Karan Vora (kv2154)
Machine Learning Assignment 4

Problem 1):

Compute the mean of the image data of shape (100, 1900). And then subtract the mean from all data points to get the de-centered data.

$$\mu = mean(X)$$

and get x from

$$x = X - \mu$$

Computing the covariance matrix

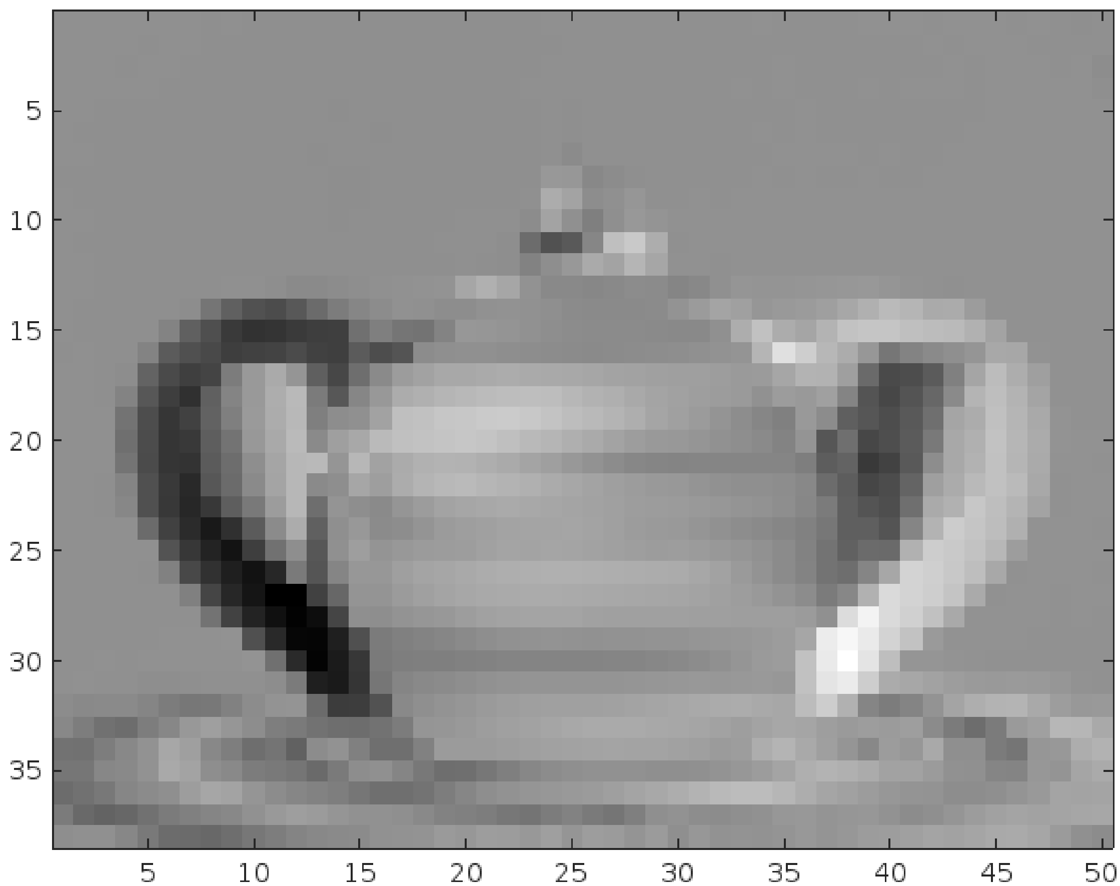
$$C = x^T x$$

Apply eigenvector Decomposition to covariance matrix

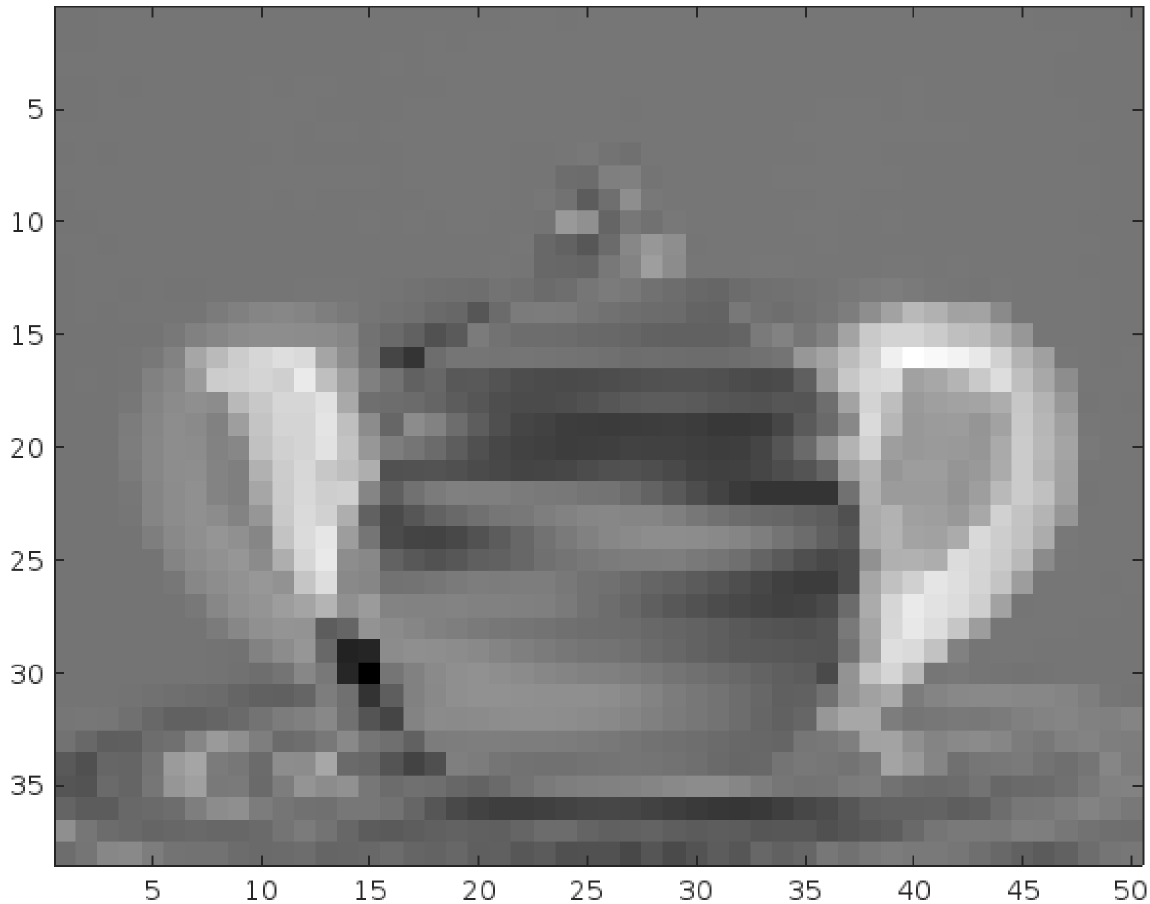
$$C = V \Lambda V^{-1}$$

The 3 most significant eigen value is

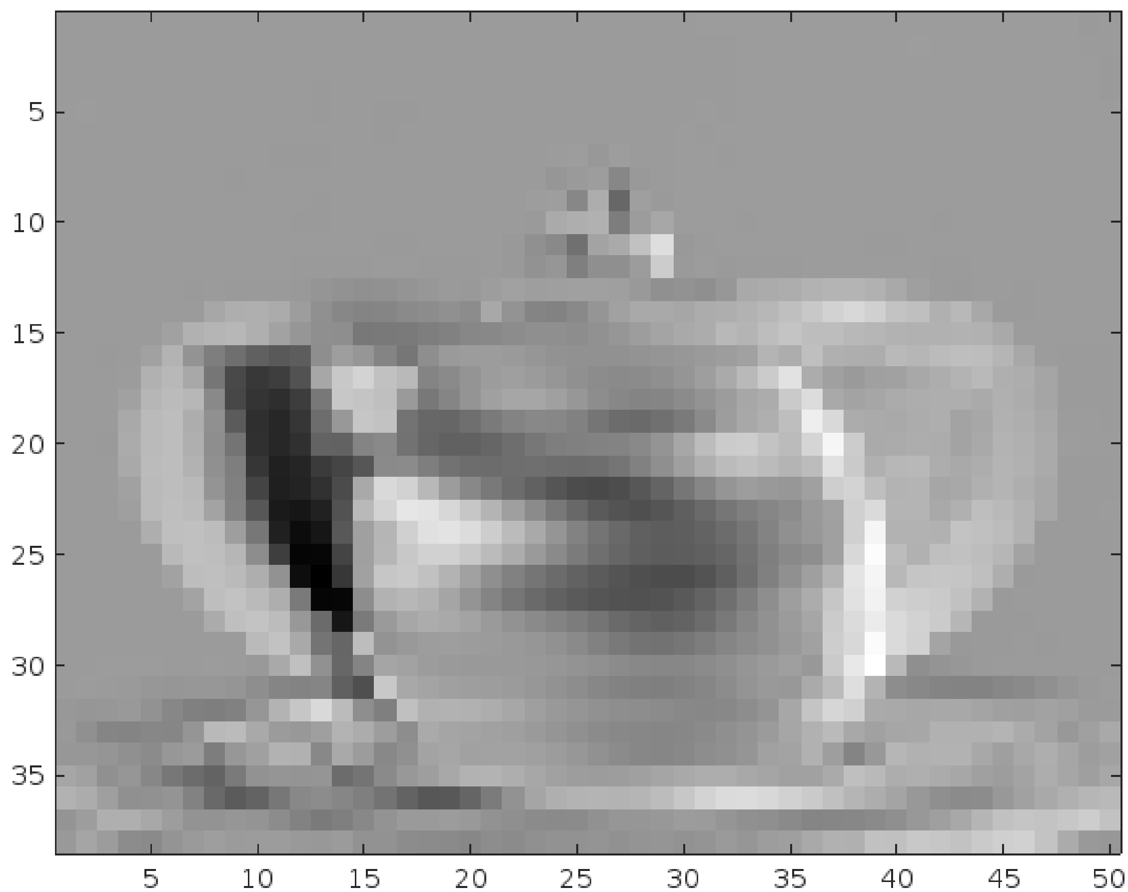
Eigen Value = 4.215022



Eigen Value = 3.016759



Eigen Value = 2.099301



The coefficient matrix is obtained by

$$c_{ij}=(X_i-\mu)^T V'_j = x_i^T V'_j$$

This is just the inner product of x and the top 3 eigenvectors

c = **xV’**

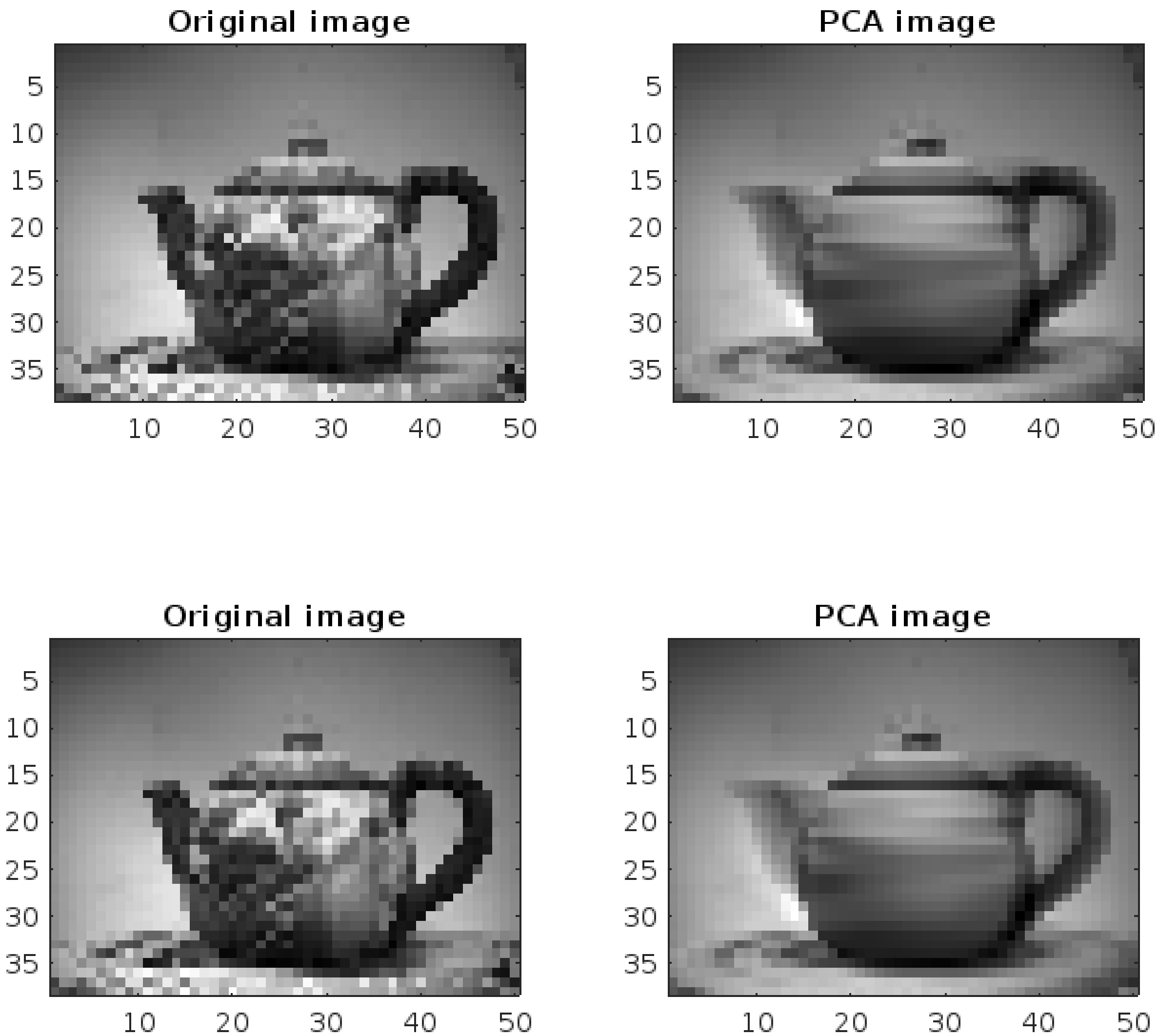
we can calculate $\hat{X}$ by

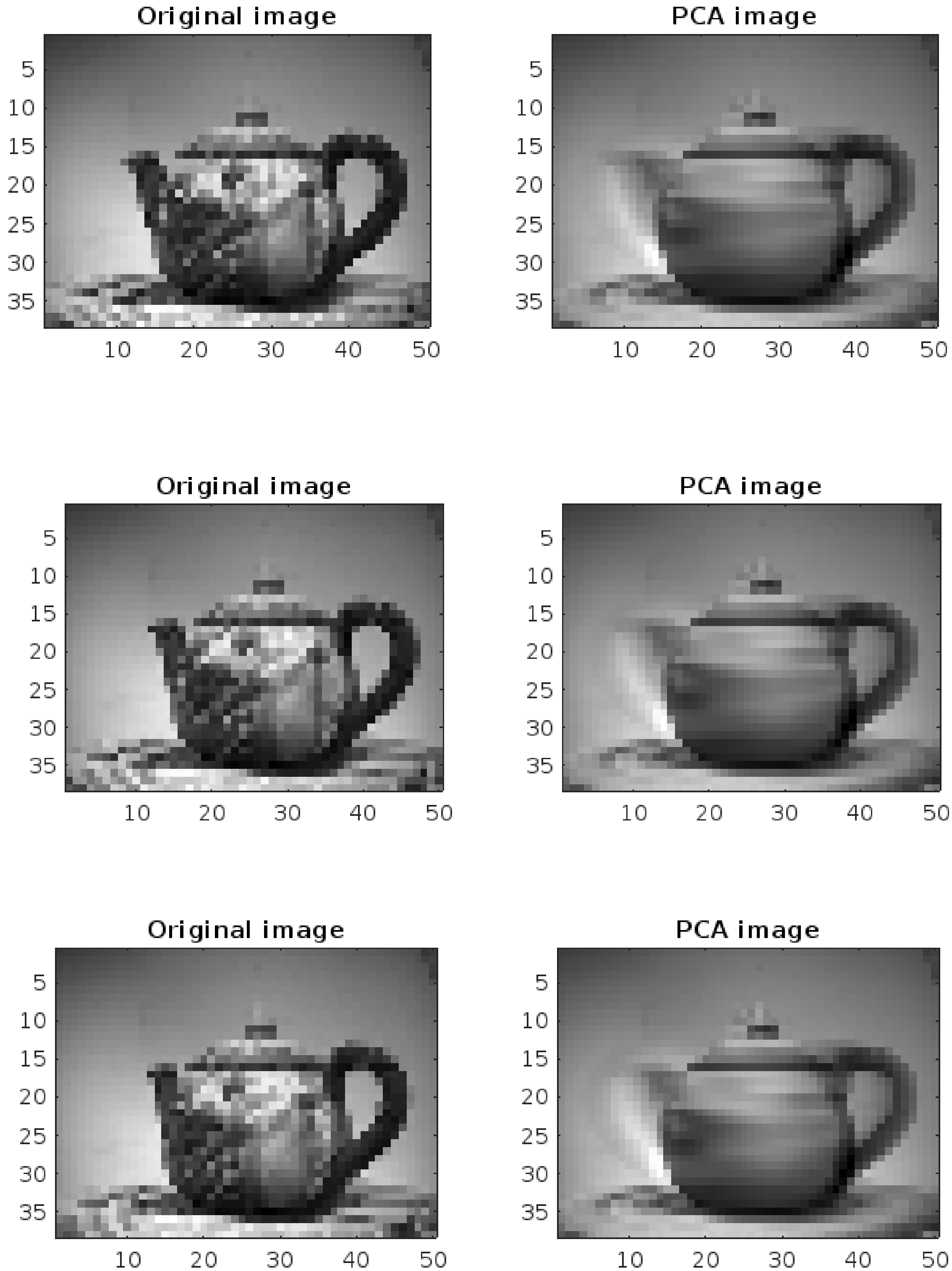
$$\hat{X}=\mu+\sum_{j=1}^{c} c_{ij} V'_j$$

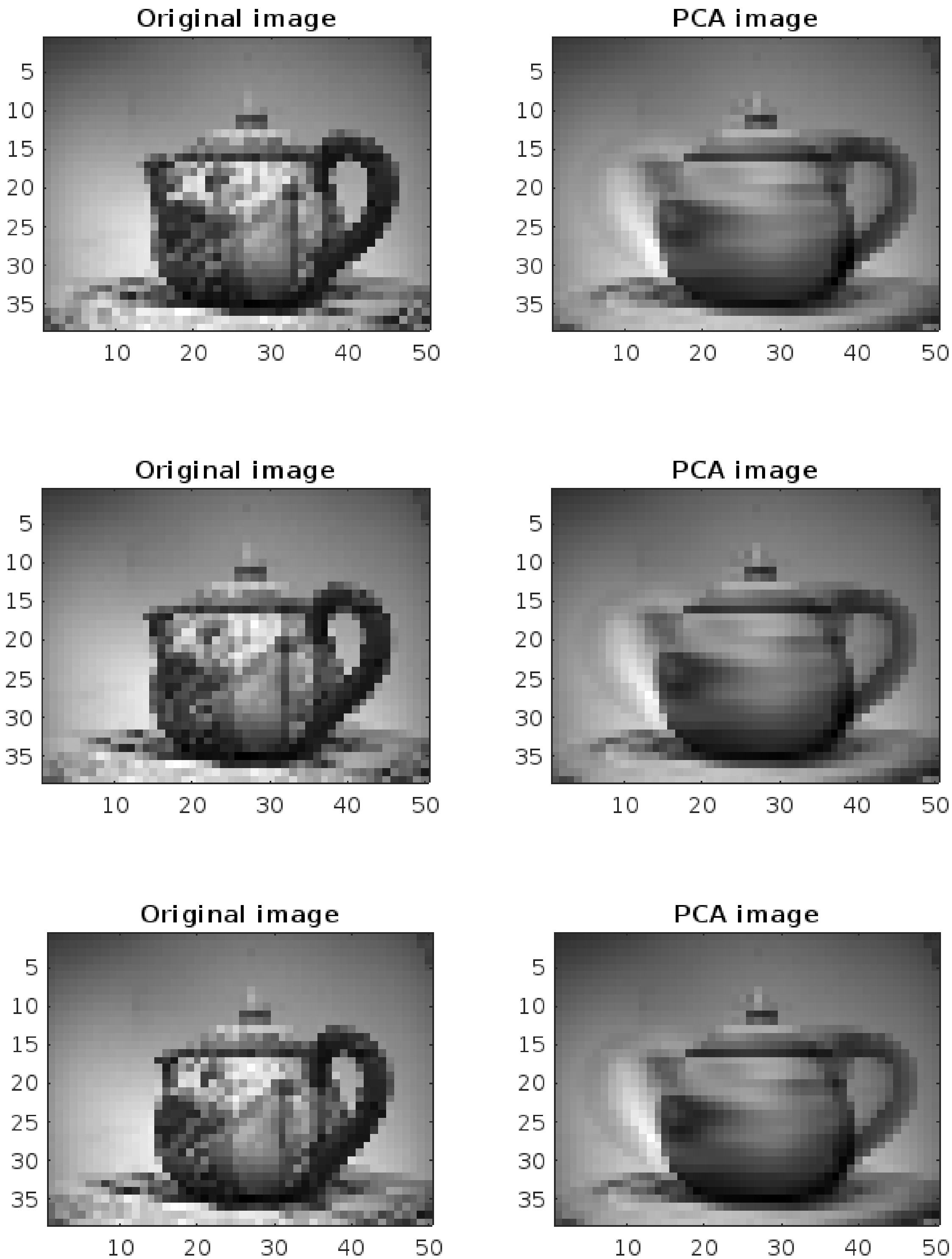
which in simple terms is

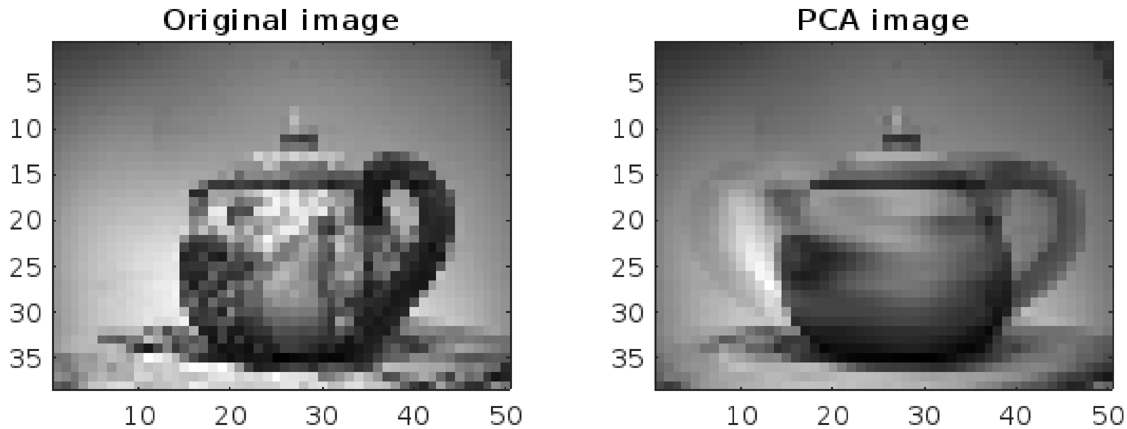
$$\hat{X}=\mu+cV'$$

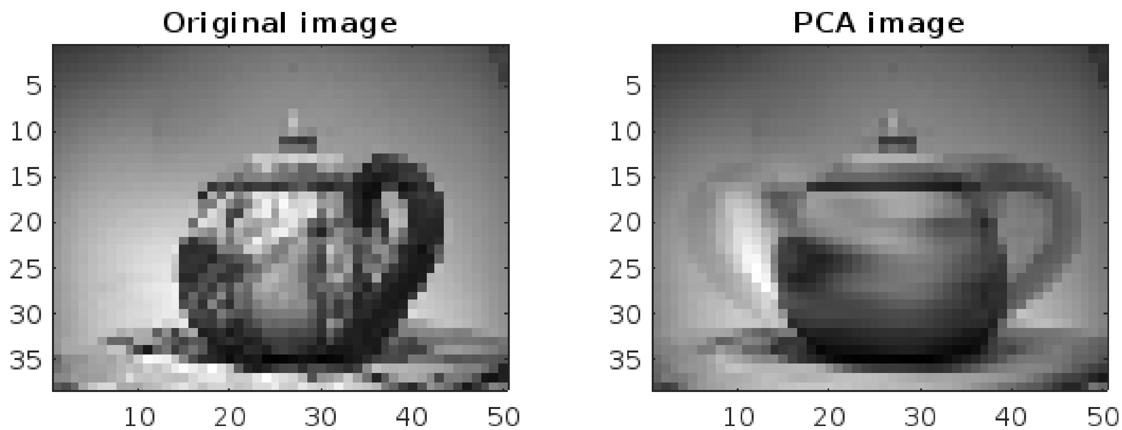
The 10 generated images are


Original image
PCA image
5
10
15
20
25
30
35
10
20
30
40
50
Original image
PCA image
5
10
15
20
25
30
35
10
20
30
40
50


Original image
PCA image
5
10
15
20
25
30
35
10
20
30
40
50
Original image
PCA image
5
10
15
20
25
30
35
10
20
30
40
50
Original image
PCA image
5
10
15
20
25
30
35
10
20
30
40
50


Original image
PCA image
5
10
15
20
25
30
35
10
20
30
40
50
Original image
PCA image
5
10
15
20
25
30
35
10
20
30
40
50
Original image
PCA image
5
10
15
20
25
30
35
10
20
30
40
50





==

Problem 2):

First box 8 Apples 4 Oranges
Second box 10 Apples 2 Oranges

Probability of selecting box = $P(E_1) = P(E_2) = \frac{1}{2}$

Select Apple from First box

$P(A/E_1) = \frac{8}{12} = \frac{2}{3}$

Select Apple from Second box

P(A/E_2) = $\frac{10}{12}=\frac{5}{6}$

Applying Bayes Theorem

$$P(E_1/A)=\frac{P(E_1)P(A/E_1)}{P(E_1)P(A/E_1)+P(E_2)P(A/E_2)}$$

$$= \frac{\frac{1}{2}x\frac{2}{3}}{(\frac{1}{2}x\frac{2}{3})+(\frac{1}{2}x\frac{5}{6})}$$

$$P(E_1/A)=\frac{4}{9}$$

==

Problem 3):

(2)

(1)

arameter, $\theta = \{\alpha, \mu_1, \Sigma_1, \mu_2, \Sigma_2\}$

$$P(y|\theta) = \alpha^y (1-\alpha)^{1-y}$$

$$P(x, y|\theta) = \mathcal{N}(x|\mu_y, \Sigma_y)$$

Using the parameter θ, we can generate the IID d
from :-

$$P(x|y, \theta) = P(y|\theta) P(x|y, \theta)$$

We will we maximum likelihood, to learns the param
from data

$$L(\theta) = \log P(\text{Data}|\theta)$$

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

According to the question there are two classes :-
Class 1 :- $y_i \in 0$
Class 2 :- $y_i \in 1$

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

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

$$= \sum_{i=1}^{N} \log P(y_i|\alpha) + \sum_{y \in 0} \log p(x_i|\mu_0 \Sigma_0) +$$

$$\sum_{y \in 1} \log p(x_i|\mu_1 \Sigma_1)$$

Estimating the likehood ~~for the object~~

$$P(y_i|\alpha) = \alpha^{y_i} (1-\alpha)^{1-y_i} \quad \text{—— } 1$$

$$P(x_i|\mu_0; \Sigma_0) = \frac{1}{2\pi^{D/2} (\sqrt{|\Sigma_0|})} \exp\left(\frac{-1}{2} (\vec{x} - \vec{\mu_0})^T \cdot \Sigma_0^{-1} (\vec{x} - \vec{\mu_0})\right)$$

$$P(x_i|\mu_1 \Sigma_1) = \frac{1}{2\pi^{D/2} (\sqrt{|\Sigma_1|})} \exp\left(\frac{-1}{2} (\vec{x} - \vec{\mu_1})^T \Sigma_1^{-1} (\vec{x} - \vec{\mu_1})\right)$$

Diff. eqn (1) wrt α

$$\frac{\partial}{\partial \alpha} (L(\theta)) = \frac{\partial}{\partial \alpha} \sum_{i=1}^{N} \log(\alpha^{y_i} (1-\alpha)^{1-y_i}) + \frac{\partial}{\partial \alpha}(\text{const}) = 0$$

$$\frac{\partial}{\partial \alpha} \sum_{i=1}^{N} (y_i \log \alpha + (1-y_i) \log(1-\alpha)) = 0$$

$$\frac{\partial}{\partial \alpha}\left(\sum_{i \in \text{class } 1} \log \alpha + \sum_{i \in \text{class } 0} \log(1-\alpha) \right) = 0$$

$$\sum_{i \in \text{class } 0} \frac{1}{\alpha} - \sum_{i \in \text{class } 1} \frac{1}{1-\alpha} = 0$$

$$\frac{N_1}{\alpha} - \frac{N_0}{1-\alpha} = 0$$

$$\frac{N_1}{\alpha} = \frac{N_0}{1-\alpha}$$

$$\alpha = \frac{N_1}{N_0 + N_1}$$

Differentiating the eqn wrt to μ_0

$$\frac{\partial(L(\theta))}{\partial \mu_0} = \frac{\partial}{\partial \mu_0}\left(\sum_{y_i \in 0} \log\left(\frac{1}{(2\pi)^{D/2} \sqrt{|\Sigma_0|}} - \exp\left(-\frac{1}{2}(\vec{x}_i - \vec{\mu}_0)^T \Sigma_0^{-1} (\vec{x}_i - \vec{\mu}_0)\right)\right.\right.$$

$$= \frac{\partial}{\partial \mu_0}\left(\sum_{y_i \in 0} \left(-\frac{D}{2}\log(2\pi) - \frac{1}{2}\log(\Sigma_0)\right) - \frac{1}{2}(\vec{x}_i - \vec{\mu}_0)^T \Sigma_0^{-1} (\vec{x}_i - \vec{\mu}_0)\right)$$

4

$$\frac{\partial}{\partial \mu_0} \sum_{y_i \in 0} \left[-\frac{1}{2} (\vec{x_i} - \vec{\mu_0})^T \Sigma_0^{-1} (\vec{x_i} - \vec{\mu_0}) \right] = 0$$

$$\sum_{y_i \in 0} (\vec{x_i} - \vec{\mu_0})^T \Sigma_0^{-1} = 0$$

$\Sigma_0^{-1} \neq 0$,

$$\sum_{y_i \in 0} (\vec{x} - \vec{\mu_0})^T = 0$$

$$\sum_{y_i \in 0} \vec{x} - N_0 \mu_0 = 0$$

$$\mu_0 = \frac{\sum_{y_i \in 0} \vec{x}}{N_0}$$

Similarly,

$$\mu_1 = \frac{\sum_{y_i \in 1} \vec{x}}{N_1}$$

Now, diff. wrt Σ_0^{-1}

$$\frac{\partial(L(\Theta))}{\partial \Sigma_0^{-1}} = \frac{\partial}{\partial \Sigma_0^{-1}} \sum_{y_i \in 0} \log P(x_i | \mu_0, \Sigma_0) + \frac{\partial}{\partial \Sigma_0^{-1}} (\text{const})$$

$$= \frac{\partial}{\partial \Sigma_0^{-1}} \sum_{y_i \in 0} \log P(x_i | \mu_0, \Sigma_0)$$

$$= \frac{\partial}{\partial \Sigma_0^{-1}} \left(\sum_{y_i \in 0} \log \left(\frac{1}{(2\pi)^{D/2} \sqrt{|\Sigma_0|}} \exp\left(-\frac{1}{2} (\vec{x}_i - \vec{\mu}_0)^T \Sigma_0^{-1} (\vec{x}_i - \vec{\mu}_0) \right) \right) \right) = 0$$

$$= \frac{\partial}{\partial \Sigma_0^{-1}} \left[\frac{N_0}{2} \log |\Sigma_0^{-1}| - \frac{1}{2} \sum_{y_i \in 0} -\frac{1}{2} (\vec{x}_i - \vec{\mu}_0)^T \Sigma_0^{-1} (\vec{x}_i - \vec{\mu}_0) \right]$$

$$= \frac{\partial}{\partial \Sigma_0^{-1}} \left[\frac{N_0}{2} \log |\Sigma_0^{-1}| - \frac{1}{2} \sum_{y \in 0} \mathrm{tr}\left[(\vec{x}_i - \vec{\mu}_0)^T \Sigma_0^{-1} (\vec{x}_i - \vec{\mu}_0) \right] \right]$$

Now, $\Sigma_0^{-1} = A$

Therefore we will differentiate the eqn wrt A

$$\frac{d}{dA}\left(\frac{N_0}{2}\log|A| - \frac{1}{2}\sum_{y_i \in 0} tr\left[(\vec{x}_i - \vec{\mu}_0)^T A (\vec{x}_i - \vec{\mu}_0)\right]\right) = 0$$

$$\frac{N_0}{2}(A^{-1})^T - \frac{1}{2}\sum_{y_i \in 0}\left[(\vec{x}_i - \vec{\mu}_0)(\vec{x}_i - \vec{\mu}_0)^T\right]^T = 0$$

$$\frac{N_0}{2}A^{-1} - \frac{1}{2}\sum_{y_i \in 0}(\vec{x}_i - \vec{\mu}_0)(\vec{x}_i - \vec{\mu}_0)^T = 0$$

~~Therefore when differentiating wrt~~

$$\frac{N_0}{2}\Sigma_0 - \frac{1}{2}\sum_{i, y_i \in 0}(\vec{x}_i - \vec{\mu}_0)(\vec{x}_i - \vec{\mu}_0)^T = 0$$

$$N_0(\Sigma_0) = \sum_{i, y_i \in 0}(\vec{x}_i - \vec{\mu}_0)(\vec{x}_i - \vec{\mu}_0)^T$$

$$\Sigma_0 = \frac{1}{N_0}\sum_{i, y_i \in 0}(\vec{x}_i - \vec{\mu}_0)(\vec{x}_i - \vec{\mu}_0)^T$$

Similarly,

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

Now, The given Bayes optimal decision is

$$y = \text{argmax}_{\hat{y} = \{0,1\}} P(\hat{y}|x)$$

For a linear decision Boundry of 2 class problem

$$P(y=1|x) = P(y=0|x) = 0.5$$

Using the conditional parameters we get :-

$$P(y=1|x) = \frac{P(x, y=1)}{P(x)}$$

$$= \frac{P(x, y=1) P(y=1)}{\sum_y P(x,y)}$$

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

(8)

$$= \frac{p(x|y=1)\,p(y=1)}{p(x|y=1)\,p(y=1) + p(x|y=0)\,p(y=0)}$$

$$p(y=1) = \alpha^1 (1-\alpha)^0$$
$$= \alpha$$

$$p(y=0) = \alpha^0 (1-\alpha)^{1-0}$$
$$= 1-\alpha$$

$\therefore$ $p(x|y=1) = N(x|\mu_1 \Sigma_1)$

$$p(y=1|x) = \frac{\alpha N(x|\mu_1 \Sigma_1)}{\alpha(N(x|\mu_1 \Sigma_1) + (1-\alpha) N(x|\mu_0, \Sigma_0))}$$

$$p(y=0|x) = \frac{p(x|y=0) \cdot p(y=0)}{p(x|y=0)\,p(y=1) + p(x|y=0)\,p(y=0)}$$

$$= \frac{(1-\alpha) N(x|\mu_0 \Sigma_0)}{\alpha N(x|\mu_1 \Sigma_1) + (1-\alpha) N(x|\mu_0, \Sigma_0)}$$

$\therefore$ To calculate decision boundry

$$p(y=1|x) = p(y=0|x)$$

$\therefore (1-\alpha)\, N(x \mid \mu_0, \Sigma_0) = \alpha\,(N(x \mid \mu_0, \Sigma_0))$ 1

Sub. N we get

$$\frac{\alpha}{(2\pi)^{D/2}\sqrt{|\Sigma|}} \exp\left(-\frac{1}{2}(\vec{x}-\vec{\mu}_1)\Sigma_1^{-1}(\vec{x}-\vec{\mu}_1)^T\right) =$$

$$\frac{(1-\alpha)}{(2\pi)^{D/2}\sqrt{|\Sigma|}} \exp\left((\vec{x}-\vec{\mu}_0)\Sigma_0(\vec{x}_1-\vec{\mu}_1)^T\right)$$

~~by~~ simplifying

$$-\frac{1}{2}(\vec{x}-\vec{\mu}_1)\Sigma_1^{-1}(\vec{x}-\vec{\mu}_1)^T + \frac{1}{2}(\vec{x}-\vec{\mu}_0)\Sigma_0^{-1}(\vec{x}-\vec{\mu}_0)^T$$
$$= \log\left(\frac{1-\alpha}{\alpha}\sqrt{\frac{|\Sigma_1|}{|\Sigma_0|}}\right)$$

Now, for linear classification

$\Sigma_0, \Sigma_1 = \Sigma$ and $\alpha = \frac{N_1}{N_0+N_1}$

$$-\frac{1}{2}(\vec{x}-\vec{\mu}_1)\Sigma^{-1}(\vec{x}-\vec{\mu}_1)^T + \frac{1}{2}(\vec{x}-\vec{\mu}_0)\Sigma^{-1}(\vec{x}-\vec{\mu}_0)$$
$$= \log\left(\frac{N_0}{N_1}\right)$$

$$\Sigma^{-1}(\mu_1-\mu_0)^T\vec{x} + \frac{1}{2}(\mu_0+\mu_1)^T\vec{\Sigma}(\mu_0-\mu_1) +$$

$$\log\left(\frac{N_1}{N_0}\right) = 0$$

$$w = \Sigma(\mu_1-\mu_0)$$

$$b = \frac{1}{2}(\mu_0+\mu_1)^T\Sigma^{-1}(\mu_0+\mu_1)^{-1}\log\frac{N_1}{N_0}$$

$$f(x) = \mathrm{sign}(w^T x + b)$$

If, $\Sigma_0 \neq \Sigma_1$

$$-\frac{1}{2}(\vec{x}-\vec{\mu_1})^T\Sigma_1^{-1}(\vec{x}-\vec{\mu_1}) +$$

$$\frac{1}{2}(\vec{x}-\mu_0)^T\Sigma_0^{-1}(\vec{x}-\vec{\mu_0}) + \frac{1}{2}\log\frac{\Sigma_0}{\Sigma_1} +$$

$$\log\left(\frac{N_1}{N_0}\right) = 0$$