ISL ASSIGNMENT 3

ANUSHA MUPPALLA

16286311

1) Given, $P(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}}$

substract 1 from both sides

$1 - p(x) = 1 - \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}} \Rightarrow \frac{1}{1+e^{\beta_0+\beta_1 x}}$

$1 - p(x) = \frac{1}{1+e^{\beta_0+\beta_1 x}}$

$\frac{1}{1-p(x)} = 1 + e^{\beta_0+\beta_1 x}$

multiply $p(x)$ on both sides

$p(x) \frac{1}{1-p(x)} = p(x)\left(1+e^{\beta_0+\beta_1 x}\right)$

$\frac{p(x)}{1-p(x)} = \frac{e^{\beta_0+\beta_1 x}}{\cancel{1+e^{\beta_0+\beta_1 x}}} \cancel{\left(1+e^{\beta_0+\beta_1 x}\right)}$

$\therefore \boxed{\frac{p(x)}{1-p(x)} = e^{\beta_0+\beta_1 x}}$

2) a) if ~~Bayes decision bound~~

Given, x_1 = IO traffic in Gbs, x_2 = cpu utilization

$\hat{\beta}_0 = -5$, $\hat{\beta}_1 = 0.05$, $\hat{\beta}_2 = 4$

$$p^{\wedge}(x) = \frac{e^{-5+0.05(\text{Io traffic})+4(\text{cpu utilization})}}{1+e^{-5+0.05(\text{Io traffic})+4(\text{cpu utilization})}}$$

$$= \frac{e^{-5+0.05(50)+4(0.8)}}{1+e^{-5+0.05(50)+4(0.8)}}$$

$$= \underline{\underline{0.64}}$$

b) utilization $= 0.8$, $p = 0.5$

$$0.5 = \frac{e^{-5+0.05(\text{Io tn})+(0.8)4}}{1+e^{-5+0.05(\text{Io})+4(0.8)}}$$

$$0.5\left(1+e^{-5+0.05(\text{Io})+4(0.8)}\right) = e^{-5+0.05(\text{Io})+4(0.8)}$$

applying log on bothsides

$$\ln(1) = -5+0.05(\text{Io})+4(0.8)$$

$$0 = -4.2 + (0.05)(\text{Io})$$

$$\text{Io traffic} = \frac{4.2}{0.05}$$

$\therefore$ Io traffic needed $= 36\%$

3) a if the Bayes decision boundary is linear, then we can expect Linear discriminant analysis to perform better for the test data because quadratic discriminant analysis to perform better on the training set because QDA could overfit the linearity on Bayes decision boundary

b) if the Bayes decision boundary is non linear then we can expect quadratic discriminant analysis to perform better both training and test datasets because common covarience matrix is inaccurate

c) As n (sample size) increases, quadratic discriminant analysis is more flexible, varience becomes less concern. The test prediction accuracy of QDA would be expected to improve as the training set increases.

d) False, as n is small and P is large the flexibility of quadratic discriminant analysis can be highly to overfitting the model whereas as P is ~~large~~ large minimizing variance is better with LDA than QDA.

4) Given,

$$P_{yes}(x) = \frac{\pi_{yes} \exp\left(-\frac{1}{2\sigma^2}(x-\mu_{yes})^2\right)}{\sum \pi_l \exp\left(\frac{-1}{2\sigma^2}(x-\mu_l)^2\right)}$$

x=4 last year (Given)

$$P_1(4) = \frac{0.6\, e^{-\left(\frac{1}{2(25)}(4-8)^2\right)}}{0.6\, e^{-\left(\frac{1}{2(25)}(4-8)^2\right)} + 0.4\, e^{-\left(1/50\,(4-0)^2\right)}}$$

$$= \frac{0.6\, e^{-\left(\frac{1}{50}(-4)^2\right)}}{0.6\, e^{-\left(\frac{1}{50}(-4)^2\right)} + 0.4\, e^{-\left(\frac{16}{50}\right)}} = \frac{0.6 e^{-\frac{16}{50}}}{0.6\, e^{-\frac{16}{50}} + 0.4 e^{-\frac{16}{50}}}$$

$\simeq 0.38$

5) Error rate on training data = 18%
Error rate on testing data = 25%
so, average = $\frac{18+25}{2}$ = 21.5%, test data error = 25%
In case knn, k=1 training error rate of 0% because
$P(Y=j \mid X=x_i) = I(y_i = j) \simeq 1$
if $y_i = j$ and 0 if not i.e 0% training error rate
average error rate of 15% which indicates test error rat = 15*2 = 30% for knn which is greater than test data error rate for logistic regression

so logistic regression is better

6) a Given, ~~P(X)~~ =

$\frac{P(x)}{1-P(x)} \Rightarrow 0.35$ ~~$P(x) = 0.35P(x)$~~ $P(x) = 0.35 - 0.35P(x)$

which transform into $P(x) = \frac{0.35}{1+0.35}$

$= 0.259$

i.e 25% of people defaulting on their credit payment

6) b Given $P(x) = 18\%$

$\frac{0.18}{1-0.18} = 0.21$

i.e 21%

so the odds that she will default is then 21%

7)

classifing a table into 2 matrixes.

$X_1 = \begin{bmatrix} -6 & 9 \\ -4 & 8 \\ 1 & 8 \\ -4 & 6 \\ -4 & 2 \\ -2 & 3 \\ 2 & 5 \\ -2 & 7 \end{bmatrix}$ $X_2 = \begin{bmatrix} 2 & -2 \\ 3 & -4 \\ -1 & -3 \\ 2 & 2 \\ -2 & -6 \\ 3 & 1 \\ 2 & -6 \end{bmatrix}$

Mean $\mu_1 = \left[-\frac{19}{8}, \frac{48}{8}\right]$ $\mu_2 = \left[\frac{9}{7} \quad -\frac{18}{7}\right]$

overal mean $= \begin{bmatrix} \dfrac{\frac{-19}{8}+\frac{19}{7}}{2} & \dfrac{\frac{48}{8}-\frac{18}{7}}{2} \end{bmatrix}$

$= \begin{bmatrix} \dfrac{-133+72}{56\times 2} & \dfrac{336-144}{56\times 2} \end{bmatrix}$

$= \begin{bmatrix} \dfrac{-61}{112} & \dfrac{192}{112} \end{bmatrix} = \begin{bmatrix} 0.54 & 1.71 \end{bmatrix}$

Normalizing everything to have mean of 0

$$x_1^0 = \begin{bmatrix} -6-(-0.5) & 9-(1.71) \\ -4-(-0.54) & 8-(1.71) \\ 1-(-0.54) & 8-(1.71) \\ -4-(-0.54) & 6-1.71 \\ -4-(-0.54) & 2-1.71 \\ -2-(-0.54) & 3-1.71 \\ 2-(-0.54) & 5-1.71 \\ -2-(-0.54) & 7-1.71 \end{bmatrix} = \begin{bmatrix} 5.44 & 7.3 \\ 3.44 & 6.3 \\ 1.56 & 6.3 \\ 3.44 & 4.3 \\ 3.44 & 0.3 \\ 1.44 & 1.3 \\ 2.56 & 3.3 \\ 1.44 & 5.3 \end{bmatrix}$$

$$x_2^0 = \begin{bmatrix} 2-(-0.54) & -2-1.71 \\ 3+0.54 & -4-1.7 \\ -1+0.54 & -3-1.7 \\ 2+0.54 & 2-1.7 \\ -2+0.54 & -6-1.7 \\ 3+0.54 & 1-1.7 \\ 2+0.54 & -6-1.7 \end{bmatrix} = \begin{bmatrix} 2.56 & -3.7 \\ 3.56 & -5.7 \\ 0.44 & -4.7 \\ 2.56 & 0.3 \\ 1.44 & -7.7 \\ 3.56 & +0.7 \\ 2.56 & -7.7 \end{bmatrix}$$

covariance matrix for each class $\quad C_i = \frac{(x_i^0)^T x_i^0}{n_i}$

$$C_2 = \begin{bmatrix} 6.75 & -9.12 \\ -9.12 & 26.78 \end{bmatrix}$$

$$C_1 = \begin{bmatrix} 9.78 & 13.12 \\ 13.12 & 24 \end{bmatrix}$$

overall covariance

$$c(r,s) = \frac{1}{n} \sum_{i=1}^{g} n_i c_i(r,s)$$

$$= \begin{bmatrix} \frac{7}{15}(6.75) + \frac{8}{15}(9.78) & \frac{7}{15}(-9.12) + \frac{8}{15}(13.12) \\ \frac{7}{15}(-9.12) + \frac{8}{15}(13.12) & \frac{7}{15}(26.78) + \frac{8}{15}(24) \end{bmatrix} = \begin{bmatrix} 8.37 & 2.74 \\ 2.74 & 25.3 \end{bmatrix}$$

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

$$P = \begin{bmatrix} 0.56 \\ 0.43 \end{bmatrix}$$

$$|C| = ad - bc$$

$$|C| = (8.37)(25.3) - (-2.741)(-2.741)$$

$$= 204.25$$

$$C^{-1} = \frac{1}{|C|} \operatorname{adj}(C)$$

$$= \frac{1}{204.25} \begin{bmatrix} 25.3 & -2.741 \\ -2.741 & 8.37 \end{bmatrix}$$

Bays rule.

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

$$f_1 = \mu_1 C^{-1} x_k^T - \frac{1}{2} \mu_1 C^{-1} \mu_1^T + \ln(P_1)$$

$$f_0 = \mu_0 C^{-0} x_k^T - \frac{1}{2} \mu_0 C^{-0} \mu_0^T + \ln(P_0)$$

<u>example</u>:-

$x_1 = 1, \ x_2 = 0$

$$f_1 = \begin{bmatrix} 0.12 & 0.08 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 0.189 \end{bmatrix} + 0.336$$

$$= 0.128 - 0.189 + 0.336$$

$$= 0.275$$

$$f_0 = \begin{bmatrix} -0.38 & 0.28 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 1.28 \end{bmatrix} - 0.629$$

$$= -2.289$$

similarly calculate for all classes i.e C_1, C_0 -- samples.

7)b

score calculation:-

$\mu_1 - \mu_2 = [3.635 \;\; 8.57]$

$\mu_1 + \mu_2 = [-1.115 \;\; 3.43]$

$[\mu_1+\mu_2][\mu_1-\mu_2]^T = 25.32$

$w_0 = \ln\left[\frac{P_1}{P_2}\right] = \frac{1}{2}\{\mu_1+\mu_2\}[\mu_1-\mu_2]^T$

$= -12.79$

$W = C^{-1}[\mu_1-\mu_2]^T = \begin{bmatrix} 0.124 & -0.0134 \\ -0.0134 & 0.0409 \end{bmatrix}\begin{bmatrix} 3.64 \\ 8.57 \end{bmatrix} = \begin{bmatrix} 0.336 \\ 0.3017 \end{bmatrix}$

$\text{score} = w^T x_k^T + w_0$

$= \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}[0.34 \;\; 0.30] - 12.79$

x_1	x_2	f_0	f_1	y
1	0	-2.28	0.27	1
1	4	-1.16	0.78	1
-3	-2	1.14	0.39	0
-1	1	-0.82	0.16	1
2	4	1.12	1.35	1