Homework 4

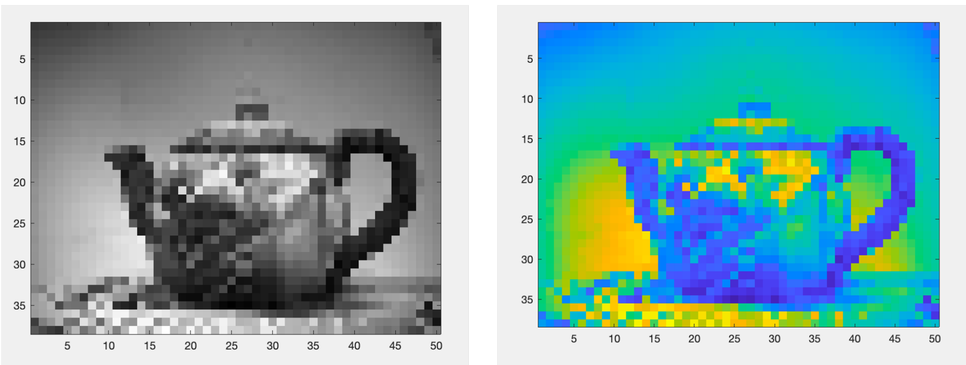
Student name: *Zihan Zeng*

Course: *ECE 6143 Machine Learning* – Professor: *Yury Dvorkin*
Due date: *Nov 18, 2021*

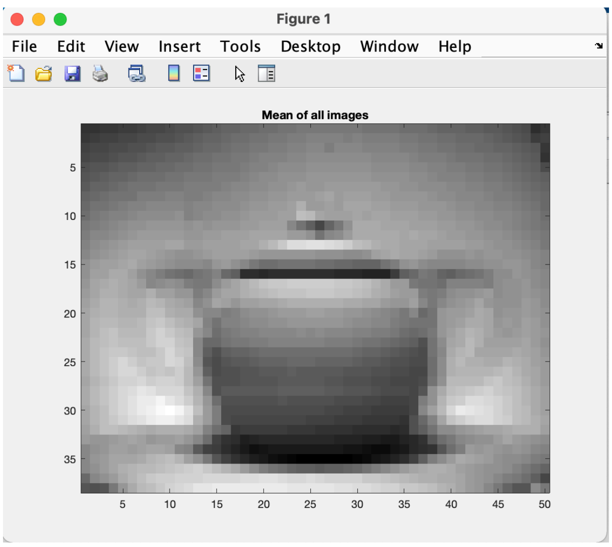
Problem 1

Solution.

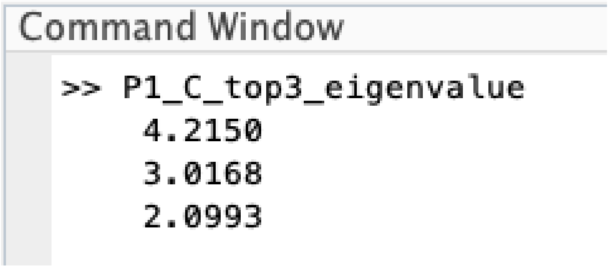
(A) To view images:
The below image randomly selects a certain image in the data set, showing its original image and grayscale image.



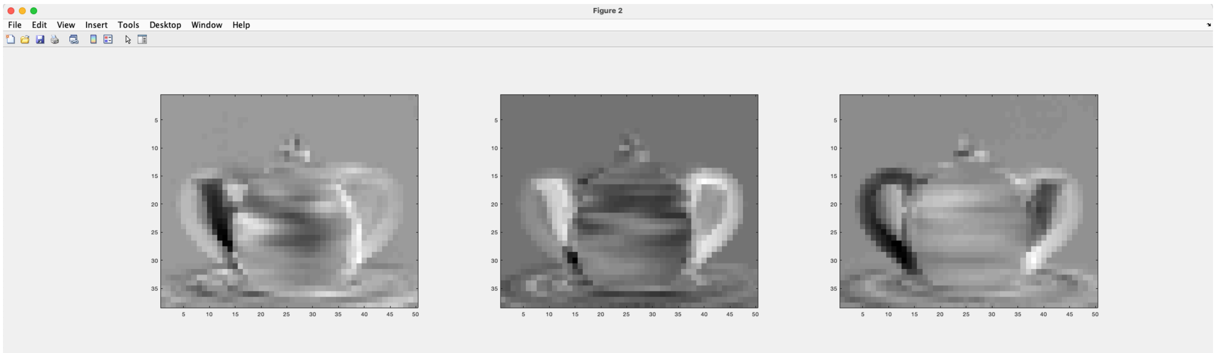
(B) Compute the data mean:
The picture below shows that the result after we compute mean.
(Source Code: *P1_B_compute_mean.m*)



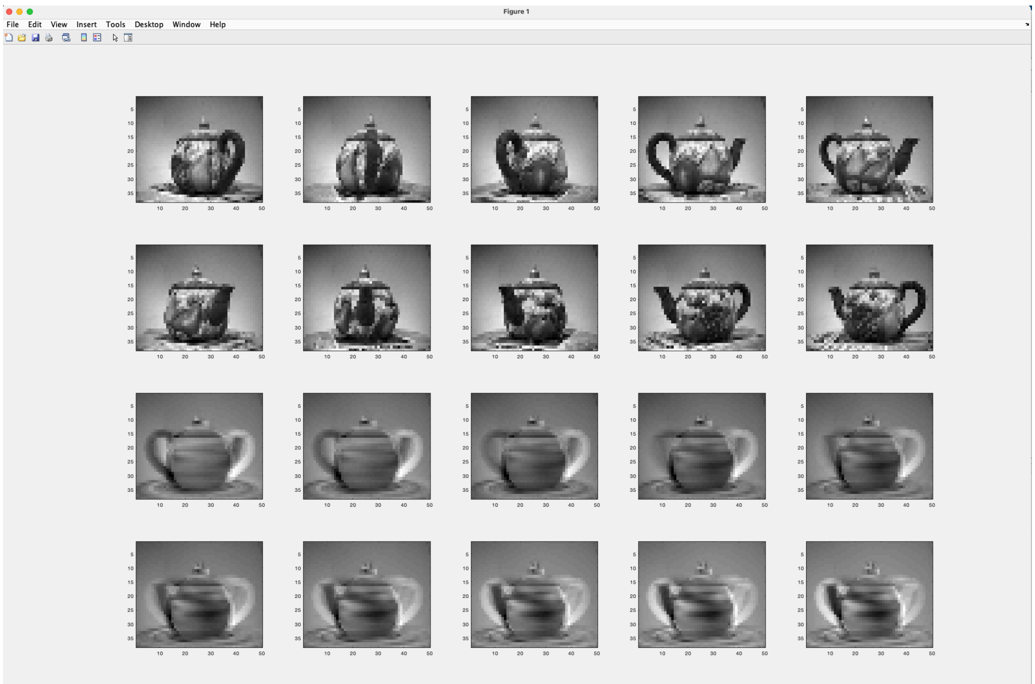
(*C*) Top3 eigenvectors of the data covariance matrix:
Through calculation, we can get the top 3 eigenvectors of the data covariance matrix. These three values are shown in the figure below.
(Source Code: *P1_C_top3_eigenvector.m*)



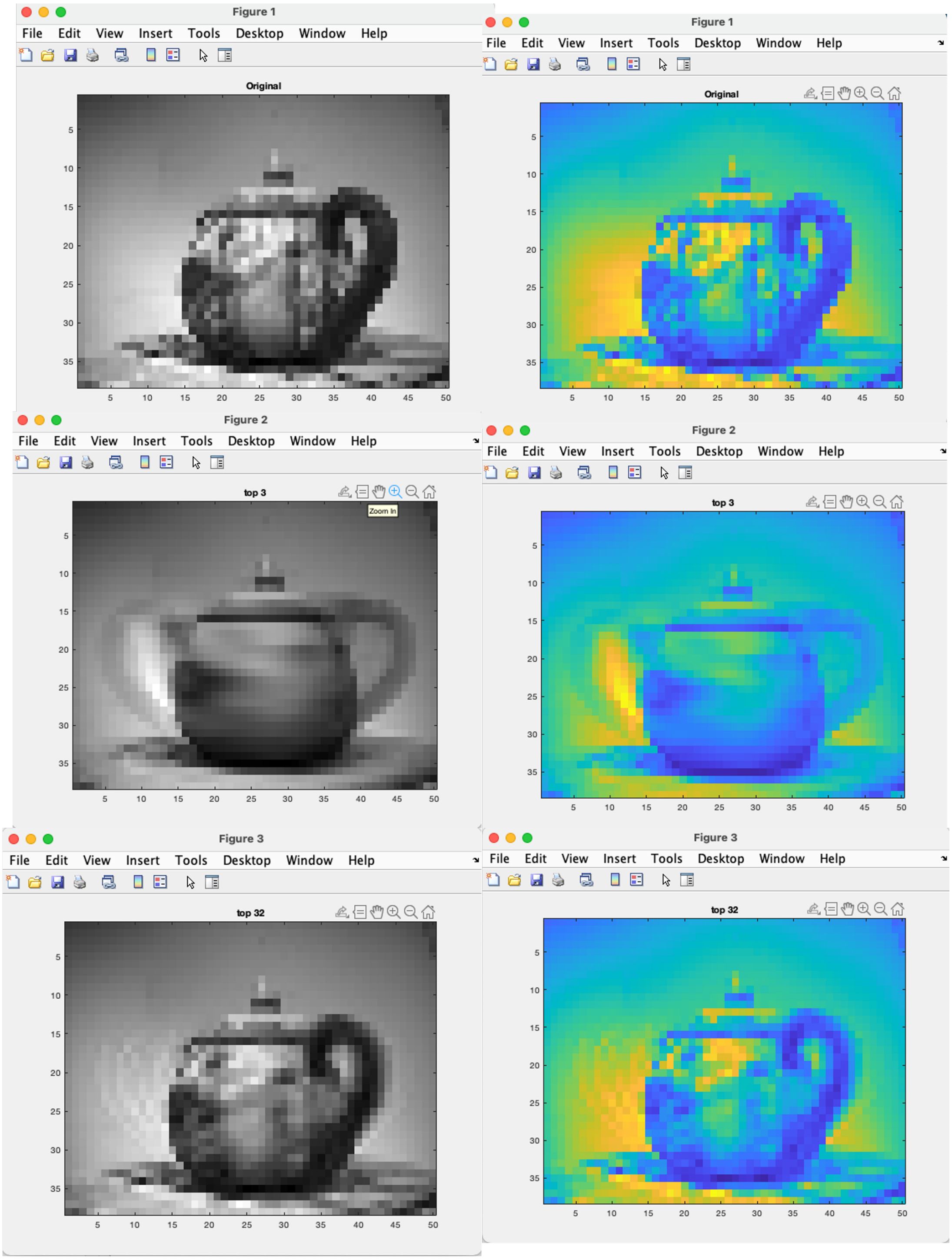
The following figure shows the picture in the case of top3 eigenvectors:
(Source Code: *P1_C_top3_eigenvector.m*)



(*D*) Using PCA:
I randomly selected 10 pictures, showing the situation before and after reconstruct.
(Source Code: *P1_D_usingPCA.m*)



(*E*) Discussion:
From the above figure, we can see that the reconstructed picture can roughly depict the outline of the teapot, but some details are ignored. According to my previous experience, we need to add more features.
(Source Code: *P1_E_discussion.m*)
So I chose top 32 eigenvectors and want to compare it, as shown in the figure below:



It can be seen that the more features, the better the reconstruction.

Problem 2

Solution. Based on the information we know, let P(X) denote the probability of choosing the first box, and P(Y) denote the probability of choosing the second box.

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

Let A denote the event of taking apple from the box. So we can get:

$$P(A|X) = \frac{8}{12} = \frac{2}{3} \qquad P(A|Y) = \frac{10}{12} = \frac{5}{6}$$

And apply Bayes' rule:

$$\begin{aligned} P(X|A) &= \frac{P(X)P(A|X)}{P(X)P(A|X) + P(Y)P(A|Y)} \\ &= \frac{\frac{1}{2} \times \frac{2}{3}}{\frac{1}{2} \times \frac{2}{3} + \frac{1}{2} \times \frac{5}{6}} = \frac{4}{9} \end{aligned}$$

Problem 3

Solution. The process of mathematical proof is as follows:
(1) $L(\theta) = P(x_1, x_2, \ldots, x_n, y_1, y_2, \ldots, y_n \mid \theta)$, and we know that all the data are *i.i.d.* So we can get:

$$L(\theta) = P(x_1, x_2, \ldots, x_n, y_1, y_2, \cdots, y_n \mid \theta) = \prod_{i=1}^{N} P(x_1, y_i \mid \theta)$$

$$\begin{aligned} L(\theta) &= \sum_{i=1}^{N} \log P(x_i, y_i \mid \theta) \\ &= \sum_{i=1}^{N} \log \left[P(y_i \mid \theta) P(x_i \mid y_i, \theta)\right] \\ &= \sum_{i=1}^{N} \log P(y_i \mid \theta) + \sum_{i=1}^{N} \log P(x_i \mid y_i, \theta) \\ &= \sum_{i=1}^{N} \log P(y_i \mid \alpha) + \sum_{y_i \in 0} \log P(x_i \mid \mu_0, \Sigma_0) + \Sigma_{y_i \in 1} \log P(x_i \mid \mu_1, \Sigma_1) \\ &= \sum_{i=1}^{N} \log \alpha^{y_i}(1-\alpha)^{1-y_i} + \Sigma_{y_i \in 0} \log P(x_i \mid \mu_0, \Sigma_0) + \sum_{y_i \in 1} \log P(x_i \mid \mu_1, \Sigma_1) \\ &= \Sigma_{y_i \in 0} \log(1-\alpha) + \Sigma_{y_i \in 1} \log \alpha + \Sigma_{y_i \in 0} \log P(x_i \mid \mu_0, \Sigma_0) + \Sigma_{y_i \in 1} \log P(x_i \mid \mu_1, \Sigma_1) \end{aligned}$$

$$\alpha^{*}=\arg\max_{\alpha} l(\theta)$$
$$\frac{\partial}{\partial \alpha}\Big|(\theta)=\frac{\partial}{\partial \alpha}\left[\Sigma_{y_i\in 0}\log(1-\alpha)+\Sigma_{y_i\in 1}\log\alpha\right]=\Sigma_{y_i\in 0}\frac{-1}{1-\alpha}+\Sigma_{y_i\in}\Big|\frac{1}{\alpha}$$

Let $N_0=$ the number of $y_i=0$ $N_1=$ the number of $y_i=1$ We have $\frac{\partial}{\partial 2}L(\theta)=\frac{N_1}{\alpha}+\frac{-N_0}{1-\alpha}$

$$\text{Compute } \frac{\partial}{\partial \alpha}L(\theta)=\frac{N_1}{\alpha}+\frac{-N_0}{1-\alpha}=0$$
$$(1-\alpha)N_1-\alpha N_0=0$$
$$\alpha=\frac{N_1}{N_0+N_1}$$

$\mu_0^{*}=\operatorname{argmax}_{\mu} L(\theta)$

$$\frac{\partial}{\partial \mu_0}L(\theta)=\frac{\partial}{\partial \mu_0}\Sigma_{y_i\in 0}\log P\left(x_i \mid \mu_0,\Sigma_0\right)=\frac{\partial}{\partial \mu_0}\Sigma_{y_i\in 0}\log\frac{1}{\left(2\pi\rho^{\rho 3}\sqrt{\Sigma_1}\right.}\exp\left(-\frac{1}{2}\left(\vec{x}_i-\vec{\mu}_0\right)^{\top}\Sigma_i^{-1}\left(\vec{x}_i\vec{\mu}_v\right)\right)$$
$$=\frac{d}{\partial \mu_0}\Sigma_{y_i\in 0}\left[-\frac{1}{2}\left(\vec{x}_i-\vec{\mu}_0\right)^{\top}\Sigma_0^{-1}\left(\vec{x}_i-\vec{\mu}_0\right)\right]$$

$$\text{Compute } \frac{\partial}{\partial \mu_0}l(\theta)=0$$
$$\frac{\partial}{\partial \mu_0}\Sigma_{y_i\in 0}\left[-\frac{1}{2}\left(\vec{x}_i-\vec{\mu}_0\right)^{\top}\Sigma_0^{-1}\left(\vec{x}_i-\vec{\mu}_0\right)\right]=0$$
$$\Sigma_{y_1\in 0}\left(\vec{x}_1-\vec{\mu}_0\right)^{\top}\Sigma_0^{-1}=0$$

Compute $\frac{\partial}{\partial \mu_0}(\omega)=0$

$$\frac{\partial}{\partial \mu_0}\sum y_i\in 0\left[-\frac{1}{2}\left(\vec{x}_i-\vec{\mu}_0\right)^{\top}\Sigma_0^{-1}\left(\vec{x}_i-\vec{\mu}_0\right)\right]=0$$
$$\sum y_i\in 0 \quad \left(\vec{x}_i-\vec{\mu}_0\right)^{\top}\Sigma_0^{-1}=0$$
$$\vec{\mu}_0=\frac{1}{N_0}\Sigma y_i\in 0\vec{x}_i$$

$\vec{\mu}_0^{*}=\frac{1}{N_0}\Sigma y_i\in 0\vec{x}_i$

So similarly, we can also apply this method into different parameters. Therefore, we can get
$$\begin{cases}\alpha^{*}=\frac{N_1}{N_0+N_1}\\ \vec{\mu}_0^{*}=\frac{1}{N_0}\Sigma_{y_i\in 0}\vec{x}_i\\ \mu_1^{*}=\frac{1}{N_1}\sum y_{i\in 1}x_i\\ \Sigma_0^{*}=\frac{1}{N_0}\Sigma_{y_i\in 0}\left(\vec{x}_i-\vec{\mu}_0\right)\left(\vec{x}_1-\vec{\mu}_0\right)^{\top}\\ \Sigma_1^{*}=\frac{1}{N_1}\Sigma_{y_1\in 1}\left(x_i-\mu_1\right)\left(x_i-\mu_1\right)^{\top}\end{cases}$$

(2) $y = \operatorname{argmax} y = \{0.j] p(\hat{y} \mid x)$

$$P(\hat{y} \mid x) = \frac{P(x,y)}{P(x)} = \frac{P(x,y)}{\Sigma_y P(x,y)} = \frac{P(x,y)}{P(x,y=0)+P(x,y=1)}$$

$$P(y=1 \mid x) = \frac{P(x,y=1)}{P(x,y=0)+P(x,y=1)}$$

$$= \frac{P\left(x,y=1 \mid \theta^*\right)}{P\left(x,y=0 \mid \theta^*\right)+P\left(x,y=1 \mid \theta^*\right)}$$

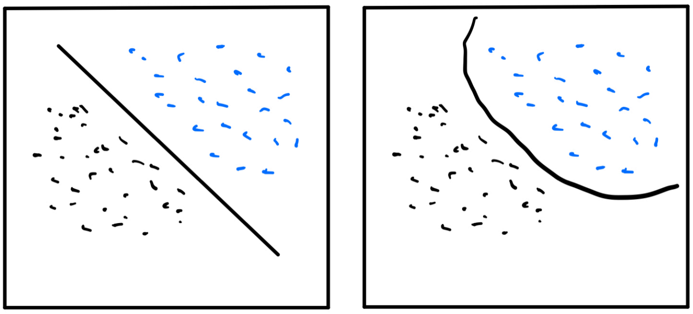
$$= \frac{P\left(x \mid y=1,\theta^*\right) P\left(y=1 \mid \theta^*\right)}{P\left(x \mid y=0,\theta^*\right) P\left(y=0 \mid \theta^*\right)+P\left(x \mid y=1,\theta^*\right) P\left(y=1 \mid \theta^*\right)}$$

$$= \frac{N\left(x \mid \mu_1^*, z_1^*\right) \alpha^*}{N\left(x \mid \mu_0^*, \Sigma_0^*\right)\left(1-^*\right)+N\left(x \mid \mu_1, \Sigma_1\right) \alpha^*}$$

$$= \frac{1}{\frac{N\left(x \mid \mu_0^*, \Sigma_0^*\right)\left(1-^*\right)}{N\left(x \mid \mu_1^*, \Sigma^*\right) \alpha^*}+1}$$

$$\text{if } \Sigma_0^* = \Sigma_1^* \quad = \frac{1}{\exp\left(-\frac{1}{2}\left(x-\mu_0\right)^{\top} \Sigma^{-1}\left(x-\mu_0\right)+\frac{1}{2}\left(x-\mu_1\right)^{\top} \Sigma^{-1}\left(x-\mu_1\right)\right)\left[\frac{1}{\alpha}-1\right]+1}$$

(3)



As shown in the figure above, the boundary is different in the case of linear and quadratic.