Homework 4

Komal Niraula (kn2505)

Problem 1:

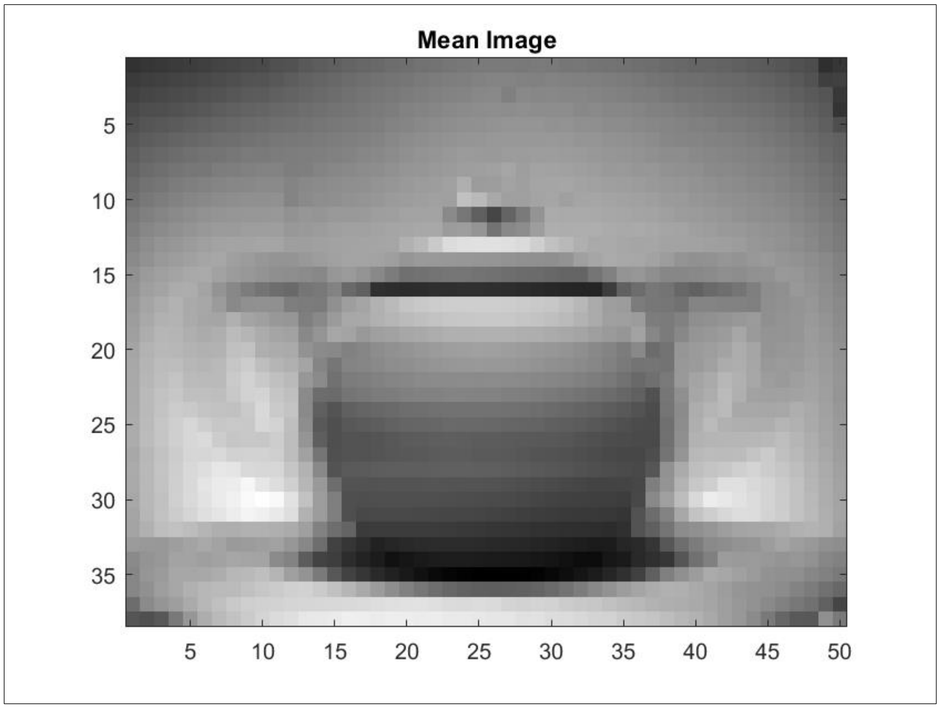

Fig 1.1: Mean Image

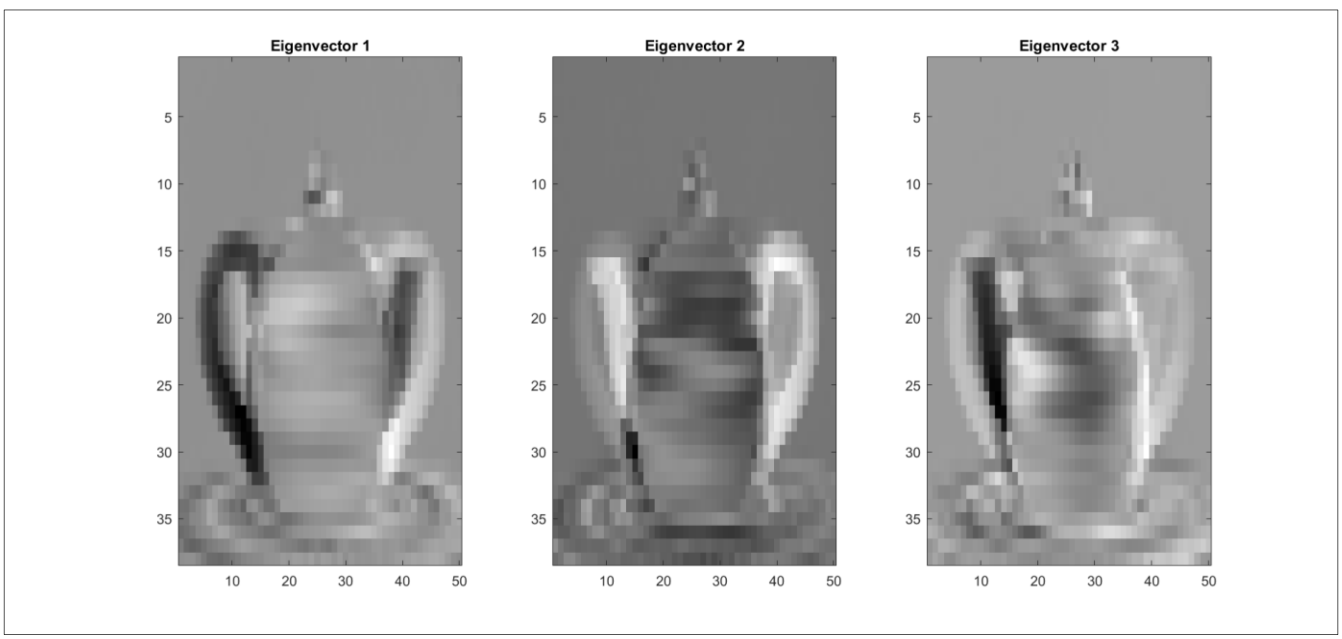

Fig 1.2: Top 3 Eigenvector Image

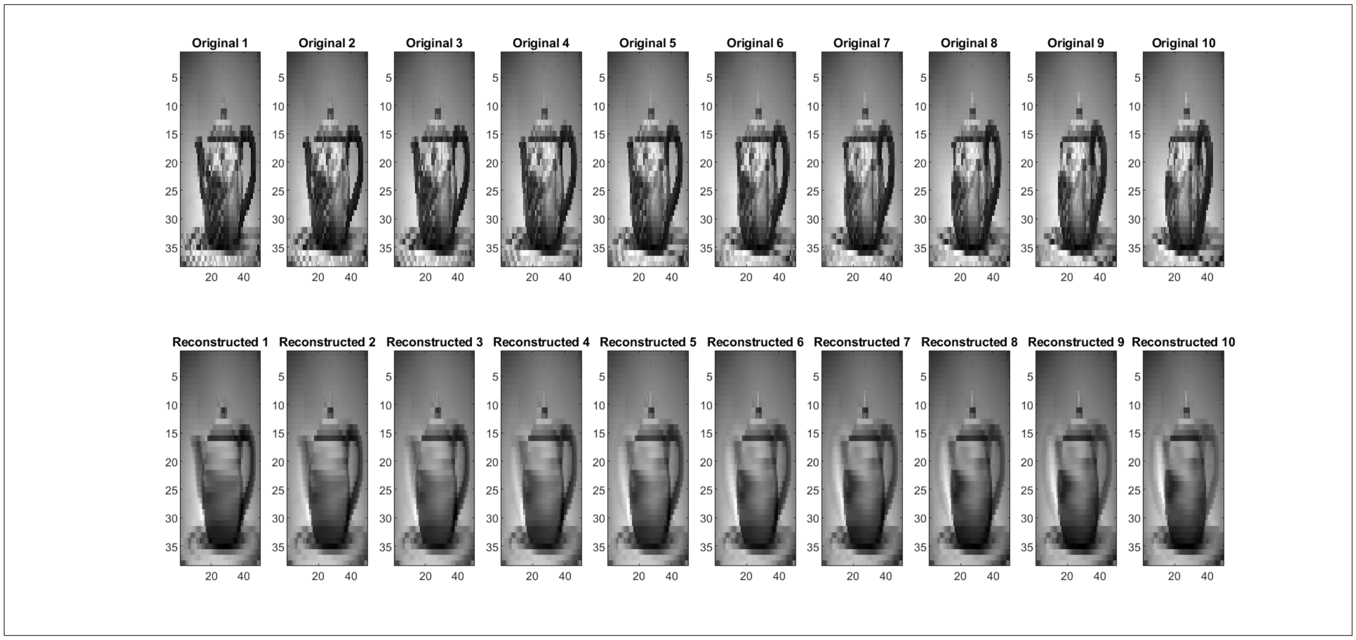

Fig 1.3: Original and Reconstructed Image

Observing the figure, we can say that the reconstructed image 1, 2 & 3 seems to be better than the 8, 9 & 10. The main challenge seems to be in reconstructing the handle and spout of the teapot. To observe, how the images differ, let's calculate mean squared error between reconstructed images and original images.

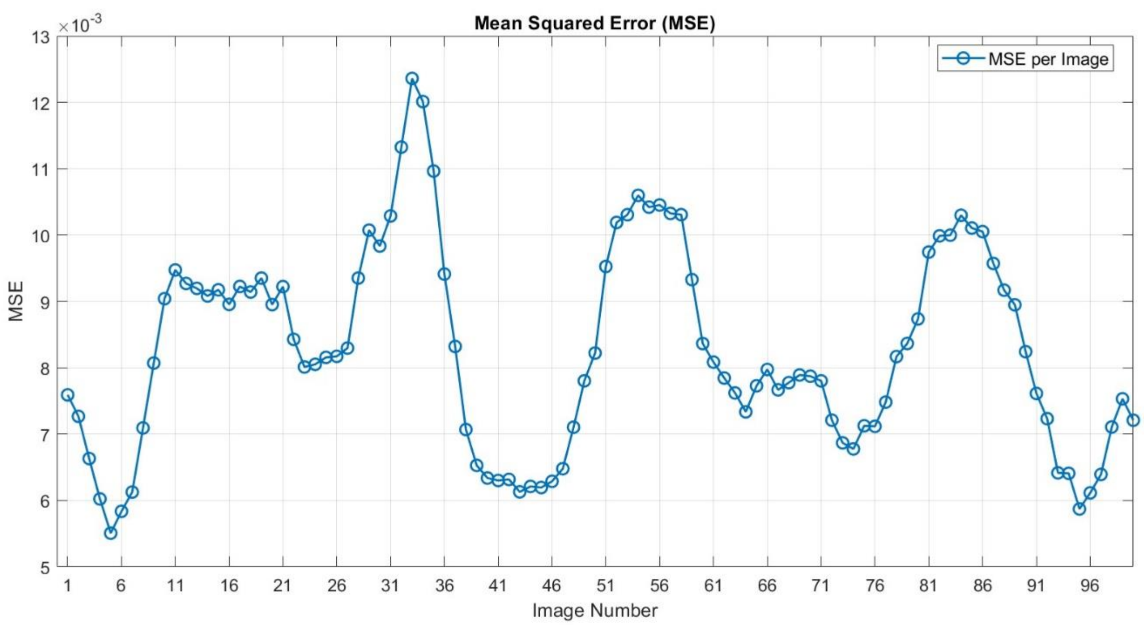

Fig 1.4: Mean Squared Error between reconstructed and original images

From above figure (1.4), it is evident that image number 5 in the dataset has the best match with its reconstructed image. In contrast, image number 33, with highest MSE, has the lowest level of similarity. Let's visualize these images to see how the best and least match images differs.

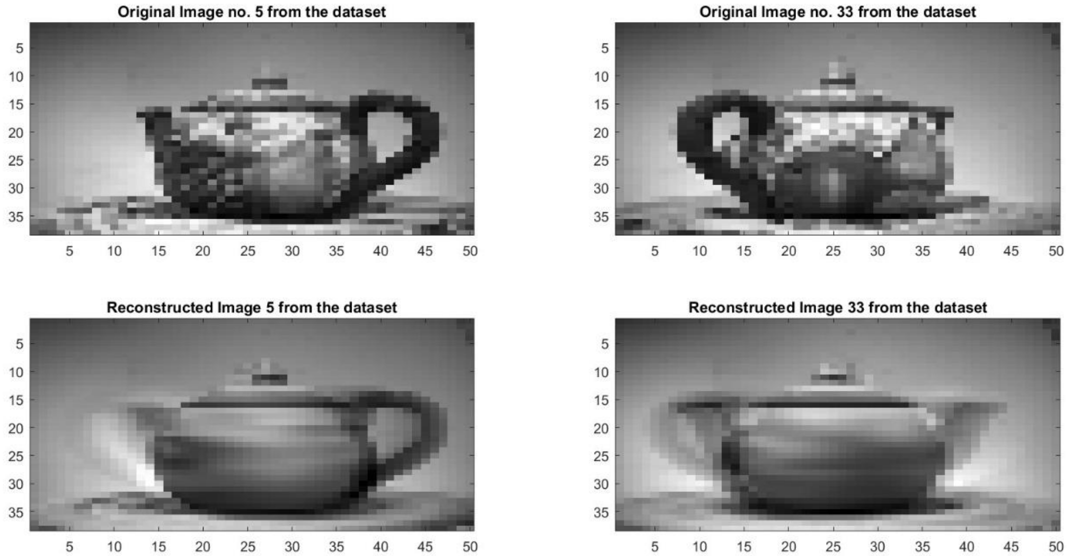

Fig 1.5: Original and reconstruction of image no. 5 & 33 from the dataset

We can see that the handle of teapot is more clear in the reconstructed image 5 than in the reconstructed image 33. Even the original image 5 seems to be more clear than the original image 33. Original image 33 has a lot of noise.

To improve the quality of reconstructed images, we can reduce the noises in the original image. We need to do data preprocessing for this. One of the method this can be done is by using regularization. Also, here we have only used 3 eigenvectors. This limits the reconstruction accuracy. We can experiment with more eigenvectors like 10 or 20 to check if the better images can be reconstructed. We can also try normalizing the data to ensure that all dimensions contribute equally.

Problem 2:

Let,

B1 = Box 1 is chosen

B2 = Box 2 is chosen

A = An apple is chosen

As there are only two boxes with no biases, the probability of choosing either box is the same i.e. half = P(B1) = P(B2) = 0.5

P(A|B1) = Number of apples in Box 1 / Total items in Box 1

= 8 / (8 + 4) = 8 / 12 = 2/3

P(A|B2) = Number of apples in Box 2 / Total items in Box 2

= 10 / (10 + 2) = 10 / 12 = 5/6

Probability of choosing an apple, P(A) = P(A|B1) × P(B1) + P(A|B2) × P(B2)

= (2/3 * 0.5) + (5/6 * 0.5)

= 3 / 4

Using Bayes Theorem,

Probability that the apple came from first box P(B1|A) = P(A|B1) × P(B1) / P(A)

= ((2/3) × 0.5) / (3/4)

= 1/3 * 4/3

= 4/9

Therefore, there is a 4/9 probability that an apple came from the first box.

Problem 3:

Answers

Given N i.i.d samples (x_i, y_i), the likelihood is:

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

where, $p(y_i|\theta) = \alpha^{y_i}(1-\alpha)^{1-y_i}$ &

$p(x_i|y_i, \theta) = \mathcal{N}(x_i|\mu_{y_i}, \Sigma_{y_i})$

Gaussian probability density function:

$$\mathcal{N}(x|\mu, \Sigma) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right)$$

Taking the log of $L(\theta)$:

$$\log L(\theta) = \sum_{i=1}^{N} \log p(y_i|\theta) + \sum_{i=1}^{N} \log p(x_i|y_i, \theta)$$

$$\log L(\theta) = \sum_{i=1}^{N} \left[y_i \log\alpha + (1-y_i)\log(1-\alpha)\right] + \sum_{i=1}^{N} \log \mathcal{N}(x_i|\mu_{y_i}, \Sigma_{y_i})$$

The class prior contributes the term:

$$\sum_{i=1}^{N} \left[y_i \log\alpha + (1-y_i)\log(1-\alpha)\right]$$

$N_1 = \sum_{i=1}^{N} y_i$, this is the total number of samples with $y=1$, and $N_2 = N - N_1$, the total number with $y=0$.

The prior term simplifies:

$$N_1 \log \alpha + N_2 \log (1-\alpha)$$

Deriving prior probability α

$$\frac{\partial}{\partial \alpha}\left(N_1 \log \alpha + N_2 \log (1-\alpha)\right) = \frac{N_1}{\alpha} - \frac{N_2}{1-\alpha}$$

Setting the derivative to zero:

$$\frac{N_1}{\alpha} = \frac{N_2}{1-\alpha}$$

$$N_1(1-\alpha) = N_2 \alpha$$

$$\alpha = \frac{N_1}{N}$$

Taking log of the Gaussian:

$$\log \mathcal{N}(x_i \mid \mu_{y_i}, \Sigma_{y_i}) = -\frac{1}{2}\left[(x_i - \mu_{y_i})^T \Sigma_{y_i}^{-1} (x_i - \mu_{y_i}) + \log |\Sigma_{y_i}| + d \log (2\pi)\right]$$

Deriving Mean

Log-likelihood for $y=1$ is:

$$\log L(\mu_1) = -\frac{1}{2}\sum_{y_i=1}\left[(x_i-\mu_1)^T\Sigma_1^{-1}(x_i-\mu_1)+\log|\Sigma_1|+d\log(2\pi)\right]$$

Taking derivative with respect to μ_1:

$$\frac{\partial \log L}{\partial \mu_1} = -\frac{1}{2}\sum_{y_i=1}\left[-2\Sigma_1^{-1}(x_i-\mu_1)\right]$$

$$= \Sigma_1^{-1}\sum_{y_i=1}(x_i-\mu_1)$$

Setting the derivative to zero:

$$\Sigma_1^{-1}\sum_{y_i=1}(x_i-\mu_1)=0$$

$$\mu_1 = \frac{1}{N_1}\sum_{y_i=1} x_i \quad , \text{ where } N_1 = \sum_{i=1}^{N} y_i$$

log-likelihood for $y=0$ is:

$$\log L(\mu_2) = -\frac{1}{2}\sum_{y_i=0}\left[(x_i-\mu_2)^T\Sigma_2^{-1}(x_i-\mu_2)+\log|\Sigma_2|+d\log(2\pi)\right]$$

Taking derivative with respect to μ_2:

$$\frac{\partial \log L}{\partial \mu_2} = -\frac{1}{2}\sum_{y_i=0}\left[-2\Sigma_2^{-1}(x_i-\mu_2)\right]$$

$$= \Sigma_2^{-1}\sum_{y_i=0}(x_i-\mu_2)$$

Setting the derivative to zero:

$$\Sigma_2^{-1}\sum_{y_i=0}(x_i-\mu_2)=0$$

$$\mu_2 = \frac{1}{N_2}\sum_{y_i=0} x_i, \text{ where } N_2 = \sum_{i=1}^{N}(1-y_i)$$

Deriving Covariance Matrix:

Using μ_1 derived previously, the log-likelihood contribution for Σ_i:

$$\log L(\Sigma_1) = -\frac{1}{2} \sum_{y_i=1} \left[(x_i-\mu_1)^T \Sigma_1^{-1} (x_i-\mu_1) + \log|\Sigma_1| + d\log(2\pi) \right]$$

Taking derivative with respect to Σ_i

$$\frac{\partial \log L}{\partial \Sigma_1} = -\frac{1}{2} \sum_{y_i=1} \left[-\Sigma_1^{-1} + \Sigma_1^{-1}(x_i-\mu_1)(x_i-\mu_1)^T \Sigma_1^{-1} \right]$$

Setting the derivative to 0,

$$\sum_{y_i=1} \left[\Sigma_1 - (x_i-\mu_1)(x_i-\mu_1)^T \right] = 0$$

$$\Sigma_1 = \frac{1}{N_1} \sum_{y_i=1} (x_i-\mu_1)(x_i-\mu_1)^T$$

Using μ_2 derived previously, the log-likelihood contribution for Σ_2:

$$\log L(\Sigma_2) = -\frac{1}{2} \sum_{y_i=0} \left[(x_i-\mu_2)^T \Sigma_2^{-1} (x_i-\mu_2) + \log|\Sigma_2| + d\log(2\pi) \right]$$

Taking derivative with respect to Σ_2:

$$\frac{\partial \log L}{\partial \Sigma_2} = -\frac{1}{2} \sum_{y_i=0} \left[-\Sigma_2^{-1} + \Sigma_2^{-1}(x_i-\mu_2)(x_i-\mu_2)^T \Sigma_2^{-1} \right]$$

Setting the derivative to 0,

$$\sum_{y_i=0} \left[\Sigma_2 - (x_i-\mu_2)(x_i-\mu_2)^T \right] = 0$$

$$\Sigma_2 = \frac{1}{N_2} \sum_{y_i=0} (x_i-\mu_2)(x_i-\mu_2)^T$$

Decision Boundary

The decision rule is:

$$y = \arg\max_{\hat{y} \in \{0,1\}} p(\hat{y}|x)$$

Using Bayes' theorem:

$$\cancel{p(y=1|x) \propto \alpha}$$

$p(y=1|x)$ is directly proportional to $\alpha \cdot \mathcal{N}(x|\mu_1, \Sigma_1)$

$p(y=0|x)$ is directly proportional to $(1-\alpha) \cdot \mathcal{N}(x|\mu_2, \Sigma_2)$

Taking log for comparison:

$$\log p(y=1|x) - \log p(y=0|x) > 0$$

Substituting Gaussian log terms:

$$\log\alpha - \log(1-\alpha) - \frac{1}{2}\left[(x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1) + \log|\Sigma_1|\right] + \frac{1}{2}\left[(x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2) + \log|\Sigma_2|\right] > 0$$

$$x^T(\Sigma_1^{-1} - \Sigma_2^{-1})x + 2x^T(\Sigma_2^{-1}\mu_2 - \Sigma_1^{-1}\mu_1) + \text{constant} = 0$$

If the covariances are equal, the quadratic term disappears:

$$x^T(\Sigma_1^{-1} - \Sigma_2^{-1})x = 0$$

So, the decision boundary is linear.

$$x^T\Sigma^{-1}(\mu_2 - \mu_1) + \text{constant} = 0$$

If the covariances are unequal, the quadratic term remains:

$$x^T(\Sigma_1^{-1} - \Sigma_2^{-1})x \neq 0$$

So, the decision boundary is quadratic.

Decision boundary for the case where covariances are not equal: