CS 613 - Machine Learning

Assignment 4 - Dimensionality Reduction
Robert Thompson

1 Theory Questions

All of the theory questions will use the following data:

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

1. Zscore the data and create a 2D plot of the datapoints, visualizing class one data as squares, and class two data as circles (5pts).

   (a) Define $X = \begin{bmatrix} -2 & 1 \\ -5 & -4 \\ -3 & 1 \\ 0 & 3 \\ -8 & 11 \\ -2 & 5 \\ 1 & 0 \\ 5 & -1 \\ -1 & -3 \\ 6 & 1 \end{bmatrix}$

   (b) Calculate the X Means and Standard Deviations

   $\mu = \begin{bmatrix} -0.9 & 1.4 \end{bmatrix}$

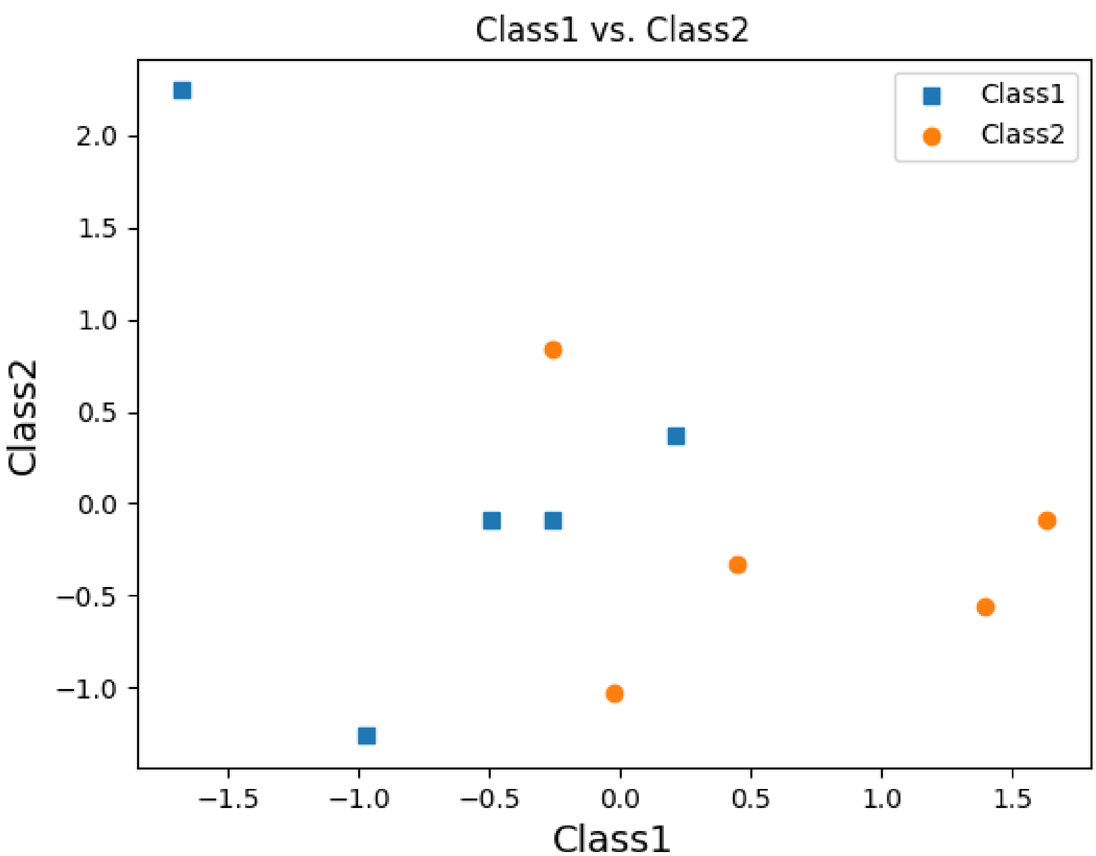
   $\sigma = \begin{bmatrix} 4.23 & 4.27 \end{bmatrix}$

(c) Z-Score our X Data

$$X_{zscored} = \begin{bmatrix} -2 & 1 \\ -5 & -4 \\ -3 & 1 \\ 0 & 3 \\ -8 & 11 \\ -2 & 5 \\ 1 & 0 \\ 5 & -1 \\ -1 & -3 \\ 6 & 1 \end{bmatrix} - \begin{bmatrix} -0.9 & 1.4 \end{bmatrix} / \begin{bmatrix} 4.23 & 4.27 \end{bmatrix}$$

$$= \begin{bmatrix} -0.26 & -0.09 \\ -0.97 & -1.26 \\ -0.50 & -0.09 \\ 0.21 & 0.37 \\ -1.68 & 2.25 \\ -0.26 & 0.84 \\ 0.45 & -0.33 \\ 1.40 & -0.56 \\ -0.02 & -1.03 \\ 1.63 & -0.09 \end{bmatrix}$$

(d) 2D Class 1 vs. Class 2 Plot



2. PCA

   (a) Find the principle components of the data. You may use an *eig* function, but show the math leading up to, and after, using that function. Make sure that your principle components are normalized to be unit length. (5pts).

      i. Transpose our $X_{zscored}$ Data

$$X^T = \begin{bmatrix} -0.26 & -0.97 & -0.50 & 0.21 & -1.68 & -0.26 & 0.45 & 1.40 & -0.02 & 1.63 \\ -0.09 & -1.26 & -0.09 & 0.37 & 2.25 & 0.84 & -0.33 & -0.56 & -1.03 & -0.09 \end{bmatrix}$$

      ii. Calculate the Covariance Matrix of our $X_{zscored}$ Data

$$\sum = \frac{X^T X}{N-1} = \begin{bmatrix} -1 & -0.41 \\ -0.41 & 1 \end{bmatrix}$$

      iii. Perform Eigendecomposition on the Covariance Matrix

$$\text{Eigenvalues} = \begin{bmatrix} 1.41 & 0.59 \end{bmatrix}$$

$$\text{Eigenvectors} = \begin{bmatrix} 0.71 & 0.71 \\ -0.71 & 0.71 \end{bmatrix}$$

      iv. Determine Largest Eigenvalue and its Associated Eigenvector

Largest Eigenvalue $= \begin{bmatrix} 1.41 \end{bmatrix}$

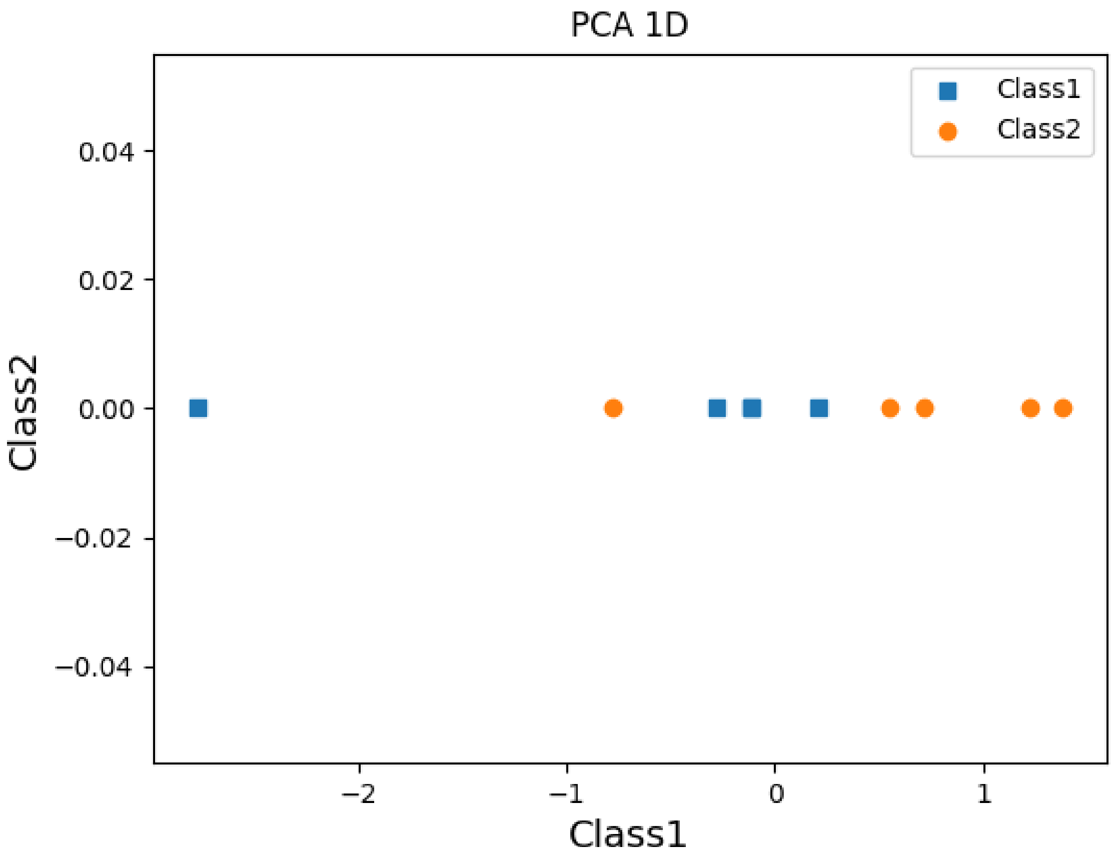
Largest Eigenvectors $= \begin{bmatrix} 0.71 \\ -0.71 \end{bmatrix}$

v. Project $X_{zscored}$ into Eigenvector Space

$$Z = XW = \begin{bmatrix} 0.26 & -0.09 \\ -0.97 & -1.26 \\ -0.50 & -0.09 \\ 0.21 & 0.37 \\ -1.68 & 2.25 \\ -0.26 & 0.84 \\ 0.45 & -0.33 \\ 1.40 & -0.56 \\ -0.02 & -1.03 \\ 1.63 & -0.09 \end{bmatrix} \begin{bmatrix} 0.71 \\ -0.71 \end{bmatrix}$$

$$= \begin{bmatrix} -0.12 \\ 0.21 \\ -0.29 \\ -0.11 \\ -2.78 \\ -0.78 \\ 0.55 \\ 1.38 \\ 0.71 \\ 1.22 \end{bmatrix}$$

(b) Project your data down to 1D using the principle component associated with the largest eigenvalue. Plot this data in 1D, visualizing class one data as squares, and class two data as circles (3pts).



(c) Based on your plot, does the PCA projection provide good class separation? Why or why not (2pts)?
Given the 1-dimensional plot of class 1 and class 2 it is extremely easy to see how the classes are separated. There is no overlap between class 1 (squares) and class 2 (circles) on the graph. Therefore, this PCA projection provides a good class separation.

3. LDA

(a) Using LDA, find the direction of projection (show your work in detail similar to the prior question). Normalize this vector to be unit length (5pts).

i. Compute the Means for Class 1 and Class 2

$$\mu^{(1)} = \begin{bmatrix} -0.64 & 0.24 \end{bmatrix}$$

$$\mu^{(2)} = \begin{bmatrix} 0.64 & -0.24 \end{bmatrix}$$

ii. Compute Scatter Matrices for Class 1 and Class 2

$$\sigma^{(1)^2} = \begin{bmatrix} 2.04 & -0.75 \\ -0.75 & 0.27 \end{bmatrix}$$

$$\sigma^{(2)^2} = \begin{bmatrix} 4.08 & -1.50 \\ -1.50 & 0.55 \end{bmatrix}$$

iii. Compute the Within and Between Class Scatter Matrices

$$S_B = (\mu^{(1)} - \mu^{(2)})^T(\mu^{(1)} - \mu^{(2)}) = \begin{bmatrix} 4.08 & -1.50 \\ -1.50 & 0.55 \end{bmatrix}$$

$$S_w = (\sigma^{(1)^2} + \sigma^{(2)^2}) = \begin{bmatrix} 4.92 & -2.18 \\ -2.18 & 8.45 \end{bmatrix}$$

$$S_w^{-1} = \frac{1}{(4.92 * 8.45) - (-2.18 - (-2.18))} \begin{bmatrix} 8.45 & 2.18 \\ 2.18 & 4.92 \end{bmatrix} = \begin{bmatrix} 0.23 & 0.06 \\ 0.06 & 0.13 \end{bmatrix}$$

iv. Perform the Eigendecomposition on $S_W^{-1} S_B$

$$\text{Eigenvalues} = \begin{bmatrix} 8.36\text{e}{-01} & 5.20\text{e}{-18} \end{bmatrix}$$

$$\text{Eigenvectors} = \begin{bmatrix} 0.99 & 0.344 \\ 0.05 & 0.94 \end{bmatrix}$$

v. Determine Largest Eigenvalue and its Associated Eigenvector

$$\text{Largest Eigenvalue} = \begin{bmatrix} 8.36\text{e}{-01} \end{bmatrix}$$

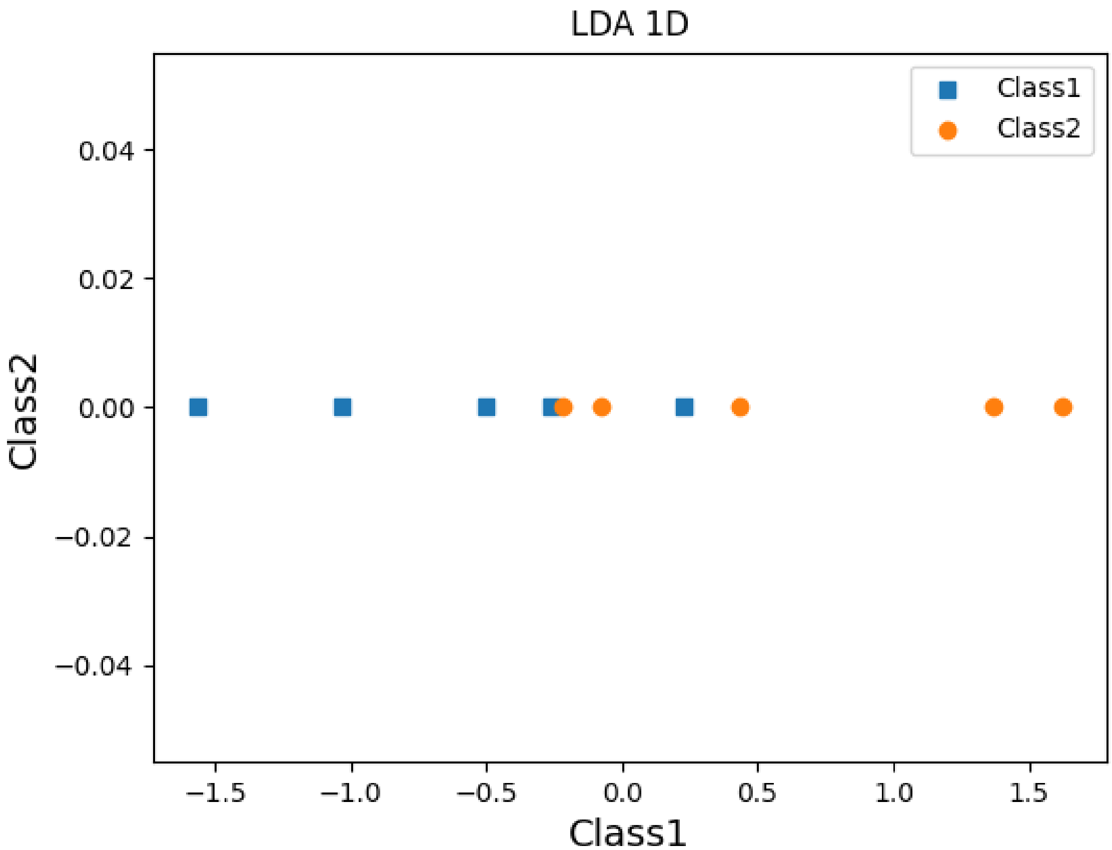
$$\text{Largest Eigenvectors} = \begin{bmatrix} 0.99 \\ 0.05 \end{bmatrix}$$

vi. Project $X_{zscored}$ into Eigenvector Space

$$Z = XW = \begin{bmatrix} 0.26 & -0.09 \\ -0.97 & -1.26 \\ -0.50 & -0.09 \\ 0.21 & 0.37 \\ -1.68 & 2.25 \\ -0.26 & 0.84 \\ 0.45 & -0.33 \\ 1.40 & -0.56 \\ -0.02 & -1.03 \\ 1.63 & -0.09 \end{bmatrix} \begin{bmatrix} 0.99 \\ 0.05 \end{bmatrix}$$

$$= \begin{bmatrix} -0.26 \\ -1.03 \\ -0.50 \\ 0.23 \\ -1.57 \\ -0.22 \\ 0.43 \\ 1.37 \\ -0.74 \\ 1.63 \end{bmatrix}$$

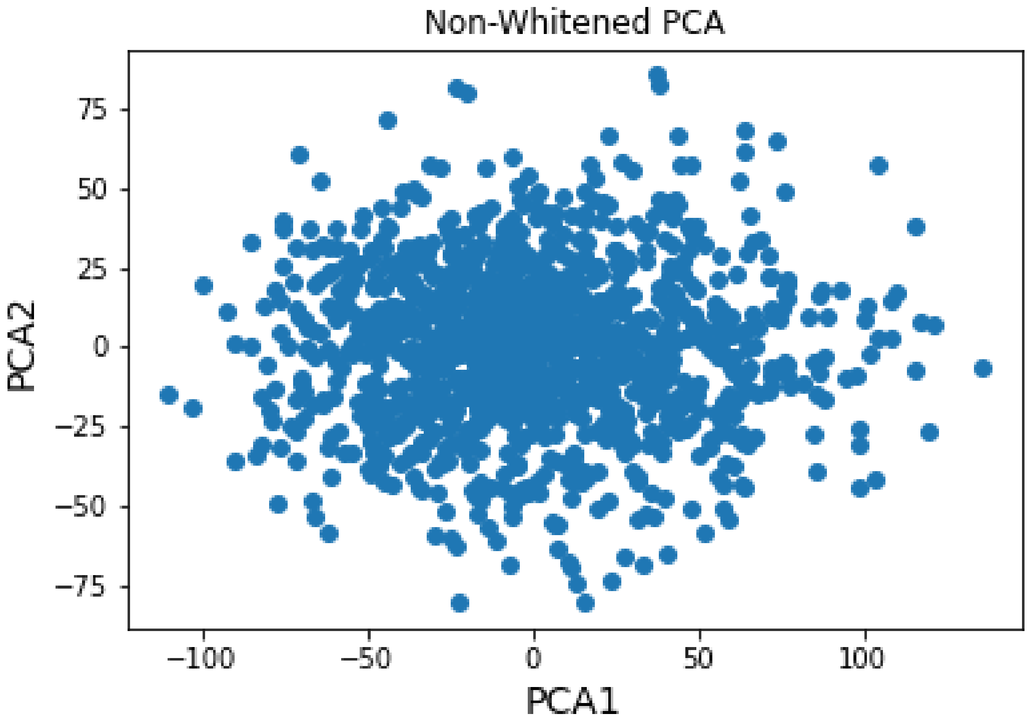
(b) Project your data down to 1D using the the direction of projection found from the LDA process. Plot this data in 1D, visualizing class one data as squares and class two data as circles (3pts).



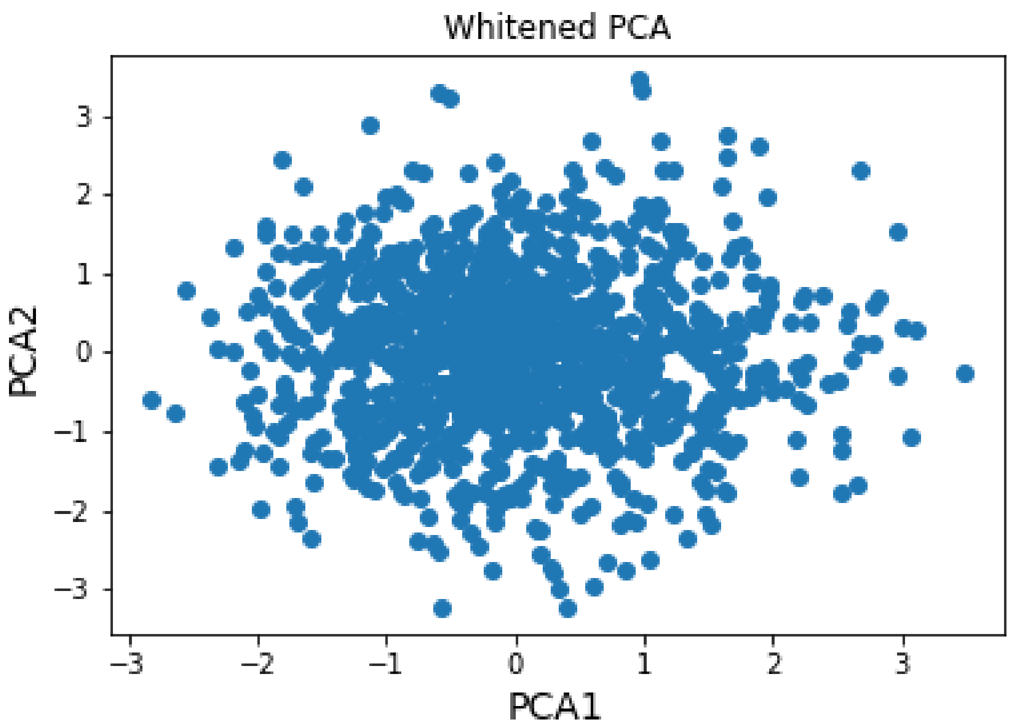
(c) Based on your plot, does the LDA projection provide good class separation? Why or why not (2pts)?
Given the 1-dimensional LDA plot of class 1 (squares) and class 2 (circles) it is extremely easy to see how the classes are not fully separated. Therefore, this LDA projection does NOT provide a good class separation.

2 Dimensionality Reduction for Visualization

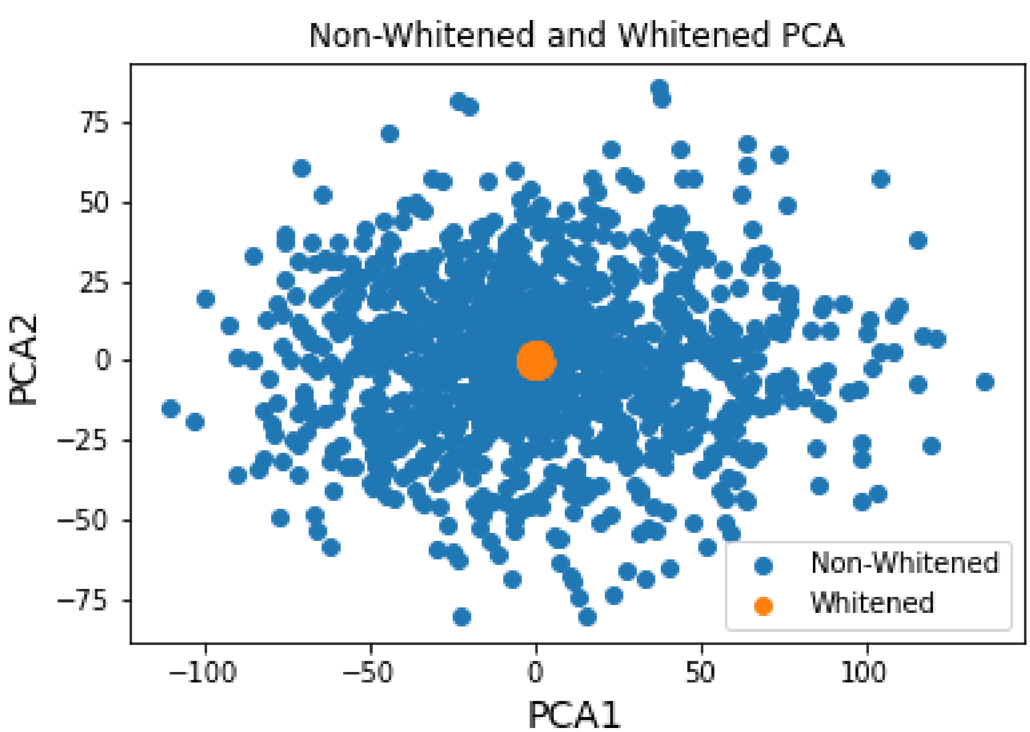
1. Non-Whitened 2D PCA Projected Data



2. Whitened 2D PCA Projected Data



3. Non-Whitened and Whitened Overlay 2D PCA Projected Data



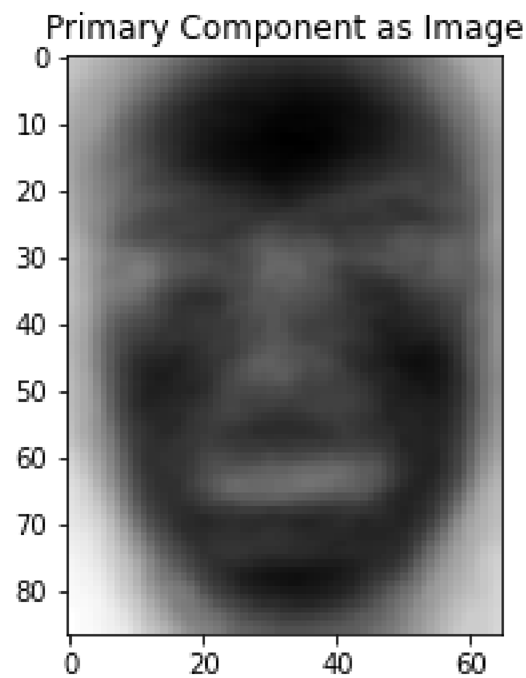
3 Dimensionality Reduction for KNNs

1. Validation Accuracy

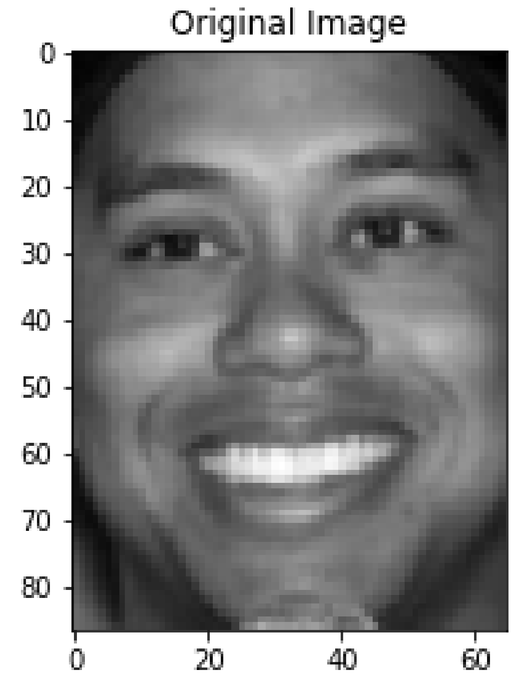
   - $k = 1$ $D = original$: 0.19128329297820823
   - $k = 1$ $D = 100$ PCA: 0.1791767554479419
   - $k = 1$ $D = 100$ PCA Whitened: 0.25181598062953997

4 Eigenfaces as Compression

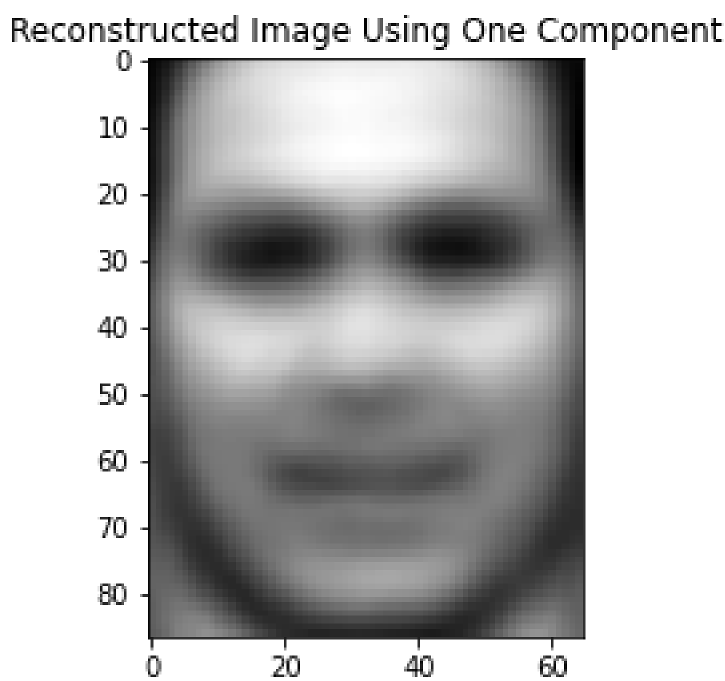
1. Primary Component Visualized as an Image



2. Image Used for Compression / Reconstruction



3. Reconstructed Image Using One Component



4. Minimum Number of Components Necessary to Perform 95% Reconstruction: 181

5. Reconstructed Image Using the Minimum Number of Components (181)