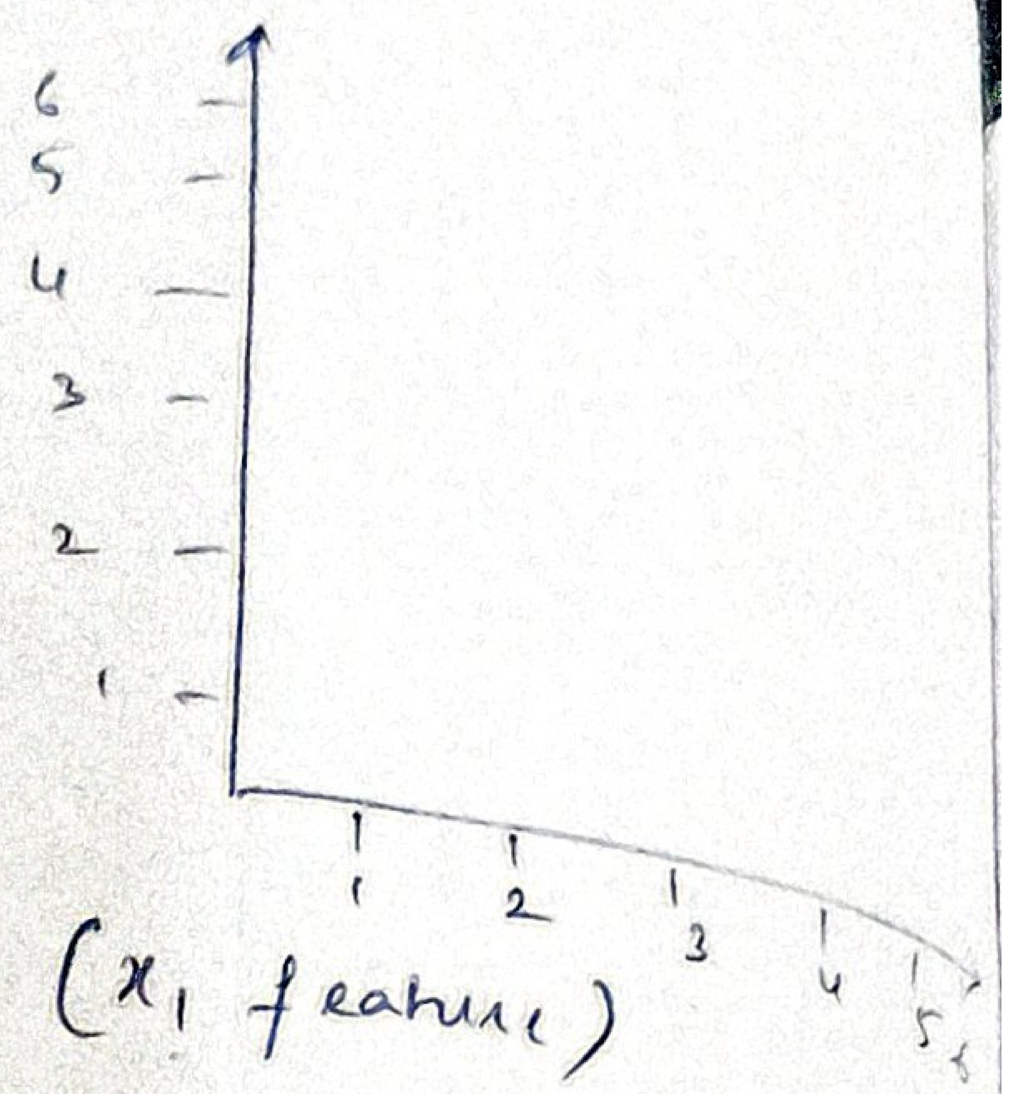
Goal D.T in Regression Setting

	x_1	x_2	y
1	1	2	1
2	3	6	5
3.	5	1	2

$I = 1, 2, 3$

$K = 2$

$$\boxed{\begin{array}{l} S_1 = 2, 4 \\ S_2 = 1.5, 4 \end{array}}$$



For $\boxed{s = 2}$, For (x_1 feature)

$E^{+}_{x_1, 2}$

I^+	I^-
$1 > 2 \rightarrow N$	$1 < 2 = T$
$3 > 2 \rightarrow T$	$3 < 2 = F$
$5 > 2 \rightarrow T$	$5 < 2 = F$

So :

$\boxed{>}$

I^+ Yes / No I^-

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

Index = 1, 2, 3

∴ Take a mean of [y] predict of 2 & 3 index.

$\hat{y}^{+} = \frac{5+2}{2} = \frac{7}{2} = 3.5$ [Avg]

$(y - \hat{y}^+)^2$ - $\hat{y}^- = 1/1 = 1$

$= (5 - 3.5)^2 + (2 - 3.5)^2 + (1-1)^2$

$= 2.25 + 2.25$

$= 4.5$

$$\boxed{E^{(+)(-)}_{x_1, 2} = 4.5}$$

For $s_1 = 4$ (x_1 feature)

$(x_2 ; 4)$

I^+	I^-
$1 > 4 = F$	$1 < 4 = T$
$3 > 4 = F$	$3 < 5 = T$
$5 > 4 = T$	$5 < 2 = F$

$\boxed{I^+ > I^-}$

Yes / No.

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

Take mean I^+ & I^-

$\hat{y}(\text{hat})^+ = 2/1 = 2$

$\hat{y}^- = 1+5 = 6/2 = 3$ } Avg

$= (2-2)^2 + (5-3)^2 + (1-3)^2$

$= 0 + 4 + 4$

$= 8.$

$$\boxed{E_{x_2, 4} = 8}$$

$S = \frac{1.5, 4}{2}$

2.5

I^+	I^-
1.5 — F	$2 < 2.5 \to$ F
$2.5 \to$ T	$6 < 2.5 \to$ F
2.5 F	$1 < 2.5 \to$ T.

$\boxed{I^+, I^-}$

Yes → $\{1, 2\}$ (I^+)

No → $\{3\}$. $I^{(-)}$

Take mean of y for each.

$\hat{y}^+ = \frac{I[1]\ [2]}{} = \frac{1 + 5}{2} = \frac{\cancel{6}\ 3}{\cancel{2}} = 3$

$\hat{y}^- = 2/1 = 2$

$E_{x, 1.5} = (3-3)^2 + (5-3)^2 + (2-2)^2.$

$= (2)^2 + (2)^2 + (0)^2.$

$= 4 + 4 + 0$

$= 8.$

$\boxed{E_{x,1.5} = 8.}$

$E_{x_2, 4}$

I^+	I^-
$2 > 4 \to$ F	$2 < 4 \to$ T
$6 > 4$ — T	$6 < 4 \to$ F
$1 > 4$ — F	$1 < 4 \to$ T

$\boxed{I^+}$ I^-

Yes → $\{2\}$

$\{1, 3\}$.

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

$\hat{y}^- = 1 + 2 = 3/2 = 1.5$.

$E_{x_2, 4} = (5-5)^2 + (1-1.5)^2 + (2-1.5)^2.$

$= 0 + (0.5)^2 + (0.5)^2.$

$= 0.25 + 0.25$

$= \underline{0.5}$

$\boxed{E_{x_2, 4} = 0.5}$

$J^*, S^* = \arg \min_{\text{find } n.} E_j, S.$

$J^* S^* = [x_2, 4)$

$\boxed{x_2 \geq 4}$

No → $\boxed{1.5}$

Yes → $\boxed{5 > 2}$

$K = 2$

$|I^-| = 2$

$|I^+| = 1$

We end up $\boxed{\text{As } K = 2}$

Eg: 6 +ve | 6 -ve. Sample
$n(+p)$ | $n(-ve)$
$p = 6$ | $n = 6.$

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

- Base two algor → bcy Two bit.

$$= \frac{-p}{p+n} \log_2 \frac{p}{p+n} - \frac{n}{p+n} \log_2 \frac{p}{p+n}.$$

$$- \sum_{i=1}^{K} -P_i \log_2 P_i$$

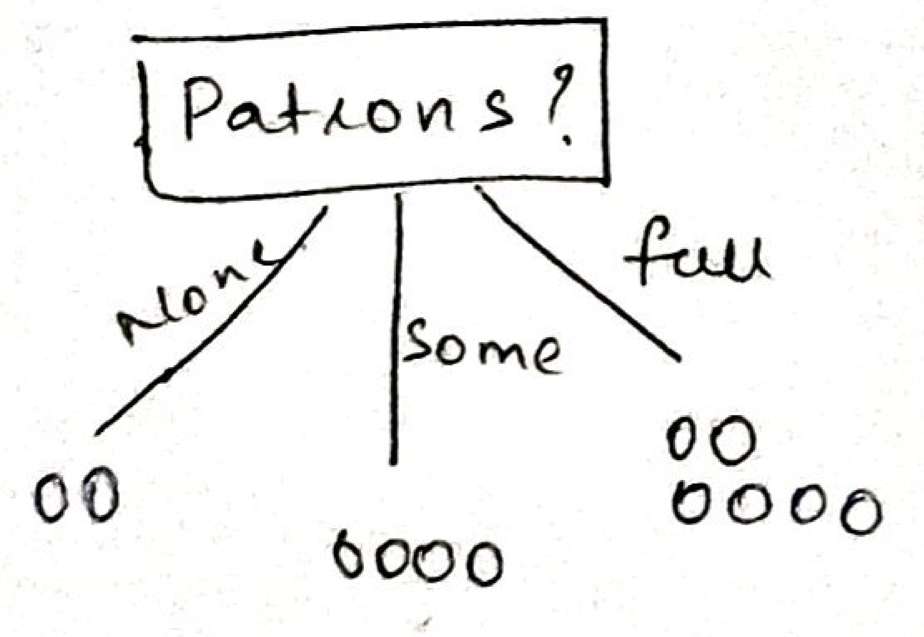
$$H\left[\frac{6}{6+6}\right] = \frac{-6}{6+(-6)} \log_2 \frac{6}{6+(-6)} - \frac{(-6)}{6+(-6)} \log_2 \frac{6}{6+6.}$$

$$= \underline{\underline{1}}$$

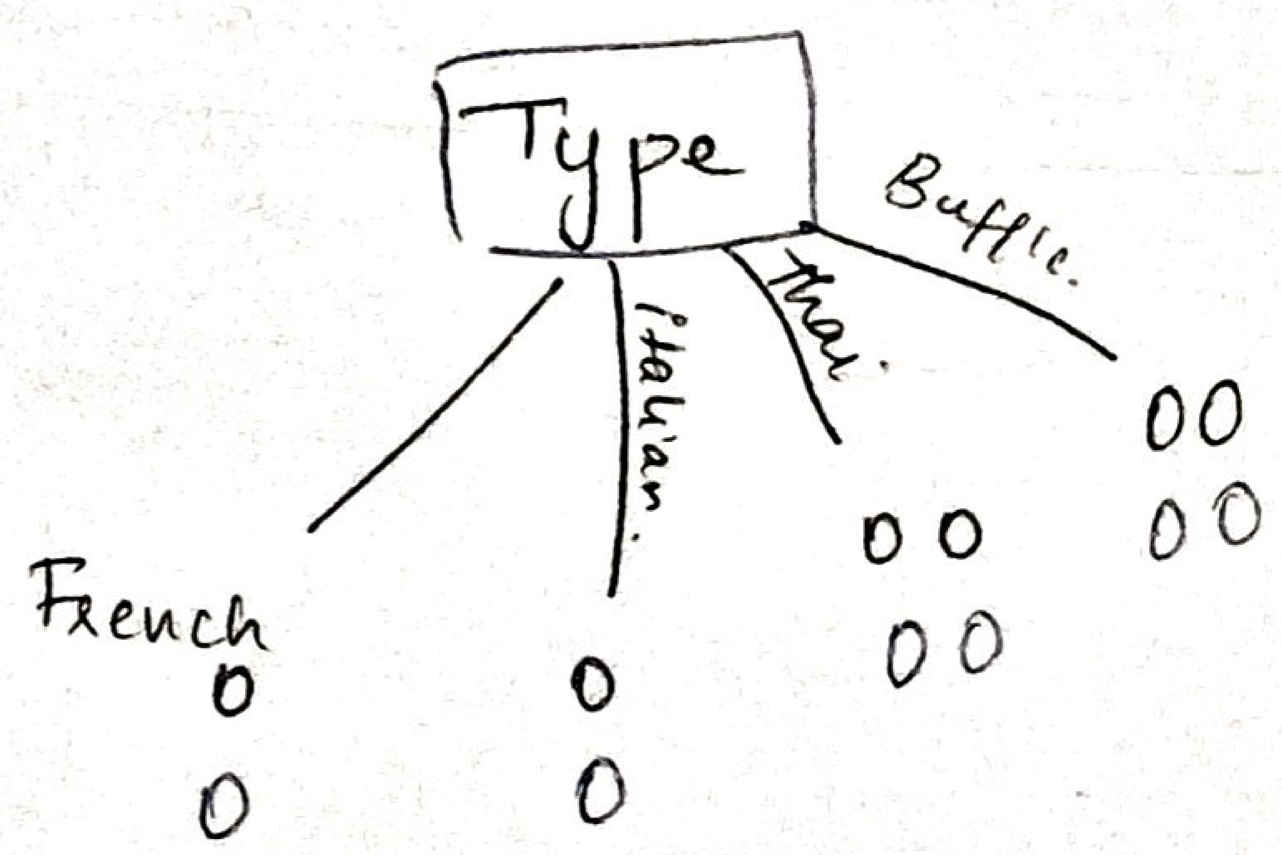
$\underline{IG = 0}$

eg:

000000 +ve
000000 -ve



×
000000
000000



Eentrop $E(A) = 1$

expected entropy.
↓
what will be the value.
Probabity

$$EH(A) = \sum_{i=1}^{K} -\left(\frac{P_i + n_i}{P+n}\right) H\left(\frac{P_i}{P_i + n_i}, \frac{n_i}{P_i + n_i}\right)$$

$$I(G) = H\left(\frac{P}{P+n}, \frac{n}{P+n} - EH(A)\right) \rightarrow \text{How much.}$$

↓
information Gain

We will maximize the $I[G]$ → is reduce [...]
if $E[A]$ → is high [purity in parent is low].

i.e. maximize the entropy gain. Feb 3 | 02 | 2025

information gain : entropy $H(\text{Patrons})$ − entropy $H(\text{Patrons})$.

$$H(\text{patrons}) = H\left(\frac{p}{p+n}, \frac{n}{p+n}\right)$$

$$= H\left(\frac{6}{12}, \frac{6}{12}\right)$$

$$EH(\text{patrons}) = \sum_{i=1}^{3}\left(\frac{p_i+n_i}{p+n}\right) H\left(\frac{p_i}{p_i+n_i}, \frac{n_i}{p_i+n_i}\right)$$

$$= \sum_{i=1}^{3}\left(\frac{p_i+n_i}{12} H\left(\frac{0}{2}, \frac{2}{2}\right) + \frac{p_2+n_2}{12} H\left(\frac{4}{4}, \frac{0}{4}\right) + \frac{p_3+n_3}{12}\right.$$

$$\frac{2}{6}, \frac{4}{6}.$$

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

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

$$\frac{6}{12}\left(1\log 3 - 2\log 3\right)$$

$$= 0.166(0-0) + 0.33 + 0.5(0.477 - 0.954).$$

$$(-0.23)$$

$$= \underline{\underline{0.46}}.$$

$$IG(P) = H(P) - EH(P)$$

$$1 - 0.466 = 5.41.$$

IG = ~~5.41~~ information gain → idea of entropy

$\underline{\underline{0.545}}$

Lecture- 28

① BOOSTING REGRESSION TREE (Ada Boost Regression)

① $f(x) = 0$ $\quad r = y$.

for iteration 1 to B.

{

① fit a model $f_b(x; \theta) \longrightarrow r$.

② update

eg: 2 Two feature.

①

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

② initialize

$F(x) = 0$

$f_0(x) = 0$

③ iteration 1

$r = y$

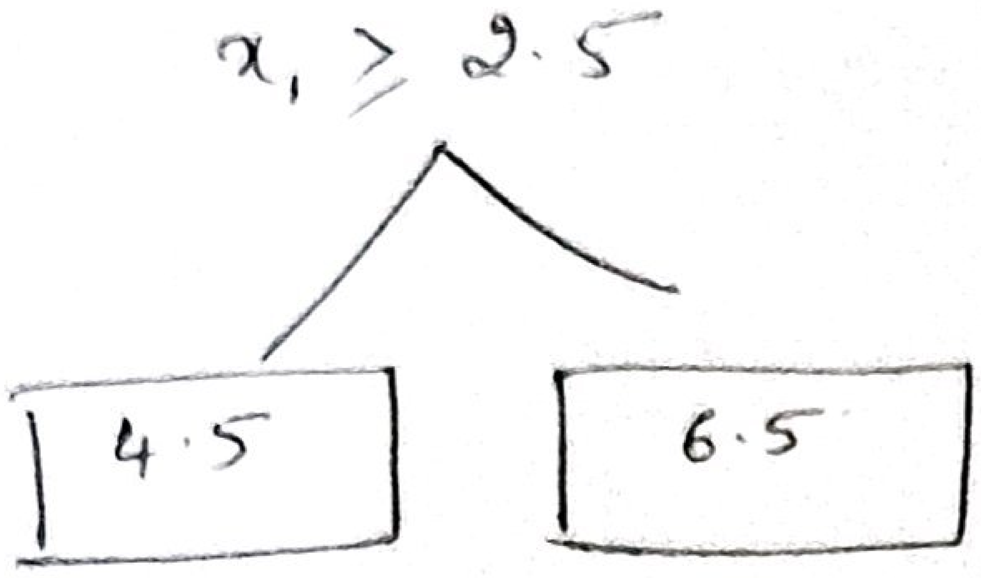
Data set

Step 1

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

$\Longrightarrow$ Fit a model

$f_1(x) \longrightarrow$ res



(b) update the $\lambda = 0.1$.

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

(c) update the residuals

$$r^{(1)} = r^{(1)} - \lambda f_1(x^{(1)}) = 4 - [0.1(4.5)] = 3.55$$
$$r^{(2)} = r^{(2)} - \lambda f_2(x^{(1)}) = 5 - [0.1(4.5)] = 4.55$$
$$r^{(3)} = r^{(3)} - \lambda f_2(x^{(1)}) = 6 - [0.1(4.5)] = 5.35$$
$$r^{(4)} = r^{(4)} - x f_2(x^{(1)}) = 7 - [0.1(4.5)] = 6.35$$

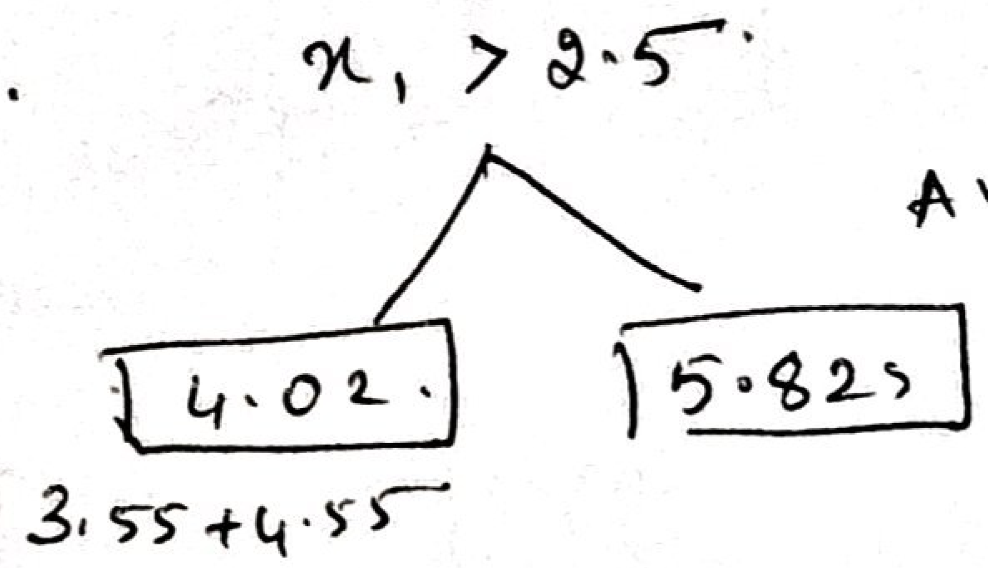
Here Residual are Reduce when compared to (r_i)

Iteration (2)

$r_1 = 3.55$
$r_2 = 4.55$
$r_3 = 5.35$
$r_4 = 6.35$

x_1	x_2	r_2
1	2	3.55
2	3	4.55
3	4	5.35
4	5	6.35

onstruct weak learner. (a) fit a model.



Avg (5.35, 6.35).

$\lambda = 0$

$$r^{(1)} = r^{(1)} - \lambda_2(x^{(1)}) = 3.145 \quad [3.55 - 0.1(4.05)]$$
$$r^{(2)} = r^{(2)} - \lambda_2(x^{(2)}) = 4.445$$
$$r^{(3)} = r^{(3)} - \lambda_2(x^{(3)}) = 4.765$$
$$r^{(4)} = r^{(4)} - \lambda_2(x^{(4)}) = 5.765$$

the r's reduce.

$$x, \theta) - f_0(x) + \lambda f_1(x) + \lambda(f_2(x) +$$
$$= 0 + 19.6 + 12.805 +$$