Problem 1

Use the matlab code

```
imagesc(reshape(teapotImages(i,:),38,50));
```

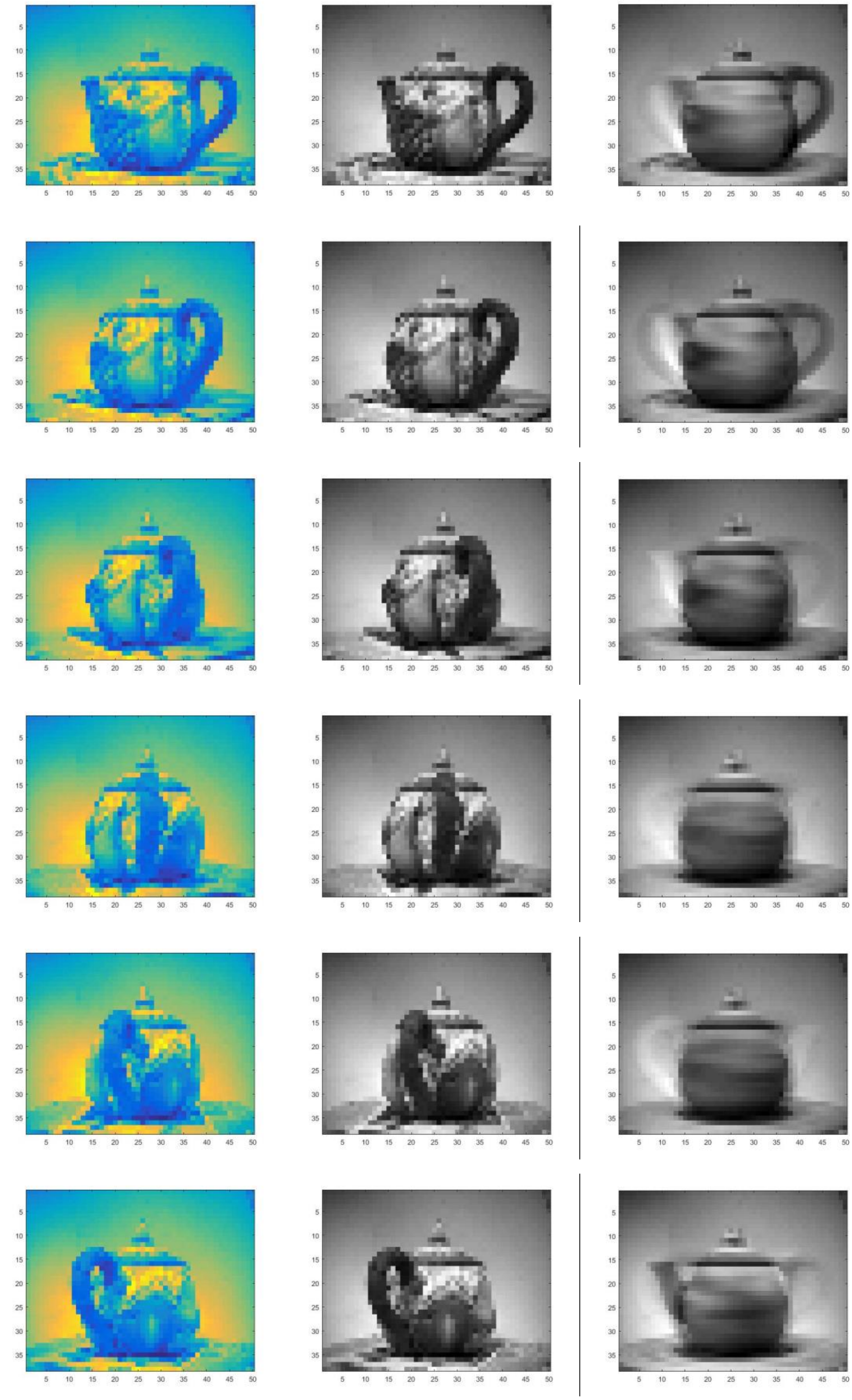
View the i = 5th, 10th, 15th, 20th, 25th, 30th, 35th, 40th, 45th and 50th images in the original data set (a total of ten images). The images are shown in the left column below.
Then map the color of these ten pictures into gray, the images are shown in the middle column below, and the matlab code is as follows:

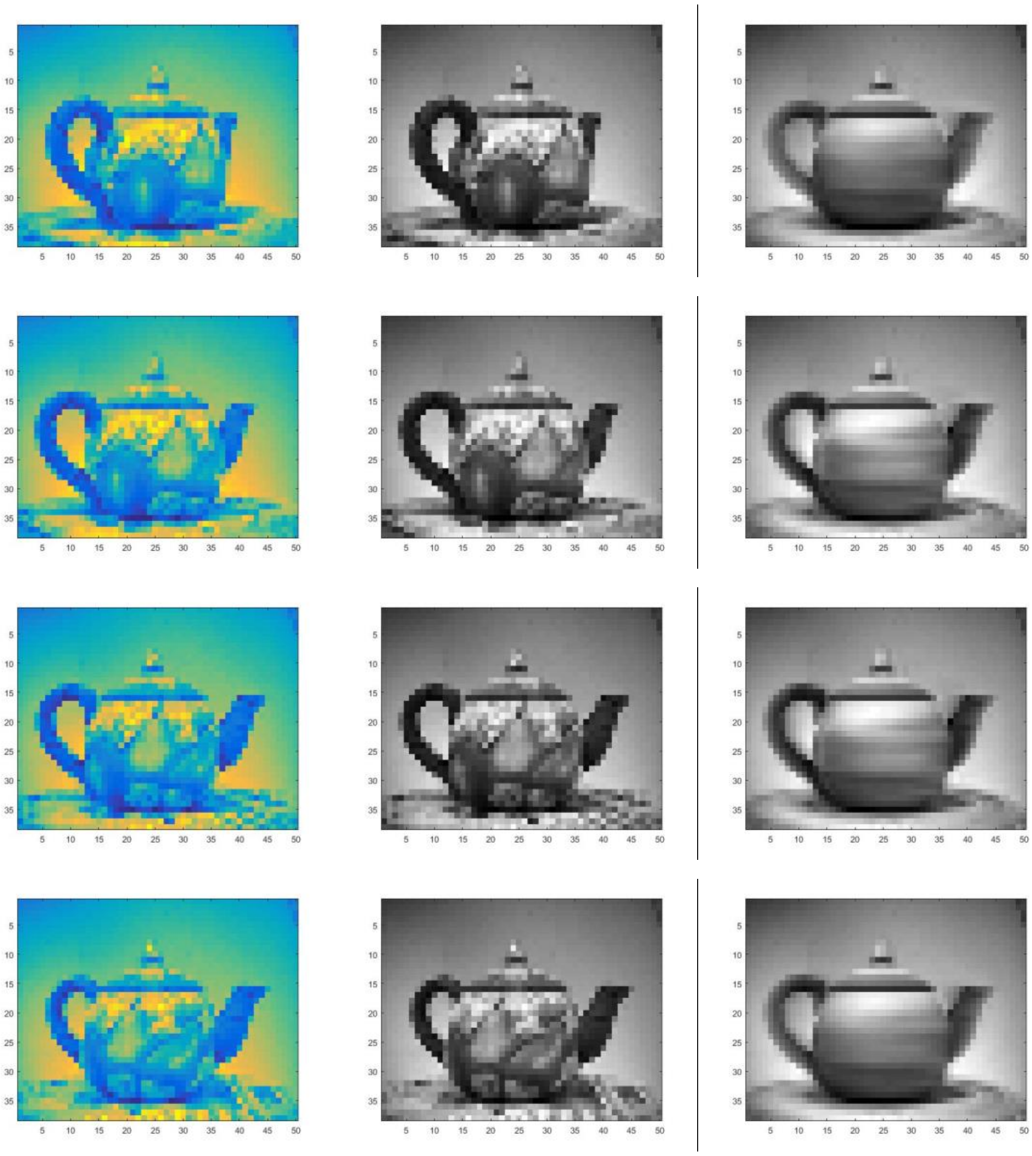
```
for i=5:5:50
figure;
colormap gray;
imagesc(reshape(teapotImages(i,:),38,50));
end
```

Firstly, each two-dimensional image in the image training library is stretched into a vector, and then the principal component analysis is carried out to obtain the transformation matrix of the principal component and the image mean vector. The compressed image is added with the image mean vector after the inverse transformation of the image matrix to obtain the approximate vector of the vector before compression. The image is displayed at the bottom, which shows that there is less information loss in the image. The images are shown in the right column below and the matlab code is as follows:

```
load('teapots.mat')
X = teapotImages;
m = mean(X,1);
X = X - ones(size(X,1),1)*m;
Cov = X'*X/size(X,1);
[E, D] = eig(Cov);
d = diag(D);
[dum,ord] = sort(-d);
E = E(:,ord);
d = d(ord);
C = X*E;
E = E';
for i=5:5:50
figure;
colormap gray;
imagesc(reshape(m+C(i,1:n)*E(1:n,:),38,50));
end
```

Here are the images:





Problem 2

Box 1	8 apples	4 oranges
Box 2	10 apples	2 oranges

Probability of selecting Box1 and Box2: $P(B_1) = P(B_2) = \frac{1}{2}$

Set A as the event: select apple from a box

$$P(A \mid B_1) = \frac{8}{12} = \frac{2}{3}$$

$$P(A \mid B_2) = \frac{10}{12} = \frac{5}{6}$$

Apply Baye's rule: $P(B_1 \mid A) = \dfrac{P(A \mid B_1)P(B_1)}{P(A \mid B_1)P(B_1) + P(A \mid B_2)P(B_2)}$

$$= \frac{\frac{1}{2} \times \frac{2}{3}}{\frac{1}{2} \times \frac{2}{3} + \frac{1}{2} \times \frac{5}{6}}$$

$$= \frac{4}{9}$$

Problem 3

Have two classes, each with their own Gaussian:

$\{(x_1, y_1), \cdots, (x_N, y_N)\} \quad x \in R^D \;\; \mathrm{y} \in \{1, 2\}$

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

We can generate iid data from $p(x, y \mid \theta) = p(y \mid \theta)\, p(x \mid y, \theta)$ by:

1) flipping a coin to get y via Bernoulli $p(y \mid \theta) = \alpha^y (1-\alpha)^{2-y}$

2) sampling an x form y'th Gaussian $p(x \mid y, \theta) = N(x \mid \mu_y, \Sigma_y)$

Recover parameters from data using maximum likelihood:

$$l(\theta) = \log p(data \mid \theta) = \sum_{i=1}^{N} \log p(x_i, y_i \mid \theta)$$
$$= \sum_{i=1}^{N} \log p(y_i \mid \theta) + \sum_{i=1}^{N} \log p(x_i \mid y_i, \theta)$$
$$= \sum_{i=1}^{N} \log p(y_i \mid \alpha) + \sum_{y_i \in 1} \log p(x_i \mid \mu_1, \Sigma_1) + \sum_{y_i \in 2} \log p(x_i \mid \mu_2, \Sigma_2)$$

Max likelihood can be done separately for the 3 terms

$$l = \sum_{i=1}^{N} \log p(y_i \mid \alpha) + \sum_{y_i \in 1} \log p(x_i \mid \mu_1, \Sigma_1) + \sum_{y_i \in 2} \log p(x_i \mid \mu_2, \Sigma_2)$$

Count # of pos & neg examples (class prior): $\alpha = \dfrac{N_2}{N_1 + N_2}$

Get mean & cov of negatives and mean & cov of positives:

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

$$\mu_2 = \frac{1}{N_2} \sum_{y_i \in 2} x_i \qquad \Sigma_2 = \frac{1}{N_2} \sum_{y_i \in 2} (x_i - \mu_2)(x_i - \mu_2)^T$$

Bayes Optimal Decision: $\hat{y} = \arg\max_{y=\{1,2\}} p(y \mid x)$

$$p(y \mid x) = \frac{p(x, y)}{p(x)} = \frac{p(x, y)}{\sum_y p(x, y)} = \frac{p(x, y)}{p(x, y=1) + p(x, y=2)}$$

Example case: $y = 2$, plotting decision boundary when $= 0.5$

$$p(y=2 \mid x) = \frac{p(x, y=2)}{p(x, y=1) + p(x, y=2)} = \frac{\alpha N(x \mid \mu_2, \Sigma_2)}{(1-\alpha) N(x \mid \mu_1, \Sigma_1) + \alpha N(x \mid \mu_2, \Sigma_2)} = 0.5$$

If covariances are equal: $\Sigma_2 = \Sigma_1 = \Sigma$

$$\alpha N(x|\mu_2,\Sigma_2) = (1-\alpha)N(x|\mu_1,\Sigma_1)$$

$$\alpha \cdot \frac{1}{(2\pi)^{\frac{D}{2}}\sqrt{|\Sigma|}}\exp(-\frac{1}{2}(x-\mu_2)^T\Sigma^{-1}(x-\mu_2)) = (1-\alpha)\cdot\frac{1}{(2\pi)^{\frac{D}{2}}\sqrt{|\Sigma|}}\exp(-\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1))$$

$$\frac{\exp(-\frac{1}{2}(x-\mu_2)^T\Sigma^{-1}(x-\mu_2))}{\exp(-\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1))} = \frac{1-\alpha}{\alpha} = C$$

$$-\frac{1}{2}(x-\mu_2)^T\Sigma^{-1}(x-\mu_2)+\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)$$

$$-\frac{1}{2}(x^T\Sigma^{-1}-\mu_2^T\Sigma^{-1})(x-\mu_2)+\frac{1}{2}(x^T\Sigma^{-1}-\mu_1^T\Sigma^{-1})(x-\mu_1)$$

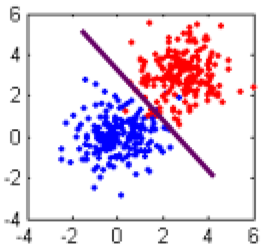
$$-\frac{1}{2}(x^T\Sigma^{-1}x-x^T\Sigma^{-1}\mu_2-\mu_2^T\Sigma^{-1}x+\mu_2^T\Sigma^{-1}\mu_2)+\frac{1}{2}(x^T\Sigma^{-1}x-x^T\Sigma^{-1}\mu_1-\mu_1^T\Sigma^{-1}x+\mu_1^T\Sigma^{-1}\mu_1)$$

and we know that $x^T\Sigma^{-1}\mu = \mu^T\Sigma^{-1}x$

$$-\frac{1}{2}(x^T\Sigma^{-1}x-2x^T\Sigma^{-1}\mu_2+\mu_2^T\Sigma^{-1}\mu_2)+\frac{1}{2}(x^T\Sigma^{-1}x-2x^T\Sigma^{-1}\mu_1+\mu_1^T\Sigma^{-1}\mu_1)$$

$$x^T\Sigma^{-1}\mu_2-x^T\Sigma^{-1}\mu_1+\frac{1}{2}\mu_1^T\Sigma^{-1}\mu_1-\frac{1}{2}\mu_2^T\Sigma^{-1}\mu_2$$

In this polynomial, we can see that the exponent of x is 1, so it's a linear decision.



If covariances are equal: $\Sigma_2 \neq \Sigma_1$

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

The polynomial will be like:

$$-\frac{1}{2}(x^T\Sigma_2^{-1}x-2x^T\Sigma_2^{-1}\mu_2+\mu_2^T\Sigma_2^{-1}\mu_2)+\frac{1}{2}(x^T\Sigma_1^{-1}x-2x^T\Sigma_1^{-1}\mu_1+\mu_1^T\Sigma_1^{-1}\mu_1)$$

In this polynomial, we can see that the exponent of x is 2, so it's a quadratic decision.