Vidyalekhan

Name: Bhosale Ratnesh Sambhajirao
Roll No: 19MF10010

Solution 1]

$n(y=1) = 53$, $n(y=2) = 59$, $n(y=3) = 88$

$$\therefore \text{Entropy of } y = -\left[\frac{53}{200}\log_2\frac{53}{200} + \frac{59}{200}\log_2\frac{59}{200} + \frac{88}{200}\log_2\frac{88}{200}\right]$$

$$= 1.548$$

Total $\quad n(x1=1) = \dfrac{105}{200}$

$$n(x1=2) = \frac{95}{200}$$

$n(x2=1) = \dfrac{96}{200} \qquad n(x2=2) = \dfrac{104}{200}$

$n(x3=A, x_3=B, x3=C, x3=D) = \dfrac{50}{200}$ (each)

Now, for $x1$:

$x1=1$	
$y=1$	31/105
$y=2$	40/105
$y=3$	34/105

$x_1=2$	
$y=1$	22/95
$y=2$	19/95
$y=3$	54/95

for $x2$:

$x2=1$	
$y=1$	25/96
$y=2$	26/96
$y=3$	45/96

$x=2$	
$y=1$	28/104
$y=2$	33/104
$y=3$	43/104

for $x3$:

$x3=A$	
$y=1$	35/50
$y=2$	14/50
$y=3$	1/50

$x3=B$	
$y=1$	15/50
$y=2$	25/50
$y=3$	10/50

$x3=C$	
$y=1$	1/50
$y=2$	15/50
$y=3$	34/50.

$x3=0$	
$y=1$	$2/50$
$y=2$	$5/50$
$y=3$	$43/50$

Entropy of $x1=1$

$$= -\left[\frac{31}{105}\log_2\frac{31}{105} + \frac{40}{105}\log_2\frac{40}{105} + \frac{34}{105}\log_2\frac{34}{105}\right]$$

$$= \underline{1.57}$$

Entropy of $x1=2$

$$= -\left[\frac{22}{95}\log_2\frac{22}{95} + \frac{19}{95}\log_2\frac{19}{95} + \frac{54}{95}\log_2\frac{54}{95}\right]$$

$$= \underline{1.41}$$

Entropy of $x2=1$

$$= -\left[\frac{22}{96}\log_2\frac{22}{96} + \frac{26}{96}\log_2\frac{28}{96} + \frac{45}{96}\log_2\frac{45}{96}\right]$$

$$= \underline{1.52}$$

Entropy of $x2=2$

$$= -\left[\frac{28}{104}\log_2\frac{28}{104} + \frac{33}{104}\log_2\frac{33}{104} + \frac{43}{104}\log_2\frac{43}{104}\right]$$

$$= \underline{1.56}$$

<u>Now, Information Gains:</u>

$$\text{Gain}(x1) = 1.548 - \frac{105}{200}\times 1.57 - \frac{95}{200}\times 1.41$$

$$= 0.054$$

$$\text{Gain}(x2) = 1.548 - \frac{96}{200}\times 1.52 - \frac{104}{200}\times 1.56$$

$$= 0.0072$$

For $x3=A$, Entropy =

$$= -\left[\frac{35}{50}\log_2\frac{35}{50} + \frac{14}{50}\log_2\frac{14}{50} + \frac{1}{50}\log_2\frac{1}{50}\right]$$

$$= 0.98$$

$x3=B$, Entropy =

$$= -\left[\frac{15}{50}\log\frac{15}{50} + \frac{25}{50}\log_2\frac{25}{50} + \frac{10}{50}\log_2\frac{10}{50}\right] = \underline{1.48}$$

$x3 = C$, Entropy:

$$= -\left[\frac{1}{50}\log_2\frac{1}{50} + \frac{15}{50}\log_2\frac{15}{50} + \frac{34}{50}\log_2\frac{34}{50}\right] = \underline{1.01}$$

$x3 = D$, Entropy:

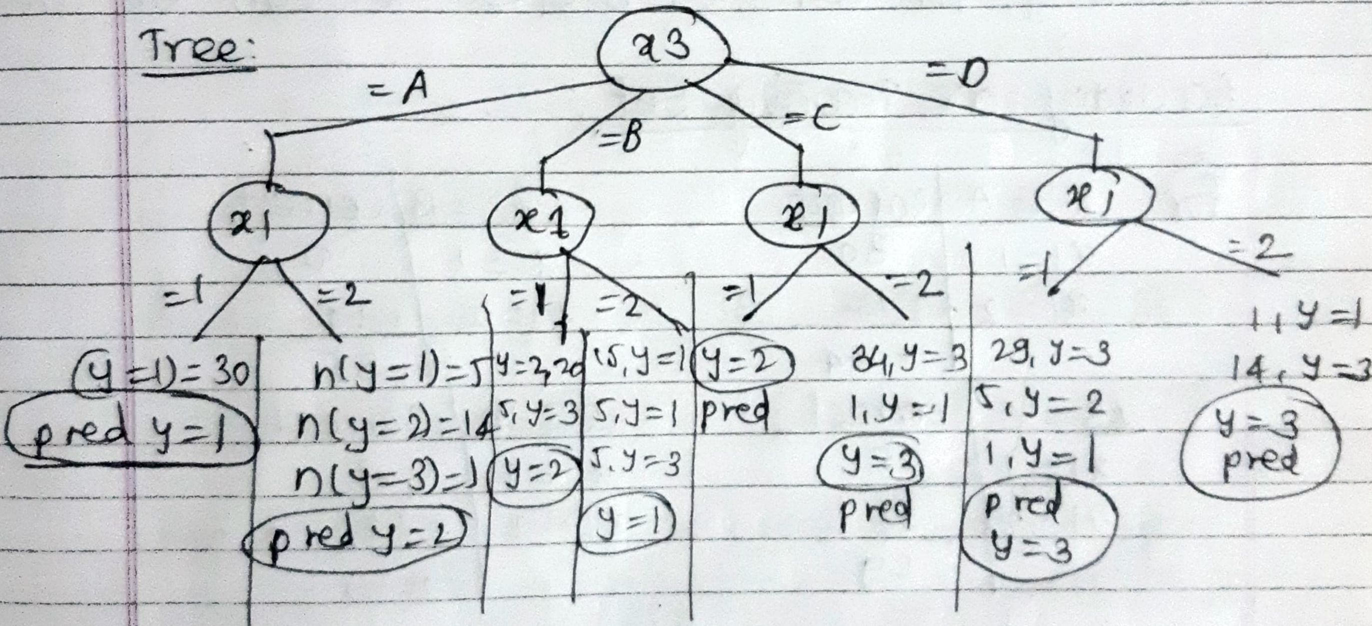
$$= -\left[\frac{2}{50}\log_2\frac{2}{50} + \frac{5}{50}\log_2\frac{5}{50} + \frac{43}{50}\log_2\frac{43}{50}\right] = \underline{0.7}$$

$$\therefore \text{Gain}(x3) = 1.548 - \frac{1}{4}\left[0.98 + 0.48 + 1.01 + 0.7\right]$$

$$= \underline{0.505}$$

Hence, we see that $x3$ has highest information gain.

Tree:



for $x3 = A$, we confine our set of consideration original entropy for $x3 = A = 0.98$

Again

$x_1 = 1$	
y=1	30/30
y=2	0
y=3	0

$x1 = 2$	
y=1	5/20
y=2	14/20
y=3	1/20

Entropy($x1=1$) $= 0$

Entropy($x1=2$) $= \underline{1.076}$

$x_2 = 1$	
$y=1$	$18/23$
$y=2$	$5/23$
$y=3$	$0/23$

$x_2 = 2$	
$y=1$	$17/27$
$y=2$	$9/27$
$y=3$	$1/27$

Entropy $(x_2=1) = 0.755$

Entropy $(x_2=2) = 1.124.$

Hence, $\text{Gain}(x_1) = 0.98 - \frac{20}{50} \times 1.076 = 0.549$

$$\text{Gain}(x_2) = 0.98 - \frac{23}{50} \times 0.755 - \frac{27}{50} \times 1.124$$

$$= \underline{0.025}$$

Hence, x_1 has more gain so, x_1 will be used to classify

Similarly, for all $x_3 = B, C, D$ x_1 can be used.

Accuracy on Training Set :

for

$x_3 = A$	correct
$x_1 = 1$	30
$x_1 = 2$	14
	+ 44

$x_3 = B$	Correct.
$x_1 = 1$	20
$x_1 = 2$	15
	35

$x_3 = C$	correct
$x_1 = 1$	15
$x_1 = 2$	34
	49

$x_3 = D$	correct
$x_1 = 1$	29
$x_1 = 2$	14
	+ 23

Total correctly classified $= 44 + 35 + 23 + 49$
$= 171.$

$$\text{Accuracy} = \frac{171}{200} \times 100 = \underline{\underline{85.5\%}}$$

Vidyalekhan

DATE
PAGE

(2) Solution :

$P(x1=2) = P(x1=1) = P(x2=1) = P(x2=2) = \frac{1}{2}$

$P(x3 = A \text{ or } B \text{ or } C \text{ or } D) = All = \frac{1}{4}$.

$P(y=1) = \frac{53}{200}$ $P(y=2) = \frac{53}{200}$ $P(y=3) = \frac{88}{200}$

Also, $P(x1=2 \mid y=1) = \frac{22}{53}$, $P(x2=1 \mid y=1) = \frac{25}{53}$

$P(x3=A \mid y=1) = \frac{35}{53}$

$P(y=1 \mid x1=2, x2=1, x3=A) = \frac{P(y=1)\, P\left(\frac{x1=2}{y=1}\right) \cdot P\left(\frac{x2=1}{y=1}\right) P\left(\frac{x3=A}{y=1}\right)}{P(x1=2)\, P(x2=1)\, P(x3=A)}$.

8^{st}

$= \frac{\frac{22}{53} \times \frac{25}{53} \times \frac{35}{53} \times \frac{53}{200}}{\frac{1}{8}} = 0.273$ (for $y=1$)

Also, $P(x1=2 \mid y=2) = \frac{19}{59}$ $P(x2=1 \mid y=2) = \frac{26}{59}$

$P(x3=1 \mid y=2) = \frac{14}{59}$

Reqd Prob $P(y=2 \mid x1=2, x2=1, x3=A)$.

$= \frac{19 \times 26 \times 14 \times 59}{(59)^3 \times 200} \times 8 = 0.023$ (for $y=2$).

Also, $P(x1=2 \mid y=3) = \frac{54}{88}$ $P(x2=1 \mid y=3) = \frac{45}{88}$

$P(x3=A \mid y=3) = \frac{1}{88}$

Reqd P that $P(y=3 \mid \text{conditions})$.

$= \frac{54 \times 45 \times 88 \times 8}{200 \times (88)^3} = 0.10$ (for $y=3$).

Hence, we predict label = $y=1$

confidence of pred = 0.03375

Vidyalekhan

DATE

PAGE

IInd $P(x3=B|y=1) = \frac{15}{53}$

for $y=1$

Reqd prob $= \frac{15 \times 22 \times 25}{(53)^3} \times \frac{53}{200} \times 8 = 0.117$

for $y=2$ $P(x3=B|y=2) = \frac{25}{59}$

Reqd Prob $= \frac{25 \times 19 \times 26 \times 59}{(53)^3 \times 200} \times 8 = 0.142$

for $y=3$ $P(x3=B|y=3) = \frac{10}{88}$

Reqd Prob $= \frac{10 \times 54 \times 45 \times 88 \times 8}{(88)^3 \times 200} = 0.128$

Predicted label $y=2$

confidence $= \underline{0.0177}$

IIIrd We need $P(y=1 | x1=1, x2=1, x3=C)$

for $y=1$ $P(x1=1|Y) = \frac{31}{53}$ $P(x2=1|Y) = \frac{25}{53}$

$P(x3=C|Y) = \frac{1}{53}$

Reqd Prob $= \frac{31 \times 25 \times 1}{(53)^3} \times 8 \times \frac{53}{200} = 0.011$

for $y=2$ $P(x1=1|Y) = \frac{40}{59}$

$P(x2=1|Y) = \frac{26}{59}$ $P(x3=C|Y) = \frac{15}{59}$

Reqd Prob $= \frac{40 \times 15 \times 26 \times 8 \times 59}{(59)^3 \times 200} = 0.179$

for $y=3$ $P(x1=1|Y) = \frac{34}{88}$ $P(x2=1|Y) = \frac{45}{88}$

$P(x3=C|Y) = \frac{34}{88}$.

Reqd $P = \frac{34 \times 45 \times 34 \times 8}{(88)^3 \times 200} \times 88 = \underline{0.968}$

Hence Predicted label $y = 3$ with confidence $= 0.0335$

IVth We need $P(y = 1, 2, 3 \mid x_1 = 2, x_2 = 1, x_3 = 0)$

for $y=1$ $\quad P(x_1 = 2 \mid Y) = \frac{22}{53} \quad P(x_2 = 1 \mid Y) = \frac{25}{53}$

$P(x_3 = 0 \mid Y) = \frac{8}{53}$

$$P = \frac{22 \times 2 \times 25 \times 8 \times 53}{(53)^3 \times 200} = 0.015$$

for $y=2$ $\quad P(x_1 = 2 \mid Y) = \frac{19}{59} \quad P(x_2 = 1 \mid Y) = \frac{26}{59}$

$P(x_3 = 0 \mid Y) = \frac{5}{59}$

$$P = \frac{19 \times 5 \times 26 \times 8 \times 59}{(59)^3 \times 200} = 0.028$$

for $y=3$

$P(x_1 = 2 \mid Y) = \frac{54}{88} \quad P(x_2 = 1 \mid Y) = \frac{45}{88}$

$P(x_3 = 0 \mid Y) = \frac{43}{88}$

$$P = \frac{54 \times 43 \times 45 \times 8 \times 88}{(88)^3 \times 200} = 0.539$$

Hence, predicted label $y = 3$ with confidence $= 0.067$

Sol^n (5) : After plotting x_2 on y-axis & x_1 in x-axis we can observe that it's x_1 that segregates the $Y=1$ & $Y=2$ label points.

Observing from the 2-D plot we can choose threshold as $x_1 > 12$ or $x_1 < 12$

for $x_1 > 12$:

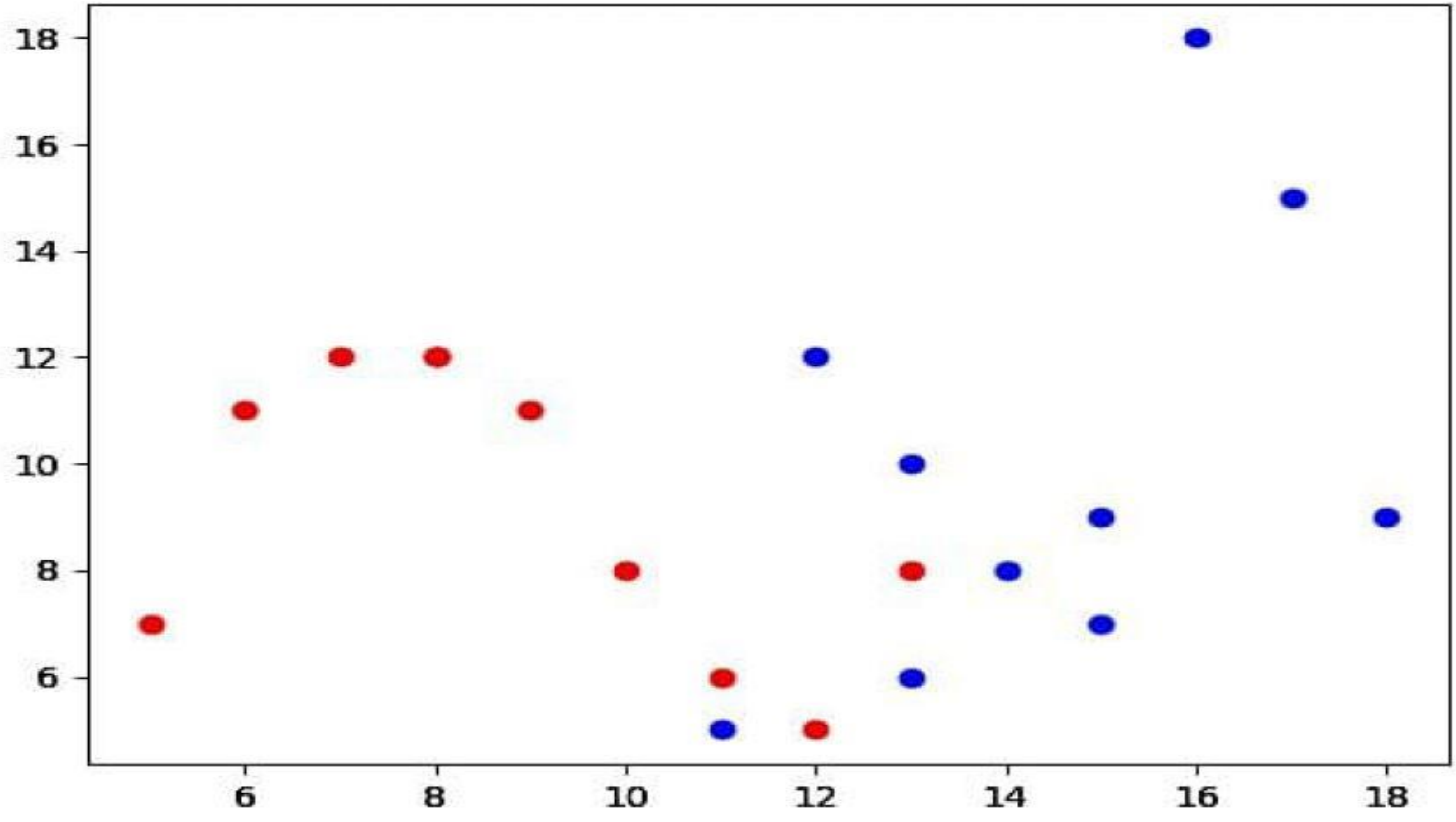
$n(y=1) = 1$

$n(y=2) = 8$

for $x_1 < 12$:

$n(y=1) = 9$

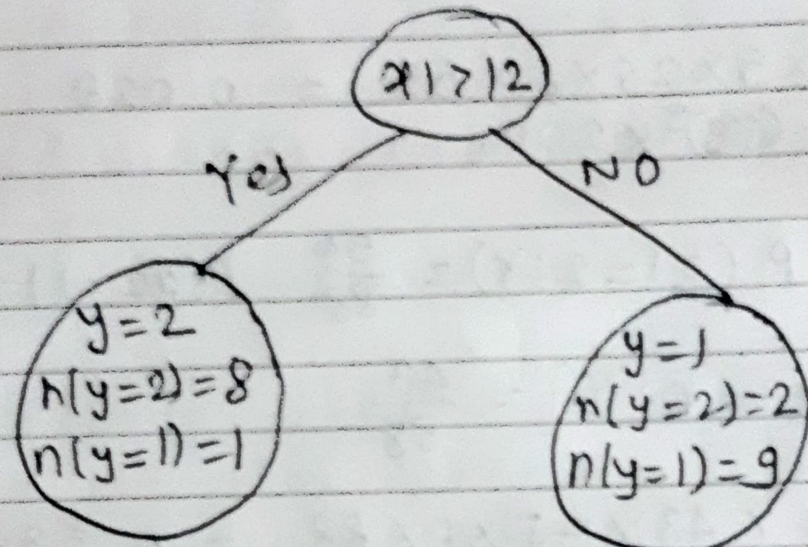
$n(y=2) = 2$


18
16
14
12
10
8
6
6
8
10
12
14
16
18

Hence stump is

Blue are $x1 > 12$

Red are $x1 < 12$



Predict $y=2$ Predict $y=1$

(4) for $y=1$, we have:

Sample mean $(x1) = 8.9$ Mean $= \frac{\Sigma x_i}{N} = \bar{x}$

Sample Variance $(x1) = 6.766$ variance $= \frac{\Sigma(x_i - \bar{x})^2}{N-1}$

Sample Mean $(x2) = 9.2$
Sample variance $(x2) = 7.288$

for $y=2$, we have:
Sample mean$(x1) = 14.4$, Sample variance $(x1) = 4.93$
Sample mean $(x2) = 9.9$, Sample " $(x2) = 16.54$.

Confidence:
for $x_1 = 5$, $x_2 = 7$ or $(5,7)$
$P(Y_1 | (5,7)) + P(y_2 | (5,7)) = 1$ → Mutually exclusive

$$P(y_i (5,7)) = \frac{P(x_1 = 5 | y_i) \cdot P(x_2 = 7 | i) \cdot P(y_i)}{K}$$

$\Rightarrow \frac{1}{K}[0.0498 \times 0.106 \times 0.5 + 0] = 1$

$\therefore P(y_i) = \frac{1}{2}$ (only) Two values of y

$\Rightarrow K = 0.0498 \times 0.106 \times 0.5$
And $[P(x_1 = 5 | y=2) \approx 0]$.

$\therefore P(y=1 | (5,7)) = 1$ confidence $= 1$
$P(y=0 | (5,7)) = 0$

(ii) for $(7,12)$ Here $\because P(x_1 = 7 | y=2) \approx 0$
Hence, $\frac{1}{K}[P(x_1 = 7 | y_1) \cdot P(x2 = 12) | y_1] + 0] = 1$

$\therefore P(y=1 | (7,2)) = 1$ [Approx. confidence $= 1$]
$P(y=0 | (7,2)) = 0$

Vidyalekhan

DATE

PAGE

(iii) for $C(12,5)$

$$P(y_i \mid C(12,5)) + P(y_2 \mid C(12,5)) = 1$$

$$\Rightarrow \frac{1}{K}\left[0.075 \times 0.0441 \times 0.5 + 0.1 \times 0.0475 \times 0.5\right] = 1$$

$$\Rightarrow K = 0.004$$

$\therefore P(y_1 \mid C(12,5)) = 0.416$

$\therefore P(y_2 \mid C(12,5)) = 0.595$

Confidence $= 0.595$

(iv) $C(10,8)$ We have,

$$\frac{1}{K}\left[0.1408 \times 0.1339 \times 0.5 + 0.02 \times 0.087 \times 0.5\right] = 1$$

$$\Rightarrow K = 0.0105$$

$P(y_1 \mid C(10,8)) = 0.894$

$P(y_2 \mid C(10,8)) = 0.105$

confidence $= 0.89$

(v) $C(6,11)$ We have,

$$\frac{1}{K}\left[0.0824 \times 0.11 \times 0.5 + 0\right] = 1$$

$P(x_1 = 6 \mid y = 0) = 0.$

$P(y_1 \mid C(6,11)) = 1$

$P(y_2 \mid C(6,11)) = 0$

Confidence $= 1$

(vi) $C(13,8)$ We have.

$$\frac{1}{K}\left[0.044 \times 0.183 \times 0.5 + 0.147 \times 0.087 \times 0.5\right] = 1$$

$$\Rightarrow K = 0.0076$$

$P(y_1 \mid C(13,8)) = 0.313$ $\quad$ $P(y_2 \mid C(13,8)) = 0.687$

confidence $= 0.687$

(vii) (8,12) we get,

$$\frac{1}{k}\left[0.14 \times 0.086 \times 0.5 + 0.0028 \times 0.085 \times 0.5\right] = 1$$

$$\Rightarrow k = 0.0063$$

$P(y_1 \mid (8,12)) = 0.99$ confidence = 0.99

$P(y_2 \mid (8,12)) = 0.01$

(viii) (11,6) we get,

$$\frac{1}{k}\left[0.11 \times 0.007 \times 0.5 + 0.055 \times 0.061 \times 0.5\right] = 1$$

$$k = 0.0021$$

$P(y_1 \mid (11,6)) = 0.193$ confidence = 0.807

$P(y_2 \mid (11,6)) = 0.807$

(ix) (13,6) we get,

$$\frac{1}{k}\left[0.044 \times 0.073 \times 0.5 + 0.147 \times 0.06 \times 0.5\right] = 1$$

$$k = 0.0061$$

$P(y_1 \mid (13,6)) = 0.262$ confidence = 0.738

$P(y_2 \mid (13,6)) = 0.738$

(x) (14,8) we have.

$$\frac{1}{k}\left[0.0224 \times 0.1339 \times 0.5 + 0.1768 \times 0.087 \times 0.5\right] = 1$$

$$\Rightarrow k = 0.0092.$$

$P(y_1 \mid (14,8)) = 0.162$ confidence = 0.838

$P(y_2 \mid (14,8)) = 0.838$

(xi) (17,15) we have

$$\frac{1}{k}\left[0.0012 \times 0.014 \times 0.5 + 0.09 \times 0.044 \times 0.5\right] = 1$$

$$\Rightarrow k = 0.0212$$

$P(y_1 \mid (17,15)) = 0$ confidence = 1

$P(y_2 \mid (17,15)) = 1$

(xii) C(15, 9) we get,

$\frac{1}{K}[0.009 \times 0.1474 \times 0.5 + 0.1733 \times 0.095 \times 0.5] = 1$

$\Rightarrow K = 0.009$

$P(y_1 | C(15,9)) = 0.08$ confidence = 0.92

$P(y_2 | C(15,9)) = 0.92$

(xiii) C(13, 10) we get

$\frac{1}{K}[0.11 \times 0.044 \times 0.5 + 0.0556 \times 0.0475 \times 0.5] = 1$

$\Rightarrow K = 0.01030$

$P(y_1 | C(13,10)) = 0.303$ confidence = 0.697

$P(y_2 | C(13,10)) = 0.697$

(xiv) C(11, 5) we get

$\frac{1}{K}[0.1107 \times 0.044 \times 0.5 + 0.055 \times 0.0475 \times 0.5] = 1$

$\Rightarrow K = 0.00376$

$P(y_1 | C(11,5)) = 0.649$ confidence = 0.649

$P(y_2 | C(11,5)) = 0.351$

(xv) C(16, 18) we get,

since $P(x_2 = 18 | y=1) \approx 0$.

$P(y_1 | C(16,18)) = 0$ $P(y_2 | C(16,18)) = 1$

confidence = 1

(xvi) C(15, 7) we get,

$\frac{1}{K}[0.0098 \times 0.106 \times 0.5 + 0.173 \times 0.07 \times 0.5] = 1$

$\Rightarrow K = 0.0071$

$P(y_1 | C(15,7)) = 0.073$ confidence = 0.727

$P(y_2 | C(15,7)) = 0.927$

(xviii) $(12,12)$ we get.

$$\frac{1}{K}\left[0.0754 \times 0.086 \times 0.5 + 0.1 \times 0.085 \times 0.5\right] = 1$$

$$K = 0.0075$$

$P(y_1 | (12,12)) = 0.431$ $\quad$ $\underline{\text{Confidence} = 0.569}$

$P(y_2 | (12,12)) = 0.569$

(xix) $(18,9)$ we get,

As $P(y=1 | x_1 = 18) \simeq 0$

Hence, $P(y_1 | (18,9)) = 0$ $\quad$ $P(y_2 | (18,9)) \simeq 1$

$\underline{\text{confidence} = 1}$

Hence, we see for $x_1 = 12$, $x_2 = 12$ we have least confidence $\underline{= 0.569}$

Solution

(5) We have (x_i, y_i, w_i) such N points where x_i is D dimensional vector. we have weighted losses in MSE. $\therefore$ Our loss function

$$L = \sum_{i=1}^{N} w_i (y_{p_i} - y_i)^2$$

Where,

$$y_{p_i} = a^T x_i + b$$

$$\Rightarrow L = \sum_{1}^{N} w_i (a^T x_i + b - y_i)^2$$

To minimize the loss function we need to do $\frac{\partial L}{\partial a} = \frac{\partial L}{\partial b} = 0$.

$$\frac{\partial L}{\partial b} = 2\sum_{1}^{N} w_i (a^T x_i + b - y_i) = 0$$

$$\Rightarrow a^T \sum_{1}^{N} w_i x_i + b \sum_{1}^{N} w_i = \sum_{1}^{N} y_i w_i$$

Vidyalekhan

DATE

PAGE

$$\Rightarrow b = \frac{\sum^{N} W_i y_i}{\sum W_i} - a^T \frac{\sum W_i x_i}{\sum W_i}$$

$$\Rightarrow b = \bar{y}_w - a^T \bar{x}_w$$

Where, $\bar{y}_w = \frac{\sum^{N} W_i y_i}{\sum W_i}$, $\bar{x}_w = \frac{\sum W_i x_i}{\sum W_i}$

weighted means of y & x.

Now, $\frac{\partial L}{\partial a} = 2 \sum W_i x_i (a^T x_i + b - y_i) = 0$

$$\Rightarrow \sum W_i a^T x_i x_i^T + \sum W_i x_i b = \sum W_i x_i y_i$$

Putting value of b here, we get

$$\Rightarrow \sum W_i a^T x_i x_i^T + \sum (W_i x_i \bar{y}_w - W_i x_i a^T \bar{x}_w) = \sum W_i x_i y_i$$

$$\Rightarrow a^T \sum (W_i x_i x_i^T - W_i x_i \bar{x}_w) = \sum W_i x_i y_i - \bar{y}_w \sum W_i x_i$$

$$\Rightarrow a^T = \frac{\sum_{i}^{N} (W_i x_i y_i - \bar{y}_w W_i x_i)}{\sum_{i}^{N} (W_i x_i x_i^T - \bar{x}_w W_i x_i)}$$

and $b = \bar{y}_w - a^T \bar{x}_w$, where $\bar{y}_w$ & $\bar{x}_w$ are weighted means.

Bhosale Ratnesh Sambhajirao
19MF10010