CS 383 - Machine Learning

Assignment 1 - Dimensionality Reduction
Summer 2017
Amir Omidi

1 Theory Questions

1.1 Question 1

1.1.1 Part A

Our starting data for this section is the following:

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

We need to standardize our data before we get started with PCA. To do this we need the average and standard deviation of each column.

Lets first calculate the average of each column:

- Column 1 average:

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

- Column 2 average:

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

Now we should calculate the standard deviation of each column using the average we calculated above.

- Column 1 standard deviation:

  - Find variance squared of each row:

    1. $(-2+0.9)^2 = 1.21$
    2. $(-5+0.9)^2 = 16.81$
    3. $(-3+0.9)^2 = 4.41$
    4. ...

- Add them all up:

$$1.21 + 16.81 + 4.41 + 0.81 + 50.41 + 1.21 + 3.61 + 34.81 + 0.01 + 47.61 = 160.90$$

- Since we're calculating a sample standard deviation, we divide the sum by 9 ($samplesize - 1$) and then take a square root of it.

$$\sqrt{\frac{160.90}{9}} = 4.2282$$

- Column 2 standard deviation:

  - Find variance squared of each row:

    1. $(1 - 1.4)^2 = 0.16$
    2. $(-4 - 1.4)^2 = 29.16$
    3. $(1 - 1.4)^2 = 0.16$
    4. ...

  - Add them all up:

$$0.16 + 29.16 + 0.16 + 2.56 + 92.16 + 12.96 + 1.96 + 5.76 + 19.36 + 0.16 = 164.40$$

  - Since we're calculating a sample standard deviation, we divide the sum by 9 ($samplesize - 1$) and then take a square root of it.

$$\sqrt{\frac{160.40}{9}} = 4.27395$$

Now we can start standardizing our matrix.

- Standardization 1 - We can now begin standardizing our data. The first step is to subtract the mean of each column from the elements of each column:

$$\begin{bmatrix} -1.1 & -0.4 \\ -4.1 & -5.4 \\ -2.1 & -0.4 \\ 0.9 & 1.6 \\ -7.1 & 9.6 \\ -1.1 & 3.6 \\ 1.9 & -1.4 \\ 5.9 & -2.4 \\ -0.1 & -4.4 \\ 6.9 & -0.4 \end{bmatrix}$$

- Standardization 2 - The second step is to divide each element of each column by the standard deviation of that column:

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

The next step is to calculate the covariance of the matrix.

We will be using the following formula: $X'X/(N-1)$

$$\begin{bmatrix} -0.2602 & -0.9697 & -0.4967 & 0.2129 & -1.6792 & -0.2602 & 0.4494 & 1.3954 & -0.0237 & 1.6319 \\ -0.0936 & -1.2635 & -0.0936 & 0.3744 & 2.2462 & 0.8423 & -0.3276 & -0.5615 & -1.0295 & -0.0936 \end{bmatrix} \times$$
$$\begin{bmatrix} -0.2602 & -0.0936 \\ -0.9697 & -1.2635 \\ -0.4967 & -0.0936 \\ 0.2129 & 0.3744 \\ -1.6792 & 2.2462 \\ -0.2602 & 0.8423 \\ 0.4494 & -0.3276 \\ 1.3954 & -0.5615 \\ -0.0237 & -1.0295 \\ 1.6319 & -0.0936 \end{bmatrix} \times \frac{1}{9} = \begin{bmatrix} 1.0000 & -0.4083 \\ -0.4083 & 1.0000 \end{bmatrix}$$

Now that we have the covariance matrix, finding the eigenvalues is an easy task.

$$\begin{bmatrix} 1-\lambda & -0.4083 \\ -0.4083 & 1-\lambda \end{bmatrix} \Rightarrow \lambda^2 - 2\lambda + (1-(-0.4083)^2) = \lambda^2 + 2\lambda + 0.8333$$
$$\lambda = 0.5917, 1.4083$$

Now that we have the eigenvalues, we need to find the eigenvectors.

$$\begin{bmatrix} 1.0000 & -0.4083 \\ -0.4083 & 1.0000 \end{bmatrix} - \begin{bmatrix} 0.5917 & 0 \\ 0 & 0.5917 \end{bmatrix} = \begin{bmatrix} 0.4083 & -0.4083 \\ -0.4083 & 0.4083 \end{bmatrix}$$
$$\begin{bmatrix} 0.4083 & -0.4083 \\ -0.4083 & 0.4083 \end{bmatrix} \times \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$
$$\begin{matrix} 0.4083a - 0.4083b = 0 \\ -0.4083a + 0.4083b = 0 \end{matrix} \Rightarrow b = t$$

Eigenvector is any scalar multiple of $\begin{bmatrix} 1 & 1 \end{bmatrix}^T$

And for the other eigenvalue, doing the above steps we see that it is any scalar multiple of: $\begin{bmatrix} -1 & 1 \end{bmatrix}^T$

1.1.2 Part B

The largest eigenvalue is clearly 1.4083. The corresponding eigenvector for that eigenvalue is mentioned above. By multiplying the data with that eigenvector we would successfully project the data onto the PCA of the largest eigenvalue and reduced the conventionality of the data from $D = 2$ to $D = 1$.

$$\begin{bmatrix} -0.2602 & -0.0936 \\ -0.9697 & -1.2635 \\ -0.4967 & -0.0936 \\ 0.2129 & 0.3744 \\ -1.6792 & 2.2462 \\ -0.2602 & 0.8423 \\ 0.4494 & -0.3276 \\ 1.3954 & -0.5615 \\ -0.0237 & -1.0295 \\ 1.6319 & -0.0936 \end{bmatrix} \times \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.1667 \\ -0.2938 \\ 0.4031 \\ 0.1615 \\ 3.9254 \\ 1.1025 \\ -0.7769 \\ -1.9569 \\ -1.0058 \\ -1.7255 \end{bmatrix}$$

1.2 Question 2

1.2.1 Part A

We have two classes of data:

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

These are the labels that we have:

$$Labels \in \{-8, -5, -4, -3, -2, -1, 0, 1, 3, 5, 6, 11\}$$

Let's start with the process of information gain.
Feature one

$$\begin{aligned}
p_{-8} &= 1, & n_{-8} &= 0 \\
p_{-5} &= 1, & n_{-5} &= 0 \\
p_{-4} &= 0, & n_{-4} &= 0 \\
p_{-3} &= 1, & n_{-3} &= 0 \\
p_{-2} &= 1, & n_{-2} &= 1 \\
p_{-1} &= 0, & n_{-1} &= 1 \\
p_{0} &= 1, & n_{0} &= 0 \\
p_{1} &= 0, & n_{1} &= 1 \\
p_{3} &= 0, & n_{3} &= 0 \\
p_{5} &= 0, & n_{5} &= 1 \\
p_{6} &= 0, & n_{6} &= 1 \\
p_{11} &= 0, & n_{11} &= 0
\end{aligned}$$

Now we plug them into the remainder formula:

$$
\begin{aligned}
&Reamainder(1) = \\
&\qquad \frac{1}{10} \times [(\frac{-1}{1}\log_2 \frac{1}{1}) + 0] && \text{This is 0} \\
&\qquad +0 && \text{This is for label -5 which is also 0.} \\
&\qquad +0 && \text{Same as above.} \\
&\qquad +0 \\
&\qquad +\frac{2}{10} \times [(\frac{-1}{2}\log_2 \frac{1}{2}) + (\frac{-1}{2}\log_2 \frac{1}{2})] && \text{This is } \frac{1}{5} \\
&\qquad +0 \\
&\qquad +0 \\
&\qquad +0 \\
&\qquad + \ldots \\
&\qquad = \frac{1}{5}
\end{aligned}
$$

$$Remainder(1) = \frac{1}{5} = 0.2000$$

Feature two

$$
\begin{aligned}
p_{-8} &= 0, & n_{-8} &= 0 \\
p_{-5} &= 0, & n_{-5} &= 0 \\
p_{-4} &= 1, & n_{-4} &= 0 \\
p_{-3} &= 0, & n_{-3} &= 1 \\
p_{-2} &= 0, & n_{-2} &= 0 \\
p_{-1} &= 0, & n_{-1} &= 1 \\
p_{0} &= 0, & n_{0} &= 1 \\
p_{1} &= 2, & n_{1} &= 1 \\
p_{3} &= 1, & n_{3} &= 0 \\
p_{5} &= 0, & n_{5} &= 1 \\
p_{6} &= 0, & n_{6} &= 0 \\
p_{11} &= 1, & n_{11} &= 0
\end{aligned}
$$

Now we plug them into the remainder formula:

$$Reamainder(2) =$$

$$\frac{3}{10} \times [(\frac{-2}{3}\log_2\frac{2}{3}) + (\frac{-1}{3}\log_2\frac{1}{3})] \qquad = 0.2755$$

$$Remainder(2) = 0.275$$

The information gain for each feature

$$IG(1) = [2 \times (\frac{-5}{10}\log_2\frac{5}{10})] - 0.2000 = 0.8000 \tag{1}$$

$$IG(2) = [2 \times (\frac{-5}{10}\log_2\frac{5}{10})] - 0.2755 = 0.7245 \tag{2}$$

1.2.2 Part B

As you can clearly see from equations 1 and 2, the more discriminating feature is feature 1.

1.2.3 Part C

Let's start by standardizing our entire data set. The standardized matrix was already found in Question 1 part a.

$$standardize(\begin{bmatrix} -2 & 1 \\ -5 & -4 \\ -3 & 1 \\ 0 & 3 \\ -8 & 11 \\ -2 & 5 \\ 1 & 0 \\ 5 & -1 \\ -1 & -3 \\ 6 & 1 \end{bmatrix}) = \begin{bmatrix} -0.2602 & -0.0936 \\ -0.9697 & -1.2635 \\ -0.4967 & -0.0936 \\ 0.2129 & 0.3744 \\ -1.6792 & 2.2462 \\ -0.2602 & 0.8423 \\ 0.4494 & -0.3276 \\ 1.3954 & -0.5615 \\ -0.0237 & -1.0295 \\ 1.6319 & -0.0936 \end{bmatrix}$$

We should get the mean of each class now:

$$\mu_1 = mean\left(\begin{bmatrix} -0.2602 & -0.0936 \\ -0.9697 & -1.2635 \\ -0.4967 & -0.0936 \\ 0.2129 & 0.3744 \\ -1.6792 & 2.2462 \end{bmatrix}\right) = \begin{bmatrix} -0.6386 & 0.2340 \end{bmatrix} \tag{3}$$

$$\mu_2 = mean\left(\begin{bmatrix} -0.2602 & 0.8423 \\ 0.4494 & -0.3276 \\ 1.3954 & -0.5615 \\ -0.0237 & -1.0295 \\ 1.6319 & -0.0936 \end{bmatrix}\right) = \begin{bmatrix} 0.6386 & -0.2300 \end{bmatrix} \tag{4}$$

We can use this data to calculate S_B:

$$\mu_1 - \mu_2 = \begin{bmatrix} -1.2771 & 0.4680 \end{bmatrix}$$

$$S_B = \begin{bmatrix} -1.2771 \\ 0.4680 \end{bmatrix} \times \begin{bmatrix} -1.2771 & 0.4680 \end{bmatrix} = \begin{bmatrix} 1.6311 & -0.5976 \\ -0.5976 & 0.2190 \end{bmatrix} \tag{5}$$

Now, we need to calculate the scatter matrices for each class. Before we do that, we need to calculate the covariance matrix for each class. In order to do this, we need to standardize each class to be able to find the covariance easier.

$$\begin{bmatrix} -2 & 1 \\ -5 & -4 \\ -3 & 1 \\ 0 & 3 \\ -8 & 11 \end{bmatrix} - \left(\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix} \times \begin{bmatrix} -2 & 1 \\ -5 & -4 \\ -3 & 1 \\ 0 & 3 \\ -8 & 11 \end{bmatrix} * \frac{1}{5} \right) = \begin{bmatrix} 2.9327 & -1.2635 \\ 2.2232 & -2.4333 \\ 2.6962 & -1.2635 \\ 3.4057 & -0.7955 \\ 1.5136 & 1.0763 \end{bmatrix}$$

$$cov(1) = \begin{bmatrix} 2.9327 & 2.2232 & 2.6962 & 3.4057 & 1.5136 \\ -1.2635 & -2.4333 & -1.2635 & -0.7955 & 1.0763 \end{bmatrix} \begin{bmatrix} 2.9327 & -1.2635 \\ 2.2232 & -2.4333 \\ 2.6962 & -1.2635 \\ 3.4057 & -0.7955 \\ 1.5136 & 1.0763 \end{bmatrix} \times \frac{1}{4} =$$

$$\begin{bmatrix} 0.5202 & -0.4123 \\ -0.4123 & 1.6314 \end{bmatrix}$$

Following these same steps for the second class we finally get:

$$cov(1) = \begin{bmatrix} 0.5202 & -0.4123 \\ -0.4123 & 1.6314 \end{bmatrix}$$

$$cov(2) = \begin{bmatrix} 0.7104 & -0.1328 \\ -0.1328 & 0.4818 \end{bmatrix}$$

Now we use these variance-covariance matrices to find the scatter matrix:

$$\sigma_1^2 = \begin{bmatrix} 2.0808 & -1.6490 \\ -1.6490 & 6.5255 \end{bmatrix} \tag{6}$$

$$\sigma_2^2 = \begin{bmatrix} 2.8415 & -0.5312 \\ -0.5312 & 1.9270 \end{bmatrix} \tag{7}$$

Now that we have the scatter matrices 6 and 7 we must calculate the within class scatter matrix.

$$S_w = 6 + 7 = \begin{bmatrix} 4.9223 & -2.1803 \\ -2.1803 & 8.4526 \end{bmatrix} \tag{8}$$

Now we need to find the inverse of 8.

$$det(S_w) = 36.8525$$

$$S_w^{-1} = \frac{1}{36.8525} \times \begin{bmatrix} 8.4526 & 2.1803 \\ 2.1803 & 4.9223 \end{bmatrix} = \begin{bmatrix} 0.2294 & 0.0592 \\ 0.0592 & 0.1336 \end{bmatrix}$$

The corresponding non-zero eigenvector for $S_w^{-}1S_B$ is:

$$W = \begin{bmatrix} 0.9998 & 0.0492 \end{bmatrix}^T \tag{9}$$

This vector is already a unit vector since:

$$\parallel W \parallel = 1$$

1.2.4 Part D

Lets project our data using this principal component.

$$\begin{bmatrix} -0.2602 & -0.0936 \\ -0.9697 & -1.2635 \\ -0.4967 & -0.0936 \\ 0.2129 & 0.3744 \\ -1.6792 & 2.2462 \\ -0.2602 & 0.8423 \\ 0.4494 & -0.3276 \\ 1.3954 & -0.5615 \\ -0.0237 & -1.0295 \\ 1.6319 & -0.0936 \end{bmatrix} \times \begin{bmatrix} 0.9998 \\ 0.0492 \end{bmatrix} = \begin{bmatrix} -0.2644 \\ -1.0306 \\ -0.5007 \\ 0.2310 \\ -1.5667 \\ -0.2184 \\ 0.4327 \\ 1.3661 \\ -0.0742 \\ 1.6253 \end{bmatrix}$$

1.2.5 Part E

This separation isn't good. If you look at the figure below, you can see that some features from the first class are getting mixed with the second class.

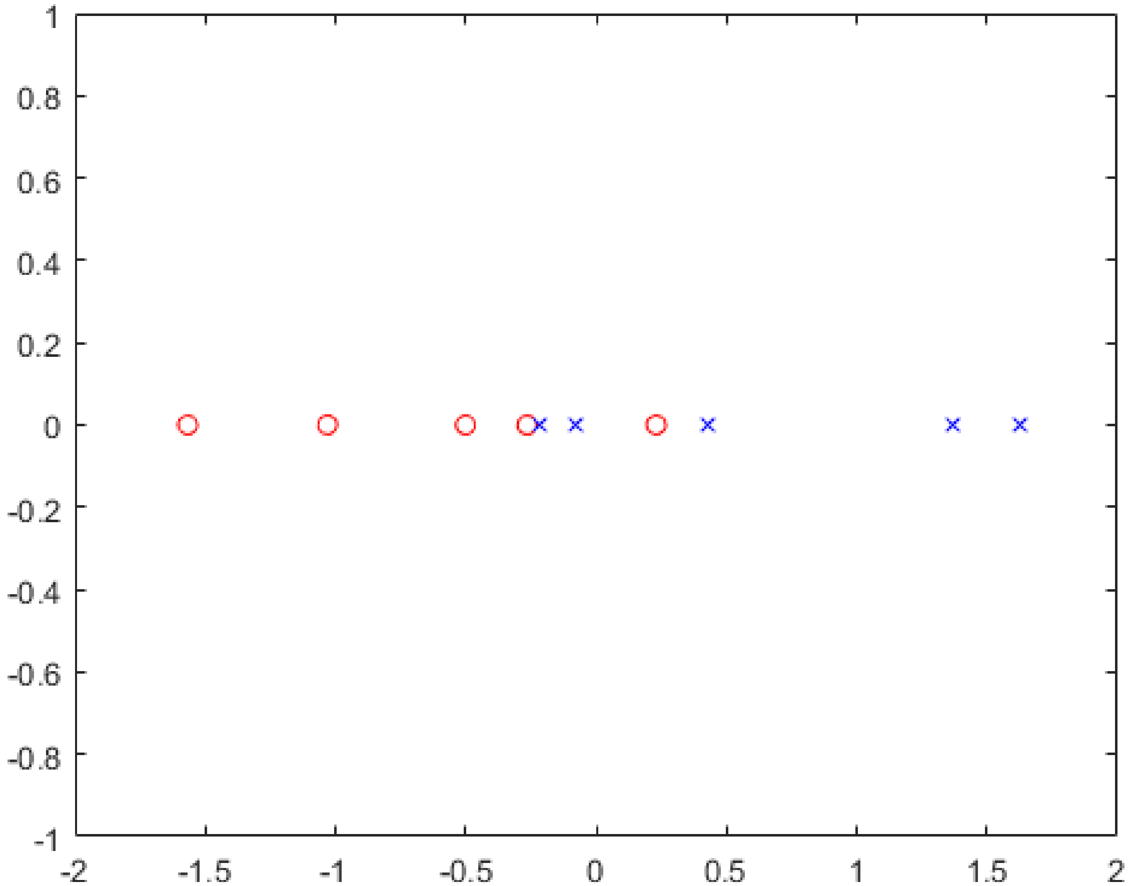

Figure 1: Projection of data onto the principal component

2 Programming Questions

2.1 Dimensionality Reduction via PCA

Visualization of PCA:

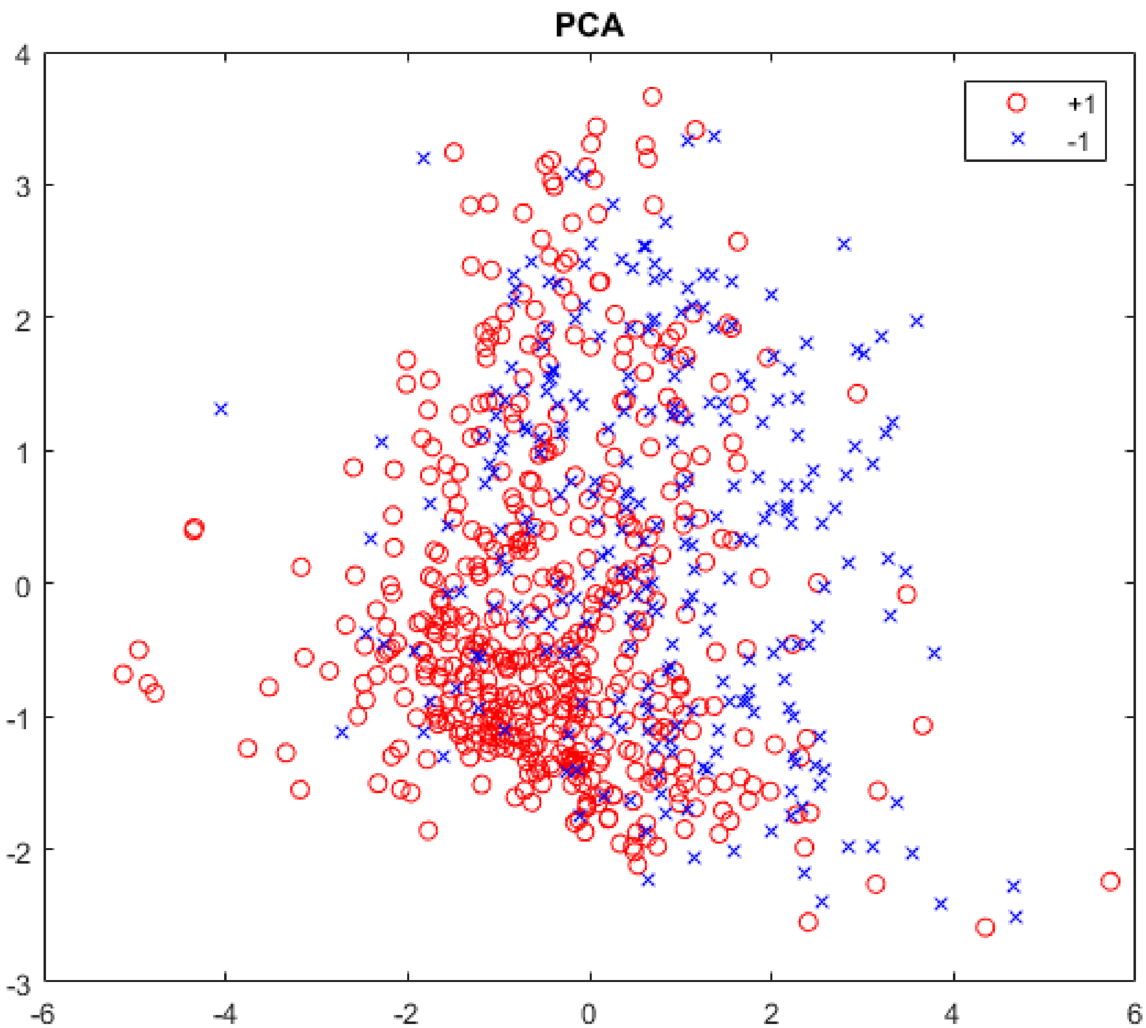

Figure 2: Visualization of PCA

2.2 Eigenfaces

2.2.1 Value of k

The value of k was found to be 37.

2.2.2 Visualization of primary principle component

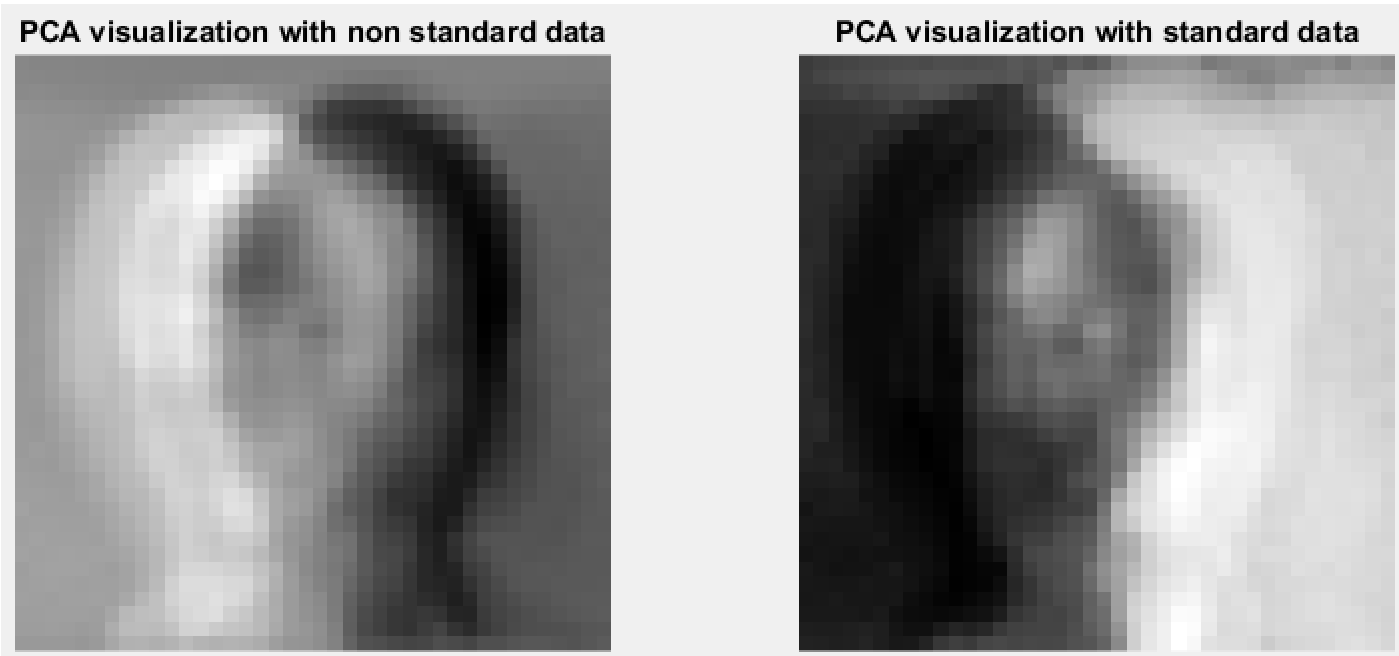

Figure 3: PCA visualization

2.2.3 Visualization of the reconstruction of the first person

The top row is the visualization with non standardized data while the bottom row is the visualization with standardized data.

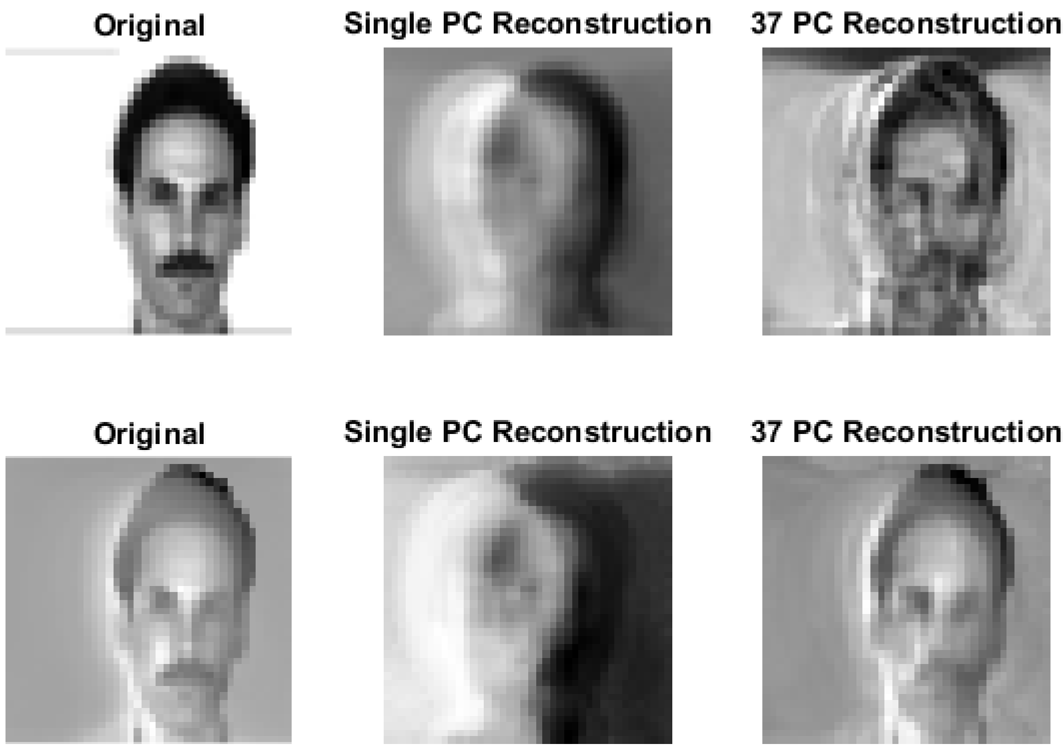

Figure 4: Visualization of the reconstruction of the first person