CSCI 5521 : INTRO TO MACHINE LEARNING

HOMEWORK - 2

(5202540) UJVAL BANGALORE UMESH (banga038)

Q 1)

(a) Given priors $P(C_1)$ and $P(C_2)$ and Bernoulli densities $p_1 = P(x=0|C_1)$ and $p_2 = P(x=0|C_2)$

Hence $1-p_1 = P(x=1|C_1)$ and $1-p_2 = P(x=1|C_2)$

Using Baye's theorem

$$P(C_1|x=0) = \frac{P(x=0|C_1) \times P(C_1)}{P(x=0)}$$

$$= \frac{P(x=0|C_1) \times P(C_1)}{P(x=0|C_1) \times P(C_1) + P(x=0|C_2) \times P(C_2)}$$

$$= \frac{p_1 \times P(C_1)}{p_1 \times P(C_1) + p_2 \times P(C_2)}$$

^{11}ly

$$P(C_2|x=0) = \frac{P(x=0|C_2) \times P(C_2)}{P(x=0)}$$

$$= \frac{P(x=0|C_2) \times P(C_2)}{P(x=0|C_1) \times P(C_1) + P(x=0|C_2) \times P(C_2)}$$

$$= \frac{p_2 \times P(C_2)}{p_1 \times P(C_1) + p_2 \times P(C_2)}$$

For

$$P(C_1|x=1) = \frac{P(x=1|C_1) \times P(C_1)}{P(x=1|C_1) \times P(C_1) + P(x=1|C_2) \times P(C_2)}$$

$$= \frac{(1-P_1) \times P(C_1)}{(1-P_1)\, P(C_1) + (1-P_2)\, P(C_2)}$$

$$P(C_2|x=1) = \frac{P(x=1|C_2) \times P(C_2)}{P(x=1|C_1) \times P(C_1) + P(x=1|C_2) \times P(C_2)}$$

$$= \frac{(1-P_2) \times P(C_2)}{(1-P_1) \times P(C_1) + (1-P_2)\, P(C_2)}$$

Q
1(b) Given

$P_{ij} = P(x_j = 0|C_i)$ for D-dimensional independent Bernoulli densities for $i=1,2$ classes and $j=1,2,\ldots D$ dimensions

As per Bayes theorem (eqn), on generalisation from Q1(a)

we get $$P(C_i|x) = \frac{P(x|C_i) \times P(C_i)}{\sum_{i=1}^{K} P(x|C_i) \times P(C_i)}$$

In this case, 2 classes

$$P(C_1|x) = \frac{P(x|C_1) \times P(C_1)}{P(x|C_1)\, P(C_1) + P(x|C_2)\, P(C_2)}$$

For $j = 1$ to D, we have For class 1

$$P(C_1|x) = \frac{\left(\prod_{j=1}^{D} P_{1j}^{(1-x_j)} (1-P_{1j})^{x_j}\right) \times P(C_1)}{P(C_1) \times \left[\prod_{j=1}^{D} P_{1j}^{(1-x_j)} (1-P_{1j})^{x_j}\right] + P(C_2) \times \left[\prod_{j=1}^{D} (1-P_{2j})^{x_j} P_{2j}^{(1-x_j)}\right]}$$

For class 2,

$$P(C_2|x) = \frac{\left[\prod_{j=1}^{D} P_{2j}^{(1-x_j)} (1-P_{2j})^{x_j}\right] \times P(C_2)}{P(C_1) \times \left[\prod_{j=1}^{D} P_{1j}^{(1-x_j)} (1-P_{1j})^{x_j}\right] + P(C_2) \times \left[\prod_{j=1}^{D} (1-P_{2j})^{x_j} P_{2j}^{(1-x_j)}\right]}$$

Q 1(c) We will have 12 cases.

Given:

$D = 2$, $P_{11} = 0.6$, $P_{12} = 0.1$, $P_{21} = 0.6$, $P_{22} = 0.9$ and

priors are $P(C_1) = 0.2, 0.6, 0.8$ & $P(C_2) = 1 - P(C_1)$

$P(C_2) = 0.8, 0.4, 0.2$

(i) $(x_1, x_2) = \{0, 0\}$ and $P(C_1) = 0.2$ & $P(C_2) = 0.8$

$$P(C_1|x) = \frac{P(C_1) \times P(x|C_1)}{P(x)} = \frac{P(C_1) \times P(x|C_1)}{P(C_1) \times P(x|C_1) + P(C_2) \times P(x|C_2)}$$

$$= \frac{0.2 \times P_{11}^{(1-x_1)} (1-P_{11})^{x_1} \times P_{12}^{(1-x_2)} (1-P_{12})^{x_2}}{0.2 \times P_{11}^{(1-x_1)} (1-P_{11})^{x_1} \times P_{12}^{(1-x_2)} (1-P_{12})^{x_2} + 0.8 \times \left((1-P_{21})^{x_1} P_{21}^{(1-x_1)} \times (1-P_{22})^{x_2} P_{22}^{(1-x_2)}\right)} \quad -(1)$$

subs in (1),

$$P(C_1|x) = \frac{0.012}{0.444} = \frac{3}{111} \qquad P(C_2|x) = 1 - P(C_1|x) = 1 - \frac{3}{111} = \frac{108}{111} \quad (2)$$

(ii) $(x_1, x_2) = \{0,0\}$, $P(C_1) = 0.6$ & $P(C_2) = 0.4$

$$P(C_1|x) = \frac{P(C_1) \times P(x|C_1)}{P(C_1) \times P(x|C_1) + P(C_2) \times P(x|C_2)}$$

$$= \frac{0.6 \times P_{11}^{(1-x_1)}(1-P_{11})^{x_1} \times P_{12}^{(1-x_2)}(1-P_{12})^{x_2}}{0.6 \times P_{11}^{(1-x_1)}(1-P_{11})^{x_1} \times P_{12}^{(1-x_2)}(1-P_{12})^{x_2} + 0.4 \times (1-P_{21})^{x_1} P_{21}^{(1-x_1)} \times (1-P_{22})^{x_2} P_{22}^{(1-x_2)}}$$

$$P(C_1|x) = \frac{0.036}{0.036 + 0.4 \times 0.54} = \frac{36}{252} = \frac{3}{21} = \frac{1}{7}$$

$$P(C_2|x) = 1 - \frac{3}{21} = \frac{18}{21} = \frac{6}{7}.$$

(iii) $(x_1 x_2) = \{0,0\}$; $P(C_1) = 0.8$, $P(C_2) = 0.2$

$$P(C_1|x) = \frac{P(C_1) \times P(x|C_1)}{P(C_1) \times P(x|C_1) + P(C_2) \times P(x|C_2)}$$

$$P(C_1|x) = \frac{0.6 \times P_{11}^{(1-x_1)}(1-P_{11})^{x_1} \times P_{12}^{(1-x_2)}(1-P_{12})^{x_2}}{0.8 \times P_{11}^{(1-x_1)}(1-P_{11})^{x_1} \times P_{12}^{(1-x_2)}(1-P_{12})^{x_2} + 0.2 (1-P_{21})^{x_1} \times P_{21}^{(1-x_2)} (1-P_{22})^{x_2} P_{22}^{(1-x_2)}}$$

$$P(C_1|x) = \frac{0.8 \times 0.06}{0.156} = \frac{4}{13} \qquad P(C_2|x) = 1 - \frac{4}{13} = \frac{9}{13}$$

(iv) $(x_1, x_2) = \{0, 1\}$ $\quad P(C_1) = 0.2 \quad P(C_2) = 0.8$.

For depiction, The numerator is written as Num & denominator = Num + other [class].

$$\text{Num} = P(C_1) \times P_{11}^{(1-x_1)}(1-P_{11})^{x_1} \times P_{12}^{(1-x_2)}(1-P_{12})^{x_2}$$

$$= 0.2 \times 0.54 = 0.108.$$

$$\text{other} = P(C_2) \times \{(1-P_{21})^{x_1} P_{21}^{(1-x_1)} \times P_{22}^{(1-x_2)}(1-P_{22})^{x_2}\}$$

$$= 0.8 \times 0.06 = 0.048.$$

$$P(C_1|x) = \frac{\text{Num}}{\text{Num+other}} = \frac{0.108}{0.156} = \frac{108}{156} = \frac{9}{13}$$

$$P(C_2|x) = 1 - 9/13 = 4/13$$

(v) when $(x_1, x_2) = \{0, 1\}$ $\quad P(C_1) = 0.6$, $P(C_2) = 0.4$

$$\text{Num} = P(C_1) \times P_{11}^{(1-x_1)}(1-P_{11})^{x_1} \times P_{12}^{(1-x_2)}(1-P_{12})^{x_2}$$

$$= 0.6 \times 0.6 \times 0.9 = 0.324$$

$$\text{other} = P(C_2) \times \{(1-P_{21})^{x_1} P_{21}^{(1-x_1)} \times P_{22}^{(1-x_2)}(1-P_{22})^{x_1}\}$$

$$= 0.4 \times 0.6 \times 0.1 = 0.024$$

$$P(C_1|x) = \frac{\text{Num}}{\text{Num+other}} = \frac{0.324}{0.348} = \frac{81}{87}$$

$$P(C_2|x) = \frac{6}{87}$$

Scanned by CamScanner

(VI) $(x_1, x_2) = (0,1)$; $P(C_1) = 0.8$ $P(C_2) = 0.2$

||ly from previous Num = $P(C_1) \times P_{11}^{(1-x_1)} (1-P_{12})^{x_2}$

Num =

$= 0.8 \times 0.6 \times 0.9$

$= 0.432$

other $= 0.012$

$P(C_1|x) = \dfrac{0.432}{0.432+0.012} = \dfrac{108}{111}$ $P(C_2|x) = 1 - \dfrac{108}{111} = \dfrac{3}{111}$

(VII) $(x_1, x_2) = \{1, 0\}$, $P(C_1) = 0.2$ $P(C_2) = 0.8$

Num = $P(C_1) \times P_{11}^{(1-x_1)} (1-P_{11})^{x_1} \times P_{12}^{(1-x_2)} (1-P_{12})^{x_2}$

$= 0.2 \times 0.04 = 0.008$

other $= P(C_2) \times \{ (1-P_{21})^{(x_1)} P_{21}^{(1-x_1)} \times (1-P_{22})^{x_2} P_{22}^{(1-x_2)})$

$= 0.8 \times 0.36 = 0.288$

$P(C_1|x) = \dfrac{0.008}{0.296} = \dfrac{1}{37}$

$P(C_2|x) = \dfrac{36}{37}$

(viii) $(x_1, x_2) = \{1, 0\}$ $P(C_1) = 0.6$ $P(C_2) = 0.4$

From previous

Num $= P(C_1) \times (1-p_{11})^{x_1} \, p_{12}^{(1-x_2)}$

$= 0.6 \times 0.4 \times 0.1 = 0.024$

other $= P(C_2) \times (1-p_{21})^{x_1} \, p_{22}^{(1-x_2)}$

$= 0.4 \times 0.36 = 0.144$

$$P(C_1/x) = \frac{0.024}{0.168} = \frac{1}{7}$$

$$P(C_2/x) = 1 - \frac{1}{7} = \frac{6}{7}$$

(ix) when $(x_1, x_2) = \{1, 0\}$, $P(C_1) = 0.8$, $P(C_2) = 0.2$

Num $= 0.8 \times 0.4 \times 0.1 = 0.032$

other $= 0.2 \times 0.4 \times 0.9 = 0.072$

$$P(C_1/x) = \frac{0.032}{0.104} = \frac{4}{13}$$

$$P(C_2/x) = 1 - \frac{4}{13} = \frac{9}{13}$$

(x) $(x_1, x_2) = (1,1)$, $P(C_1) = 0.2$, $P(C_2) = 0.8$

Num = $P(C_1) \times P_{11}^{(1-x_1)} (1-P_{11})^{x_1} \times P_{12}^{(1-x_2)} \times (1-P_{12})^{x_2}$

$= P(C_1) \times (1-P_{11})^{x_1} \times (1-P_{12})^{x_2}$

$= 0.2 \times 0.4 \times 0.9 = 0.072$

other = $P(C_2) \times \{ (1-P_{21})^{x_1} P_{21}^{(1-x_1)} (1-P_{22})^{x_2} P_{22}^{1-x_2} \}$

$= P(C_2) \times (1-P_{21}) \times (1-P_{22})$

$= 0.8 \times 0.4 \times 0.1 = 0.032$

$P(C_1|x) = \dfrac{0.072}{0.104} = \dfrac{9}{13}$

$P(C_2|x) = 1 - \dfrac{9}{13} = \dfrac{4}{13}$

(xi) $(x_1, x_2) = \{1,1\}$ $P(C_1) = 0.6$, $P(C_2) = 0.4$

Num = $0.6 \times 0.36 = 0.216$

other = $0.4 \times 0.04 = 0.016$

$P(C_1|x) = \dfrac{0.216}{0.232} = \dfrac{27}{29}$

$P(C_2|x) = 1 - \dfrac{27}{29} = \dfrac{2}{29}$

(XII) $(x_1, x_2) = \{1, 1\}$ $P(C_1) = 0.8$ $P(C_2) = 0.2$

Num = $0.8 \times 0.36 = 0.288$

Other = $0.2 \times 0.04 = 0.008$

$$P(C_1|x) = \frac{0.288}{0.296} = \frac{36}{37}$$

$$P(C_2|x) = \frac{1}{37}$$

Report Homework-2

Ujval Bangalore Umesh (5202540)

Question 1

d)

Bayes_learning.m: Bayes_Learning(training data , validation data function returns the outputs p1: learned Bernoulli parameters of the first class, p2: learned Bernoulli parameters of the second class, pc1: best prior of the first class, pc2: best prior of the second class.

The validation error rates for sigma [-5:1:5] are:

'validation error for sigma = -5 is 0.235955'

'validation error for sigma = -4 is 0.202247'

'validation error for sigma = -3 is 0.224719'

'validation error for sigma = -2 is 0.213483'

'validation error for sigma = -1 is 0.235955'

'validation error for sigma = 0 is 0.280899

'validation error for sigma = 1 is 0.280899'

'validation error for sigma = 2 is 0.325843'

'validation error for sigma = 3 is 0.325843'

'validation error for sigma = 4 is 0.325843'

'validation error for sigma = 5 is 0.314607'

Bayes_Testing: The function takes in learned parameters p1,p2,pC1 and pC2 and outputs the error rate on the test set.

The error rate on test set is 0.146067.

Question 2

a) The KNN function outputs the error rate on the test set for k=1,3,5 and 7.

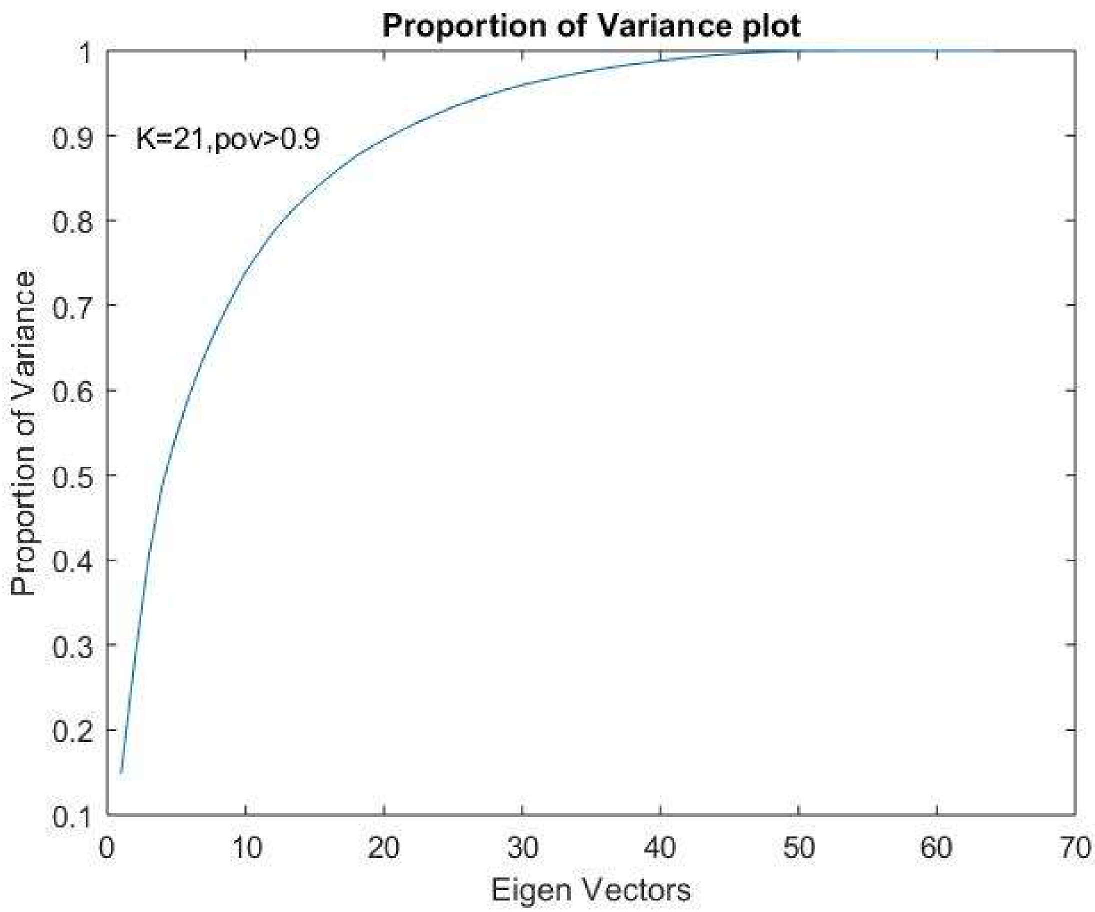
'The error rate for k =1 is 0.053872'

'The error rate for k =3 is 0.040404'

'The error rate for k =5 is 0.043771'

'The error rate for k =7 is 0.053872'

b) The PCA is implemented and a plot of proportion of variance(pov) is generated. It is depicted below:

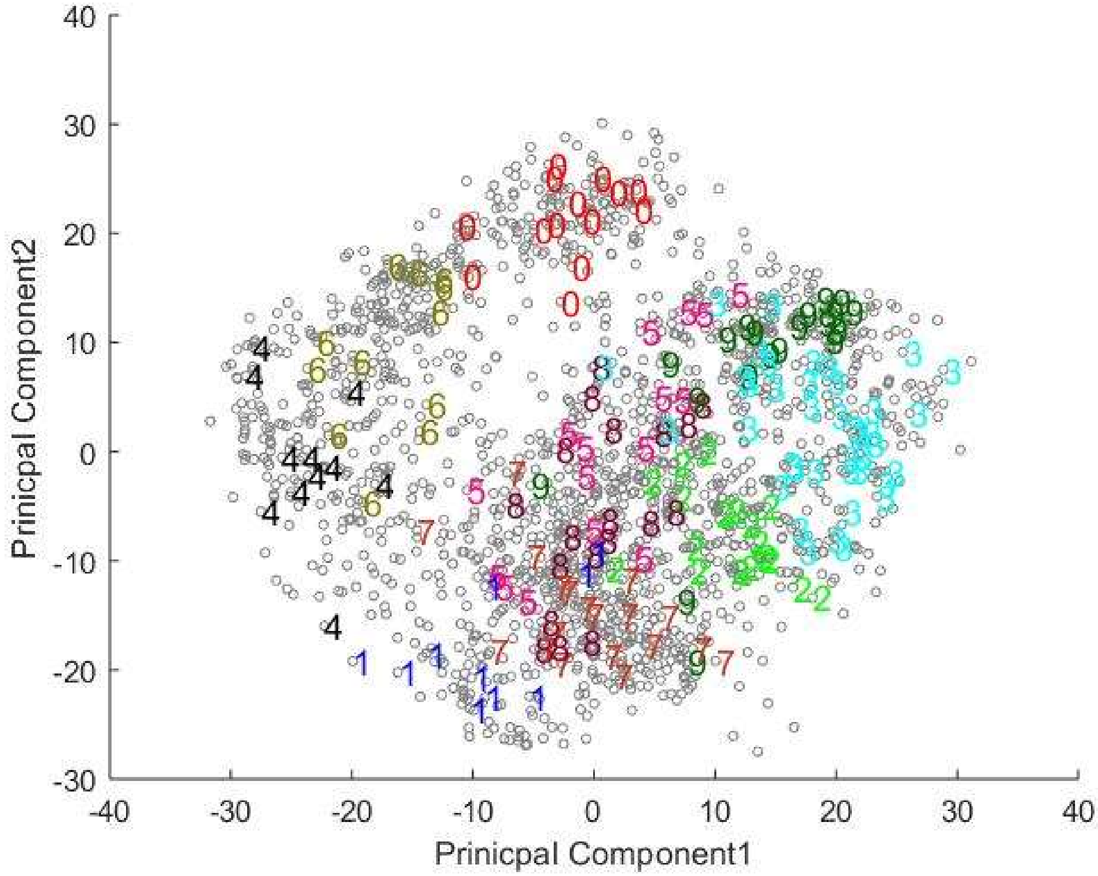


K is found to be equal to 21 which can explain more than 90% of the variance of the training data.

The error rate on the test set for different values of k are :

'The error rate for k =1 is 0.043771'
'The error rate for k =3 is 0.040404'
'The error rate for k =5 is 0.043771'
'The error rate for k =7 is 0.040404'

c) Plot is shown for the projection of training and test data in R^2 using the first two principal components.



d) The error on the test data for each combination of L and k are :

'The error rate for L=2, k=1 is 0.447811'

'The error rate for L=2, k=3 is 0.414141'

'The error rate for L=2, k=5 is 0.407407'

'The error rate for L=4, k=1 is 0.191919'

'The error rate for L=4, k=3 is 0.185185'

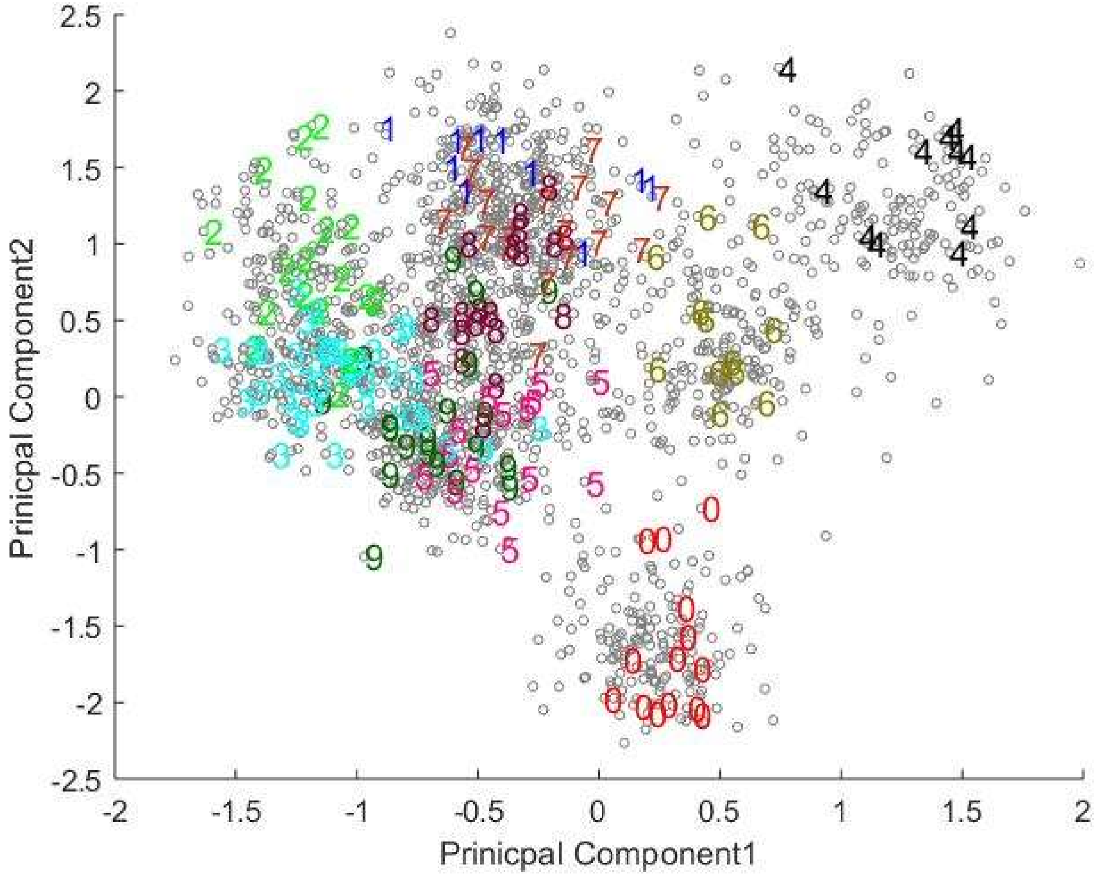
'The error rate for L=4, k=5 is 0.158249'

'The error rate for L=9, k=1 is 0.097643'

'The error rate for L=9, k=3 is 0.094276'

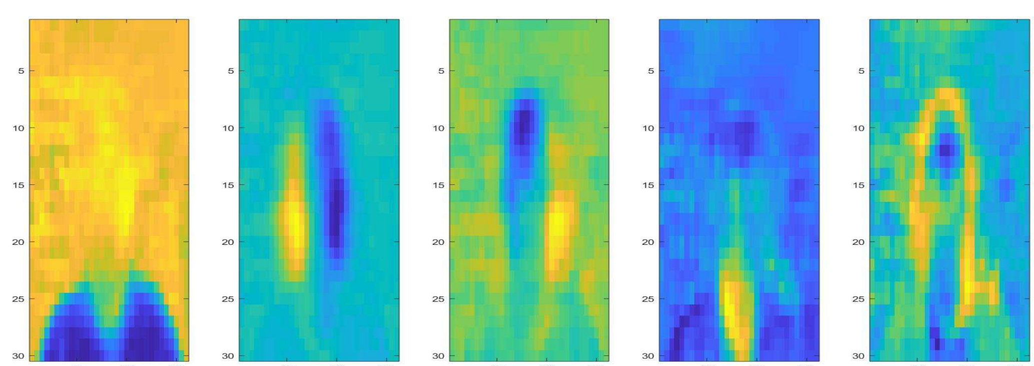
'The error rate for L=9, k=5 is 0.094276'

e) Plot is shown for the projection of training and test data in R^2 using the first two principal components.

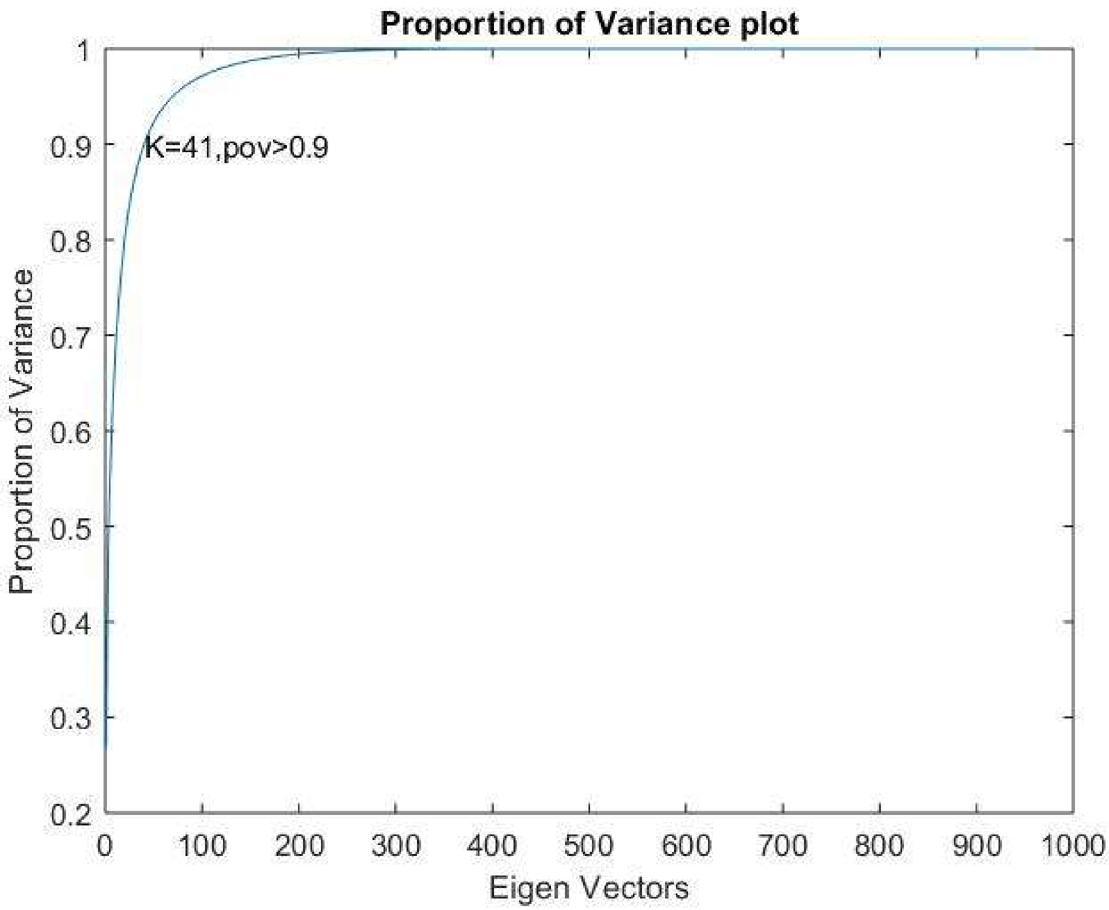


Question 3

a) Plot of the first 5 Eigen-faces of the combination of training and test data



b) The PCA is implemented and a plot of proportion of variance(pov) is generated. It is depicted below:

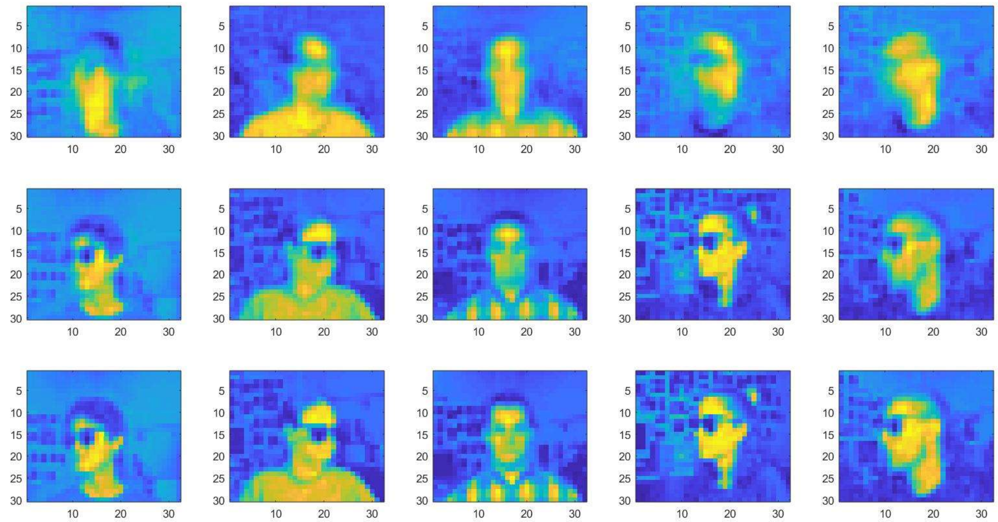


K is found to be equal to 41 which can explain more than 90% of the variance of the training data.

The error rate on the test set for different values of k are:

'The error rate for k= 1 is 0.104839'
'The error rate for k= 3 is 0.241935'
'The error rate for k= 5 is 0.395161'
'The error rate for k= 7 is 0.395161'

c)



The first row corresponds to K=10. The second row corresponds to K=50 and the third row corresponds to K=100.

The Contribution of each Eigen vector is given by its Eigen Value. From the pov, It is evident that K=41 Principal components explains 90% of the variance of the data.

For K=10, we can visually see that the visibility is poor since the 10 principal components cannot explain the variance of the image when projected. Hence, we cannot clearly distinguish whether the person in the image has sunglasses or not.

For K=50, we can see the image more clearly and can distinguish the presence of sunglasses. This is because for K=41 Principal components can explain more than 90% of the variance and K=50 could explain even more than that.

For K=100, we can see that there is not much difference between K=50 and K=100 since most of the variance is captured within 41 Principal Components.