Greg Matthews
CS 383
January 15th, 2018

Assignment 1 – Dimensionality Reduction

Part 1: Theory Questions

1. Consider the following data:

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

(a) Find the principle components of the data (you must show the math, including how you compute the eigenvectors and eigenvalues). Make sure you standardize the data first and that your principle components are normalized to be unit length. As for the amount of detail needed in your work imagine that you were working on paper with a basic calculator. Show me whatever you would be writing on that paper. (5pts).

Answer:

Step 1: Standardize Data

$Feature\ 1\text{: } \mu = -0.9,\ \sigma = 4.23 \qquad where\ \mu = mean,\ \sigma = standard\ deviation,$

$Feature\ 2\text{: } \mu = 1.4,\ \sigma = 4.27$

$$X_{new} = \frac{X_{i,j} - \mu}{\sigma} \quad \forall i, j \in X$$

Data after standardizing becomes:

$$\begin{bmatrix} -0.260 & -0.094 \\ -0.969 & -1.26 \\ -0.497 & -0.094 \\ 0.212 & 0.374 \\ -1.68 & 2.25 \\ -0.260 & 0.842 \\ 0.449 & -0.328 \\ 1.39 & -0.562 \\ -0.024 & -1.03 \\ 1.63 & -0.094 \end{bmatrix}$$

Step 2: Computing Covariance Matrix:

Computing Covariance Matrix:

$$\sum = \frac{X^T X}{N-1} \qquad where\ N\ is\ \#\ of\ observations\ /\ feature = 10$$

$$\Sigma = \begin{bmatrix} -0.260 & -0.969 & -0.497 & 0.212 & -1.68 & -0.260 & 0.449 & 1.39 & -0.024 & 1.63 \\ -0.094 & -1.26 & -0.094 & 0.374 & 2.25 & 0.842 & -0.328 & -0.562 & -1.03 & -0.094 \end{bmatrix} \begin{bmatrix} -0.260 & -0.094 \\ -0.969 & -1.26 \\ -0.497 & -0.094 \\ 0.212 & 0.374 \\ -1.68 & 2.25 \\ -0.260 & 0.842 \\ 0.449 & -0.328 \\ 1.39 & -0.562 \\ -0.024 & -1.03 \\ 1.63 & -0.094 \end{bmatrix} /9$$

$$\sum = \begin{bmatrix} 1.0 & -0.408 \\ -0.408 & 1.0 \end{bmatrix}$$

Step 3: Computing Eigen Vector/Eigen Values:

From argmax function:

$$J(w) = argmax_w(w^T \Sigma w - \alpha\ (w^T w - 1))$$

$$\frac{dJ}{dw} = 2\Sigma w\ -\ 2\alpha w\ =\ 0$$

$$\Sigma w\ =\ \alpha w$$

$$where\ \Sigma = covariance\ matrix, and\ \ \alpha = scaling\ factor$$

w can be solved using eigen-decomposition:

$$\Sigma w\ =\ \alpha w \quad \rightarrow \quad Ax\ =\ \lambda x$$

$$where\ \lambda = eigen\ value, and\ x = eigen\ vector$$

Eigen values exist iff A is a square matrix and determinant of $A - \lambda I = 0$:

$$|A - \lambda I| = 0$$

$$\left| \begin{bmatrix} 1.0 & -0.408 \\ -0.408 & 1.0 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \right| = 0$$

$$\left| \begin{bmatrix} 1.0 - \lambda & -0.408 \\ -0.408 & 1.0 - \lambda \end{bmatrix} \right| = 0$$

$$(1.0 - \lambda)^2 - (-0.408)^2 = 0$$

$$1.0 - 2\lambda + \lambda^2 - 0.166 = 0$$

$$\lambda^2 - 2\lambda + 0.833 = 0$$

Solving for eigen values using quadratic equation:

$$\lambda = \frac{2 \pm \sqrt{-2^2 - 4(1)(0.833)}}{2} = \frac{2 \pm 0.817}{2} = 1 \pm 0.409 = [1.408 \quad 0.591]$$

Finding eigen vector of largest eigenvalue:

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

$$\left[\begin{bmatrix} 1.0 & -0.408 \\ -0.408 & 1.0 \end{bmatrix} - \begin{bmatrix} 1.408 & 0 \\ 0 & 1.408 \end{bmatrix}\right] * \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} -0.408 & -0.408 \\ -0.408 & -0.408 \end{bmatrix} * \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$-0.408x - 0.408y = 0$$

$$Let\ x = 1, then\ y = -1$$

$$eigen\ vector \rightarrow\ w = \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

Normalizing w to be unit length:

$$w = \frac{w}{|w|} = \frac{\begin{bmatrix} 1 \\ -1 \end{bmatrix}}{\begin{bmatrix} \sqrt{2} \\ \sqrt{2} \end{bmatrix}} = \begin{bmatrix} \mathbf{0.707} \\ \mathbf{-0.707} \end{bmatrix}$$ **= Principal Component 1**

Finding eigen vector of second largest eigenvalue:

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

$$\left[\begin{bmatrix} 1.0 & -0.408 \\ -0.408 & 1.0 \end{bmatrix} - \begin{bmatrix} 0.591 & 0 \\ 0 & 0.591 \end{bmatrix}\right] * \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\begin{bmatrix} 0.408 & -0.408 \\ -0.408 & 0.408 \end{bmatrix} * \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$0.408x - 0.408y = 0$$

$$Let\ x = 1, then\ y = 1$$

$$eigen\ vector \rightarrow\ w = \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Normalizing w to be unit length:

$$w = \frac{w}{|w|} = \frac{\begin{bmatrix} 1 \\ 1 \end{bmatrix}}{\begin{bmatrix} \sqrt{2} \\ \sqrt{2} \end{bmatrix}} = \begin{bmatrix} \mathbf{0.707} \\ \mathbf{0.707} \end{bmatrix}$$ **= Principal Component 2**

(b) Project the data onto the principal component corresponding to the largest eigenvalue found in the previous part (3pts).

Answer:

$$New\ data: \quad Z = Xw$$

$$Z = \begin{bmatrix} -0.260 & -0.094 \\ -0.969 & -1.26 \\ -0.497 & -0.094 \\ 0.212 & 0.374 \\ -1.68 & 2.25 \\ -0.260 & 0.842 \\ 0.449 & -0.328 \\ 1.39 & -0.562 \\ -0.024 & -1.03 \\ 1.63 & -0.094 \end{bmatrix} * \begin{bmatrix} 0.707 \\ -0.707 \end{bmatrix} = \begin{bmatrix} \mathbf{-0.117} \\ \mathbf{0.205} \\ \mathbf{-0.285} \\ \mathbf{-0.115} \\ \mathbf{-2.79} \\ \mathbf{-0.779} \\ \mathbf{0.549} \\ \mathbf{1.38} \\ \mathbf{0.711} \\ \mathbf{1.22} \end{bmatrix}$$

2. Consider the following data:

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

(a) Compute the information gain for each feature. You could standardize the data overall, although it won't make a difference. (5pts).

Answer:

Feature 1:

$$IG(A) = H\left(\frac{p}{p+n}, \frac{n}{p+n}\right) - remainder(A) = H\left(\frac{5}{10}, \frac{5}{10}\right) - remainder(A)$$

$$IG(A) = \left[-\frac{5}{10}\log_2\frac{5}{10} - \frac{5}{10}\log_2\frac{5}{10}\right] - remainder(A) = 1 - remainder(A)$$

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

$$remainder(A) = \frac{2}{10} H\left(\frac{1}{2}, \frac{1}{2}\right) = \frac{2}{10} * \left[-\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2}\right] = 0.2$$

$$IG(A) = 1 - 0.2 = \mathbf{0.8}$$

Feature 2:

$$IG(A) = H\left(\frac{p}{p+n}, \frac{n}{p+n}\right) - remainder(A) = H\left(\frac{5}{10}, \frac{5}{10}\right) - remainder(A)$$

$$IG(A) = \left[-\frac{5}{10}\log_2\frac{5}{10} - \frac{5}{10}\log_2\frac{5}{10}\right] - remainder(A) = 1 - remainder(A)$$

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

$$remainder(A) = \frac{3}{10} H\left(\frac{2}{3}, \frac{1}{3}\right) = \frac{3}{10} * \left[-\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3}\right] = 0.2755$$

$$IG(A) = 1 - 0.2755 = \mathbf{0.7245}$$

(b) Which feature is more discriminating based on results in part a (1pt)?

Answer:

Based on the results, the first feature was computed to have an information gain of 0.8 whereas the second feature was found to be 0.7245. From these results, it can be concluded that the first feature is more discriminating due to a higher information gain, which correlates to an increase in class separability.

(c) Using LDA, find the direction of projection (you must show the math, however for this one you don't have to show the computation for finding the eigenvalues and eigenvectors). Normalize this vector to be unit length (5pts).

Answer:

1. Standardizing the data:

$Feature\ 1\text{: } \mu = -0.9,\ \sigma = 4.23$ $\qquad where\ \mu = mean,\ \sigma = standard\ deviation,$

$Feature\ 2\text{: } \mu = 1.4,\ \sigma = 4.27$

$$X_{stand} = \begin{bmatrix} -0.260 & -0.094 \\ -0.969 & -1.26 \\ -0.497 & -0.094 \\ 0.212 & 0.374 \\ -1.68 & 2.25 \\ -0.260 & 0.842 \\ 0.449 & -0.328 \\ 1.39 & -0.562 \\ -0.024 & -1.03 \\ 1.63 & -0.094 \end{bmatrix}$$

2. Computing means for each class:

$$Class\ 1\ mean: \mu_1 = [-0.639, 0.234]$$

$$Class\ 2\ mean: \mu_2 = [0.639, -0.234]$$

3. Computing scatter matrices for each class:

$Class$ 1:

$$\sigma_1^2 = (|C_1| - 1) * cov(C_1)$$

$$\sigma_1^2 = (5 - 1) * cov(C_1)$$

$$Cov(C_1) = \Sigma = \begin{bmatrix} E[(C_1[:,0] - \mu_1[0])(C_1[:,0] - \mu_1[0])] & E[(C_1[:,0] - \mu_1[0])(C_1[:,1] - \mu_1[1])] \\ E[(C_1[:,1] - \mu_1[1])(C_1[:,0] - \mu_1[0])] & E[(C_1[:,1] - \mu_1[1])(C_1[:,1] - \mu_1[1])] \end{bmatrix}$$

$$where\ E[(C_1[:,i] - \mu_1[i])(C_1[:,j] - \mu_1[j])] = \frac{1}{N}\sum_{k=1}^{N}(C_k[:,i] - \mu_i)\left(C_k[:,j] - \mu_j\right)$$

$$\Sigma = \begin{bmatrix} 0.520 & -0.412 \\ -0.412 & 1.63 \end{bmatrix}$$

$$\sigma_1^2 = (5 - 1) * \begin{bmatrix} 0.520 & -0.412 \\ -0.412 & 1.63 \end{bmatrix} = \begin{bmatrix} \mathbf{2.08} & \mathbf{-1.65} \\ \mathbf{-1.65} & \mathbf{6.53} \end{bmatrix}$$

$Class$ 2:

$$\sigma_2^2 = (|C_2| - 1) * cov(C_2)$$

$$\sigma_2^2 = (5 - 1) * cov(C_2)$$

$$Cov(C_1) = \Sigma = \begin{bmatrix} E[(C_2[:,0] - \mu_2[0])(C_2[:,0] - \mu_2[0])] & E[(C_2[:,0] - \mu_2[0])(C_2[:,1] - \mu_2[1])] \\ E[(C_2[:,1] - \mu_2[1])(C_2[:,0] - \mu_2[0])] & E[(C_2[:,1] - \mu_2[1])(C_2[:,1] - \mu_2[1])] \end{bmatrix}$$

$$where\ E[(C_2[:,0] - \mu_2)(C_2[:,0] - \mu_2)] = \frac{1}{N}\sum_{k=1}^{N}(C_k[:,i] - \mu_2[i])(C_k[:,j] - \mu_2[j])$$

$$\Sigma = \begin{bmatrix} 0.710 & -0.133 \\ -0.133 & 0.482 \end{bmatrix}$$

$$\sigma_2^2 = (5 - 1) * \begin{bmatrix} 0.710 & -0.133 \\ -0.133 & 0.482 \end{bmatrix} = \begin{bmatrix} \mathbf{2.84} & \mathbf{-0.531} \\ \mathbf{-0.531} & \mathbf{1.93} \end{bmatrix}$$

4. Computing within/between class scatter matrices:

$Within\ Class\ Scatter\ Matrix\colon\ S_w = \sigma_1^2 + \sigma_2^2$

$$S_w = \begin{bmatrix} 2.08 & -1.65 \\ -1.65 & 6.53 \end{bmatrix} + \begin{bmatrix} 2.84 & -0.531 \\ -0.531 & 1.93 \end{bmatrix}$$

$$S_w = \begin{bmatrix} \mathbf{4.92} & \mathbf{-2.18} \\ \mathbf{-2.18} & \mathbf{8.45} \end{bmatrix}$$

$Between\ Class\ Scatter\ Matrix\colon\ S_b = (\mu_1 - \mu_2)^T(\mu_1 - \mu_2)$

$S_b = ([-0.639, 0.234] - [0.639, -0.234])^T([-0.639, 0.234] - [0.639, -0.234])$

$$S_b = \begin{bmatrix} \mathbf{1.63} & \mathbf{-0.597} \\ \mathbf{-0.597} & \mathbf{0.219} \end{bmatrix}$$

1. Eigen Decomposition to solve for W:

$$Fischer\ Discriminant\ Function\colon J(W) = \frac{(\mu_1 W - \mu_2 W)^T(\mu_1 W - \mu_2 W)}{(\sigma_1 W)^T(\sigma_1 W) + \ (\sigma_2 W)^T(\sigma_2 W)}$$

$$J(W) = \frac{W^T S_b W}{W^T S_W W} \quad ,Taking\ directive\ wrt\ W,$$

$$S_b W - \frac{W^T S_b W S_w W}{W^T S_w W} = 0 \quad , \quad Let\ \lambda = \frac{W^T S_b W}{W^T S_w W} \quad Then,$$

$$Sw^{-1} S_b W = \ \lambda W \quad , \qquad Let\ A = Sw^{-1} S_b W$$

$$Performing\ Eigen\ Decomposition\ we\ have,$$

$$Eigen\ value = [0.333\ 0]$$

$$Eigen\ vector = \begin{bmatrix} 0.998 & 0.334 \\ 0.0491 & 0.939 \end{bmatrix}$$

$$Eigen\ vector\ with\ highest\ eigen\ value:$$

$$W = \begin{bmatrix} \mathbf{0.998} \\ \mathbf{0.0491} \end{bmatrix} = \boldsymbol{Direction\ of\ Projection}$$

(d) Project the data onto the principal component found in the previous part (3pts).

Answer:

$$Z = XW$$

$$Z = \begin{bmatrix} -0.260 & -0.094 \\ -0.969 & -1.26 \\ -0.497 & -0.094 \\ 0.212 & 0.374 \\ -1.68 & 2.25 \\ -0.260 & 0.842 \\ 0.449 & -0.328 \\ 1.39 & -0.562 \\ -0.024 & -1.03 \\ 1.63 & -0.094 \end{bmatrix} \begin{bmatrix} 0.964 \\ 0.267 \end{bmatrix} = \begin{bmatrix} \mathbf{-0.264} \\ \mathbf{-1.03} \\ \mathbf{-0.500} \\ \mathbf{0.231} \\ \mathbf{-1.57} \\ \mathbf{-0.218} \\ \mathbf{0.432} \\ \mathbf{1.37} \\ \mathbf{-0.074} \\ \mathbf{1.63} \end{bmatrix}$$

(d) Does the projection you performed in the previous part seem to provide good class separation? Why or why not (1pt)?

Answer:

The projection of the 2D space to 1D using Linear Discrimination Analysis performed adequately well, however was not able to perfectly separate the 2 classes, as shown in fig 2. We can see that the group of Class 1 and Class 2 points reside within a general area that is overlapping between one another. The reason for this is that the algorithms maximization of the difference in mean between classes and minimization in variation between points within a class reached its optimal separation, however this doesn't mean the algorithm will perfectly separate the classes.

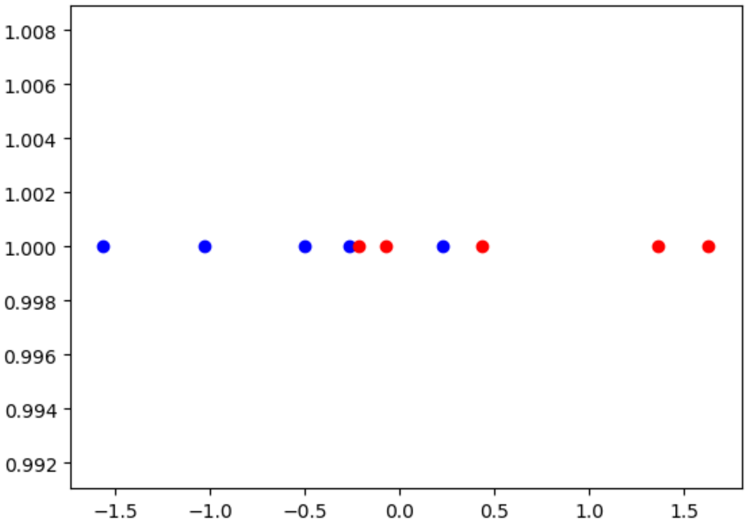


Figure 1. LDA performed on input data (Class1 = Blue, Class2 = Red)

Part 2: Dimensionality Reduction via PCA

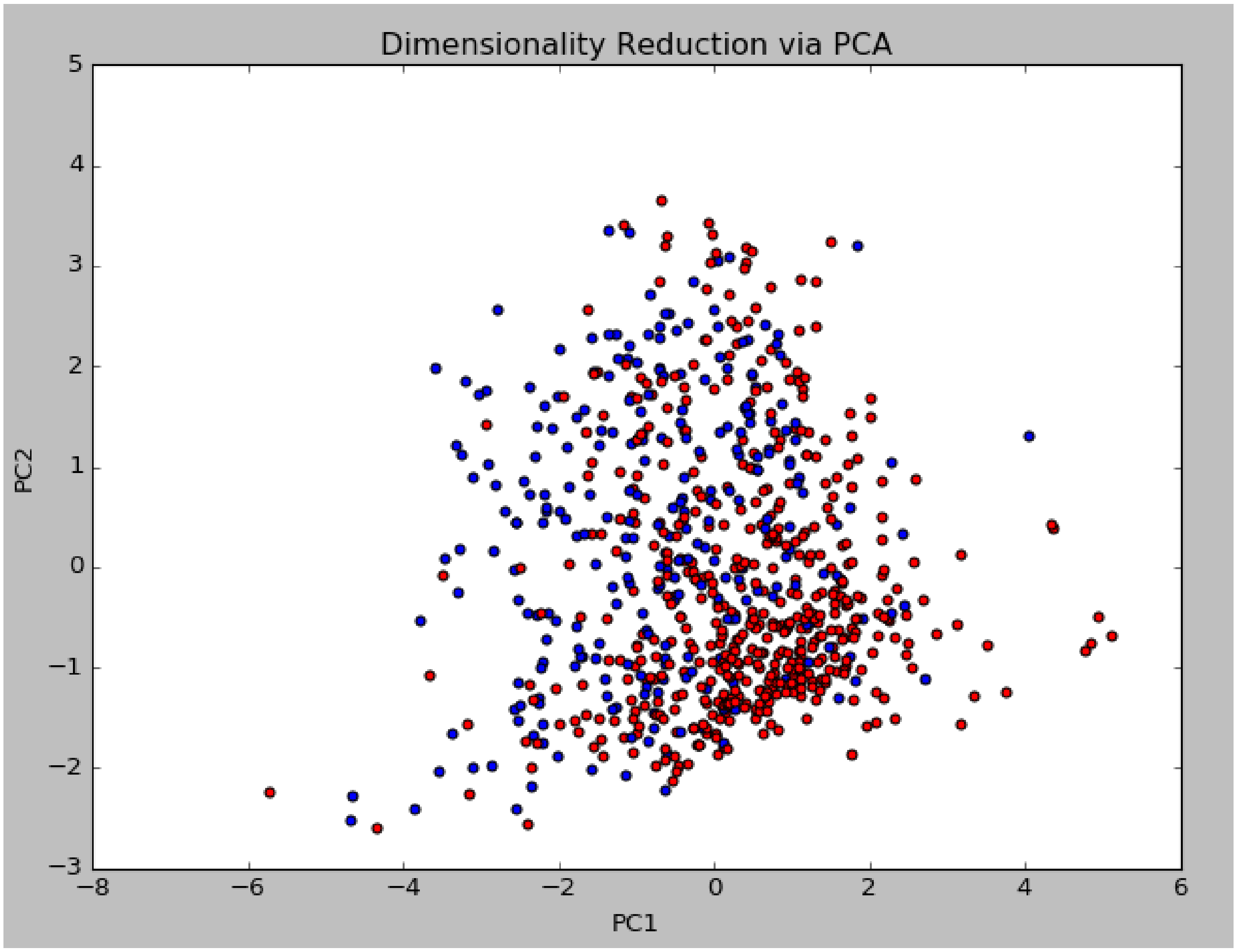


Figure 2. 2 Dimensional PCA projection of Cancer data

Note: (Figure is flipped horizontally due to different ordering)

Part 3: Eigen Faces

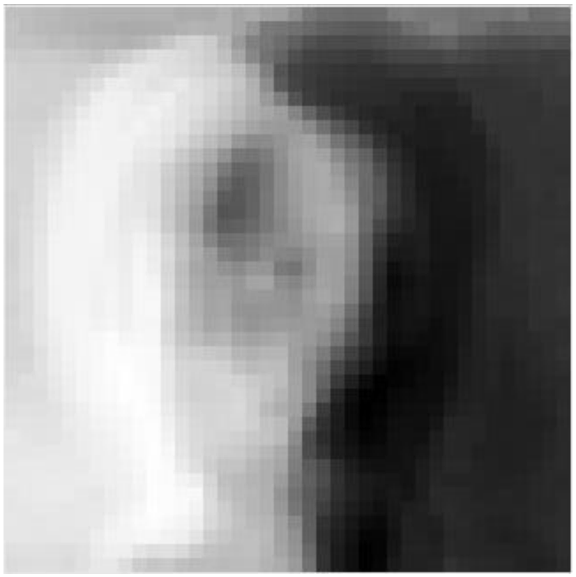
(a) Number of principle components needed to represent 95% of information, k.

Answer:

$$\boldsymbol{k = 33}$$

(b) Visualization of primary principle component

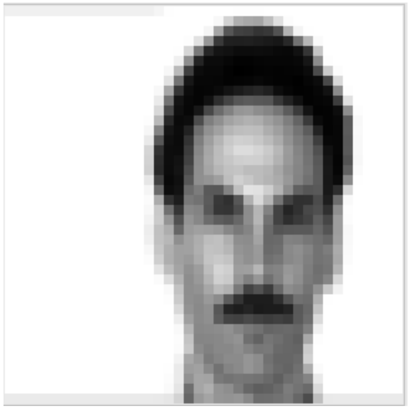
Answer:



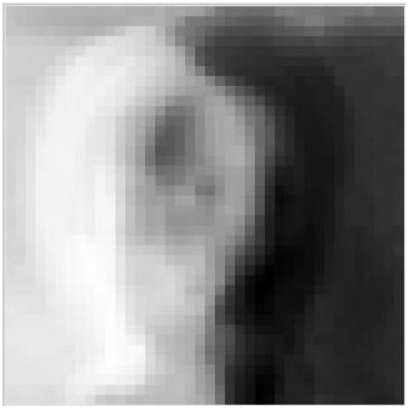
(c) Visualization of the reconstruction of the first person

Answer:

i. Original Image



ii. Single Principle Component



iii. k Principle Components