CS 613 - Machine Learning

Assignment 3 - Dimensionality Reduction & Clustering

Introduction

In this assignment you'll work on visualizing data, reducing its dimensionality and clustering it. You may not use any functions from machine learning library in your code, however you may use statistical functions. For example, if available you **MAY NOT** use functions like

- pca
- k-nearest neighbors functions

Unless explicitly told to do so. But you **MAY** use basic statistical functions like:

- std
- mean
- cov
- eig

Grading

Part 1 (Theory)	23pts
Part 2 (PCA)	25pts
Part 3 (Eigenfaces)	20pts
Part 3 (Clustering)	25pts
Report	7pts
TOTAL	100pts

Table 1: Grading Rubric

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) Co-variance Matrix:

$$\begin{bmatrix} -0.1 & -1.367 \\ -1.367 & -0.1 \end{bmatrix}$$

(b) For Eigen Values:

$$Det \begin{bmatrix} -0.1-\lambda & 1.367 \\ -1.367 & -0.1-\lambda \end{bmatrix} = 0$$

$$\text{(-0.1 - }\lambda)(-0.1-\lambda) - (-1.367)(-1.367) = 0$$
$$\text{(-0.1 - }\lambda)(-0.1-\lambda) - 1869 = 0$$
$$\text{(-0.1 - }\lambda)(-0.1-\lambda) = 1869$$
$$\text{(-0.1 - }\lambda) = \sqrt{1869} \text{ or } (-0.1-\lambda) = -\sqrt{1869}$$
$$\lambda = \sqrt{1869} - 0.1 \text{ or } \lambda = -\sqrt{1869} - 0.1$$
$$\lambda = 1.267 \text{ or } \lambda = -1.467$$

(c) For Eigen Vector:

$$\begin{bmatrix} -0.1-\lambda & 1.367 \\ -1.367 & -0.1-\lambda \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\text{(-0.1 - }\lambda)x_1 - (1.367)x_2 = 0$$
$$\text{-1.367x}_1 - (0.1-\lambda)x_2 = 0$$
$$\text{(-0.1 - 1.267)x}_1 - 1.367x_2 = 0$$
$$\text{-1.367x}_1 - (0.1-1.267)x_2 = 0$$
$$\text{x}_1 = 0.707, x_2 = -0.707$$

$$(-0.1+1.467)x_1 - 1.367x_2 = 0$$

$$\text{-}1.367\mathrm{x}_1 - (0.1 + 1.467)x_2 = 0$$
$$\mathrm{x}_1 = 0.707, x_2 = 0.707$$

(d) For Projection with the eigen vectors with highest eigen value:

$$Z = X_(standardized)\ @\ v[0] \text{ (Eigen vector associated with Eigen value = 1.267)}$$
$$Z =$$
$$[[-0.118], [0.208], [-0.285], [-0.114], [-2.776], [-0.780], [0.549], [1.384], [0.711], [1.220]]$$

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) $E(H(A) = \sum i = 1 \text{ to } k\ (p_i + n_i)/(p + n)\ H((p_i/(p_i + n_i), (n_i/(p_i + n_i)))$
Postitive Samples(1) = 4, Negative samples(1) = 6
$p_1 = 1$, $n_1 = 1$
$p_2 = 0$, $n_2 = 2$
$p_3 = 0$, $n_3 = 0$
$p_4 = 0$, $n_4 = 0$
$p_5 = 1$, $n_5 = 1$
$p_6 = 1$, $n_6 = 0$
$p_8 = 0$, $n_8 = 0$
$p_(11) = 0$, $n_(11) = 0$
$E(H(1) =$
$2/10(-1/2 * \log_2(1/2) + -1/2 * \log_2(1/2)$
$+ 2/10(-0/2 * \log_2(0/2) + -2/2 * \log_2(2/2)$
$+ 2/10(-1/2 * \log_2(1/2) + -1/2 * \log_2(1/2)$
$+ 1/10(-1/1 * \log_2(1/1) + -0/2 * \log_2(0/1)$
$E(H(1) = (2/10) + (2/10) = 0.4$
$IG(1) = (-4/10 * \log_2(4/10) + -6/10 * \log_2(6/10) - E(H(1) = 0.571$
Postitive Samples(2) = 7, Negative samples(2) = 3
$p_1 = 3$, $n_1 = 1$
$p_2 = 0$, $n_2 = 0$
$p_3 = 1$, $n_3 = 1$
$p_4 = 0$, $n_4 = 1$
$p_5 = 1$, $n_5 = 0$
$p_6 = 0$, $n_6 = 0$
$p_8 = 0$, $n_8 = 0$
$p_(11) = 1$, $n_(11) = 0$
$E(H(2) =$
$4/10(-3/4 * \log_2(3/4) + -1/4 * \log_2(1/4)$
$+ 2/10(-1/2 * \log_2(1/2) + -1/2 * \log_2(1/2)$

$+ 1/10(-0/1 * \log_2(0/1) + -1/1 * \log_2(1/1)$
$+ 1/10(-1/1 * \log_2(1/1) + -0/2 * \log_2(0/1)$
$+ 1/10(-1/1 * \log_2(1/1) + -0/2 * \log_2(0/1)$
$E(H(2) = 0.324 + 0.2 = 0.524$
$IG(2) = (-7/10 * \log_2(7/10) + -3/10 * \log_2(3/10) - E(H(2) = 0.310$

(b) Feature 1 has higher Information Gain according to part a

(c) For LDA:

$(S_W)^-1 \ @ \ S_b \ @ \ W = \lambda @ W$
$(S_W) = (\sigma_1)^2 + (\sigma_2)^2$
$(S_W) =$

$$\begin{bmatrix} 16 & -7.16 \\ -7.16 & 16 \end{bmatrix} + \begin{bmatrix} 16 & -3.63 \\ -3.63 & 16 \end{bmatrix} = \begin{bmatrix} 32 & -10.79 \\ -10.79 & 32 \end{bmatrix}$$

$(S_W)^1 =$

$$\begin{bmatrix} 0.035 & 0.012 \\ 0.012 & 0.035 \end{bmatrix}$$

$S_B = (\mu_! - \mu_2)^T (\mu_! - \mu_2)$
$S_B =$

$$\begin{bmatrix} 0.0 & 0.0 \\ 0.0 & 3.08e-35 \end{bmatrix}$$

$Eigenvalues = [0, 1.087e-36]$

W (Eigen vector with non-zero eigen value)=

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

(d) For Projection with the eigen vectors with highest eigen value:

$$Z1 = X1_{(standardized)} \ @ \ W[1] \text{ (Eigen vector associated with Eigen value} = 1.087e-36)$$
$$Z1 = [[-0.256], [-1.172], [-0.256], [0.109], [1.575]]$$

$$Z2 = X2_{(standardized)} \ @ \ W[1] \text{ (Eigen vector associated with Eigen value} = 1.087e-36)$$
$$Z2 = [[1.550], [-0.135], [-0.471], [-1.146], [0.202]]$$

(e) The projection in the previous class provides a better seperation as it keeps the negatives and positive values seperated.

2 Dimensionality Reduction via PCA

Once you have your setup complete, write a script to do the following:

1. KNN accuracy: 0.232
2. 100D, K = 1, accuracy: 0.256
3. 100D Whitened data, K = 1, accuracy: 0.351
4. Please find a 2D graph named A3P2, if it does not appear below

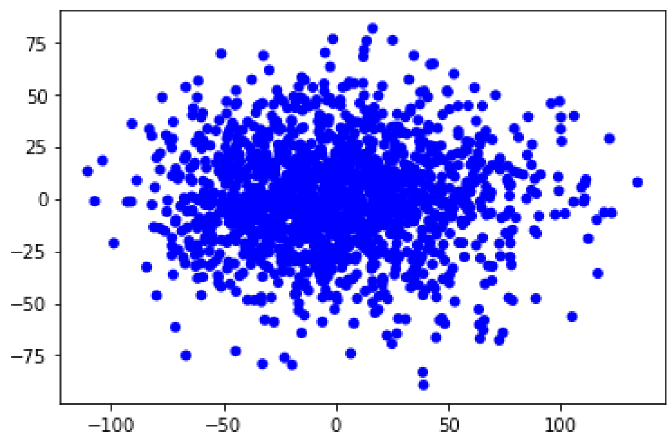

Figure 1: 2D PCA Projection of data

3 Eigenfaces

Write a script that:

1. Max image on PC1's Axis:

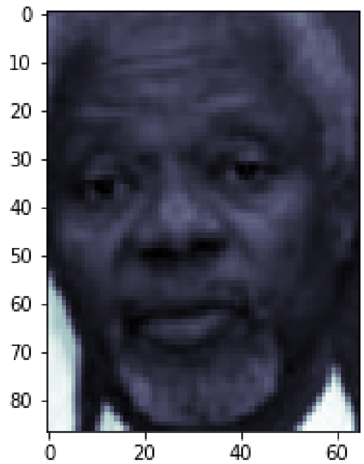

Figure 2: Max image on PC1's Axis

2. Min image on PC1's Axis:

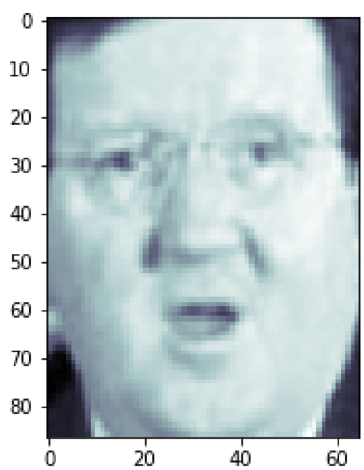

Figure 3: Min image on PC1's Axis

3. Max image on PC2's Axis:

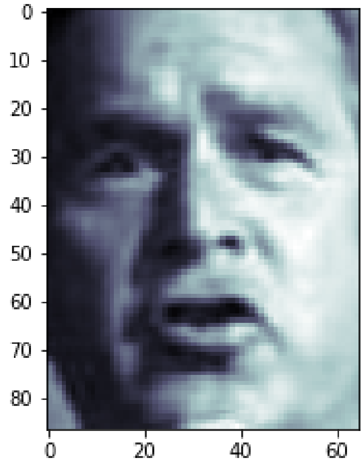

Figure 4: Max image on PC2's Axis

4. Min image on PC2's Axis:

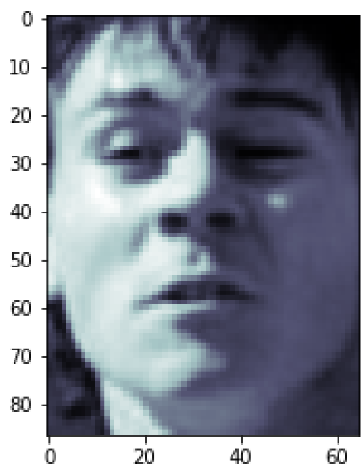

Figure 5: Min image on PC2's Axis

5. Visualizes the most important principle component as a 87x65 image:

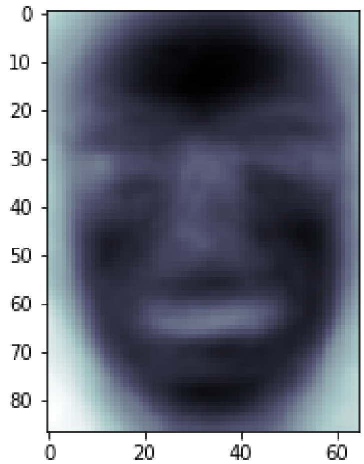

Figure 6: PC1

6. Reconstructs the *X_train[0,:]* image using the primary principle component.

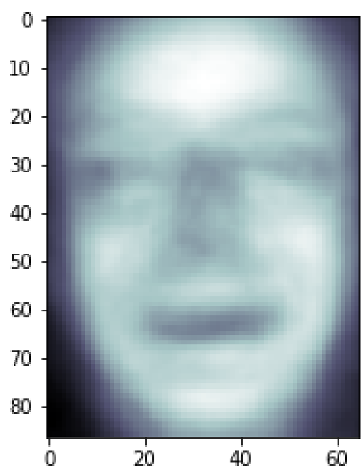

Figure 7: Reconstructed Image

7. the number of principle components necessary to encode at least 95% of the information, 1470.

8. Reconstructs the *X_train[0,:]* image using the k most significant eigen-vectors (found in the previous step, see Figure **??**).

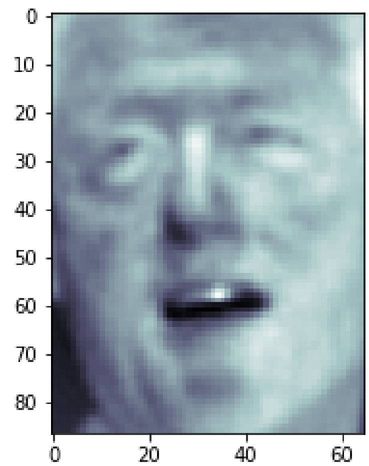

Figure 8: Reconstructed 95percent with eigen vectors

9. Please find the images in the folder as A3P341, A3P342, A3P343, A3P344, A3P36, A3P35, A3P38 if it do no appear above

4 Clustering

1. Number of Images in each Cluster:

Cluster 1 = 179
Cluster 2 = 211
Cluster 3 = 143
Cluster 4 = 92
Cluster 5 = 224
Cluster 6 = 75
Cluster 7 = 739
Cluster 8 = 10
Cluster 9 = 281
Cluster 10 = 109

2. Reconstruct the cluster centers for each of the K clusters:

Figure 9: Reconstructed Cluster Centers

3. Images closest to the cluster centers:

Figure 10: Images closest to the cluster centers

4. Images furthest from the cluster centers:

Figure 11: Images furthest from the cluster centers

5. Please find the images in the folder as A3P46, A3P47c, A3P47f if it do not appear above.

Submission

For your submission, upload to Blackboard a single zip file containing:

1. A LaTeX typeset PDF containing:
   (a) Part 1: Your answers to the theory questions.
   (b) Part 2: The visualization of the PCA result, KNN accuracies
   (c) Part 3:
      i. Visualization of primary principle component
      ii. Number of principle components needed to represent 95% of information, k.
      iii. Visualization of the reconstruction of the first person using
         A. Original image
         B. Single principle component
         C. k principle components.
   (d) Part 4: The visualization of k-means cluster centers, and the min and max images (a total of 30 images, 10 cluster centers, 2 extrema per cluster)
   (e) Source Code - python notebook