2/1/25 (Lec-02)

LEARNING

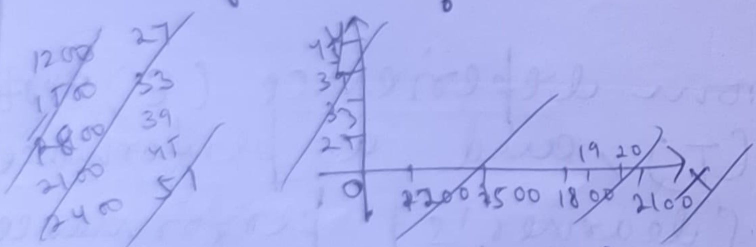
e.g:- 1)

Input (X) (House area) sq. ft.	Target (y) (House Price)
Instance-1 ⇐ 1200	27 lacs
1500	33 "
2400	51 "
3600	75 "

What is the price of a house for 2000 sq. ft. area?

$\boxed{\text{Ans:- 43 lakhs}}$

⇓



$$m = \frac{y_2 - y_1}{x_2 - x_1} = \frac{51 - 33}{2400 - 1500} = 0.02$$

$$y - y_1 = m(x - x_1) = y - 33 = 0.02\,(2000 - 1500)$$

$$y - 33 = 0.02 \times 500$$

$$y = 10 + 33 = 43$$

Input (X_1) (Area)	Input (X_2) (bedrooms)	Target (y) (Price)
12	2	34
15	2	40
24	3	61
36	3	85

what is the price with 2000 sq ft with 3 bedrooms?

Ans :- 53 lakhs

12	34
15	40
18	46
21	52
24	61
36	

$\Rightarrow \frac{\text{change in price}}{\text{change in area}} = \frac{61-40}{24-15} = 2.33 \text{ lacs}$

Price increase $= 2000 - 1500 = 5 \times 2.33 = 11.67$ lacs

2000 + 2 bedrooms $= 11.67 + 40 = 51.67$ lacs $+ 3 = \underline{53 \text{ lacs}}$

(Training data)

Q1 H.W:

$\left\{ \begin{array}{l} i = \text{rows} = \text{samples} \\ j = \text{features} = \text{columns} \end{array} \right.$

	x_1	x_2	y
Sample-1	1	2	3
Sample-2	2	1	4
Sample-3	3	3	5

Use batch GDA —

compute θ_0, θ_1 & θ_2?

Assume $x_0 = 1$

Initialise $\theta_0 = \theta_1 = \theta_2 = 0$.

$\alpha = 0.1$

Ans :- 1) compute prediction ($h_\theta(x)$ or $\hat{y}$)

$$h_\theta(x^{(1)}) = \theta_0 x_0^{(1)} + \theta_1 x_1^{(1)} + \theta_2 x_2^{(1)}$$
$$= 0\times1 + 0\times1 + 0\times2 = 0$$

$$h_\theta(x^{(2)}) = 0\times1 + 0\times2 + 0\times1 = 0$$
$$h_\theta(x^{(3)}) = 0\times1 + 0\times3 + 0\times3 = 0$$

$$\theta_0 = \theta_0 - \alpha \cdot \frac{\partial J}{\partial \theta_0} \;??$$

$$\theta_1 = \theta_1 - \alpha \cdot \frac{\partial J}{\partial \theta_1}$$

$$\theta_2 = \theta_2 - \alpha \cdot \frac{\partial J}{\partial \theta_2}$$

Before 1st itr. —

$$J(\theta) = \frac{1}{2} \times (J_1 + J_2 + J_3)$$

$$J_1 = \underbrace{(0-3)^2}_{(h_\theta(x) - y)^2} = 9$$

$$J_2 = (0-4)^2 = 16$$

$$J_3 = (0-5)^2 = 25$$

$$J(\theta) = \frac{1}{2}(9+16+25) = \frac{1}{\cancel{2}} \times \cancel{50}^{25}$$

$$\boxed{J(\theta) = 25}$$

But, we need to find partial derivative — (gradient)

$$\boxed{\frac{\partial J}{\partial \theta_j} = (h_\theta(x)^{(i)} - y^{(i)}) \cdot x_j^{(i)}}$$

ITR.-1

For θ_0; $\frac{\partial J}{\partial \theta_0} = (0-3)\times \cancel{x_0}^{1} + (0-4)\times1 + (0-5)\times1$

$$\frac{\partial J}{\partial \theta_0} = -3-4-5 = -12$$

For θ_1; (1st feature) $\frac{\partial J}{\partial \theta_1} = (0-3)\times \cancel{x_1}^{1} \, (x_1^{(1)}) + (0-4)\times 2 \, (x_1^{(2)}) + (0-5)\times 3 \, (x_1^{(3)})$

$$\frac{\partial J}{\partial \theta_1} = -3 + (-8) - 15 = -26$$

15
11
26

For θ_2 (2nd feature): $\frac{\partial J}{\partial \theta_2} = (0-3)\times \overset{(x_2^{(1)})}{2} + (0-4)\times \underset{(x_2^{(2)})}{1} + (0-5)\times \underset{(x_2^{(3)})}{3}$

$\theta_2 = -25$

Now;

$\overset{(\theta_0)}{\theta_0} = 0 - \overset{(\alpha)}{0.1} \times \overset{(\frac{\partial J}{\partial \theta_0})}{(-12)}$

$\boxed{\theta_0 = 1.2}$

$\theta_1 = 0 - 0.1\times(-26) = \underline{2.6}$

$\theta_2 = 0 - 0.1\times(-25) = \underline{2.5}$

$h_\theta(x^{(1)}) = 1.2\times 1 + 2.6\times 1 + 2.5\times 2$

$h_\theta(x^{(1)}) = 8.8$

$h_\theta(x^{(2)}) = 1.2\times 1 + 2.6\times 2 + 2.5\times 1$

$h_\theta(x^{(2)}) = 8.9$

$h_\theta(x^{(3)}) = 1.2\times 1 + \overbrace{2.6\times 3}^{(7.8)} + \underbrace{2.5\times 3}_{(7.5)} = 16.5$

$J(\theta) = \frac{1}{2}\times\left((8.8-3)^2 + (8.9-4)^2 + (16.5-5)^2\right)$

$J(\theta) = \frac{1}{2}\times(33.64 + 24.01 + 132.25)$

$\underline{J(\theta)} = \frac{1}{2}\times 189.9 = \underline{94.95}$

ITR. - 2

For θ_0; $\frac{\partial J}{\partial \theta_0} = \overset{h_\theta(x^{(1)})}{(8.8-3)}\times \overset{(x_0^{(1)})}{1} + (8.9-4)\times 1 + (16.5-5)\times 1$

$\frac{\partial J}{\partial \theta_0} = 5.8 + 4.9 + 11.5 = 22.2$

For θ_1; $\frac{\partial J}{\partial \theta_1} = (8.8-3)\times \overset{(x_1^{(1)})}{1} + (8.9-4)\times \overset{(x_1^{(2)})}{2} + (16.5-5)\times \overset{(x_1^{(3)})}{3}$

$\frac{\partial J}{\partial \theta_1} = 5.8 + 9.8 + 34.5 = 50.1$

For θ_2; $\frac{\partial J}{\partial \theta_2} = (8.8-3)\times \overset{(x_2^{(1)})}{2} + (8.9-4)\times \underset{(x_2^{(2)})}{1} + (16.5-5)\times \underset{(x_2^{(3)})}{3}$

$\frac{\partial J}{\partial \theta_2} = 11.6 + 4.9 + 34.5 = 51$

$\theta_0 = 1.2 - 0.1\times 22.2 = \underline{-1.02}$

$\theta_1 = 2.6 - 0.1\times 50.1 = \underline{-2.41}$

$\theta_2 = 2.5 - 0.1\times 51 = \underline{-2.6}$

$J(\theta) = \frac{1}{2}\times\left((-8.62-3)^2 + (-8.42-4)^2 + (-16.02-5)^2\right)$

$\boxed{J(\theta) = 365.55}$

$h_\theta(x^{(1)}) = (-1.02\times \overset{(x_0)}{1}) + (-2.41\times \underset{(x_1)}{1}) + (-2.6\times \underset{(x_2)}{2}) = -8.62$

$h_\theta(x^{(2)}) = (-1.02\times 1) + (-2.4\times 2) + (-2.6\times 1) = -8.42$

$h_\theta(x^{(3)}) = (-1.02\times 1) + (-2.4\times 3) + (-2.6\times 3)$

$h_\theta(x^{(3)}) = -16.02$

ITR. - 3

$\frac{\partial J}{\partial \theta_0} = (-8.62-3)\times \overset{(x_0^{(1)})}{1} + (-8.42-4)\times 1 + (-16.02-5)\times 1 = -45.06$

$\frac{\partial J}{\partial \theta_1} = (-8.62-3)\times \overset{(x_1^{(1)})}{1} + (-8.42-4)\times \overset{(x_1^{(2)})}{2} + (-16.02-5)\times \overset{(x_1^{(3)})}{3}$

$\frac{\partial J}{\partial \theta_1} = -99.52$

$\frac{\partial J}{\partial \theta_2} = (-8.62-3)\times \overset{(x_2^{(1)})}{2} + (-8.42-4)\times \overset{(x_2^{(2)})}{1} + (-16.02-5)\times 3$

$\frac{\partial J}{\partial \theta_2} = -98.72$

$$\theta_0 = -1.02 - (0.1 \times -45.06) = -1.02 + 4.506 = \underline{3.48}$$

$$\theta_1 = -2.41 - (0.1 \times -99.52) = \underline{7.54}$$

$$\theta_2 = -2.6 - (0.1 \times -98.72) = \underline{7.27}$$

$$h_\theta(x^{(1)}) = (3.48 \times 1) + (7.54 \times 1) + (7.27 \times 2)$$

$$h_\theta(x^{(1)}) = 25.56$$

$$h_\theta(x^{(2)}) = (3.48 \times 1) + (7.54 \times 2) + (7.27 \times 1)$$

$$h_\theta(x^{(2)}) = 25.83$$

$$h_\theta(x^{(3)}) = (3.48 \times 1) + \underbrace{(7.54 \times 3)}_{(22.62)} + \underbrace{(7.27 \times 3)}_{(21.81)}$$

$$h_\theta(x^{(3)}) = 47.91$$

$$J(\theta) = \frac{1}{2} \times \left((25.56 - 3)^2 + (25.83 - 4)^2 + (47.91 - 5)^2\right)$$

$$J(\theta) = \frac{1}{2} \times (508.95 + 476.54 + 1841.2) = \frac{1}{2} \times 2826.69$$

$$\boxed{J(\theta) = 1413.34}$$

24/1/25
[Lec.–15]

e.g.:– (K-fold cross validation)

	x_1	x_2	y
S-1	2	−1	1
S-2	0.5	1.2	0
S-3	1	2	1
S-4	−3	−2	1
S-5	4	0.1	0

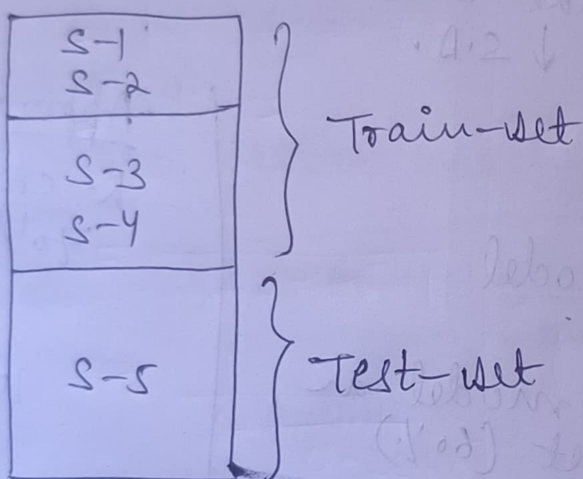
→ Do 3-fold C.V.
→ Find accuracy & S.D.

Fold-1
$\{\theta_1, \theta_2\} = \{1.8, 2.8\}$

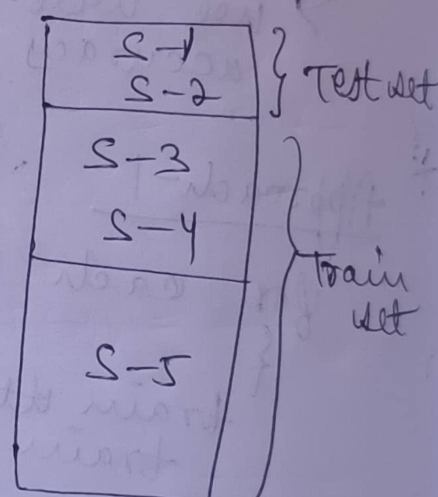
Fold-2 $\{\theta_1, \theta_2\} = \{2.1, 3.1\}$

Fold-3 $\{\theta_1, \theta_2\} = \{1.9, 4\}$

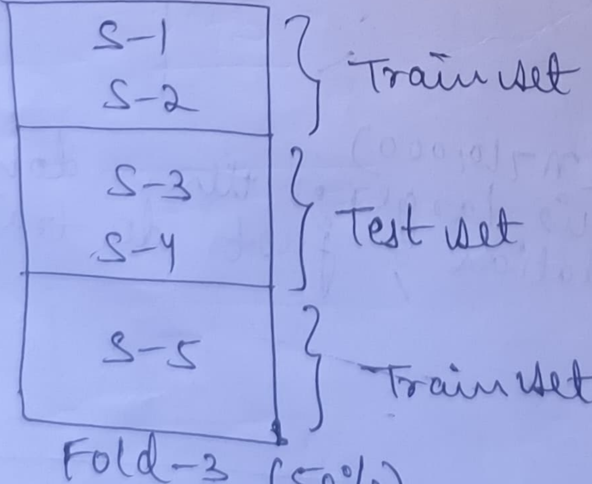
Ans:–



Fold-1 (100%)



Fold-2 (50%)



Fold-3 (50%)

For fold-1 —
(5-1, 2, 3, 4)

$h_\theta(x)^{(1)} = \theta_0 x_0^{(1)} + \theta_1 x_1^{(1)} + \theta_2 x_2^{(1)}$

$h_\theta(x)^{(1)} = 0 + ((-1.8) \times 2) + (2.8 \times -1)$

$h_\theta(x)^{(1)} = -3.6 - 2.8$

$h_\theta(x)^{(1)} = -6.4$

$h_\theta(x)^{(2)} = 0 + ((-1.8) \times 0.5) + (2.8 \times 1.2)$

$h_\theta(x)^{(2)} = -0.9 + 3.36$

$h_\theta(x)^{(2)} = 2.46$

$h_\theta(x)^{(3)} = 0 + (-1.8 \times 1) + (2.8 \times 2)$

$h_\theta(x)^{(3)} = -1.8 + 5.6 = 3.8$

$h_\theta(x)^{(4)} = 0 + (-1.8 \times -3) + (2.8 \times -2)$

$h_\theta(x)^{(4)} = 5.4 - 5.6 = -0.2$

$g(\theta^T x) = \dfrac{1}{1 + e^{\underbrace{-\theta^T \cdot x}_{(h_\theta(x))}}}$

$g(\theta^T x)^{(1)} = \dfrac{1}{1+e^{6.4}} = \dfrac{1}{1+601.84}$

$g(\theta^T x)^{(1)} = \dfrac{1}{602.84} = 1.65 \times 10^{-3}$

$g(\theta^T x)^{(1)} = 0.0016$

$g(\theta^T x)^{(2)} = \dfrac{1}{1+e^{-2.46}} = \dfrac{1}{1+0.085}$

$g(\theta^T x)^{(2)} = \dfrac{1}{1.085} = 0.92$

$g(\theta^T x)^{(3)} = \dfrac{1}{1+e^{-3.8}}$

$= \dfrac{1}{1+0.022}$

$= \dfrac{1}{1.022}$

$g(\theta^T x)^{(3)} = 0.97$

$g(\theta^T x)^{(4)} = \dfrac{1}{1+e^{0.2}}$

$= \dfrac{1}{1+1.221}$

$= \dfrac{1}{2.221}$

$g(\theta^T x)^{(4)} = 0.45$

Test (Fold-1) —

$h_\theta(x)^{(5)} = 0 + (-1.8 \times 4) + (2.8 \times 0.1)$

$h_\theta(x)^{(5)} = -7.2 + 0.28$

$h_\theta(x)^{(5)} = -6.92$

$g(\theta^T x)^{(5)} = \dfrac{1}{1+e^{6.92}}$

$= \dfrac{1}{1+1012.31}$

$= \dfrac{1}{1013.31}$

$= 9.86 \times 10^{-4}$

$g(\theta^T x)^{(5)} = 0.00098$

$\Downarrow$

Since, it is less than the threshold (0.5), so it should be labeled as 0; i.e.

matching with 'y' of test (s-5), so the accuracy of fold-1 is 100%.

For fold-2 —

(Test ⇒ s-1, s-2)

$h_\theta(x)^{(1)} = 0 + (2.1 \times 2) + (3.1 \times -1)$
$= 4.2 - 3.1 = 1.1$

$h_\theta(x)^{(2)} = 0 + (2.1 \times 0.5) + (3.1 \times 1.2)$

$h_\theta(x)^{(2)} = 1.05 + 3.72$

$h_\theta(x)^{(2)} = 4.77$

$g(\theta^T x)^{(1)} = \frac{1}{1+e^{-1.1}}$

$= \frac{1}{1+\cancel{0.0149}\ 0.3328}$

$= \frac{1}{\cancel{1.0149}\ 1.3328}$

$g(\theta^T x)^{(1)} = 0.75$

$g(\theta^T x)^{(2)} = \frac{1}{1+e^{-4.77}}$

$= \frac{1}{1+0.00848}$

$= \frac{1}{1.00848}$

$g(\theta^T x)^{(2)} = 0.99$

Since both are greater than threshold (0.5), so labelled as 1.
So, accuracy of fold-2 will be 50%

For fold-3 —

(Test ⇒ s-3, s-4)

$h_\theta(x)^{(3)} = 0 + (1.9 \times 1) + (4 \times 2) = 1.9 + 8$
$= 9.9$

$h_\theta(x)^{(4)} = 0 + (1.9 \times (-3)) + (4 \times (-2))$

$h_\theta(x)^{(4)} = -5.7 - 8$

$h_\theta(x)^{(4)} = -13.7$

$g(\theta^T (x)^{(3)}) = \frac{1}{1+e^{-9.9}}$

$= \frac{1}{1+0.00005}$

$= \frac{1}{1.00005}$

$g(\theta^T x^{(3)}) = 0.99$

$g(\theta^T x^{(4)}) = \frac{1}{1+e^{13.7}}$

$= \frac{1}{1+890911.6}$

$= \frac{1}{890912.6}$

$= 0.0000011$

$g(\theta^T x^{(4)}) \approx 0$

$g(\theta^T x^{(3)}) > 0.5$ — Predicted as 1 ✓

$g(\theta^T x^{(4)}) < 0.5$ — Predicted as 0 ✗

So, fold-3 has also 50% accuracy.

Mean or avg. accuracy $= \frac{100+50+50}{3}$

$\boxed{\text{Avg. accuracy} = 66.66\%}$

Standard deviation —

$$\sigma = \sqrt{\frac{(x_1-\mu)^2 + (x_2-\mu)^2 + (x_3-\mu)^2}{N}}$$

$x_1 = 100\%$, $x_2 = 50\%$, $x_3 = 50\%$, $\mu = 66.66\%$

$N = 3$

$$\sigma = \sqrt{\frac{(100-66.66)^2 + (50-66.66)^2 + (50-66.66)^2}{3}}$$

$$\sigma = \sqrt{\frac{(33.34)^2 + (-16.66)^2 + (-16.66)^2}{3}}$$

$$\sigma = \sqrt{\frac{1111.55 + 277.55 + 277.55}{3}}$$

$$\sigma = \sqrt{\frac{1666.65}{3}}$$

$$\sigma = \sqrt{555.55}$$

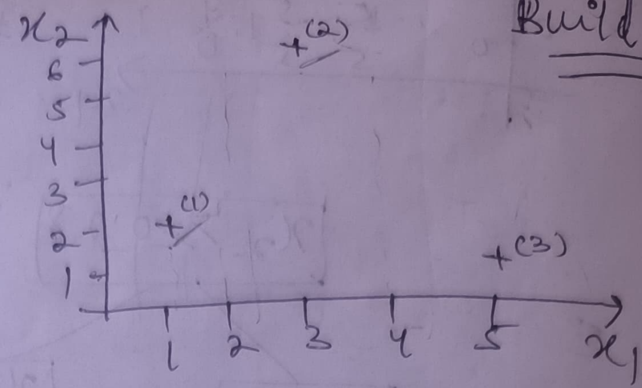
$\boxed{\sigma = 23.57\% \approx 0.23}$

6/2/25

e.g. :-

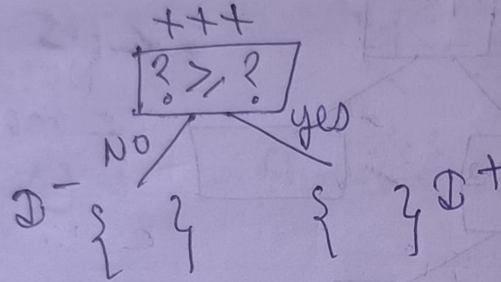
Build tree

	x_1	x_2	y
(1)	1	2	1
(2)	3	6	5
(3)	5	1	2



$I = \{1, 2, 3\} \Rightarrow$ indices of samples

$K = 2$ (min no. of nodes)



Ans :- $j = x_1, x_2$

for x_1, $s = \{2, 4\}$ | for x_2, $s = \{1.5, 4\}$

$(x_1, 2)$

Sample-1 :- I^+ | I^-

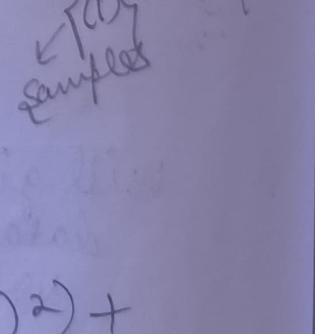
(x_1) If $s = 2 \Rightarrow 1 > 2$ NO | $1 < 2$ ✓

sample-2

If $s = 2 \Rightarrow 3 > 2$ ✓ | $3 < 2$ ✗

sample-3

If $s = 2 \Rightarrow 5 > 2$ ✓ | $5 < 2$ ✗



$$\hat{y}^+ = \frac{7}{2} = 3.5 \qquad \hat{y}^- = 1$$

$$E = \left((5-3.5)^2 + (2-3.5)^2\right) + (1-1)^2$$

$$E = 2.25 + 2.25 = 4.5$$

$(x_1, 4)$

I^+

(x_1) If $s=4 \Rightarrow 1>4$ ×

If $s=4 \Rightarrow 3>4$ ×

If $s=4 \Rightarrow 5>4$ ✓

I^-

$1<4$ ✓

$3<4$ ✓

$5<4$ ×

$\boxed{?\geq?}$ No → I^- $\{1,2\}$; yes → I^+ $\{3\}$

$$\hat{y}^+ = \frac{2}{1} = 2 \qquad \hat{y}^- = \frac{6}{2} = 3$$

$$E = \big((\underbrace{2-2}_{0})^2 + ((1-3)^2 + (5-3)^2)\big)$$

$$E = 0 + (4+4) = 8$$

$(x_2, 1.5)$

I^+

(x_2) If $s=1.5 \Rightarrow 2>1.5$ ✓

$\Rightarrow 6>1.5$ ✓

$\Rightarrow 1>1.5$ ×

I^-

$2<1.5$ ×

$6<1.5$ ×

$1<1.5$ ✓

$\boxed{?\geq?}$ No → I^- $\{3\}$; yes → I^+ $\{1,2\}$

$$\hat{y}^+ = \frac{6}{2} = 3 \qquad \hat{y}^- = \frac{2}{1} = 2$$

$$E = ((1-3)^2 + (5-3)^2) + (\underbrace{2-2}_{0})^2$$

$$E = 4+4 = 8$$

$(x_2, 4)$

I^+

(x_2) If $s=4 \Rightarrow 2>4$ ×

$6>4$ ✓

$1>4$ ×

I^-

$2<4$ ✓

$6<4$ ×

$1<4$ ✓

$\boxed{?\geq?}$ No → I^- $\{1,3\}$; yes → I^+ $\{2\}$

$$\hat{y} = \frac{1+2}{2} = \frac{3}{2} = 1.5$$

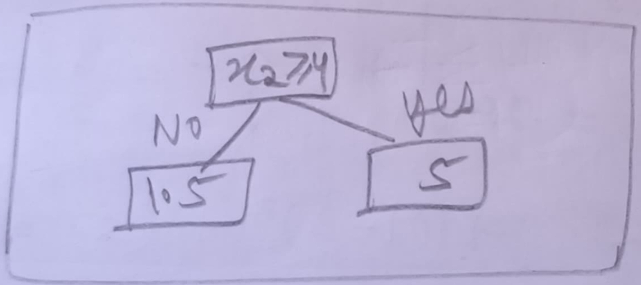
$$\hat{y}^+ = \frac{5}{1} = 5$$

$$E = (\underbrace{5-5}_{0})^2 + ((1-1.5)^2 + (2-1.5)^2)$$

$$E = 0.25 + 0.25 = 0.5$$

$$j^*, s^* = \underset{j,s}{\arg\min}\; E_{j,s}$$

$$j^*, s^* = [x_2, 4]$$



$K = 2$

$|I^-| = 2$

$|I^+| = 1$

$|I^-| = K \Rightarrow \hat{y}^- = 1.5$

$|I^+| < K \Rightarrow \hat{y}^+ = 5$

e.g:- $p = 6$ (+ve cases), $n = 6$ (-ve cases)

$$\boxed{H\left(\frac{p}{p+n}, \frac{n}{p+n}\right) = -\frac{p}{p+n}\log_2\frac{p}{p+n} - \frac{n}{p+n}\log_2\frac{p}{p+n}}$$

Entropy (of parent node)

$$= -\sum_{i=1}^{K}\left(p_i \log_2 p_i\right)$$

Ans:- $H = \left(\frac{6}{12}\log_2\frac{6}{12}\right) - \left(\frac{6}{12}\log_2\frac{6}{12}\right) = 1$

(Patrons)

(Patrons) $EH = \left[\frac{p_1+n_1}{12}\cdot H\left(\frac{0}{2}, \frac{2}{2}\right) + \frac{p_2+n_2}{12}\cdot H\left(\frac{4}{4}, \frac{0}{4}\right) + \frac{p_3+n_3}{12}\cdot H\left(\frac{2}{6}, \frac{4}{6}\right)\right]$

For children

Expected entropy (EH (a)) —

$$= \sum_{i=1}^{K}\left[\left(\frac{p_i+n_i}{p+n}\right)\cdot H\left(\frac{p_i}{p_i+n_i}, \frac{n_i}{p_i+n_i}\right)\right]$$

$$\boxed{(I)\ \text{Info. gain }(a) = H\left(\frac{p}{p+n}, \frac{n}{p+n}\right) - EH(a)}$$

It should be high/max. — It should be low

∴ Pick that feature which gives you high 'I', and less 'E'.

$$EH = \left[\frac{2}{12}H(0,1) + \frac{4}{12}H(1,0) + \frac{6}{12}H\left(\frac{2}{6}, \frac{4}{6}\right)\right]$$

$$EH = \frac{2}{12}\left(-0\log_2 0 - 1\log_2 1\right) + \frac{4}{12}\left(-1\log_2 1 - 0\log_2 0\right) + \frac{6}{12}\left(-\frac{2}{6}\log_2\frac{2}{6} - \frac{4}{6}\log_2\frac{4}{6}\right)$$

$EH = 0.455$

$I = 1 - 0.455 = 0.545$

$H = 1$

$$EH = \frac{2}{12} \cdot H\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{2}{12} \cdot H\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{4}{12} \cdot H\left(\frac{2}{4}, \frac{2}{4}\right) + \frac{4}{12} \cdot H\left(\frac{2}{4}, \frac{2}{4}\right)$$

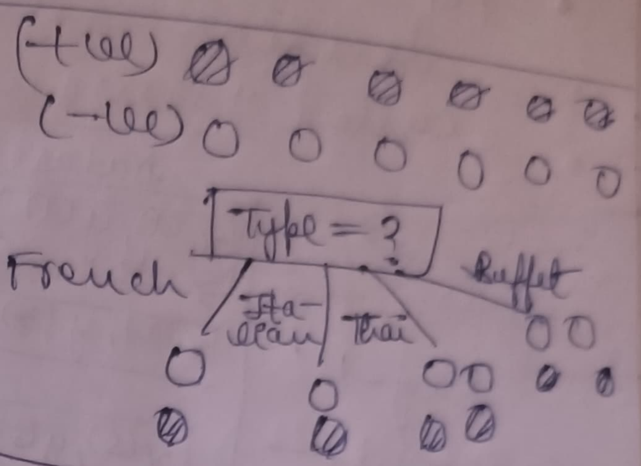
$$EH = \frac{1}{6}\left(-\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2}\right) + \frac{1}{6}\left(-\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2}\right) + \frac{1}{3}\left(-\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2}\right) + \frac{1}{3}\left(-\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2}\right)$$

$$EH = \frac{1}{6} \times (0.5 + 0.5) + \frac{1}{6} + \frac{1}{3} \times (1) + \frac{1}{3}$$

$$EH = 0.32 + 0.66$$

$$EH = 0.98$$

$$I = H - EH = 1 - 0.98 = 0.02$$



So, $I_{(Patron)} > I_{(Type)}$ and $EH(Patron) < EH(Type)$. Thus, we will select Patron as feature.

Boosted Regression Tree (GBM)

or

AdaBoost Regression

~~1) Initialize $f(x) = 0$~~
~~$r = y$~~

~~2) for iteration 1 to B~~

eg:-

x_1	x_2	y
1	2	4
2	3	5
3	4	6
4	5	7

original dataset ⇒

Ans:- $F(X) = 0$
$f_0(X) = 0$

Iteration-1 :- $r = y$

x_1	x_2	r
1	2	4
2	3	5
3	4	6
4	5	7

a) Fit a model, $f_1(X) \rightarrow r$

$\boxed{x_1 \geq 2.5}$

No → $\boxed{4.5}$ Yes → $\boxed{6.5}$

// Select / Construct tree / node using D. Tree algo. ⇐

b) Update the model, $\lambda = 0.1$

$$F(X) = f_0(X) + \lambda f_1(X)$$

c) Update the residuals —

$$r(1) = r(1) - \lambda f_1(X^{(1)})$$

$$r(1) = 4 - (0.1)(4.5) \quad // \text{ bec; } x_1 \le 2.5 \text{ so, } 4.5$$

$$= 3.55$$

$$r(2) = 5 - (0.1)(4.5) = 4.55$$

$$r(3) = 6 - (0.1)(6.5) = 5.35$$

$$r(4) = 7 - (0.1)(6.5) = 6.35$$

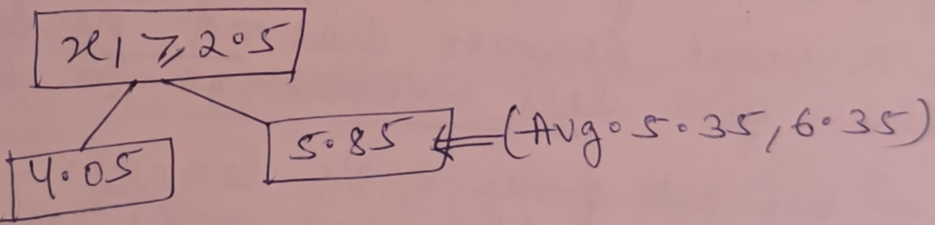
// residuals (r) is reduced

Iteration-2 :-

$$r = \begin{bmatrix} 3.55 \\ 4.55 \\ 5.35 \\ 6.35 \end{bmatrix}$$

x_1	x_2	r
1	2	3.55
2	3	4.55
3	4	5.35
4	5	6.35

a) Fit a model —



b) Update the model, $\lambda = 0.1$

$$F(X) = f_0(X) + \lambda f_1(X) + \lambda f_2(X)$$

c) Update the residuals —

$$r(1) = r(1) - \lambda f_2(X^{(1)}) = 3.55 - (0.1)(4.05)$$

$$= 3.145$$

$$r(2) = 4.55 - (0.1)(4.05) = 4.145$$

$$r(3) = 5.35 - (0.1)(\cancel{6.35}\, 5.85) = 4.765$$

$$r(4) = 6.35 - (0.1)(5.85) = 5.765$$

AdaBoost Classn. — continued - - -

19/2/25

e.g.:-

Step-1:-

x_1	x_2	y	w
1	2	+1	1
2	3	+1	1
3	3	+1	1
4	5	-1	1
5	5	-1	1
6	6	-1	1

(N=6)

Step-2:-

$\boxed{x_1 < 2.5}$ — Assume for now (But follow D.T. algo)

NO → $\boxed{-1}$; yes → $\boxed{+1}$ $\Rightarrow x_1 = 1,2$ ~~3,4,5,6~~

$x_1 = 3,4,5,6 \Leftarrow \boxed{-1}$

$$\epsilon = \frac{\sum_{i=1}^{N} w^{(i)} I\{y^{(i)} \neq f_b(x^{(i)}, \theta_b)\}}{\sum_{i=1}^{N} w^{(i)}}$$

Iteration-1

(sample-1)

$$\epsilon_1 = \frac{1 \times I\underbrace{\{+1 \neq +1\}}_{(0)}}{6} = \frac{0}{6} = 0$$

$$\epsilon_2 = \frac{1 \times I\underbrace{\{+1 \neq +1\}}_{0}}{6} = 0$$

$$\epsilon_3 = \frac{1 \times I(+1 \neq -1\}}{6} = \frac{1}{6} = 0.1666$$

$$\epsilon_4 = \frac{1 \times I\underbrace{(-1 \neq -1\}}_{0}}{6} = 0$$

$$\epsilon_5 = \frac{1 \times I\{\underbrace{-1 \neq -1}_{0}\}}{6} = 0$$

$$\epsilon_6 = \frac{1 \times I\{\underbrace{-1 \neq -1}_{0}\}}{6} = 0$$

$$\lambda_b = \frac{1}{2} \log_e \left(\frac{1-\epsilon}{\epsilon}\right)$$

$$\lambda_b = \frac{1}{2} \log_e \left(\frac{1-0.1666}{0.1666}\right)$$

$$\lambda_b = \frac{1}{2} \log_e \left(\frac{0.8334}{0.1666}\right) \Rightarrow \frac{1}{2}\left[\log_e \cancel{5.002} (0.8334) - \log_e(0.1666)\right]$$

$$\cancel{\lambda_b = \frac{1}{2} \times 2.3225} = \cancel{1.16125}$$

$$\cancel{w(n) = w(n) e^{\pm\lambda_b}}$$

$$\cancel{w(1) = w_{(1)}^{(1)} e^{-1.16125}} \; \cancel{= 3.156} \; \cancel{0.313}$$

$$\cancel{w(3) = 1 \times e^{(+1.16125)}} = \cancel{3.193}$$

$$\lambda_b \Rightarrow \frac{1}{2}\left(\ln(0.8334) - \ln(0.1666)\right)$$

$$\lambda_b \Rightarrow \frac{1}{2}(1.6099) = 0.80495$$

$w(n) = w(n)\, e^{-\lambda_b}$	$w(n) = w(n)\, e^{+\lambda_b}$
⇓	⇓
For correctly classified	For misclassn.
⇓	⇓
Sample-1, 2, 4, 5, 6	Sample-3
$W = 1 \times e^{-0.80495}$	$W = 1 \times e^{0.80495}$
$W = 0.44$	$W = 2.23$
Given less wt.	given high wt.

e.g. :- $x = (x_1, x_2, x_3)$

$y = (y_1, y_2, y_3)$

NA a target variable

$$\phi(x) = \begin{bmatrix} x_1 x_1 \\ x_1 x_2 \\ x_1 x_3 \\ x_2 x_1 \\ x_2 x_2 \\ x_2 x_3 \\ x_3 x_1 \\ x_3 x_2 \\ x_3 x_3 \end{bmatrix}$$

$x = [1, 2, 3]$

$y = [4, 5, 6]$

$< \phi(x), \phi(y) > = ?$

Ans :- $\phi(x) = \begin{bmatrix} 1\times1=1 \\ 1\times2=2 \\ 1\times3=3 \\ 2\times1=2 \\ 2\times2=4 \\ 2\times3=6 \\ 3\times1=3 \\ 3\times2=6 \\ 3\times3=9 \end{bmatrix}$, $\phi(y) = \begin{bmatrix} 16 \\ 20 \\ 24 \\ 20 \\ 25 \\ 30 \\ 24 \\ 30 \\ 36 \end{bmatrix}$

$$\langle \phi(x), \phi(y) \rangle = (1\times16) + (2\times20) + 72 + 40 + 100$$
$$+ 180 + 72 + 180 + 324$$

$$\boxed{\langle \phi(x), \phi(y) \rangle \Rightarrow 1024}$$

OR

* $K(x, y) = (\langle x, y \rangle)^2 \Rightarrow$ $\boxed{\text{Kernel - 2}}$

e.g. :- $x = [1, 2, 3]$
$y = [4, 5, 6]$

$$\langle x, y \rangle = (1\times4) + (2\times5) + (3\times6)$$
$$\langle x, y \rangle^2 = (4 + 10 + 18)^2$$
$$\boxed{\langle x, y \rangle^2 = 1024}$$

∴ This and kernel do this operation in lower-dimensional space only, i.e. equivalent to the transformation into higher-dimensional space. So, it is computationally effective as compared to kernel-1.

e.g:- Kernel :-
↳ Do kernel trick (Kernel linearization)

x_1	x_2	y
1	2	0
3	4	0

$n = 2$

, $\phi(x) = \begin{bmatrix} x_1^2 \\ x_2^2 \\ x_1 x_2 \end{bmatrix}$

⇓ higher-dim. space (p)

$X = [x_1, x_2]^T$

$X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \Rightarrow 2\text{-D}$

$X \in \mathbb{R}^2$

, $\phi(x) \Rightarrow 3\text{-D}$

$\phi(x) \Rightarrow \mathbb{R}^3$

, $\theta \in \mathbb{R}^{3} \rightarrow p'$ (3D)

, $\beta \in \mathbb{R}^{2} \rightarrow n'$ (2D)

$\theta = \sum_{i=1}^{n} \beta_i \phi(x^{(i)})$

$\theta = [\theta_1, \theta_2, \theta_3]^T$

$\theta = \begin{bmatrix} ? \\ ? \\ ? \end{bmatrix}$

Step-1 :- Transform into higher dimensional space —

Sample-1 :- $\phi(x^{(1)}) = \begin{bmatrix} x_1^2 \\ x_2^2 \\ x_1 x_2 \end{bmatrix} = \begin{bmatrix} (1)^2 \\ (2)^2 \\ 1\times 2 \end{bmatrix} = \begin{bmatrix} 1 \\ 4 \\ 2 \end{bmatrix}_{3\times 1}$

$\phi(x^{(2)}) = \begin{bmatrix} 9 \\ 16 \\ 12 \end{bmatrix}_{3\times 1}$

Step-2 :- $\phi = \beta_1 \phi(x^{(1)}) + \beta_2 \phi(x^{(2)})$

$$\begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \end{bmatrix} = \begin{bmatrix} \beta_1 + 9\beta_2 \\ 4\beta_1 + 16\beta_2 \\ 2\beta_1 + 12\beta_2 \end{bmatrix}$$

Normalization happening

$$\boxed{d = \frac{|\beta_0 + \beta_1 x_1 + \beta_2 x_2|}{\sqrt{\beta_1^2 + \beta_2^2}}}$$

e.g. :- Three lines (hyperplanes)

Find the optimum hyperplane?

$$\Downarrow$$

$$2x_1 + 3x_2 - 5 = 0 \rightarrow (1)$$

$$-x_1 + 4x_2 + 7 = 0 \rightarrow (2)$$

$$5x_1 - 12x_2 + 10 = 0 \rightarrow (3)$$

	x_1	x_2	y
S-1	3	4	+1
S-2	2	3	+1
S-3	1	-1	-1
S-4	-2	+1	-1

Ans :- 1st hyperplane —

S-1 :- $d_{11} = \dfrac{|-5 + 2\times 3 + 3\times 4|}{\sqrt{(2)^2 + (3)^2}}$

$$d_{11} = \frac{13}{\sqrt{13}} = \sqrt{13} = 3.6$$

S-2 :- $d_{12} = \dfrac{|-5 + 2\times 2 + 3\times 3|}{\sqrt{(2)^2 + (3)^2}}$

$$d_{12} = 2.21$$

$\underline{S-3}$:- $d_{13} = \dfrac{|-5 + 2\times 1 + 3\times(-1)|}{\sqrt{13}}$

$d_{13} = \dfrac{6}{\sqrt{13}} = \cancel{3.46}\ 1.66$

$\underline{S-4}$:- $d_{14} = \dfrac{|-5 + 2\times(-2) + 3\times(+1)|}{\sqrt{13}}$

$d_{14} = \cancel{3.46}\ 1.66$

So, 1st hyperplane's margin 'M_1' is $\underline{1.66}$ (~~S-3~~ S-4)

$\underline{\text{2nd hyperplane :-}}$

$\underline{S-1}$:- $d_{21} = \dfrac{|7 + (-1)\times(3) + 4\times 4|}{\sqrt{(-1)^2 + (4)^2}}$

$d_{21} = \dfrac{20}{\sqrt{17}} = 4.85$

$\underline{S-2}$:- $d_{22} = \dfrac{|7 + (-1)\times(2) + 4\times(3)|}{\sqrt{17}}$

$d_{22} = \dfrac{17}{\sqrt{17}} = 4.12$

$\underline{S-3}$:- $d_{23} = \dfrac{|7 + (-1)\times(1) + 4\times(-1)|}{\sqrt{17}}$

$d_{23} = \dfrac{2}{\sqrt{17}} = 0.48$

$\underline{S-4}$:- $d_{24} = \dfrac{|7 + (-2)\times(-1) + 1\times 4|}{\sqrt{17}}$

$d_{24} = \dfrac{13}{\sqrt{17}} = 3.15$

$\underline{M\text{-}2 = 0.48 \ (s\text{-}3)}$

3rd hyperplane :-

S-1 :- $d_{31} = \dfrac{|10 + 5\times 3 + (-12)(4)|}{\sqrt{(5)^2 + (-12)^2}}$

$d_{31} = \dfrac{23}{13} = 1.76$

S-2 :- $d_{32} = \dfrac{|10 + 5\times 2 + (-12)(3)|}{13}$

$d_{32} = \dfrac{16}{13} = 1.23$

S-3 :- $d_{33} = \dfrac{|10 + 5\times 1 + (-12)(-1)|}{13}$

$d_{33} = \dfrac{27}{13} = 2.07$

S-4 :- $d_{34} = \dfrac{|10 + 5\times(-2) + (-12)(1)|}{13}$

$d_{34} = \dfrac{12}{13} = 0.92$

$\underline{M\text{-}3 = 0.92 \ (s\text{-}4)}$

∴ Out of all the three margins, M-1 is the highest, so 1st hyperplane is the optimum one.