CS 613 - Machine Learning

Assignment 3 - Dimensionality Reduction
Clustering - Abishek S Kumar

Solutions

This is the solutions pdf file of the Assignment 2

1 Answers to Question 1

1. The original matrix:

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

   (a) The matrix after standardization:

$$\begin{bmatrix} -0.27422 & -0.0986 \\ -1.0221 & -1.3318 \\ -0.5235 & -0.0986 \\ 0.2243 & 0.3946 \\ -1.7700 & 2.3676 \\ -0.2742 & 0.8878 \\ 0.4736 & -0.3452 \\ 1.4708 & -0.5919 \\ -0.0249 & -1.0851 \\ 1.7201 & -0.09865 \end{bmatrix}$$

      i. The means along the columns of the matrix, μ_1 = -0.9, $\mu_2 = 1.4$
      ii. The standard deviations along columns of the matrix, $\sigma_1 = 4.0112$, $\sigma_2 = 4.0546$

   The covariance matrix is calculated using the formula, $(X^TX)/N(no.ofrows) - 1$ Also the eigen values and vectors are calculated as so:

$$[\Sigma - \lambda * I]X = 0 \text{ (eq. 1) };$$
$$\Sigma = cov(X)$$
$$\lambda = eigenvalues[\lambda_1\lambda_2]$$
$$I = IdentityMatrix \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$
$$\text{Covariance matrix } \Sigma = \begin{bmatrix} 1.111 & -0.4536 \\ -0.4536 & 1.111 \end{bmatrix}$$
$$\text{From eq.1: } \begin{bmatrix} 1.111 & -0.4536 \\ -0.4536 & 1.111 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix}$$
$$\begin{bmatrix} 1.111-\lambda & -0.4536 \\ -0.4536 & 1.111-\lambda \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{ (eq.2)}$$

   Next calculating the determinant and equating it to zero, $det|A|$=0 (1.111$-\lambda$)(1.111-λ) - (-0.4536)(-0.4536) $= 0$

$$(1.234 - 1.111\lambda \text{ -}1.111\lambda + \lambda^2 - 0.2057) = 0$$
$$(\lambda^2 - 2.222\lambda + 1.0283) = 0$$

   Factorizing and calculating the values of λ using $(-b +_{-} sqrt(b^2 - 4ac))/2a$

$$\lambda_1 = 1.564, \lambda_2 = 0.657$$

Substituting the values of λ into eq.2 we get two equations to solve for $X_1 and X_2$

$$\text{Substituting } \lambda_1 in eq.2 : \begin{bmatrix} 1.111 - 1.5647 & -0.4536 \\ -0.4536 & 1.111 - 1.5647 \end{bmatrix}$$

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

The eigen vectors after normalization are :

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

(b) After projecting the eigen vector of he largest eigen value $\lambda_1 = 1.5467, P = xW$:

$$\begin{bmatrix} -0.1241 \\ 0.2189 \\ -0.3004 \\ -0.1203 \\ -2.9257 \\ -0.8217 \\ 0.5790 \\ 1.4586 \\ 0.7497 \\ 1.2861 \end{bmatrix}$$

2. This part divides the original matrix into two halves

$$\text{A} = \begin{bmatrix} -2 & 1 \\ -5 & -4 \\ -3 & 1 \\ 0 & 3 \\ -8 & 11 \end{bmatrix}$$

$$\text{B} = \begin{bmatrix} -2 & 5 \\ 1 & 0 \\ 5 & -1 \\ -1 & -3 \\ 6 & 1 \end{bmatrix}$$

The total entropy of the system first:

$$H = -(6/12) * log_2(6/12) - (6/12) * log_2(6/12) = 1$$

Information gain for every feature:

$$E(f1) = (2/12) * (-1/2 * log_2(1/2) - 1/2 * log_2(1/2)) + (1/12) * (-1/1 log_2 * 1) + ... = 1/6$$

$$IG(f1) = 1 - 1/6 = 0.8333$$

Similarly,

$$E(f2) = (3/12) * (-2/3 * log_2(2/3) - 1/3 * log_2(1/3)) + (1/12) * (-1/1 log_2 * 1) + ... = 0.2295$$

$$IG(f2) = 1 - 0.2295 = 0.7704$$

Class A provides more information, hence more discriminating.

To calculate the direction of projection, we first calculate the scatter matrices for each of class, as:

$$\sigma_A^2 = \begin{bmatrix} 37.2 & -29.8 \\ -29.8 & 119.2 \end{bmatrix}$$

$$\sigma_B^2 = \begin{bmatrix} 50.8 & -9.6 \\ -9.6 & 35.2 \end{bmatrix}$$

Within class scatter matrix :

$$\mathrm{S}_W = \begin{bmatrix} 22 & -9.85 \\ -9.85 & 38.6 \end{bmatrix}$$

Inverse :

$$\mathrm{S}_W^{-1} = \begin{bmatrix} 0.051 & 0.013 \\ 0.013 & 0.029 \end{bmatrix}$$

Between class scatter matrix as:

$$\mathrm{S}_B = \begin{bmatrix} 29.16 & -10.8 \\ -10.8 & 4 \end{bmatrix}$$

The eigenvalues and eigenvectors;

$$\lambda_1 and \lambda_2 =$$

$$\begin{bmatrix} 1.3305 & -1.7347e^{-}17 \end{bmatrix}$$

$$\mathrm{X}_1 and X_2 =$$

$$\begin{bmatrix} 0.94452 & 0.32843 \\ 0.05179 & 0.99857 \end{bmatrix}$$

The non-zero eigenvalue gives the direction of projection eigen vector:

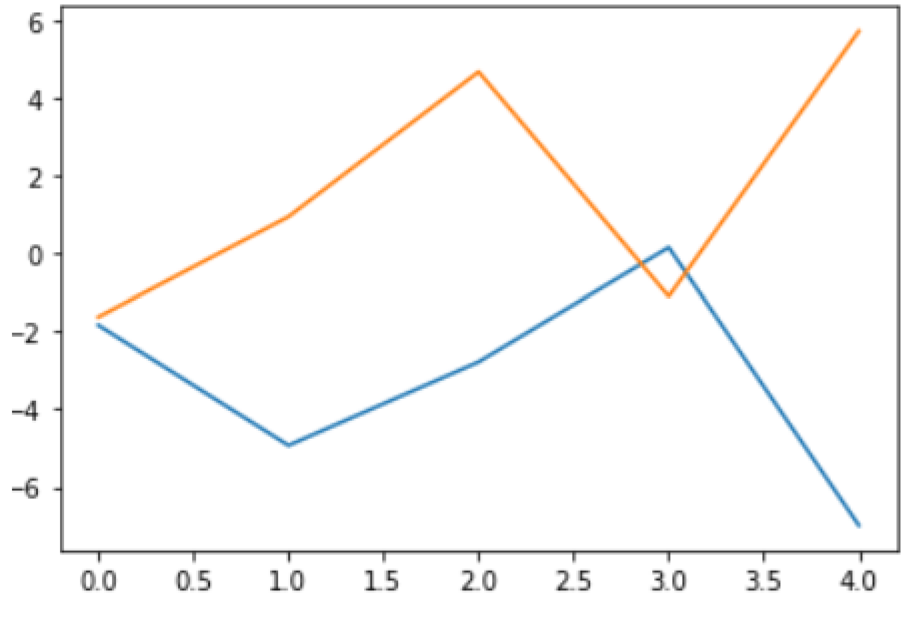
$$\mathrm{X}= \begin{bmatrix} 0.94452 \\ 0.05179 \end{bmatrix}$$

Projected data along the X1:

$$\text{For class A: } \begin{bmatrix} -1.83725 \\ -4.92980 \\ -2.78178 \\ 0.15537 \\ -6.9864 \end{bmatrix}$$

$$\text{For class B: } \begin{bmatrix} -1.63008 \\ 0.94452 \\ 4.67083 \\ -1.09998 \\ 5.718949 \end{bmatrix}$$

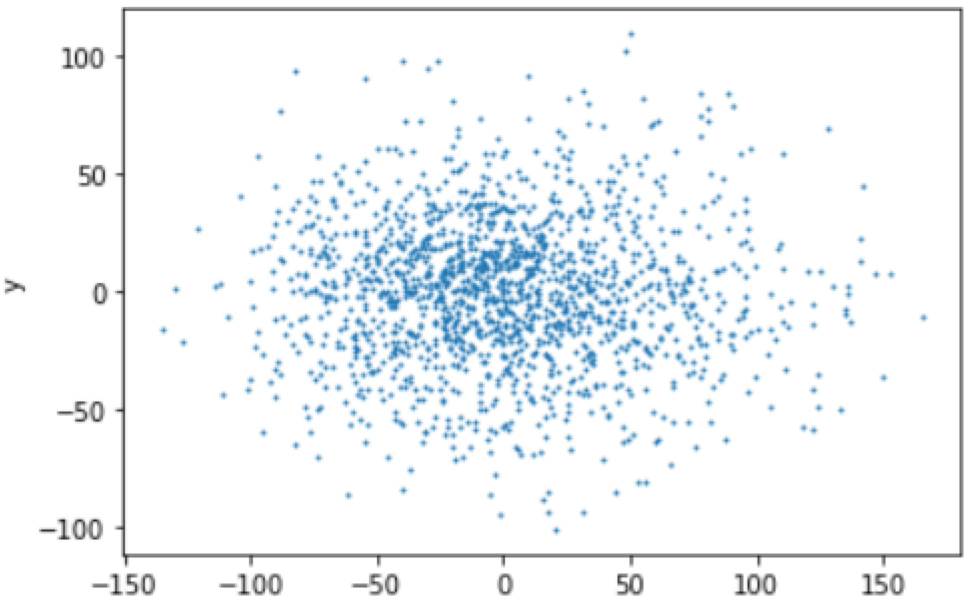
A few values are overlapping in the classifications, the and hence don't provide a perfect binary classification. Values such as -1.837 from Class A and -1.63008 from Class B are overlapping.



6
4
2
0
-2
-4
-6
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0

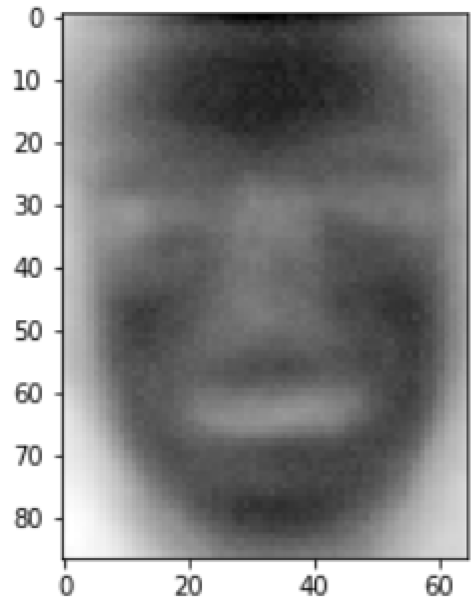
2 Answers to Question 2

1. KNN Accuracy (original) = 0.232558
2. KNN Accuracy (100D data) = 0.23872
3. KNN Accuracy (100D whitened data) = 0.211240
4. Plot:



3 Answers to Question 3

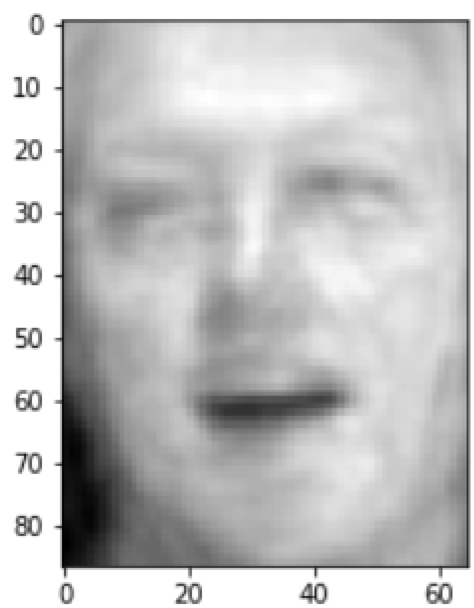
1. Visualizing Primary (Best) Principal Component:



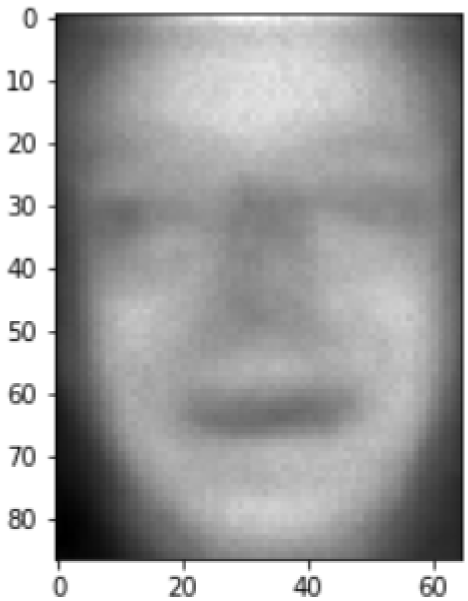
2. Eigenvalues for 95 information gain: 188 values

3. Visualization of reconstruction of first person:

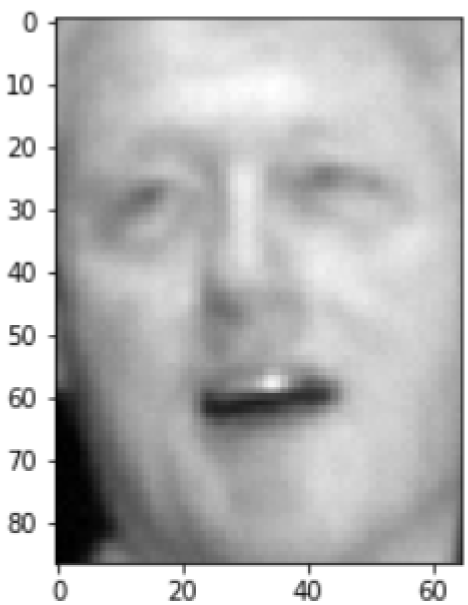
   (a) Original Image:

   (b) One Principal Component:

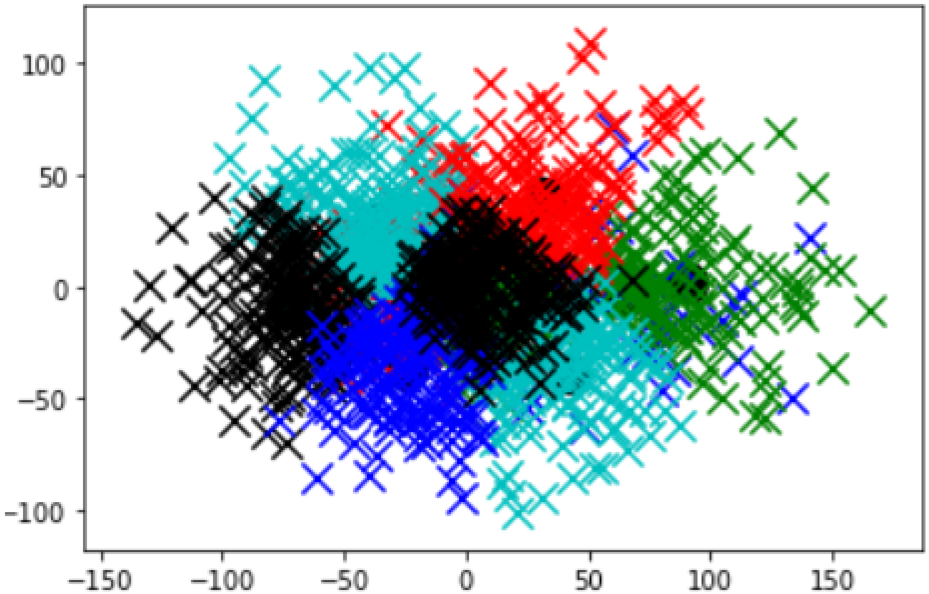


(c) k- Principal Components:



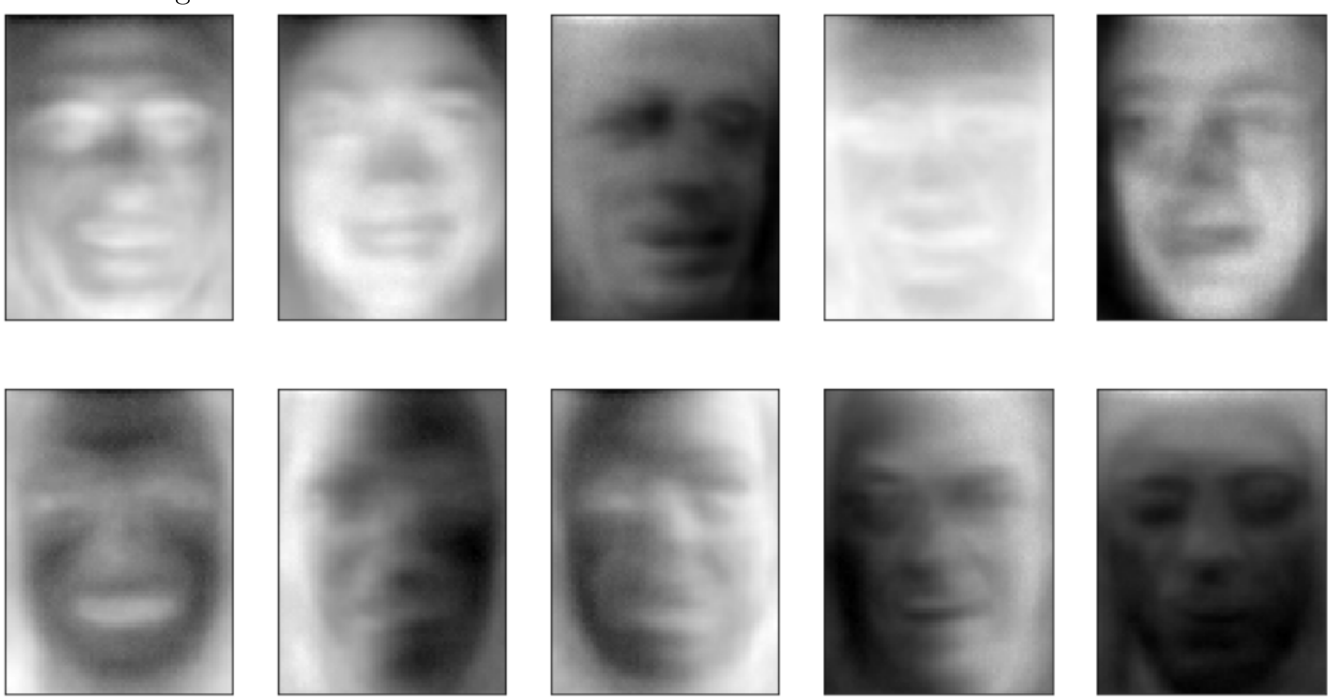
4 Answers to Question 4

1. Visualization of k-means clusters:



2. K-means clustering:

   (a) All mean images:

   (b) Max, min images for each cluster: