CS 5565, ECE 5590CI, CS 465R, HW4(LDA and Resampling) 80 pts. (50 pts. UG)

Name : Samhitha Tummanapalli
Student ID: 16246162

1. (a) Use LDA to build a classier for the following data. To get full credit you must show all work and all discriminant values.

1>

(a) grouping them to classes 1 and 0.

X_1	Y_1	Y	
4	1	0	
-11	-4	0	
8	1	0	class 0
-6	-3	0	
0	-7	0	
-3	5	0	
0	3	0	
6	3	1	
-3	7	1	
4	12	1	
10	11	1	class 1
5	5	1	
-7	8	1	
13	8	1	
5	8	1	

Solution

$$X_1 = \begin{bmatrix} 4 & 1 \\ -11 & -4 \\ 8 & 1 \\ -6 & -3 \\ 0 & -7 \\ -3 & 5 \\ 0 & 3 \end{bmatrix}$$

mean $\mu_1 = [-1.1428 \quad -0.5714]$

$\underline{\underline{X_2}}$

$$X_2 = \begin{bmatrix} 6 & 3 \\ -3 & 7 \\ 4 & 12 \\ 10 & 11 \\ 5 & 5 \\ -7 & 8 \\ 13 & 8 \\ 5 & 8 \end{bmatrix} \quad \text{mean } \mu_2 = [4.125 \quad 7.75]$$

g = number of groups in y in our case it is 2 [class 0 & class 1]

μ = it is mean of whole set = $[1.4911 \quad 3.5893] \simeq [1.5 \quad 3.6]$

$$X_1^0 = \begin{bmatrix} \frac{40-15}{10} & \frac{10-15}{10} \\ \frac{-110-15}{10} & \frac{-40-15}{10} \\ \frac{80-15}{10} & \frac{10-15}{10} \\ \frac{-60-15}{10} & \frac{-30-15}{10} \\ \frac{-15}{10} & \frac{-70-15}{10} \\ \frac{-30-15}{10} & \frac{50-15}{10} \\ \frac{-15}{10} & \frac{30-15}{10} \end{bmatrix} = \begin{bmatrix} 2.5 & -0.5 \\ -12.5 & -5.5 \\ 6.5 & -0.5 \\ -7.5 & -4.5 \\ -1.5 & -8.5 \\ -4.5 & 3.5 \\ -1.5 & 1.5 \end{bmatrix}$$

$$X_2^0 = \begin{bmatrix} \frac{60-36}{10} & \frac{30-36}{10} \\ \frac{-30-36}{10} & \frac{70-36}{10} \\ \frac{40-36}{10} & \frac{120-36}{10} \\ \frac{100-36}{10} & \frac{110-36}{10} \\ \frac{50-36}{10} & \frac{50-36}{10} \\ \frac{-70-36}{10} & \frac{80-36}{10} \\ \frac{130-36}{10} & \frac{80-36}{10} \\ \frac{50-36}{10} & \frac{80-36}{10} \end{bmatrix} = \begin{bmatrix} 2.4 & -0.6 \\ -6.6 & 3.4 \\ 0.4 & 9.4 \\ 6.4 & 7.4 \\ 1.4 & 1.4 \\ -10.6 & 4.4 \\ 9.4 & 4.4 \\ 1.4 & 4.4 \end{bmatrix}$$

$$C_i = \frac{(X_i^0)^T (X_i^0)}{n_i}$$ This we calculated in the R-programing.

Mean of the Matrix [M_mean]

$$D = M - M\_mean$$

$$C = (n-1)^{-1} \times D^T \times D$$

$$C_1 = \begin{bmatrix} 39.47 & 9.07 \\ 9.07 & 17.95 \end{bmatrix} \qquad C_2 = \begin{bmatrix} 41.8392 & 1.7321 \\ 1.7321 & 9.8392 \end{bmatrix}$$

C1

```
> #create vectors -- these will be our columns
> a <- c(2.5,-12.5,6.5,-7.5,-1.5,-4.5,-1.5)
> b <- c(-0.5,-5.5,-0.5,-4.5,-8.5,3.5,1.5)
>
>
> #create matrix from vectors
> M <- cbind(a,b)
>
> k <- ncol(M) #number of variables
> n <- nrow(M) #number of subjects
>
> #create means for each column
> M_mean <- matrix(data=1, nrow=n) %*% cbind(mean(a),mean(b))
>
> #creates a difference matrix
> D <- M - M_mean
>
> #creates the covariance matrix
> C <- ((7-1)^-1)* t(D) %*% D
> C
          a         b
a 39.476190  9.071429
b  9.071429 17.952381
>
> ## b using the covariance function
> cov(M)
          a         b
a 39.476190  9.071429
b  9.071429 17.952381
>
```

C2

```
> #create vectors -- these will be our columns
> a <- c(2.4,-6.6,0.4,6.4,1.4,-10.6,9.4,1.4)
> b <- c(-0.6,3.4,9.4,7.4,1.4,4.4,4.4,4.4)
>
>
> #create matrix from vectors
> M <- cbind(a,b)
>
> k <- ncol(M) #number of variables
> n <- nrow(M) #number of subjects
>
> #create means for each column
> M_mean <- matrix(data=1, nrow=n) %*% cbind(mean(a),mean(b))
>
> #creates a difference matrix
> D <- M - M_mean
>
> #creates the covariance matrix
> C <- ((8-1)^-1)* t(D) %*% D
> C
          a        b
a 41.839286 1.732143
b  1.732143 9.839286
>
> ## b using the covariance function
> cov(M)
          a        b
a 41.839286 1.732143
b  1.732143 9.839286
```

(4)

$$C(r,s) = \frac{1}{n} \sum_{n=1}^{g} n_i \cdot C_i(r,s)$$

$$= \begin{bmatrix} \frac{7}{15}(39.47) + \frac{8}{15}(41.8392) & \frac{7}{15}(9.07) + \frac{8}{15}(1.732) \\ \frac{7}{15}(9.07) + \frac{8}{15}(1.732) & \frac{7}{15}(17.95) + \frac{8}{15}(9.8392) \end{bmatrix}$$

$$C = \begin{bmatrix} 40.73 & 5.16 \\ 5.16 & 8.37 \end{bmatrix}$$

$$P = \begin{bmatrix} 7/15 \\ 8/15 \end{bmatrix} = \begin{bmatrix} 0.467 \\ 0.533 \end{bmatrix}$$

we can get inverse function by ginv() function which is present in 'MASS' library.

$$C^{-1} = \frac{1}{40.37 \times 8.37 - 5.16 \times 5.16} \begin{bmatrix} 8.37 & -5.16 \\ -5.16 & 40.73 \end{bmatrix} = \begin{bmatrix} 0.0265 & -0.0157 \\ -0.0157 & 0.1238 \end{bmatrix}$$

Discriminant function:-

$$f_i = \mu_i C^{-1} x_k^T - \frac{1}{2} \mu_i C^{-1} \mu_i^T + \ln(P_i)$$

let us compare the sample $[4 \ 1] = x_K$, which is of class 0

$\mu_1 = [-1.1428 \ -0.5714]$, $P_1 = 7/15$

$$f_1 = [-1.1428 \ -0.5714] \begin{bmatrix} 0.0265 & -0.0157 \\ -0.0157 & 0.1238 \end{bmatrix} \begin{bmatrix} 4 \\ 1 \end{bmatrix} - \left(\frac{1}{2}\right) [-1.1428 \ -0.5714] \begin{bmatrix} 0.0265 & -0.0157 \\ -0.0157 & 0.1238 \end{bmatrix} \begin{bmatrix} -1.1428 \\ -0.5714 \end{bmatrix} + \ln(7/15)$$

$f_1 = -1.1112$

$f_2 = 0.0055$

Score Calculation [used similarly from excel sheet]

$$M_1 - M_2 = \begin{bmatrix} -5.2678 \\ -8.3214 \end{bmatrix}$$

$$M_1 + M_2 = \begin{bmatrix} 2.9822 & 7.1786 \end{bmatrix}$$

$$W_0 = -\ln\left(\frac{0.467\,(P_1)}{0.533\,(P_2)}\right) - \frac{1}{2}\begin{bmatrix} 2.9822 & 7.1786 \end{bmatrix}\begin{bmatrix} -5.2678 \\ -8.3214 \end{bmatrix}$$

$$= 0.1322 - \tfrac{1}{2}\begin{bmatrix} -15.7096 - 59.736 \end{bmatrix}$$

$$= 37.8550.$$

$$W = C^{-1} \times [M_1 - M_2]$$

$$= \begin{bmatrix} 0.0265 & -0.0157 \\ -0.0157 & 0.1238 \end{bmatrix}\begin{bmatrix} -5.2678 \\ -8.3214 \end{bmatrix} = \begin{bmatrix} -0.009 \\ -0.9475 \end{bmatrix}$$

$$\text{Score} = x_k^T \times W^T + W_0$$

$$= \begin{bmatrix} 4 \\ 1 \end{bmatrix}\begin{bmatrix} -0.009 - 0.9475 \end{bmatrix} + 37.8550$$

$$= 36.8715.$$

X_1	X_2	Y	F_1	F_2	Score
4	1	0	-1.1112	0.0055	36.8715
-11	-4	0	-2.2195	-5.9742	41.744
8	1	0	-1.769586	-0.6171	36.8355
-6	-3	0	-1.227458	-3.989996	40.7515
0	-7	0	-3.425571	-9.924521	44.4875
-3	5	0	-2.864564	1.979666	35.1445
0	3	0	-1.47767	1.498354	35.0125
6	3	1	-1.799822	1.229779	34.9585
-3	7	1	-4.549983	2.189241	31.2495
4	12	1	-9.853027	1.686191	26.449
10	11	1	-8.643869	2.001429	27.3425
5	5	1	-2.6189	2.296766	33.0725
-7	8	1	-6.525625	1.125379	30.338
13	8	1	-6.029466	1.800129	30.158
5	8	1	-4.95593	2.802229	30.23

So by this we can assume that all the values of class 0 have the score above 35 (class0 if(score>35)) and we see that if the scores are below 35 then it is class1

(b)

X_1	X_2	F_1	F_2	score	Expected Y
1	2	-1.118502	0.9189538	35.951	0
1	4	-1.935522	1.996929	34.056	1
-6	6	-4.221847	1.543091	32.224	1
-4	-1	-0.8352621	-1.684946	38.8385	0
3	5	-2.521316	2.376491	33.0905	1
0	5	-2.57369	2.297329	33.1175	1
6	3	-1.799822	1.229779	34.9585	1
-1	-2	-0.8646782	-2.635071	39.759	0
15	-4	-9.306446	-5.783931	41.51	0
-8	-2	-1.330584	-3.163484	39.822	0

2.

2.

We know that

$$Var(X+Y) = Var(X) + Var(Y) + Cov(X,Y)$$

$$Var(cX) = c^2 Var(X)$$

$$Cov(cX, Y) = Cov(X, cY) = c \cdot Cov(X,Y)$$

We have,

$$Var(\alpha X + (1-\alpha)Y) = \alpha^2 \sigma_X^2 + (1-\alpha)\sigma_Y^2 + 2\alpha(1-\alpha)\sigma_{XY}.$$

Taking the derivative with respect to α,

$$\frac{\partial}{\partial \alpha} Var(\alpha X + (1-\alpha)Y)$$

$$= 2\alpha\sigma_X^2 - 2\sigma_Y^2 + 2\alpha\sigma_Y^2 + 2\sigma_{XY} - 4\alpha\sigma_{XY}$$

To find critical points, we equate it to 0.

$$2\alpha\sigma_X^2 - 2\sigma_Y^2 + 2\alpha\sigma_Y^2 + 2\sigma_{XY} - 4\alpha\sigma_{XY} = 0.$$

$$\Rightarrow \alpha = \frac{\sigma_Y^2 - \sigma_{XY}}{\sigma_X^2 + \sigma_Y^2 - 2\sigma_{XY}}$$

To tell this point is a minimum, we should prove that the second derivative is positive.

$$\frac{\partial^2}{\partial \alpha^2} Var(\alpha X + (1-\alpha)Y) = 2\sigma_X^2 + 2\sigma_Y^2 - 4\sigma_{XY}.$$

$$= 2 Var(X-Y) \geq 0.$$

3.

3) a) There are n observations in the sample. Since bootstrap sampling draws items with replacement, we are sampling from the same pool with same probability every time. There are $(n-1)$ items in the n observations that are not in j. Hence the

$$\text{Probability} = \frac{(n-1)}{n} = 1-\frac{1}{n}.$$

b) Since we draw with replacement, it is same as the one above.

$$\therefore \text{Probability} = \frac{(n-1)}{n} = 1-\frac{1}{n}$$

c) As bootstrapping sample with replacement, we have the probability that the jth observation is not in the bootstrap sample is the product of the probabilities that each bootstrap observation is not the jth observation from the original sample

$$(1-1/n)\cdots(1-1/n) = (1-1/n)^n.$$

as these probabilities are independent.

d) This 1 minus the probability that the jth observation is not in the bootstrap sample.

P(jth obs in bootstrap sample)

$$= 1 - (1 - 1/5)^5$$

$$= 0.672 //$$

e) This is similar to the above one. ie 1 minus the probability that the jth observation is not in the bootstrap sample.

P(jth obs. in bootstrap sample)

$$= 1 - (1 - 1/100)^{100} = 0.634 //$$

f) P(jth observation in the bootstrap sample)

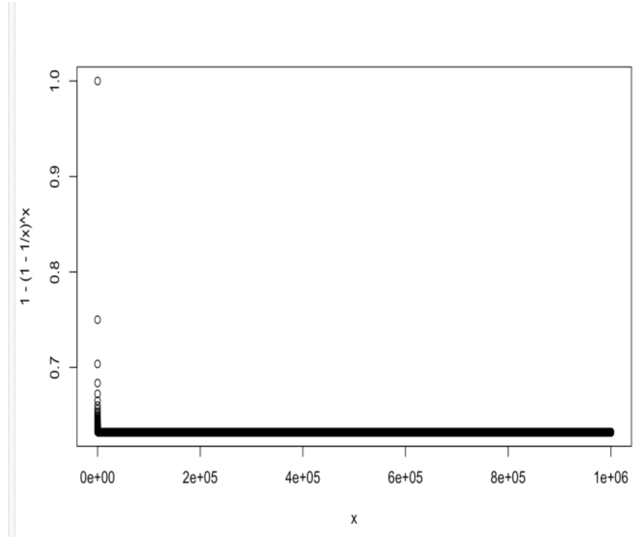
$$= 1 - \left(1 - \frac{1}{10{,}000}\right)^{10000}$$

$$= 0.632 //$$

g)

```
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> x<-1:1000000
> plot(x,1-(1-1/x)^x)
> plot(x,1-(1-1/x)^x)
```



The probability seems to converge on something around 0.63 fairly quickly, around n=100, and then stay there!
That is very odd that there is always a 63% chance that any particular thing will be in the bootstrap sample even with large datasets.

h)

h) We know that

$$\underset{n\to\infty}{Lt}\left(1+x/n\right)^n = e^x.$$

Applying this in our case, we get that the probability that a bootstrap sample of size n contains the jth observation converges to $1-1/e = 0.632$ as $n\to\infty$.

4.

4) a) The k-fold cross validation is implemented by taking the n observations and randomly splitting it into k non-overlapping groups of length of n/k. These groups act as a validation set, and the remainder of length $(n - n/k)$ acts as the training set. The test error is then estimated by averaging the k resulting MSE estimates.

b)

(I) The validation set approach has 2 main drawbacks compared to k-fold cross validation. First, the validation estimate of the test error rate can be highly variable. Second, only a subset of the observations are used to fit the model. Since statistical methods tend to perform worse when trained on fewer observations, this suggests that ~~that~~ the validation set error rate may tend to overestimate th test error rate for the model fit on the entire data set.

ii) LOOCV:

The LOOCV cross-validation approach is a special case of k-fold cross-validation in which $k = n$. This approach has 2 drawbacks compared to k-fold cross validation.

1) It requires fitting the potentially computationally expensive model n times compared to k-fold cross validation which requires the model to be fitted only k times

2) The LOOCV cross-validation approach may give approximately unbiased estimates of the test error, since each training set contains $(n-1)$ observations. However, this approach has higher variance than k-fold cross validation.

So, there is a bias-variance trade-off associated with the choice of k in k-fold cross-validation; typically using $k=5$ or $k=10$ yield test error rate estimates that suffer neither from excessively high bias nor from very high variance.

5.

5. One way to do this would be with the bootstrap. We can train on a bunch of different random samplings of the original data and see how much the estimates change.

We may estimate the standard deviation of our prediction by using the bootstrap method. In this case, rather than obtaining new independent data sets from the population and fitting our model on those data sets we instead obtain repeated random samples from the original data set. In this case, we perform sampling with replacement B times and then find the corresponding estimates and the standard deviation of those B.