CS 613 - Assignment 1

Name - Shaoshu Xu

Part 1 Theory Question

Data, X =

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

Standardize the data:

$$\mu_1 = (-2-5-3+0-8-2+1+5-1+6)/10 = -0.9$$

$$\mu_2 = (1-4+1+3+11+5+0-1-3+1)/10 = 1.4$$

$$\begin{aligned}\sigma_1 =& (((-2+0.9)^2+(-5+0.9)^2+(-3+0.9)^2+(0+0.9)^2+(-8+0.9)^2+(-2+0.9)^2+(1+0.9)^2 \\ &+(5+0.9)^2+(-1+0.9)^2+(6+0.9)^2)/9)^0.5 = 4.2282\end{aligned}$$

$$\begin{aligned}\sigma_2 =& (((1-1.4)^2+(-4-1.4)^2+(1-1.4)^2+(3-1.4)^2+(11-1.4)^2+(5-1.4)^2+(0-1.4)^2 \\ &+(-1-1.4)^2+(-3-1.4)^2+(1-1.4)^2)/9)^0.5 = 4.2740\end{aligned}$$

Subtract the means from each observation and Divide each (centered) observation by the standard deviation:

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

Compute covariance matrix:
Because the data's columns have zero mean, then we can compute the covariance matrix by this equation

$$\Sigma(X) = \frac{X^T X}{N-1}$$

$$= \begin{bmatrix} 1 & -0.4083 \\ -0.4083 & 1 \end{bmatrix}$$

$$\lambda^2 - \lambda(1+1) + (1-0.167) = \lambda^2 - 2\lambda + 0.833 = 0$$

Compute eigenvalues and eigenvectors:

$$\lambda 1 = (2 - (1 - 4 * 1 * 0.83329111)^0.5)/2 = 0.5917$$

$$\lambda 2 = (2 + (1 - 4 * 1 * 0.83329111)^0.5)/2 = 1.4083$$

$$(\Sigma(X) - \lambda I)x = 0 \tag{1}$$

$$(\begin{bmatrix} 1 & -0.4083 \\ -0.4083 & 1 \end{bmatrix} - \begin{bmatrix} 0.5917 & 0 \\ 0 & 0.5917 \end{bmatrix}) \begin{bmatrix} x \\ y \end{bmatrix} = 0, for \lambda = 0.5917 \tag{2}$$

when x =1, y =1

$$e1 = \begin{bmatrix} -0.7071 & -0.7071 \end{bmatrix}^T, normalized$$

$$(\begin{bmatrix} 1 & -0.4083 \\ -0.4083 & 1 \end{bmatrix} - \begin{bmatrix} 1.4083 & 0 \\ 0 & 1.4083 \end{bmatrix}) \begin{bmatrix} x \\ y \end{bmatrix} = 0, for \lambda = 1.4083 \tag{3}$$

when x = 1, y = -1

$$e2 = \begin{bmatrix} -0.7071 & 0.7071 \end{bmatrix}^T, normalized$$

1.(b)

Project the data onto the principal component: Because max(0.5917,1.4083) = 1.4083, so we choose the second eigenvector: V = [-0.7071; 0.7071]

Project the points onto the vector V, Z = X * V

Therefore,

$$\begin{bmatrix} -0.2602 & -0.0936 \\ -0.9697 & -1.2635 \\ -0.4967 & -0.0936 \\ 0.2129 & 0.3744 \\ -1.6792 & 2.2462 \\ -0.2602 & 0.8423 \\ 0.4494 & -0.3276 \\ 1.3954 & -0.5615 \\ -0.0237 & -1.0295 \\ 1.6319 & -0.0936 \end{bmatrix} \times \begin{bmatrix} -0.7071 \\ 0.7071 \end{bmatrix} = \begin{bmatrix} 0.1178 \\ -0.2077 \\ 0.2850 \\ 0.1142 \\ 2.7756 \\ 0.7796 \\ -0.5494 \\ -1.3837 \\ -0.7112 \\ -1.2201 \end{bmatrix}$$

2.(a)

$$Class1 = \begin{bmatrix} -2 & 1 \\ -5 & -4 \\ -3 & 1 \\ 0 & 3 \\ -8 & 11 \end{bmatrix}$$

$$Class2 = \begin{bmatrix} -2 & 5 \\ 1 & 0 \\ 5 & -1 \\ -1 & -3 \\ 6 & 1 \end{bmatrix}$$

Feature 1:

$$p_0 = 1, p_{-2} = 1, p_{-3} = 1, p_{-5} = 1, p_{-8} = 1,$$

$$n_{-2} = 1, n_{-1} = 1, n_1 = 1, n_5 = 1, n_6 = 1,$$

We only need to calculate the -2 term, the rest terms are 0:

$$Entropy(1) = \frac{2}{10} * (-\frac{1}{2} * \log_2(\frac{1}{2}) - \frac{1}{2}\log_2(\frac{1}{2})) = 0.2$$

$$IG(1) = -\frac{5}{10} * \log_2(\frac{5}{10}) - \frac{5}{10} * \log_2(\frac{5}{10}) - 0.2 = 1 - 0.2 = 0.8$$

Feature 2:

$$p_1 = 2, p_{-4} = 1, p_3 = 1, p_{11} = 1,$$

$$n_1 = 1, n_5 = 1, n_0 = 1, n_{-1} = 1, n_{-3} = 1,$$

We only need to calculate the 1 term, the rest terms are 0:

$$Entropy(1) = \frac{3}{10} * (-\frac{2}{3} * \log_2(\frac{2}{3}) - \frac{1}{3}\log_2(\frac{1}{3}))0.2755$$

$$IG(1) = -\frac{5}{10} * \log_2(\frac{5}{10}) - \frac{5}{10} * \log_2(\frac{5}{10}) = 1 - 0.2772 = 0.7245$$

2.(b)

Because IG$_1 > IG_2, so feature 1 is more discriminating.$

2.(c)

$$Class1 = \begin{bmatrix} -2 & 1 \\ -5 & -4 \\ -3 & 1 \\ 0 & 3 \\ -8 & 11 \end{bmatrix}$$

$$Class2 = \begin{bmatrix} -2 & 5 \\ 1 & 0 \\ 5 & -1 \\ -1 & -3 \\ 6 & 1 \end{bmatrix}$$

Get mean of each feature for each classes:

$$\mu_1 = \begin{bmatrix} (-2-5-3+0-8)/5 & (1-4+1+3+11)/5 \end{bmatrix} = \begin{bmatrix} -3.6 & 2.4 \end{bmatrix}$$

$$\mu_1 = \begin{bmatrix} (-2+1+5-1+6)/5 & (5+0-1-3+1)/5 \end{bmatrix} = \begin{bmatrix} 1.8 & 0.4 \end{bmatrix}$$

Compute covariance matrix:

$$\Sigma(C1) = \frac{C1^T C1}{N-1}$$

$$= \begin{bmatrix} 9.30 & -7.45 \\ -7.45 & 29.80 \end{bmatrix}$$

$$\Sigma(C2) = \frac{C2^T C2}{N-1}$$

$$= \begin{bmatrix} 12.70 & -2.40 \\ -2.40 & 8.80 \end{bmatrix}$$

Compute the scatter matrices:

$$\sigma_1^2 = (5-1) * \Sigma(C1) = 4 * \begin{bmatrix} 9.30 & -7.45 \\ -7.45 & 29.80 \end{bmatrix} = \begin{bmatrix} 37.20 & -29.80 \\ -29.80 & 119.20 \end{bmatrix}$$

$$\sigma_2^2 = (5-1) * \Sigma(C2) = 4 * \begin{bmatrix} 12.70 & -2.40 \\ -2.40 & 8.80 \end{bmatrix} = \begin{bmatrix} 50.80 & -9.60 \\ -9.60 & 35.20 \end{bmatrix}$$

$$S_b = (\mu_1 - \mu 2)^T * (\mu 1_- \mu_2) = (\begin{bmatrix} 3.6 & 2.4 \end{bmatrix} - \begin{bmatrix} 1.8 & 0.4 \end{bmatrix})^T \times (\begin{bmatrix} -3.6 & 2.4 \end{bmatrix} - \begin{bmatrix} 1.8 & 0.4 \end{bmatrix}) = \begin{bmatrix} -5.4 \\ 2.0 \end{bmatrix} \times \begin{bmatrix} -5.4 & 2.0 \end{bmatrix}$$

$$= \begin{bmatrix} 29.16 & -10.80 \\ -10.80 & 4.0 \end{bmatrix}$$

$$S_w = \Sigma^2 + \Sigma^2 = \begin{bmatrix} 37.20 & -29.80 \\ -29.80 & 119.20 \end{bmatrix}^{\cancel{2}} + \begin{bmatrix} 50.80 & -9.60 \\ -9.60 & 35.20 \end{bmatrix}^{\cancel{2}} = \begin{bmatrix} 4944.7 & -5486.3 \\ -5486.3 & 16428 \end{bmatrix}$$

$$S_w^{-1} * S_b = \begin{bmatrix} 4944.7 & -5486.3 \\ -5486.3 & 16428 \end{bmatrix}^{-} 1 \times \begin{bmatrix} 29.16 & -10.80 \\ -10.80 & 4.0 \end{bmatrix} = \begin{bmatrix} 0.0082 & -0.0030 \\ 0.0021 & -0.0008 \end{bmatrix}$$

Perform eigen-decomposition on inv(Sw) * Sb:

The eigenvalue matrix = $\begin{bmatrix} 0.0074 & 0 \\ 0 & 0 \end{bmatrix}$

The eigenvector matrix = $\begin{bmatrix} 0.9692 & 0.3473 \\ 0.2461 & 0.9377 \end{bmatrix}$

The eigenvector pertaining to the only noneigenvalue is our projection matrix. So the direction of projection:
$\begin{bmatrix} 0.9692 \\ 0.2461 \end{bmatrix}$

2.(d)

Project the data onto the principal component:
Z1 = C1*W

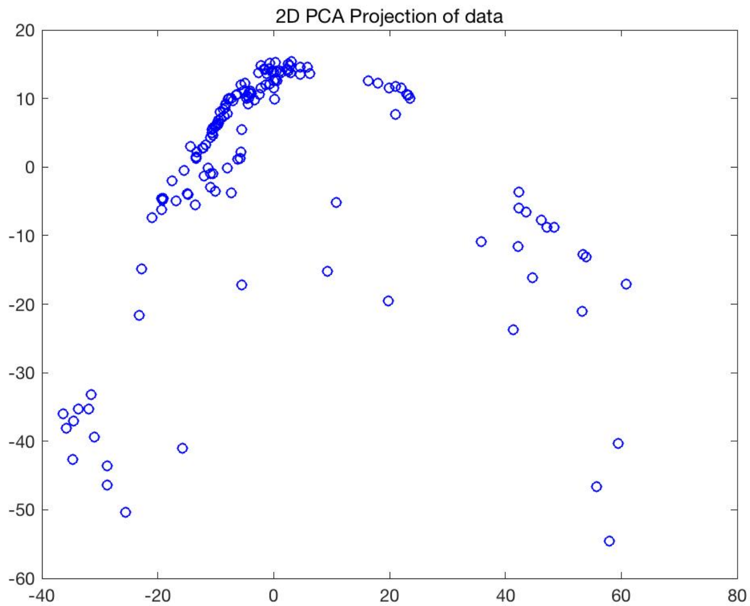
$$\begin{bmatrix} -2 & 1 \\ -5 & -4 \\ -3 & 1 \\ 0 & 3 \\ -8 & 11 \end{bmatrix} \times \begin{bmatrix} 0.9692 \\ 0.2461 \end{bmatrix} = \begin{bmatrix} -1.6923 \\ -5.8304 \\ -2.6615 \\ 0.7383 \\ -5.0465 \end{bmatrix}$$

Z2 = C2*W

$$\begin{bmatrix} -2 & 5 \\ 1 & 0 \\ 5 & -1 \\ -1 & -3 \\ 6 & 1 \end{bmatrix} \times \begin{bmatrix} 0.9692 \\ 0.2461 \end{bmatrix} = \begin{bmatrix} -0.7079 \\ 0.9692 \\ 4.5999 \\ -1.7075 \\ 6.0613 \end{bmatrix}$$

2.(e) It performas not bad. In Class1, 4 out of the 5 points are on the negative scale, and in Class2, 3 out of the 5 points are on the positive scale. However, the point 0.7383 in Class1, and the points -0.7079 and -1.7075 in Class 2 are mixed. Hence, I think the data cannot be clearly separable in this method and we could do better.

Part 2 The visualization of the PCA result



visualization.jpg

Figure 1: PCA visualization

Part 3 Eigenfaces results

i. Number of principle components needed to represent 95 percent of information, k

k = 37

ii. Visualization of primary principle component

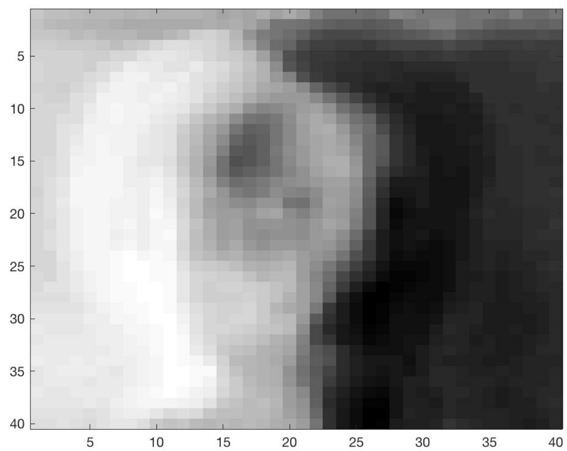


Figure 2: Primary principle component

iii. Visualization of the reconstruction of the first person

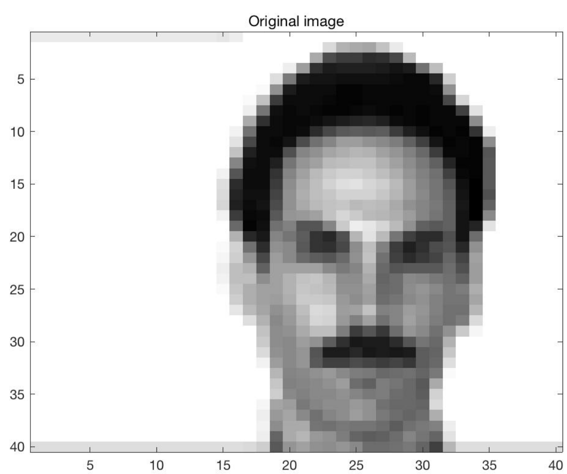


Figure 3: Reconstruction of first person - original

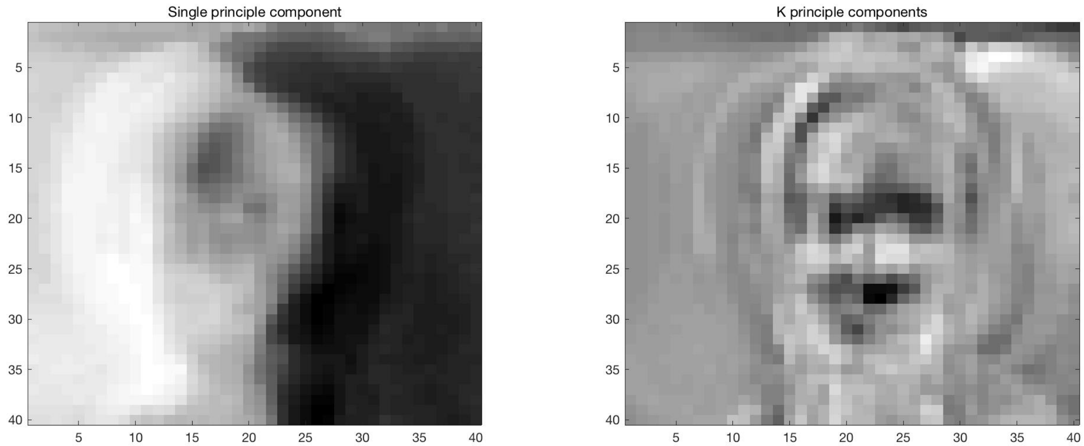


Figure 4: Reconstruction of first person - single PC and k PC

Part 4 Clustering results

i. The initial setup visualization

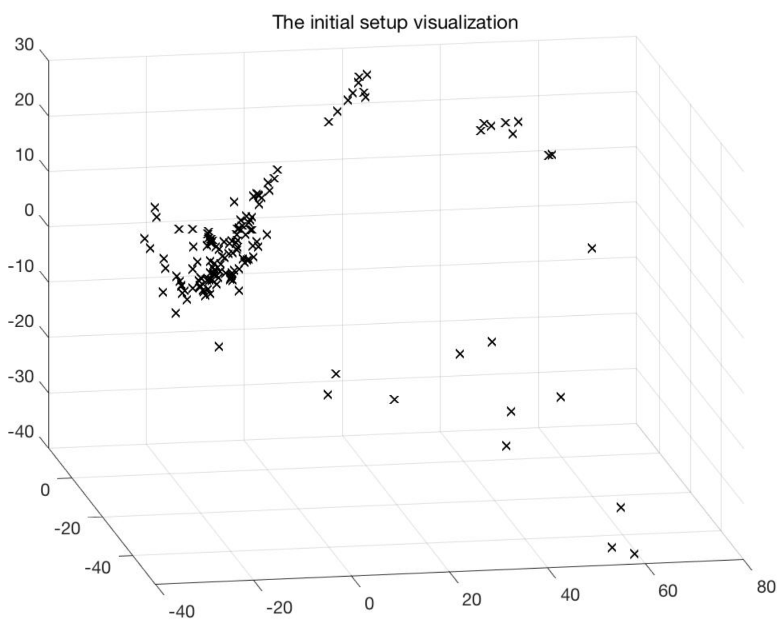


Figure 5: The initial setup visualization

ii. The initial cluster assignment visualization

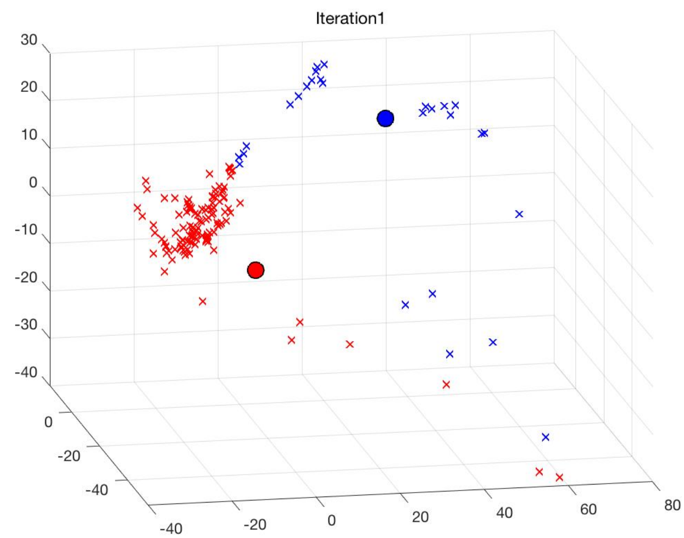


Figure 6: The initial cluster assignment visualization

iii. The final cluster assignment visualization

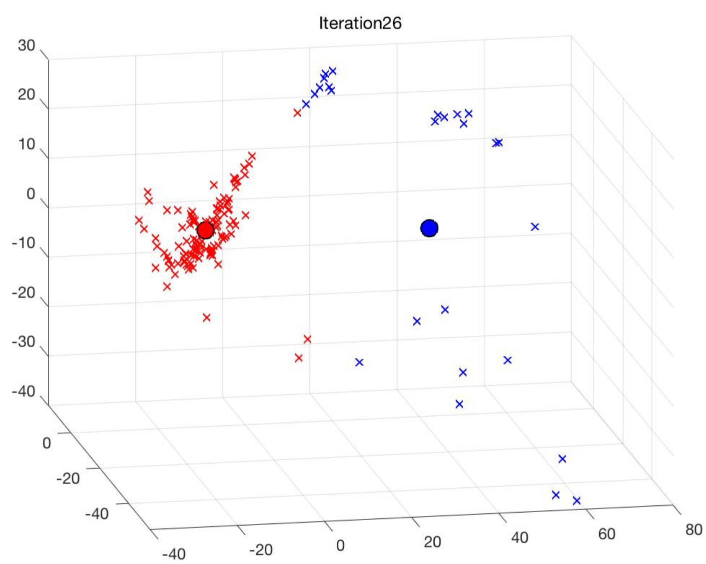


Figure 7: The final cluster assignment visualization