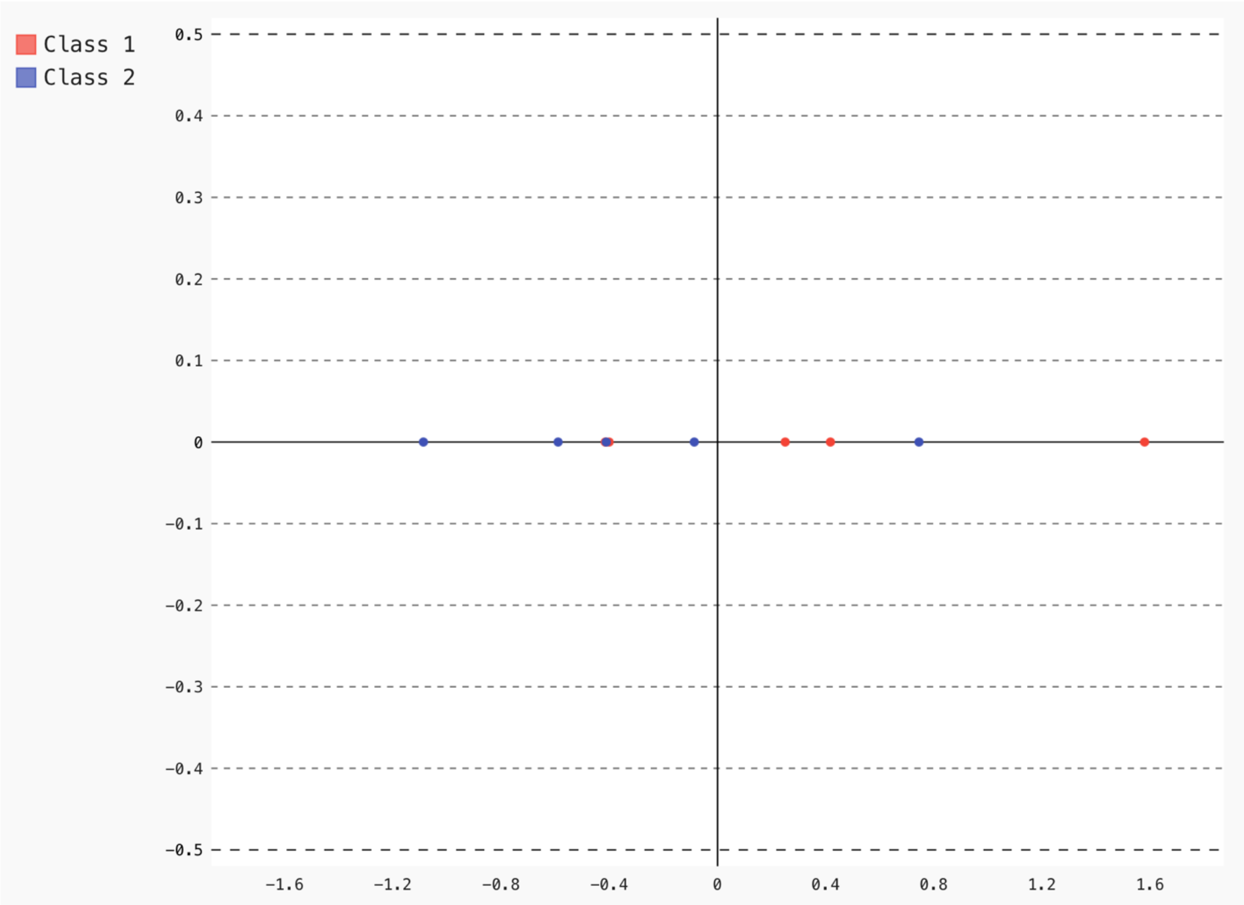

Figure 2: LDA Plot

e) **In my opinion, I don't think the LDA process of dimensionality reduction worked as well as I had initially thought. The reason for this can be explained through the figure above. There is somewhat a separation between the data, but this is not an absolute separation. In the sense that there is no strict line of demarcation stating that one one side of this line would just be class1 features and on the other side class 2 features. However, I do have to make a note that class 1 features are more predominant on the positive end of the x axis while class 2 features are more predominant on the positive side of the x axis.**

Visualization for programming part (PCA)

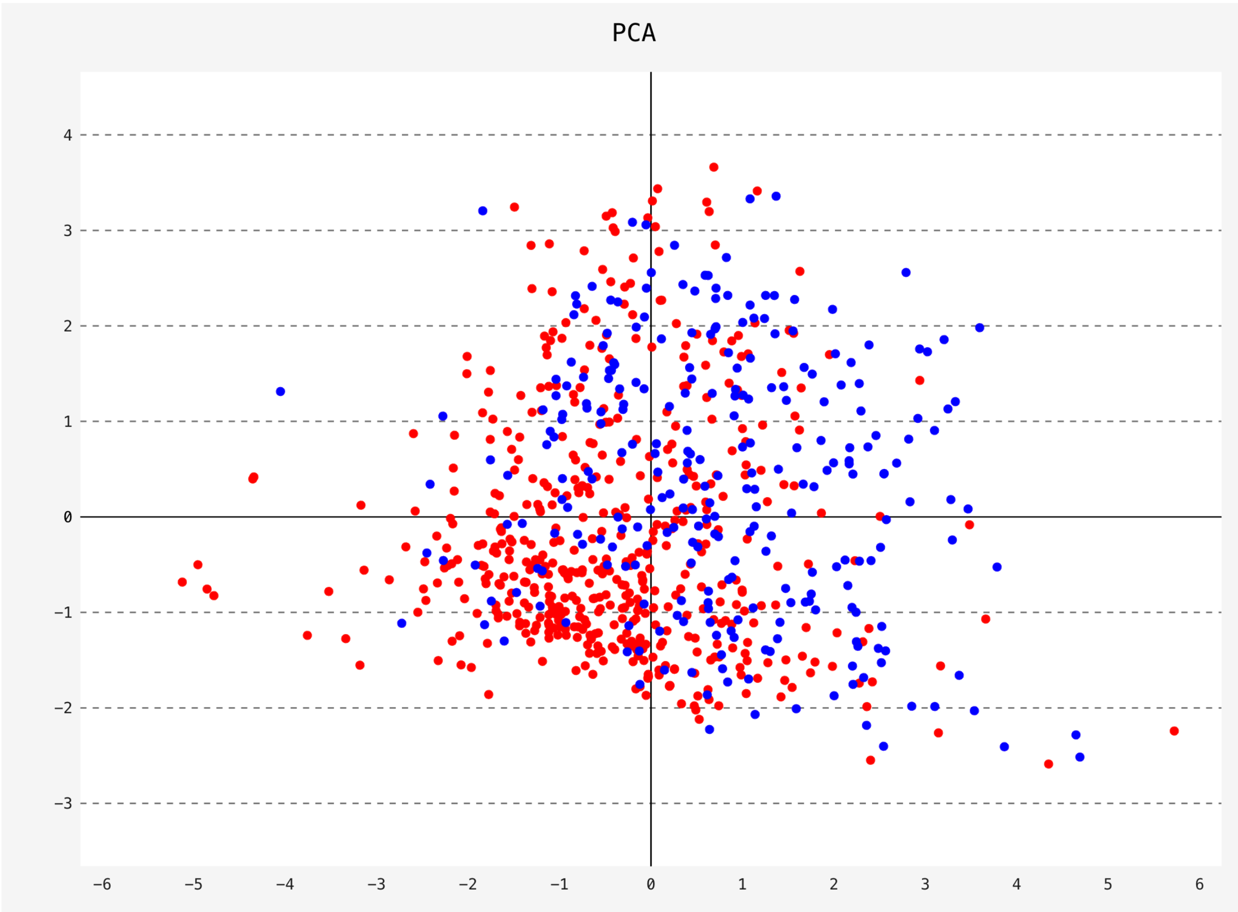

Figure 3: PCA Visualization