ECE-GY 6143 INTRO TO MACHINE LEARNING

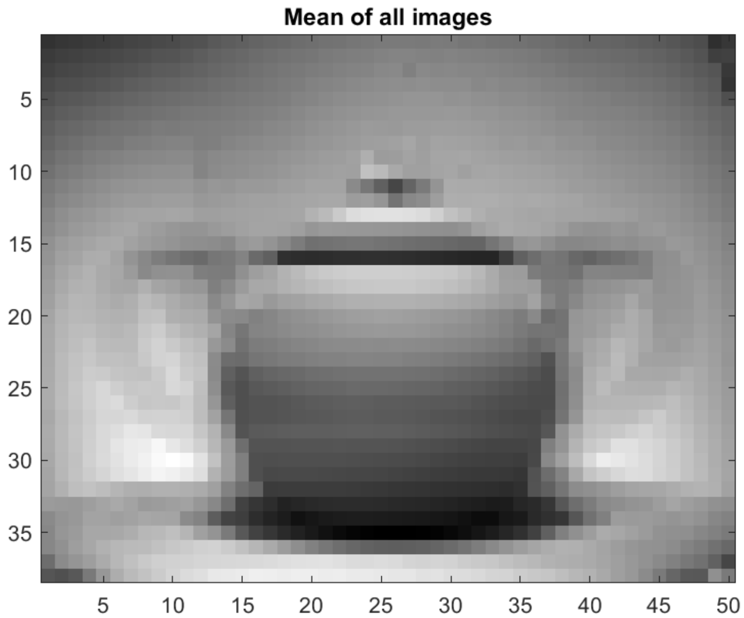
HOMEWORK-4

Pratyush Shukla (ps4534)

Question 1:

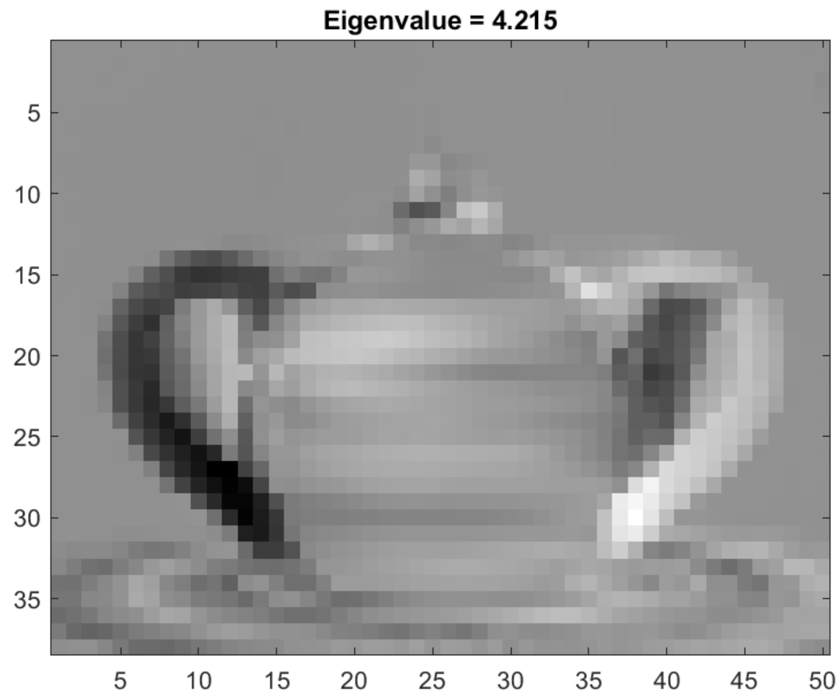
The data is given with shape (100, 1900) where the no. of images is 100, with shape (38, 50) flattened to 1900. We take the mean of the data and subtract from every data value to get the decentered data.

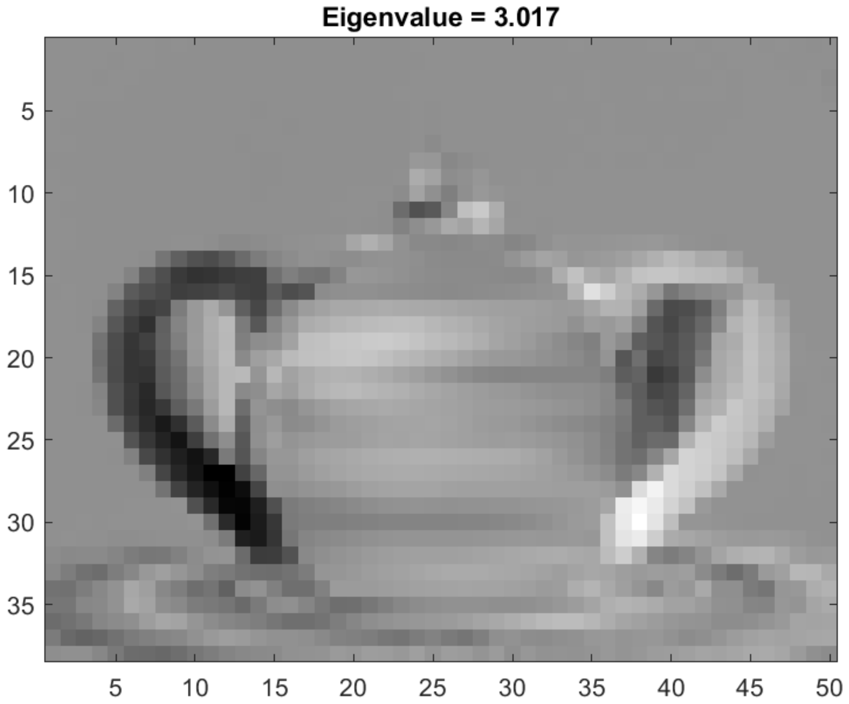
The mean of the data is shown as image below -

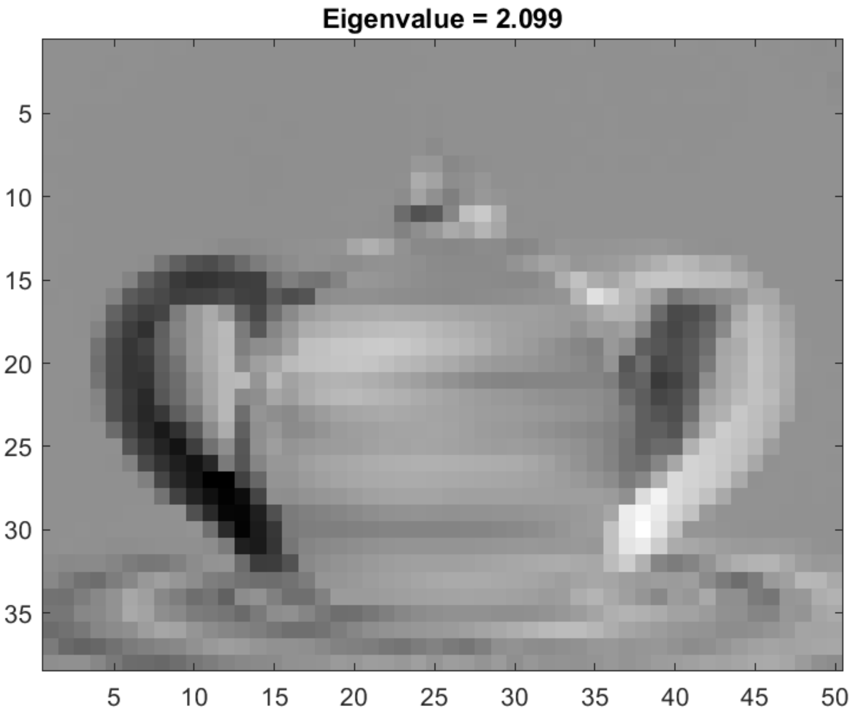


To calculate the eigenvalues, the covariance matrix is calculated, which is positive and symmetric. Hence, the eigenvalues are all non-negative. We find the top 3 eigenvalues which are - 4.2150, 3.0168, 2.0993, with respective eigenvectors.

The eigenvectors are shown as images below –

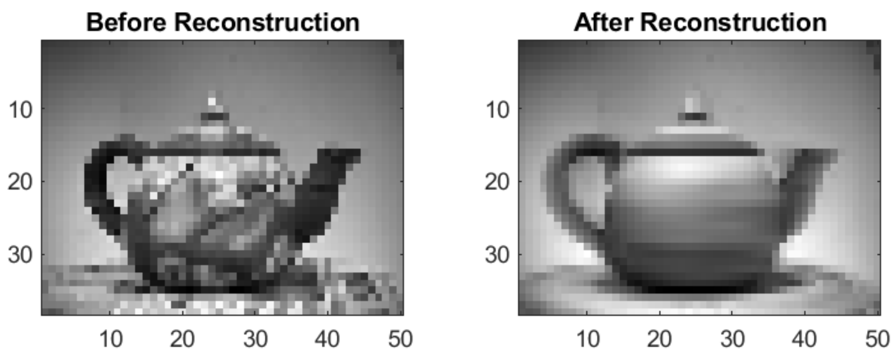


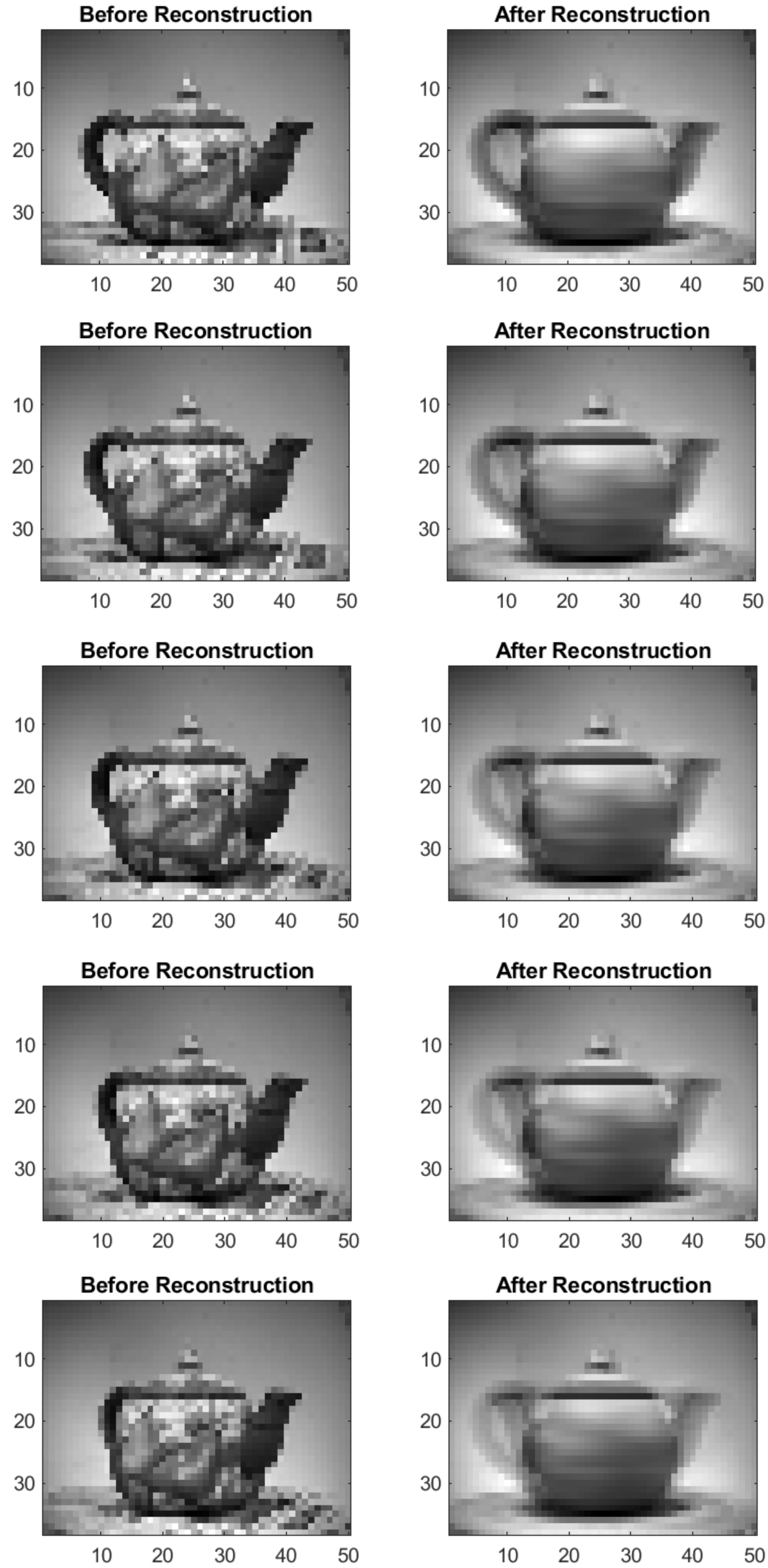


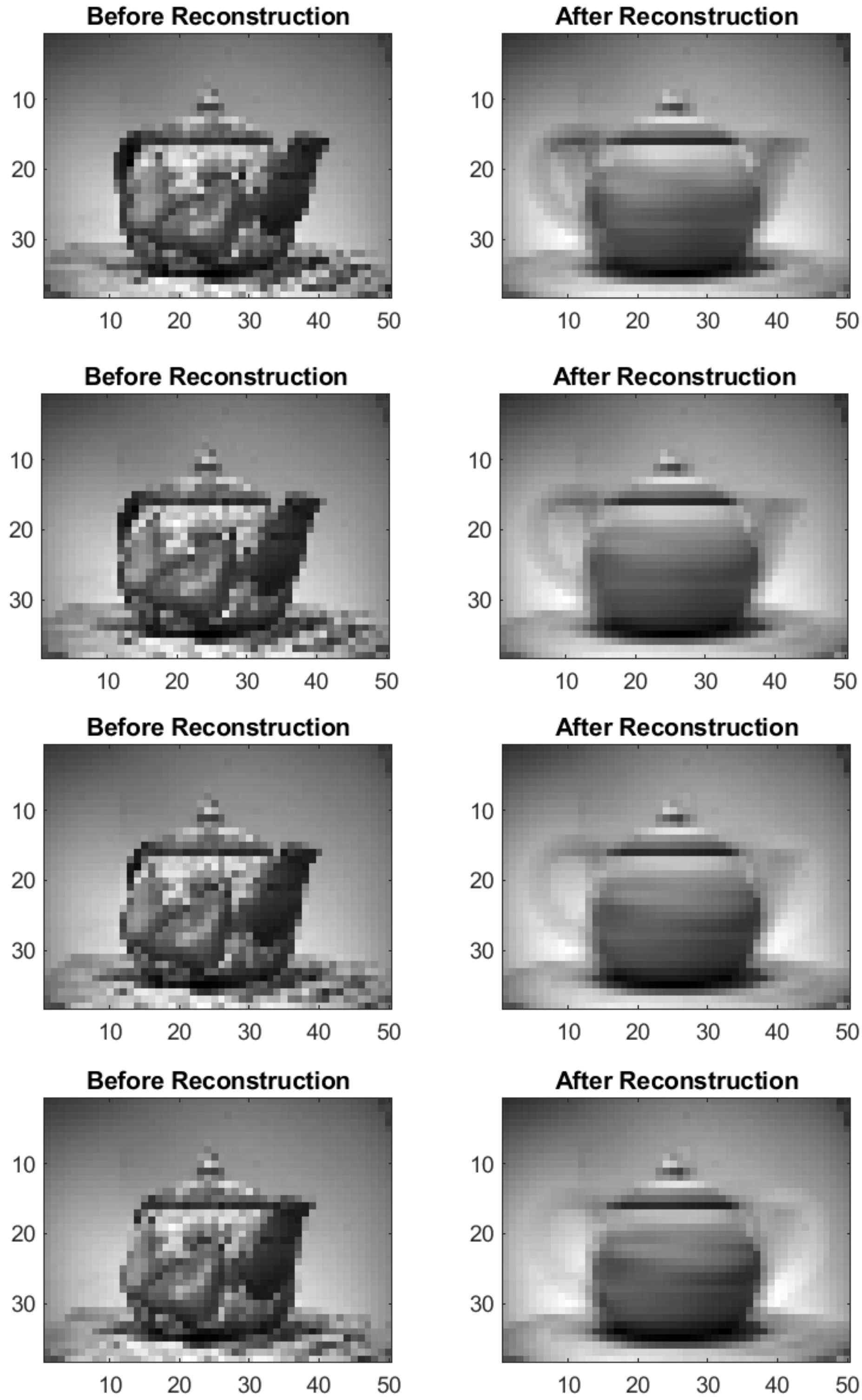


To reconstruct the images, the new coefficient matrix is calculated from the first 3 components of the eigenvectors.

The images before and after reconstruction are shown below –




Before Reconstruction
After Reconstruction
10
20
30
10
20
30
40
50
Before Reconstruction
After Reconstruction
Before Reconstruction
After Reconstruction
Before Reconstruction
After Reconstruction
Before Reconstruction
After Reconstruction



It is seen from the images that some are reconstructed properly and distinguishable while some are reconstructed poorly. The reconstruction can be improved by decoding with more components.

The PCA performance is evaluated based on the L2 normalization factor.

For the top-3 components, the result is 13.6262.

For the top-6 components, the result is 9.8303.

For the top-12 components, the result is 6.8428.

For the top-64 components, the result is 1.9002.

Hence, the more components we include in the reconstruction, the better the results.

Question 2:

2)

B1	B2
8 A	10 A
4 O	2 O

B = 'Choosing a box' ; $P(B) = \frac{1}{2}$

A = 'Apple is picked'

O = 'Orange is picked'

$$P(A) = P(A|B) = \frac{1}{2} \times \frac{8}{8+4} + \frac{1}{2} \times \frac{10}{10+2}$$

$$P(A) = \frac{1}{2} \times \frac{8+10}{12} = \frac{18}{2 \times 12} = \frac{3}{4}$$

$$P(B_1|F) = \frac{P(A|B_1)\,P(B_1)}{P(A)}$$

$$P(B_1|F) = \frac{\frac{8}{12} \cdot \frac{1}{2}}{\frac{3}{4}} = \frac{4}{9}$$

Question 3:

3) Let y_i indicate the class i for $i \in \{1, 2\}$.

$$L(\theta) = \prod_{i=1}^{N} P(x_i | \theta) = \prod_{i=1}^{N} P(x_i | y_i, \theta) P(y_i | \theta)$$

$$LL(\theta) \propto \sum_{i=1}^{N} \log P(x_i | y_i, \theta) + \log P(y_i | \theta)$$

$$LL(\theta) \propto \sum_{i=1}^{N} -\frac{1}{2} \log \left[\det(\Sigma_{y_i})\right] - \frac{1}{2} (x_i - \mu_{y_i})^T \Sigma_{y_i}^{-1} (x_i - \mu_{y_i})$$

$$+ y_i \log \alpha + (1 - y_i) \log (1 - \alpha)$$

$$\alpha^* = \arg\max L(\theta)$$

$$\Rightarrow \frac{1}{\alpha} \sum_{i=1}^{N} y_i - \frac{1}{1-\alpha} \sum_{i=1}^{N} 1 - y_i = 0$$

$$\Rightarrow \alpha = \frac{1}{N} \sum_{i=1}^{N} y_i$$

$$\mu_1^* = \arg\max L(\theta)$$

$$\Rightarrow \sum_{i, s, t, y_i = 1}^{\mu_1} (x_1 - \mu_1)^T \Sigma_1^{-1} = 0$$

$$\Rightarrow \mu_1^* = \frac{1}{N_1} \sum_{i, s, t, y_i = 1} x_i$$

$$\mu_2^* = \frac{1}{N_1} \sum_{i, s.t.\, y_i = 2} x_i$$

$$\Sigma_1^* = \arg\max L(\theta) = \frac{1}{N_1} \sum_{i, s.t.\, y_i = 1} (x_i - \mu_1)(x_i - \mu_1)^T$$

$$\Sigma_2^* = \arg\max L(\theta) = \frac{1}{N_1} \sum_{i, s.t.\, y_i = 2} (x_i - \mu_2)(x_i - \mu_2)^T$$

$$d(x) = \frac{P(Y=1 \mid X=x)}{P(Y=2 \mid X=x)}$$

If $d > 1$ then classified class 1 & vice-versa

$$\log d(x) = \log P(Y=1 \mid X=x) - \log(Y=2 \mid X=x)$$

Using Baye's Rule:

$$P(Y=j \mid X=x) = \frac{P(X=x \mid Y=j) P(Y=j)}{P(X=x)} \quad ; \pi_j = P(Y=j)$$

$$\log d(x) = \log P(X=x \mid Y=2) + \pi_2 - \log(P=x \mid Y=1) - \pi_1$$

$$\log d(x) = -(x-\mu_2)' \Sigma_2^{-1} (x-\mu_2) + \pi_2 + (x-\mu_1)' \Sigma_1^{-1} (x-\mu_1)$$

Decision boundary is quadratic.

$\Sigma_1 = \Sigma_2$ means linear equation.