Name: Harsha Vardhan Koneru

USCID: 8296272558

DSCI - 552 Midterm

Name: Harsha Vardhan Koneru

USCID: 8296272558

1) Given simple regression model,

$$Y = \beta_0 + \beta_1 X + \epsilon$$

$R^2 = 25\% = 0.25$

$n = 15, \ p = 1$

$\alpha = 0.01$

We know that

$$R^2 = 1 - \frac{RSS}{TSS}$$

$$\frac{RSS}{TSS} = 1 - R^2 = 1 - 0.25 = 0.75$$

$$\frac{RSS}{TSS} = 0.75 \longrightarrow (1)$$

We can find if given coefficient is statistically significant or not by using F-statistic.

if $|F_{obs}| > F_{p,n-p-1,\alpha}$ we reject null hypothesis and make coefficient as significant.

$$F_{obs} = \frac{TSS - RSS / p}{RSS / (n-p-1)}$$

$$= \frac{TSS - RSS}{RSS} \times \frac{n-p-1}{p}$$

$$= \left(\frac{TSS}{RSS} - 1\right) \times \frac{n-p-1}{p}$$

From 1 $\frac{RSS}{TSS} = 0.75 = \frac{3}{4}$

$$F_{obs} = \left(\frac{4}{3} - 1\right) \times \left(\frac{15-1-1}{1}\right)$$

$$= 1/3 \times 13 = 4.33$$

$$F_{p, n-p-1, \alpha} = F_{1, 13, 0.01} = 9.074$$

Here $F_{obs} < F_{p, n-p-1, \alpha}$. So, we don't have enough evidence to reject Null hypothesis.

$\therefore$ β_1 is not statistically significant.

2) Given there are three classes and three features.

classes $(k) = \{1, 2, 3\}$

features $= \{f_1, f_2, f_3\}$

$\mu_{jk} = jk$, $\sigma_{jk} = k$

$\mu_{11} = 1$	$\mu_{12} = 2$	$\mu_{13} = 3$
$\mu_{21} = 2$	$\mu_{22} = 4$	$\mu_{23} = 6$
$\mu_{31} = 3$	$\mu_{32} = 6$	$\mu_{33} = 9$

$\sigma_{11} = \sigma_{21} = \sigma_{31} = 1$

$\sigma_{12} = \sigma_{22} = \sigma_{32} = 2$

$\sigma_{13} = \sigma_{23} = \sigma_{33} = 3$

We know that

$$f_k(x_i) = Pr(X_i = x_i \mid Y = k)$$

$$= \frac{1}{\sqrt{2\pi\sigma_{ik}^2}} e^{-\frac{(x_i - \mu_{ik})^2}{2\sigma_{ik}^2}}$$

We know that in Naive Bayes classifier

$$f_k(x_1, x_2, x_3) = f_k(x_1) \cdot f_k(x_2) \cdot f_k(x_3)$$

Given $(x_1, x_2, x_3) = (1, 5, 9)$

$$f_k(1, 5, 9) = f_k(1) \cdot f_k(5) \cdot f_k(9)$$

$$f_k(x_i) = Pr(X_i = x_i \mid Y = k)$$

$$= \frac{1}{\sqrt{2\pi\sigma_{ik}^2}} e^{-\frac{(x_i - \mu_{ik})^2}{2\sigma_{ik}^2}}$$

$$f_k(1) = \frac{1}{\sqrt{2\pi}\,\sigma_{1k}} e^{-\frac{(1 - \mu_{1k})^2}{2\sigma_{1k}^2}}$$

$$f_1(1) = \frac{1}{\sqrt{2\pi}\,\sigma_{11}} e^{-\frac{(1 - \mu_{11})^2}{2\sigma_{11}^2}}$$

$$= \frac{1}{\sqrt{2\pi}} e^{-0} = \frac{1}{\sqrt{2\pi}}$$

$$f_1(5) = \frac{1}{\sqrt{2\pi}\,\sigma_{21}} e^{-\frac{(5 - \mu_{21})^2}{2\sigma_{21}^2}}$$

$$f_1(5) = \frac{1}{\sqrt{2\pi}\,\sigma_{21}} e^{-\frac{(5-\mu_{21})^2}{2\sigma_{21}^2}}$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{(5-2)^2}{2(1)}}$$

$$= \frac{1}{\sqrt{2\pi}} e^{-9/2}$$

$$f_1(9) = \frac{1}{\sqrt{2\pi}\,\sigma_{31}} e^{-\frac{(9-\mu_{31})^2}{2\sigma_{31}^2}}$$

$$= \frac{1}{\sqrt{2\pi}} e^{-18}$$

$$\therefore f_1(1,5,9) = f_1(1) \times f_1(5) \times f_1(9)$$

$$= \frac{1}{\sqrt{2\pi}} \times \frac{1}{\sqrt{2\pi}} \times e^{-9/2} \times \frac{1}{\sqrt{2\pi}} e^{-18}$$

$$= \frac{1}{2\pi\sqrt{2\pi}} e^{-45/2}$$

$$= 0.107 \times 0.000045$$

$$= 0.00000485$$

$$f_2(1) = \frac{1}{\sqrt{2\pi}\,\sigma_{12}} e^{-\frac{(1-\mu_{12})^2}{2\sigma_{12}^2}}$$

$$= \frac{1}{\sqrt{2\pi}(2)} e^{-\frac{(1-2)^2}{2(4)}}$$

$$= \frac{1}{2\sqrt{2\pi}} e^{-1/8}$$

$$f_2(5) = \frac{1}{\sqrt{2\pi}\,\sigma_{22}} e^{-\frac{(5-\mu_{22})^2}{2\sigma_{22}^2}}$$

$$= \frac{1}{2\sqrt{2\pi}} e^{-1/8}$$

$8 \times 2\pi$

$$f_2(9) = \frac{1}{\sqrt{2\pi}\,\sigma_{32}} e^{-\frac{(9-\mu_{32})^2}{2\sigma_{32}^2}}$$

$$= \frac{1}{2\sqrt{2\pi}} e^{-9/8}$$

$$\therefore f_2(1,5,9) = \left(\frac{1}{2\sqrt{2\pi}}\right)^3 \cdot e^{-\frac{1}{8}-\frac{1}{8}-\frac{9}{8}}$$

$$= \frac{1}{16\pi\sqrt{2\pi}} \cdot e^{-11/8} = 0.002$$

$$f_3(1) = \frac{1}{\sqrt{2\pi}\,\sigma_{13}} e^{-\frac{(1-\mu_{13})^2}{2\sigma_{13}^2}}$$

$$= \frac{1}{3\sqrt{2\pi}} e^{-4/18}$$

$$f_3(5) = \frac{1}{\sqrt{2\pi}\,\sigma_{23}} e^{-\frac{(5-\mu_{23})^2}{2\sigma_{23}^2}}$$

$$= \frac{1}{3\sqrt{2\pi}} e^{-1/18}$$

$$f_3(9) = \frac{1}{\sqrt{2\pi}\,\sigma_{33}} e^{-\frac{(9-\mu_{33})^2}{2\sigma_{33}^2}}$$

$$= \frac{1}{3\sqrt{2\pi}} e^{-0} = \frac{1}{3\sqrt{2\pi}}$$

$$\frac{27\times 2\pi}{54}$$

$$f_3(1,5,9) = \left(\frac{1}{3\sqrt{2\pi}}\right)^3 \cdot e^{-5/18}$$

$$= \frac{1}{54\pi\sqrt{2\pi}} \cdot e^{-5/18}$$

$$= 0.00178$$

Given $\pi_1 = \pi_2 = \pi_3 = 1/3$

$$\Pr(Y=1, X=x) = \frac{\pi_1 f_1(1,5,9)}{\sum_{i=1}^{3} \pi_i f_i(1,5,9)}$$

$$\Pr(Y=2, X=x) = \frac{\pi_2 f_2(1,5,9)}{\sum_{i=1}^{3} \pi_i f_i(1,5,9)}$$

$$\Pr(Y=3, X=x) = \frac{\pi_3 f_3(1,5,9)}{\sum_{i=1}^{3} \pi_i f_i(1,5,9)}$$

As $\pi_1 = \pi_2 = \pi_3$, then largest $\Pr(Y=k, X=x)$ will have largest $f_i(1,5,9)$

$f_1(1,5,9) = 0.00000485$

$f_2(1,5,9) = 0.002$

$f_3(1,5,9) = 0.00178$

$\therefore$ feature vector $(1,5,9)$ classifies into $K=2$

3) Given

$$\Pr(Y=1 \mid x_1, x_2) = \frac{x_1^2 x_2^2}{1 + x_1^2 x_2^2}$$

a) Odds of $Y=1$ given x_1 and x_2

$$\Pr(Y=1 \mid x_1, x_2) = \frac{x_1^2 x_2^2}{1 + x_1^2 x_2^2}$$

$$\text{odds} = \frac{\Pr(Y=1 \mid x_1, x_2)}{1 - \Pr(Y=1 \mid x_1, x_2)}$$

$$\text{odds} = \frac{\dfrac{x_1^2 x_2^2}{1 + x_1^2 x_2^2}}{1 - \dfrac{x_1^2 x_2^2}{1 + x_1^2 x_2^2}}$$

$$= \frac{\dfrac{x_1^2 x_2^2}{\cancel{1 + x_1^2 x_2^2}}}{\dfrac{1}{\cancel{1 + x_1^2 x_2^2}}} = x_1^2 x_2^2$$

$\therefore$ Odds of $Y=1$ given x_1 and x_2 is $x_1^2 x_2^2$

b) For find decision boundary

$$Pr(Y=1 \mid x_1, x_2) = 1/2$$

$$Pr(Y=1 \mid x_1, x_2) = \frac{x_1^2 x_2^2}{1 + x_1^2 x_2^2} = \frac{1}{2}$$

$$2x_1^2 x_2^2 = 1 + x_1^2 x_2^2$$

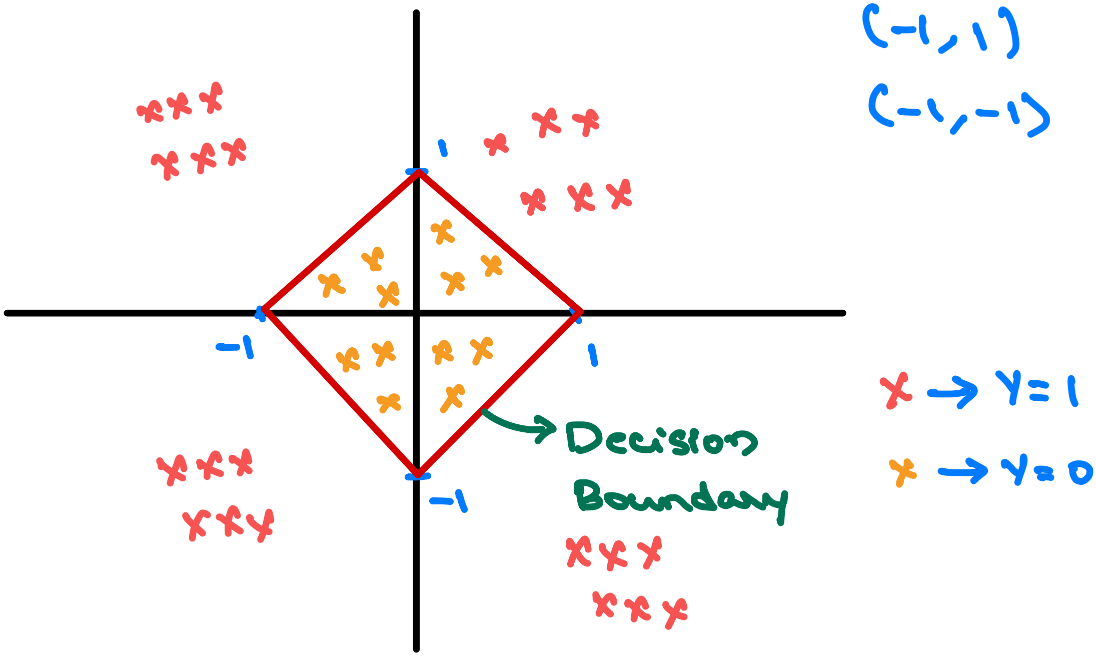
$$\boxed{x_1^2 x_2^2 = 1 \Rightarrow x_1 x_2 = \pm 1}$$

points of decision boundary = $(1, 1)$

$(1, -1)$

$(-1, 1)$

$(-1, -1)$



4) Given training set = $\{1, 2, 3\}$

Number of bootstrap samples = 10

Bootstrap samples:

$\{1,1,1\}$	$\{1,3,3\}$
$\{1,2,1\}$	$\{2,2,2\}$
$\{1,3,1\}$	$\{2,2,3\}$
$\{1,2,2\}$	$\{2,3,3\}$
$\{1,3,2\}$	$\{3,3,3\}$

$\{1,1,1\} \Rightarrow 3/3$

$\{1,2,1\} \Rightarrow 4/3$

$\{1,3,1\} \Rightarrow 5/3$

$\{1,2,2\} \Rightarrow 5/3$

$\{1,3,2\} \Rightarrow 6/3$

$\{1,3,3\} \Rightarrow 7/3$

$\{2,2,2\} \Rightarrow 6/3$

$\{2,2,3\} \Rightarrow 7/3$

$\{2,3,3\} \Rightarrow 8/3$

$\{3,3,3\} \Rightarrow 9/3$

So means of samples are

$3/3, 4/3, 5/3, 5/3, 6/3, 7/3, 6/3, 7/3, 8/3, 9/3$

Sort the data

$$3/3, \underbrace{4/3, 5/3, 5/3, 6/3, 6/3, 7/3, 7/3, 8/3}, 9/3$$

80% CI $\Rightarrow$ 8 middle points out of 10

80% CI for given data $= \left[4/3, 8/3\right]$

5)

a) Given test points

$x_1 = 5,\ y_1 = +$

$x_2 = 4,\ y_2 = -$

<u>K=1</u>

When $x_1 = 5$

Nearest neighbor $\Rightarrow$ $X = 4.9,\ Y = +$

$\hat{y}_1 = +,\ y_1 = +$

When $x_2 = 4$

Nearest neighbor $\Rightarrow$ $X = 4.5,\ Y = +$

$\hat{y}_2 = +,\ y_2 = -$

test_error for $K=1$ $= \frac{1}{2} = 0.5$

<u>K=3</u>

When $x_1 = 5$

	x	y
Nearest neighbors $\Rightarrow$	4.9	+
	5.2	−
	5.3	−

Based on majority vote,

$\hat{y}_1 = -$, $y_1 = +$

When $x_2 = 4$

Nearest neighbors $\Rightarrow$

x	y
4.5	+
4.6	+
4.9	+

Based on majority vote

$\hat{y}_2 = +$, $y_2 = -$

test_error for $K=3$ $= \frac{2}{2} = 1$

$\underline{K=5}$

When $x_1 = 5$

Nearest neighbors $\Rightarrow$

x	y
4.9	+
5.2	−
5.3	−
4.6	+
4.5	+

Based on majority vote,

$\hat{y}_1 = +$, $y_1 = +$

When $x_2 = 4$

Nearest neighbors $\Rightarrow$

x	Y
4.5	+
4.6	+
4.9	+
3.0	−
5.2	−

Based on majority vote

$$\hat{y}_2 = + \,, \; y_2 = -$$

test_error for $k = 5 = \frac{1}{2} = 0.5$

<u>$k = 9$</u>

When $x_1 = 5$

Nearest neighbors $\Rightarrow$

x	Y
4.9	+
5.2	−
5.3	−
4.6	+
4.5	+
5.5	+
3.0	−
7.0	−
0.5	−

Based on majority vote

$$\hat{y}_1 = - \,, \; y_1 = +$$

When $x_2 = 4$

Nearest neighbors $\Rightarrow$

x	y
4.5	+
4.6	+
4.9	+
3.0	−
5.2	−
5.3	−
5.5	+
7.0	−
0.5	−

Based on majority vote

$\hat{y}_2 = -$, $y_2 = -$

Test error for $K=9$ $= 1/2 = 0.5$

b) Test error for $K=1 \Rightarrow 1/2 = 0.5$

Test error for $K=3 \Rightarrow 2/2 = 1$

Test error for $K=5 \Rightarrow 1/2 = 0.5$

Test error for $K=9 \Rightarrow 1/2 = 0.5$

$\therefore$ When $K = 1, 5, 9$ we get best test results.