CS 613 - Machine Learning

1. PART-1: -

1.

(a)

Calculating the mean:

$$\mu_1 = \frac{-2-5-3+0-8-2+1+5-1+6}{10} = -0.9$$

$$\mu_2 = \frac{1-4+1+3+11+5+0-1-3+1}{10} = 1.4$$

Calculating the Standard deviation:

$$\sigma_1 = \sqrt{\frac{1609}{90}} \approx 4.2282$$

$$\sigma_2 = \sqrt{\frac{274}{15}} \approx 4.2740$$

Standardizing the data:

$$C = \begin{bmatrix} -0.260157 & -0.093590 \\ -0.969676 & -1.263467 \\ -0.496663 & -0.093590 \\ +0.212855 & +0.374360 \\ -1.679196 & +2.246164 \\ -0.260157 & +0.842311 \\ +0.449362 & -0.327565 \\ +1.395388 & -0.561541 \\ -0.023650 & -1.029492 \\ +1.631895 & -0.093590 \end{bmatrix}$$

Calculating covariance matrix:

$$cov = C^T C \div (N-1) = \begin{bmatrix} 9 & -3.674359 \\ -3.674359 & 9 \end{bmatrix} \div 9 = \begin{bmatrix} 1 & -0.408262 \\ -0.408262 & 1 \end{bmatrix}$$

Calculating eigenvalues:

$$\begin{bmatrix} 1 & -0.408262 \\ -0.408262 & 1 \end{bmatrix} - \lambda I = 0$$

$$(1-\lambda)^2 - (-0.408262)^2 = 0 \Rightarrow \lambda_1 = 0.5917, \lambda_2 = 1.4083$$

$$\lambda = \begin{bmatrix} 0.5917 \\ 1.4083 \end{bmatrix}$$

Calculating eigenvectors:

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

for $\lambda = 0.5917$:

$$\begin{bmatrix} 1-0.5917 & -0.408262 \\ -0.408262 & 1-0.5917 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0$$
$$V_1 = \begin{bmatrix} -0.7071 \\ -0.7071 \end{bmatrix}$$

for $\lambda = 1.4083$:

$$\begin{bmatrix} 1-1.4083 & -0.408262 \\ -0.408262 & 1-1.4083 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 0$$
$$V_2 = \begin{bmatrix} -0.7071 \\ 0.7071 \end{bmatrix}$$

(b)

Project the data onto the principal component corresponding to the largest eigen value:

$$C \times V_2 = \begin{bmatrix} -0.260157 & -0.093590 \\ -0.969676 & -1.263467 \\ -0.496663 & -0.093590 \\ +0.212855 & +0.374360 \\ -1.679196 & +2.246164 \\ -0.260157 & +0.842311 \\ +0.449362 & -0.327565 \\ +1.395388 & -0.561541 \\ -0.023650 & -1.029492 \\ +1.631895 & -0.093590 \end{bmatrix} \times \begin{bmatrix} -0.7071 \\ 0.7071 \end{bmatrix} = \begin{bmatrix} 0.117780 \\ -0.207741 \\ 0.285016 \\ 0.114201 \\ 2.775649 \\ 0.779563 \\ -0.549371 \\ -1.383758 \\ -0.711237 \\ -1.220102 \end{bmatrix}$$

2.

(a) The information gain of each feature can be calculated as follows: -

Calculating IG for Class 1:

$$Entropy = \frac{1+1}{5+5}H\left(\frac{1}{1+1},\frac{1}{1+1}\right) + 4 \times \frac{1+0}{5+5}H\left(\frac{1}{1+0},\frac{0}{1+0}\right) + 4 \times \frac{0+1}{5+5}H\left(\frac{0}{0+1},\frac{1}{0+1}\right) = 0.2$$

$$H\left(\frac{1}{2},\frac{1}{2}\right) = 1$$

$$H(1,0) = -1\log 1 - 0\log 0 = 0$$

$$IG(x1) = 1 - 0.2 = 0.8$$

Calculating IG for Class 2:

$$Entropy = 0.2755$$

$$H\left(\frac{2}{3},\frac{1}{3}\right) = -\frac{2}{3}\log\frac{2}{3} - \frac{1}{3}\log\frac{1}{3} = 0.9183$$

$$IG(x2) = 1 - 0.2755 = 0.7245$$

(b)

Since, $IG(x1) > IG(x2)$**.** That means IG(x1) is more discriminating.

(c)

Standardizing all the data:

The dataset mean and standard deviation: -

$\mu_1 = -0.9$, $\mu_2 = 1.4$, $\sigma_1 = 4.2282$, $\sigma_2 = 4.2740$

Then, the data becomes

$$C_1 = \begin{bmatrix} -0.260157 & -0.093590 \\ -0.969676 & -1.263467 \\ -0.496663 & -0.093590 \\ +0.212855 & +0.374360 \\ -1.679196 & +2.246164 \end{bmatrix} \text{ and } C_2 = \begin{bmatrix} -0.260157 & +0.842311 \\ +0.449362 & -0.327565 \\ +1.395388 & -0.561541 \\ -0.023650 & -1.029492 \\ +1.631895 & -0.093590 \end{bmatrix}$$

Calculating the mean for each class:

$$\mu_1 = [-0.6386 \quad 0.2340], \quad \mu_2 = [0.6386 \quad -0.2340]$$

Calculating scatter matrices for each class:

Using formula $\sigma^2 = (N-1) \times cov(C)$,

$$\sigma_1^2 = (5-1) \times cov(C_1) = \begin{bmatrix} 2.0808 & -1.6490 \\ -1.6490 & 6.5255 \end{bmatrix}$$

$$\sigma_2^2 = (5-1) \times cov(C_2) = \begin{bmatrix} 2.8415 & -0.5312 \\ -0.5312 & 1.9270 \end{bmatrix}$$

Within class scatter matrix:

$$S = \sigma_1^2 + \sigma_2^2 = \begin{bmatrix} 4.9223 & -2.1803 \\ -2.1803 & 8.4526 \end{bmatrix}$$

$$S_W^{-1} = \begin{bmatrix} 0.2294 & 0.0592 \\ 0.0592 & 0.1336 \end{bmatrix},$$

Eigen-decomposition:

$$S_B = (\mu_1 - \mu_2)^T(\mu_1 - \mu_2) = \begin{bmatrix} 1.6311 & -0.5976 \\ -0.5976 & 0.2190 \end{bmatrix}$$

$$S_W^{-1}S_B = \begin{bmatrix} 0.2294 & 0.0592 \\ 0.0592 & 0.1336 \end{bmatrix}\begin{bmatrix} 1.6311 & -0.5976 \\ -0.5976 & 0.2190 \end{bmatrix} = \begin{bmatrix} 0.3388 & -0.1241 \\ 0.0167 & -0.0061 \end{bmatrix}$$

Eigen-values and eigen-vector:

For eigen-values, I got the result $\lambda = \begin{bmatrix} 0.3327 \\ 0 \end{bmatrix}$.

Eigen-vector by using non-zero eigen-value 0.3327:- $W = \begin{bmatrix} 0.9988 \\ 0.0493 \end{bmatrix}$.

(d)

Projecting the data (Class 1):

$$\begin{bmatrix} -0.260157 & -0.093590 \\ -0.969676 & -1.263467 \\ -0.496663 & -0.093590 \\ +0.212855 & +0.374360 \\ -1.679196 & +2.246164 \end{bmatrix}\begin{bmatrix} 0.9988 \\ 0.0493 \end{bmatrix} = \begin{bmatrix} -0.2645 \\ -1.0308 \\ -0.5007 \\ 0.2311 \\ -1.5664 \end{bmatrix}$$

Projecting the data (Class 2):

$$\begin{bmatrix} -0.260157 & +0.842311 \\ +0.449362 & -0.327565 \\ +1.395388 & -0.561541 \\ -0.023650 & -1.029492 \\ +1.631895 & -0.093590 \end{bmatrix}\begin{bmatrix} 0.9988 \\ 0.0493 \end{bmatrix} = \begin{bmatrix} -0.2184 \\ 0.4327 \\ 1.3660 \\ -0.0744 \\ 1.6253 \end{bmatrix}$$

(e)

On taking the feature 3 as the projected data and computing the information gain on that.

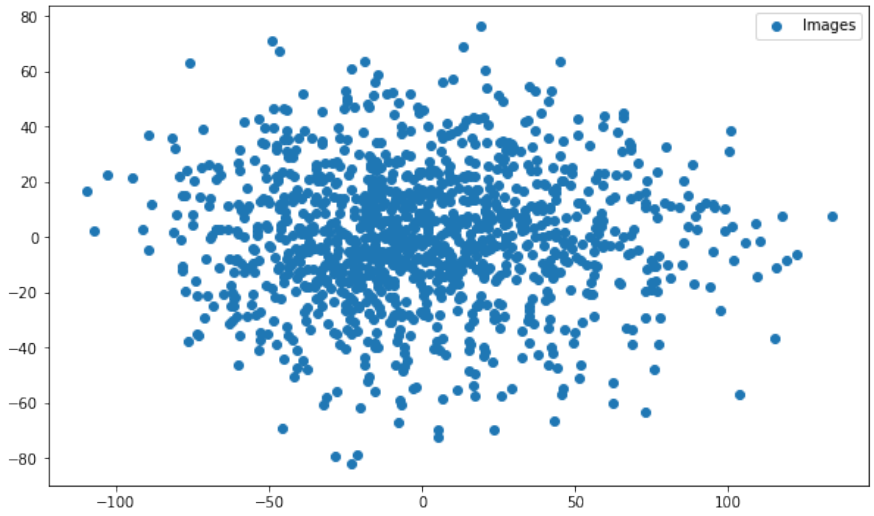
$r(3) = 10 \times \frac{1+0}{5+5} H\left(\frac{1}{1+0}, \frac{0}{1+0}\right) = 0, IG(3) = 1 - 0 = 1$

The value we are getting is greater than $IG(1)$.

For this feature, we conclude that most data in class 1 are smaller and most data in class 2 are larger. From this we can conclude that the projection we performed gives us a good class separation.

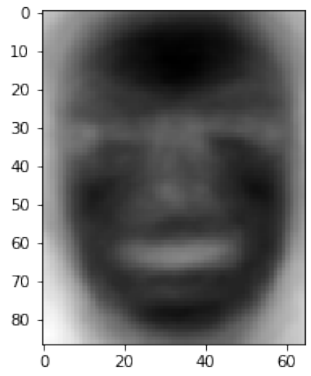
2. PART-2: -

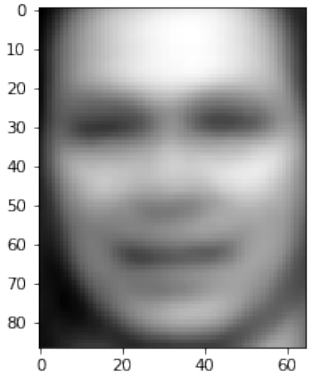
(i) Accuracy of KNN from original dataset = 0.2784
(ii) Accuracy of KNN :- 0.2784
(iii) Accuracy of KNN from PCA 100D = 0.3012
(iv) Accuracy of KNN from PCA whitened 100D = 0.3468
(v) Scatter Plot for 2D projection of data: -



3. PART 3:-

(i) Visualization of Primary principal component:-





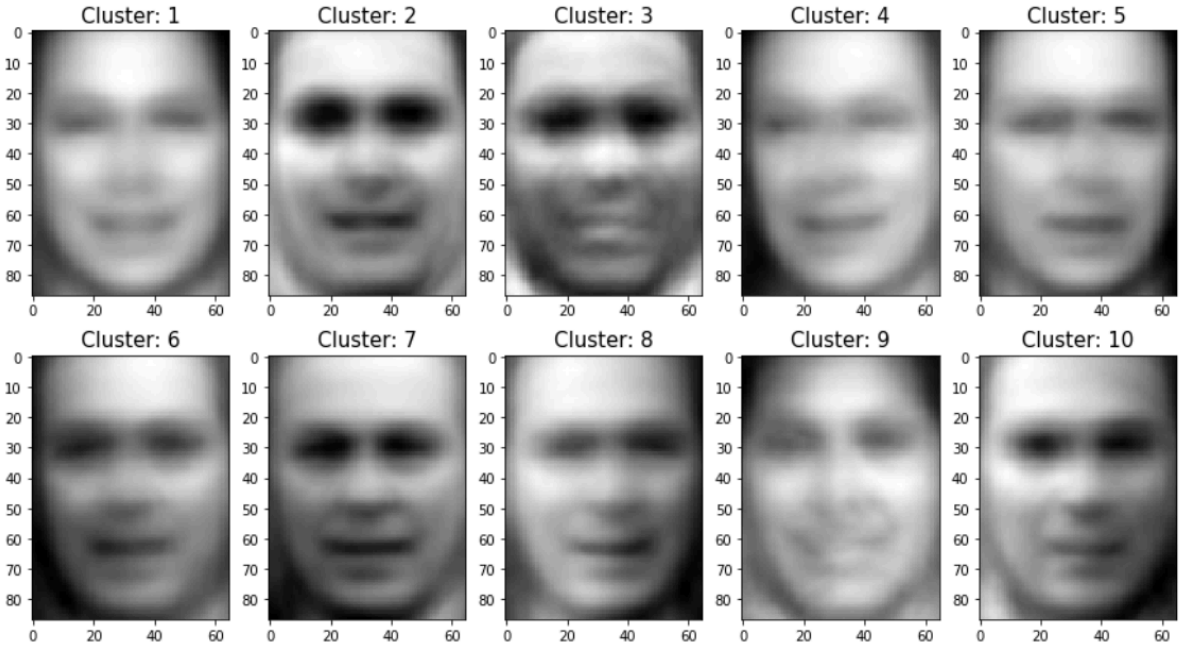
(ii) Number of principal components needed to represent 95% of information, k:-

4. PART 4:-

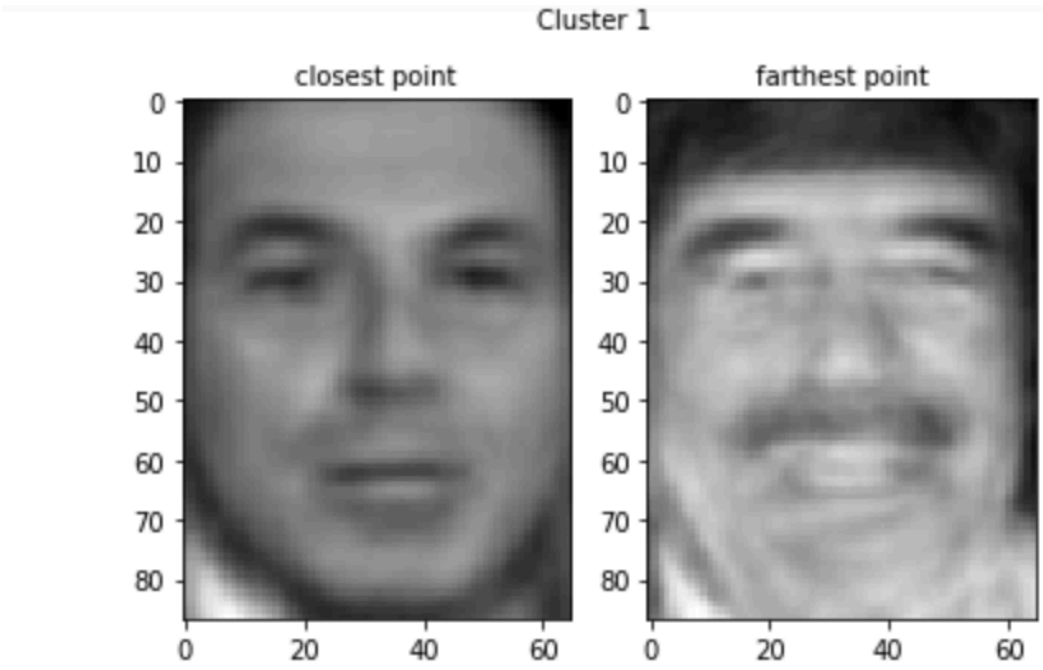
(i) The visualization of k-means cluster centers:-

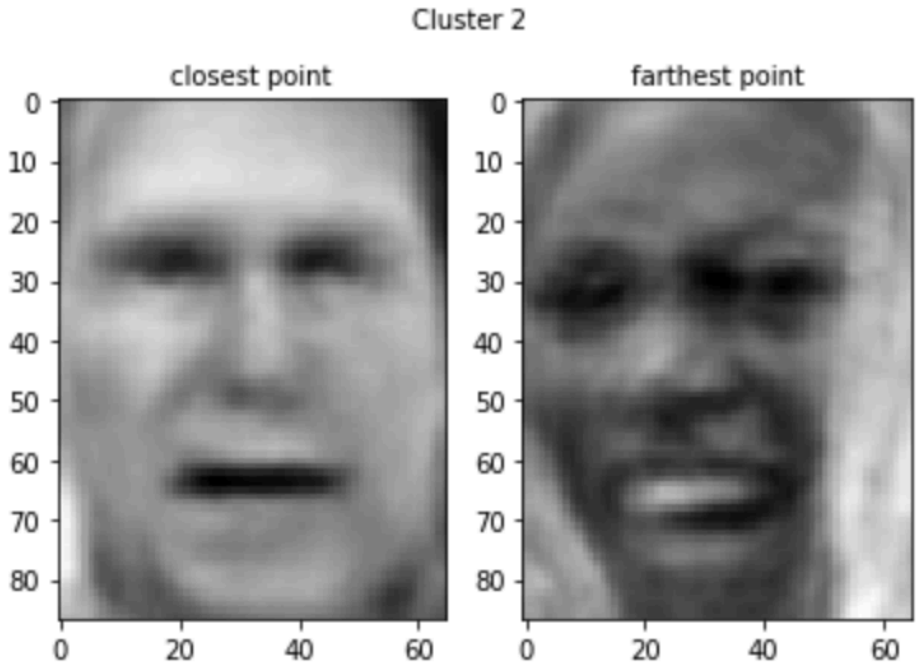
Cluster 1:- 213
Cluster 1:- 117
Cluster 1:- 70
Cluster 1:- 128
Cluster 1:- 148
Cluster 1:- 134
Cluster 1:- 129
Cluster 1:- 112
Cluster 1:- 64
Cluster 1:- 67

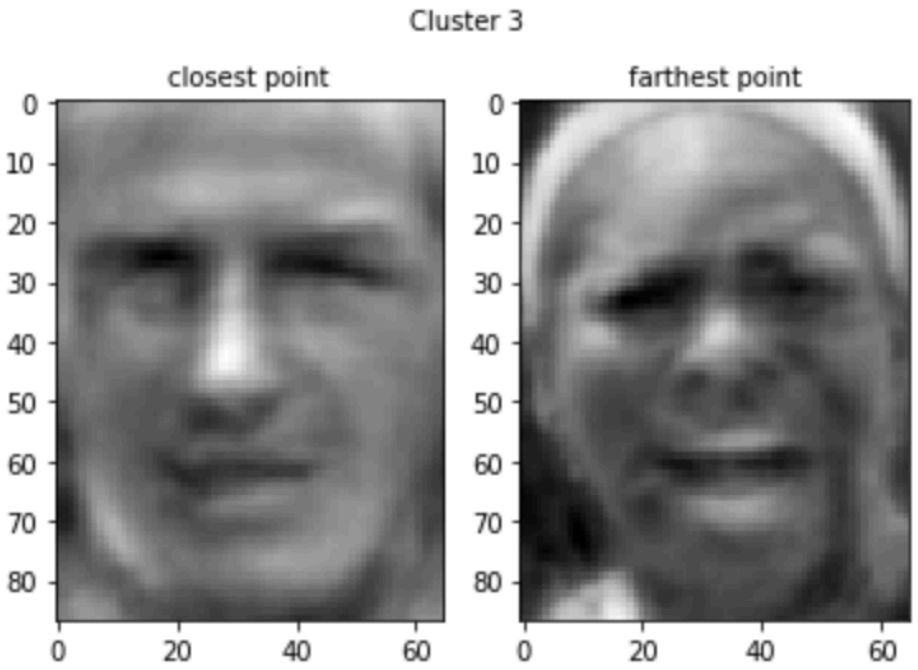
(ii)

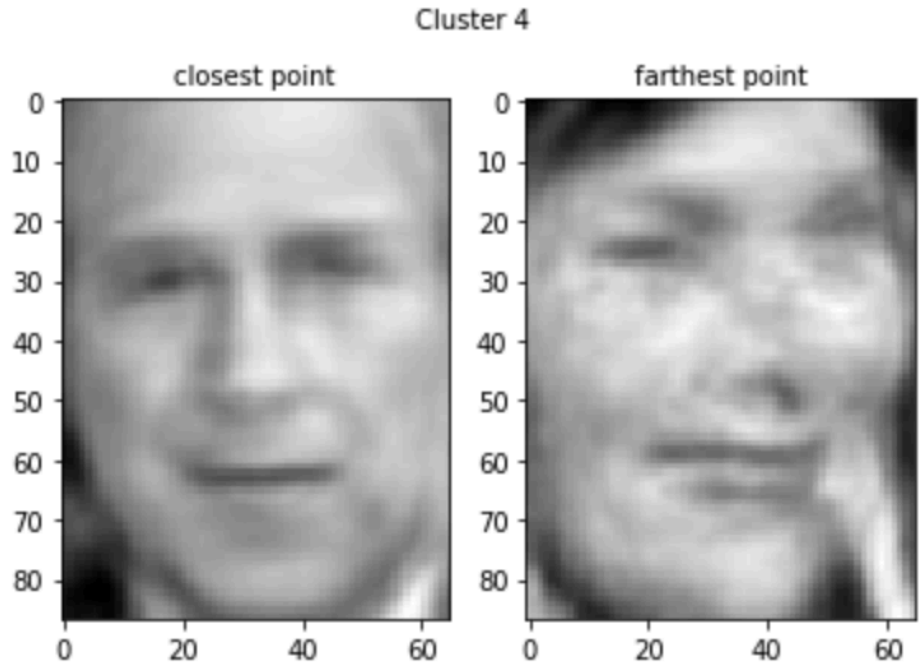


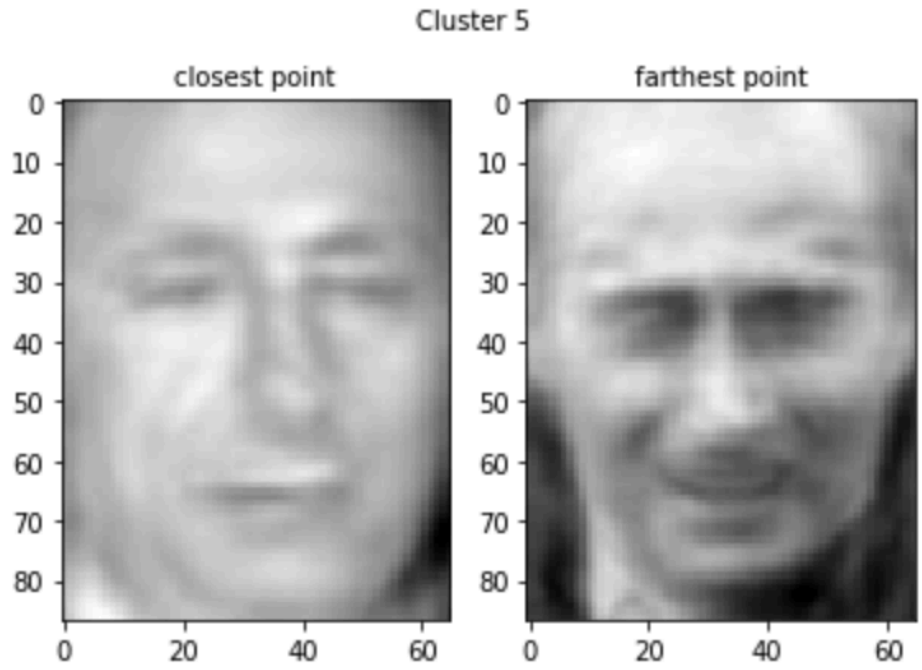
(iii) The minimum and maximum images:-

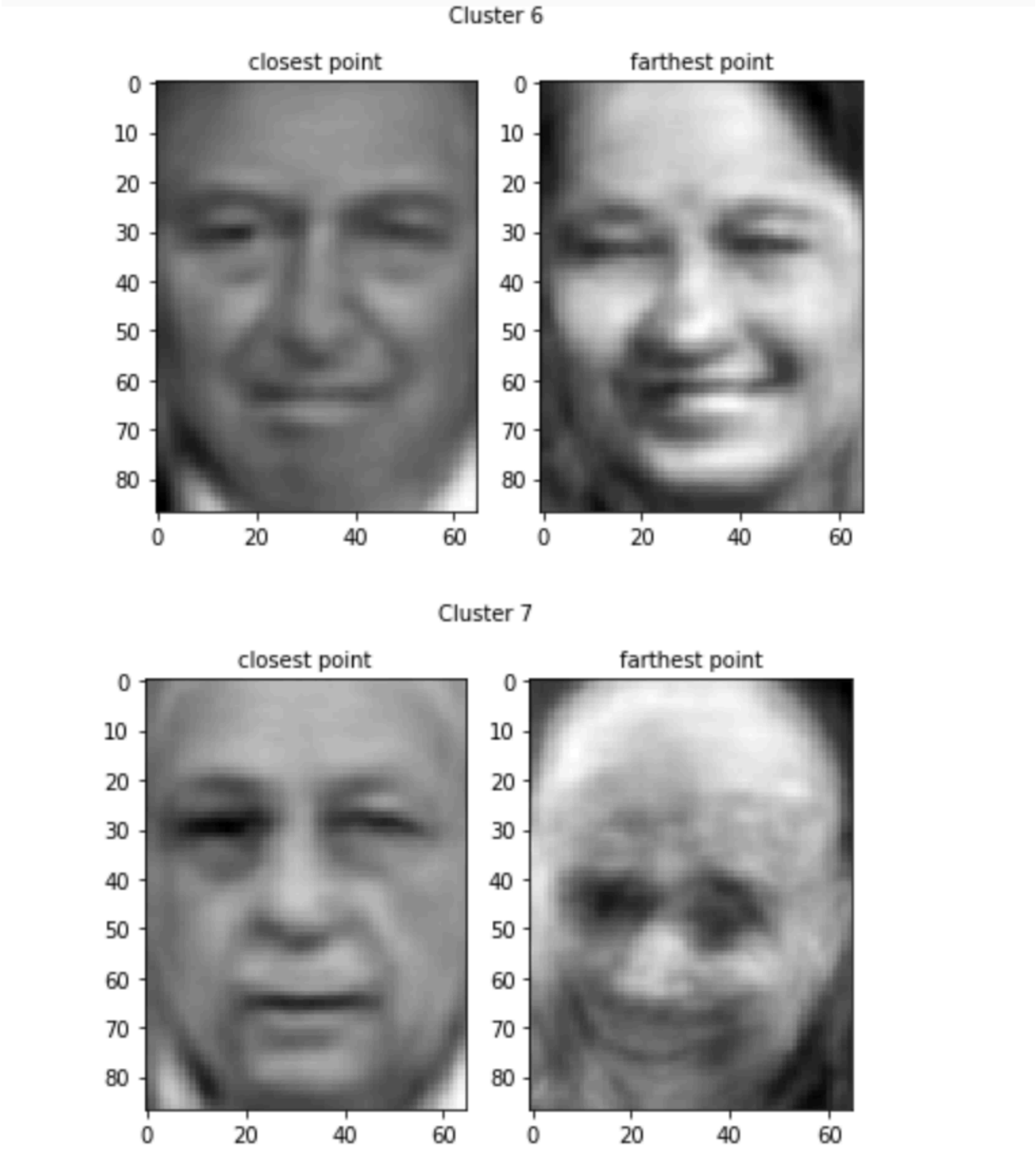

Cluster 1
closest point
farthest point
0
10
20
30
40
50
60
70
80
0
20
40
60

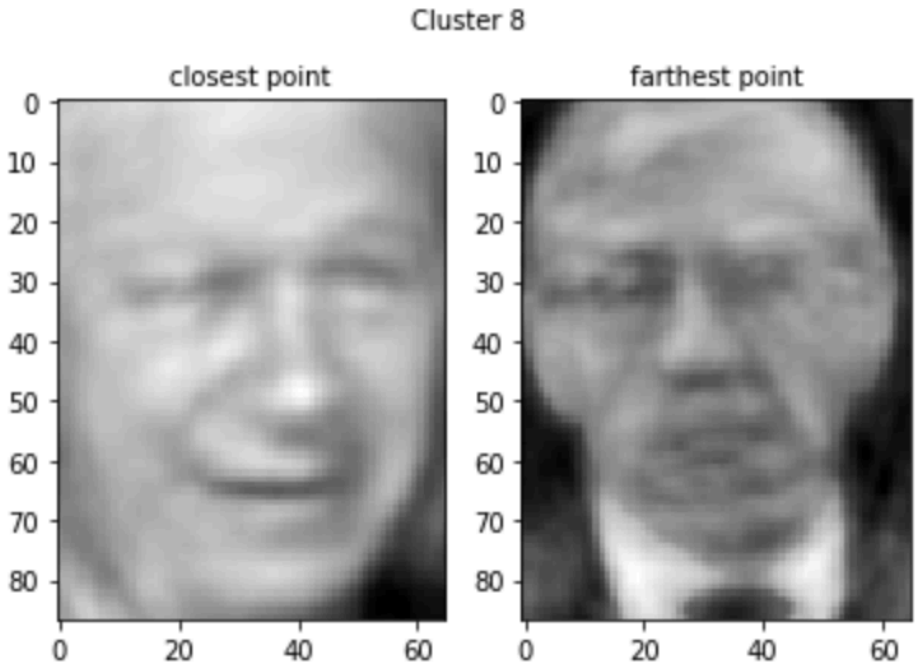

Cluster 2
closest point
farthest point
0
10
20
30
40
50
60
70
80
0
20
40
60

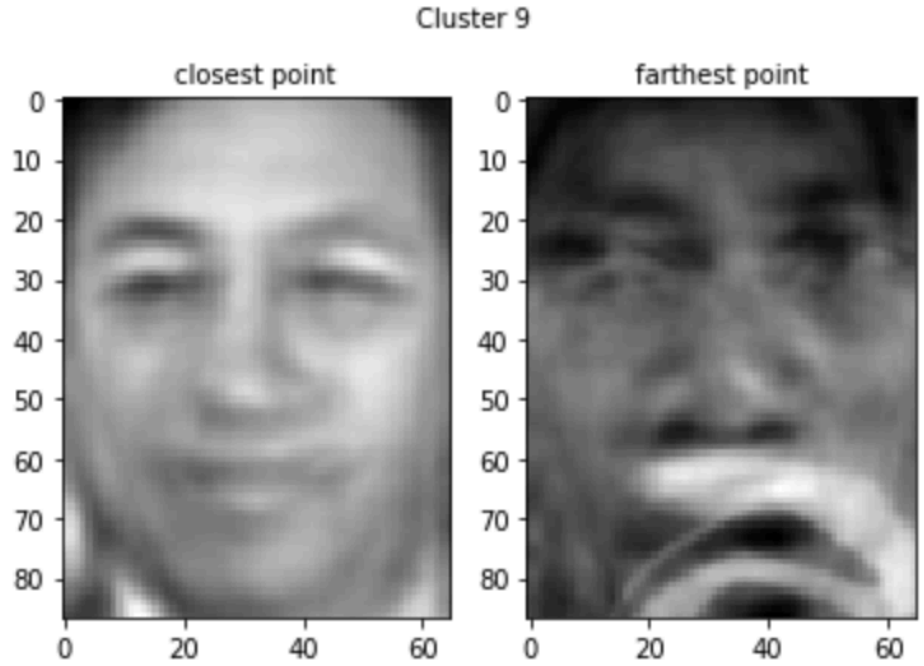

Cluster 3
closest point
farthest point
0
10
20
30
40
50
60
70
80
0
20
40
60

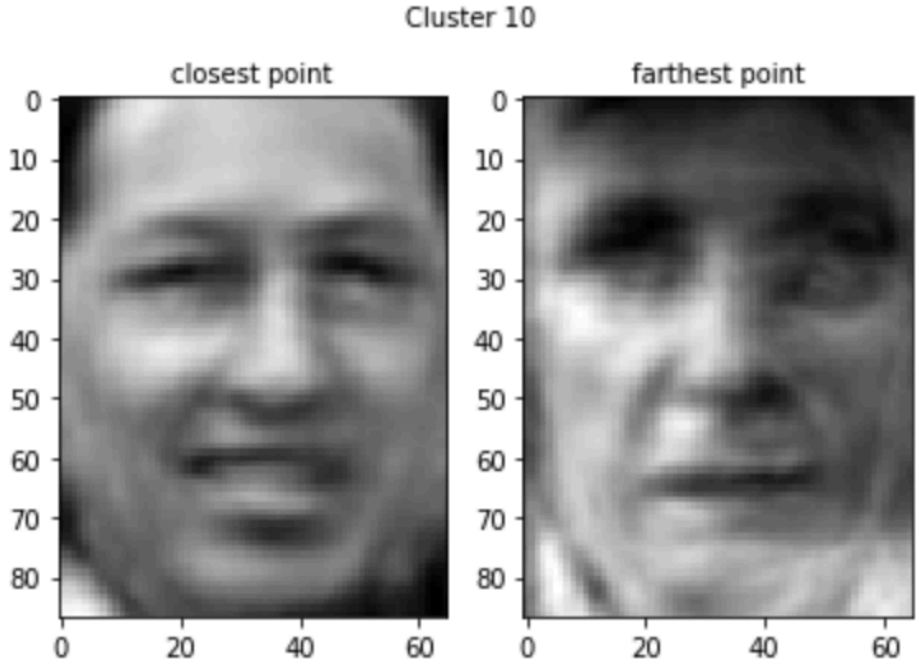

Cluster 4
closest point
farthest point
0
10
20
30
40
50
60
70
80
0
20
40
60


Cluster 5
closest point
farthest point
0
10
20
30
40
50
60
70
80
0
20
40
60


Cluster 6
closest point
farthest point
Cluster 7
closest point
farthest point


Cluster 8
closest point
farthest point
0
10
20
30
40
50
60
70
80
0
20
40
60


Cluster 9
closest point
farthest point
0
10
20
30
40
50
60
70
80
0
20
40
60


Cluster 10
closest point
farthest point
0
10
20
30
40
50
60
70
80
0
20
40
60