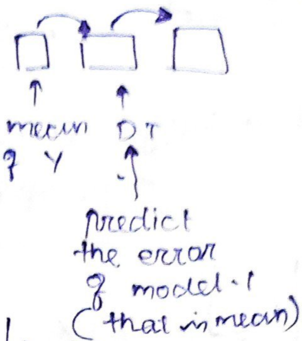
Gradient Boosting

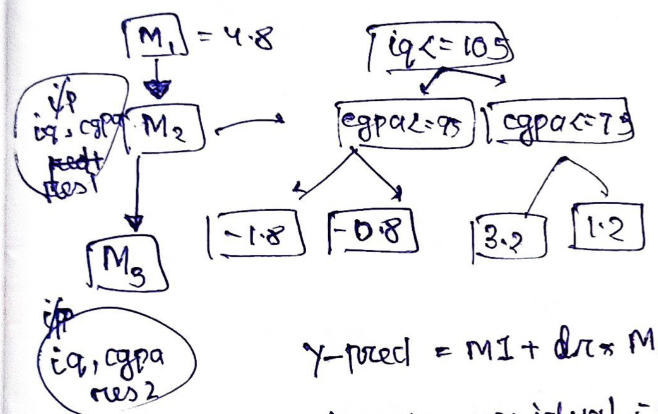
1:29 am
FRI Jun 20

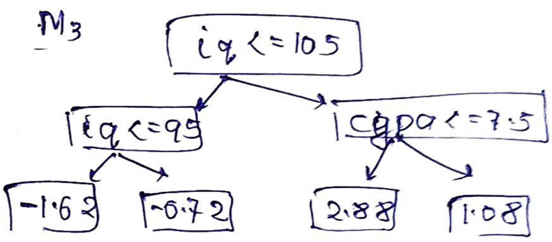


iq	cgpa	salary	pred1	res1	pred2	res2	pred3
90	8	3	4.8	-1.8	-1.8	-1.62	-1.62
100	7	4	4.8	-0.8	-0.8	0.72	-0.72
110	6	8	4.8	3.2	3.2	2.88	2.88
120	9	6	4.8	1.2	1.2	1.08	1.08
80	5	3	4.8	-1.8	-1.8	-1.62	-1.62

Actual - pred

$3 - \{4.8 + 0.1(-1.8)\} = -1.62$





y_pred = M1 + lr * M2

pseudo-residual = actual - pred

After M3

$M_1 + (0.1 \times M_2) + (0.1 \times M_3)$

$4.8 - 0.18 - 0.162 \approx 4.5$

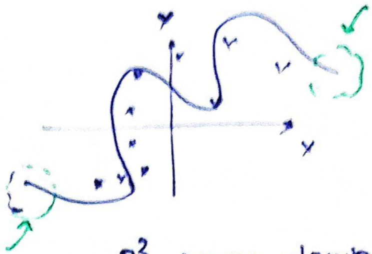
max_leaf = [8, 32]
Range give best result

Adaboost Vs Gradient Boosting

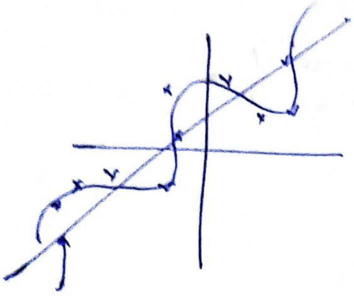
	Adaboost	Gradient Boosting
1. max leaf node	2	[8, 32]
2. learning rate	$w_1 M_1 + w_2 M_2 \ldots$ different weight	$lr\, M_1 + lr\, M_2 \ldots$ same lr

Additive modeling $\quad y = f(x)$

polynomial → Runge's phenomenon
edge up & down



R^2 score down



$y = x$
$y = \sin(x)$ } → composite function
stage wise
function creation

$f(x) = x + \sin(x)$

$$f_0(x) = \arg\min_{\gamma} \sum_{i=1}^{n} L(y_i - \gamma)$$

$$f_0(x) = \arg\min_{\gamma} \frac{1}{2}\sum_{i=1}^{n} (y_i - \gamma)^2$$ ← find a minimum value for γ

$$\frac{d}{d\gamma} f_0(x) = \frac{d}{d\gamma} \frac{1}{2} \sum_{i=1}^{n} (y_i - \gamma)^2$$

$$= \frac{1}{2} \sum_{i=1}^{n} \frac{d}{d\gamma} (y_i - \gamma)^2$$

$$= 2 \times \frac{1}{2} \sum (y_i - \gamma) \frac{d}{d\gamma} (y_i - \gamma)$$

$$= - \sum_{i=1}^{n} (y_i - \gamma) = 0$$

$$= \sum_{i=1}^{n} (\gamma - y_i)$$

Apply on data

$$\sum_{i=1}^{3} (\gamma - y_i) = 0 \Rightarrow (\gamma - 192) + (\gamma - 144) + (\gamma - 91)$$

$$3\gamma = 192 + 144 + 91$$

$$\gamma = \frac{192 + 144 + 91}{3}$$

gamma is nothing but mean for least square our error function

Pseudo Residual

$$r_{im} = -\left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}\right]_{f = f_{m-1}}$$

row

model DT

$m = 1$

3 row in our data

r_{11} $\quad$ r_{21} $\quad$ r_{31}

for first row, first DT, pseudo residual

$$r_{11} = -\left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}\right]_{f=f_0}$$

in our case $f(x_i)$ is $\hat{y}_i$ (pred)

$$= -\left[\frac{\partial}{\partial \hat{y}_i} L(y_i, \hat{y}_i)\right]_{f=f_0} = \left[+\frac{\partial}{\partial \hat{y}_i} \frac{1}{2}(y_i - \hat{y}_i)^2\right] \qquad L = \frac{1}{2}\sum_{i=1}^{n}(y_i - \hat{y})^2$$

$$= \cancel{2} \frac{\partial}{\partial \hat{y}} \left[(y_i - \hat{y}_i)\right]_{f=f_0}$$ After differation we got this

$$= \left[y_i - f(x_i)\right]_{f=f_0}$$

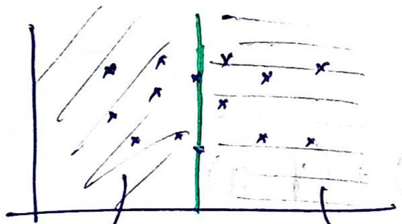
$$r_{i1} = y_i - f_0(x_i)$$

for 1st startup

$r_{11} = y_i - f_0(x_1) =$

$r_{21} = y_2 - f_0(x_2) =$

$r_{31} = y_3 - f_0(x_3) =$



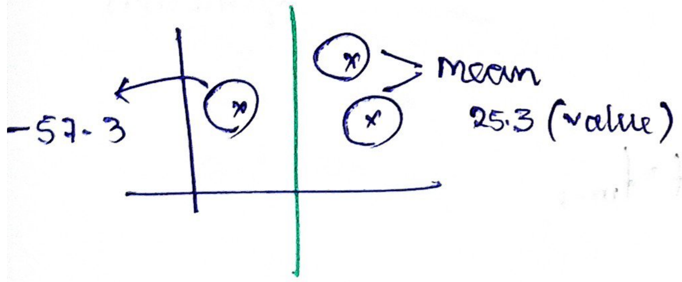
terminal-1 $\quad$ Terminal-2

R_{11} $\quad$ R_{21}

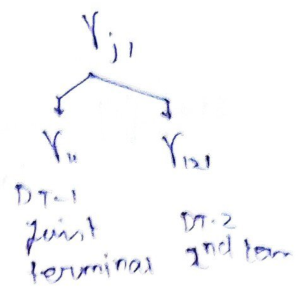
$m=1$

first decision T-1 $\quad$ 2nd DT_2 का

first terminal region $\quad$ 2nd terminal region-2



$$\gamma_{jm} = \arg\min_{\gamma} \sum_{x_i \in R_{jm}} L(y_i, f_{m-1}(x_i) + \gamma)$$



$m=1$

$$\gamma_{11} = \arg\min_{\gamma} \frac{1}{2}(y_i - (f_0(x_i) + \gamma))^2$$

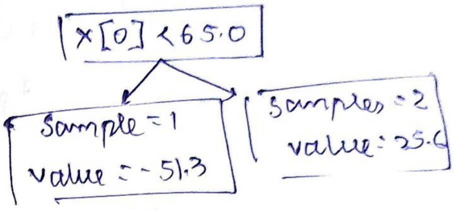
$$\frac{\partial L}{\partial \gamma} = \frac{1}{2} \times 2(y_i - (f_0(x) + \gamma)) \frac{d}{d\gamma}(y_i - f_0(x) - \gamma) = 0$$

$$= -(y_i - f_0(x) - \gamma) = 0$$

$$\gamma_{11} = y_i - f_0(x) - \gamma$$

$$= 91 - 142 - \gamma = 0$$

$$\gamma = 91 - 142 = -51$$



$$\gamma_{21} = \arg\min_{\gamma} \sum_{x_i \in R_{21}} L(y_i, f_0(x_i) + \gamma)$$

$$= \arg\min_{\gamma} \frac{1}{2}\sum_{i=1}^{2} (y_i - f_0(x_i) + \gamma)^2$$

$$= -\sum_{i=1}^{2} (y_i - f_0(x_i) - \gamma) = 0$$

$$= \sum_{i=1}^{2} (y_i - f_0(x_i) - \gamma) = 0$$

$$\Rightarrow y_1 - f_0(x_1) - \gamma + y_2 - f_0(x_2) - \gamma = 0$$

$$\Rightarrow 192 - 142 - \gamma + 144 - 142 - \gamma = 0$$

$$\Rightarrow 336 - 284 + 2\gamma = 0$$

$$\Rightarrow \gamma = \frac{52}{2} \simeq 26$$

$$f_m(x) = f_{m-1}(x) + \sum_{j=1}^{j_m} \gamma_{jm} I(x \in R_{jm})$$

$$f_1(x) = f_0(x) + DT_{output-1}$$

$$f_2(x) = f_1(x) + DT_{o/p-2}$$

3. $$\hat{f}(x) = f_M(x)$$

output check

Gradient Boosting for classification

P-3
737 pm
Sat, Jun 21 2025

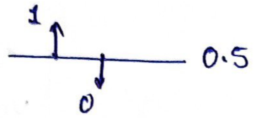
stage[1] simple model

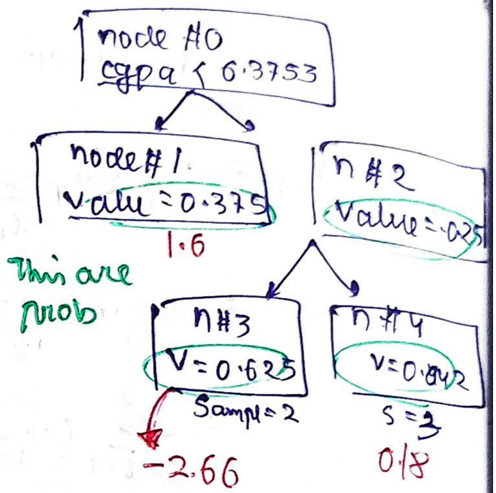
↳ $\log(\text{odds}) = \log_e\left(\frac{\#1}{\#0}\right)$

$\log_e\left(\frac{5}{3}\right) = 0.51 \qquad \log_{10}^{x}$

Probability

$$P = \frac{1}{1+e^{-\log_e(\text{odds})}} = \frac{1}{1+e^{-0.51}} = 0.62$$





pseudo residual $= y_i - f_0(x_i)$

we used regression tree

max leaf nodes = 3

Model-1 $f_0(x)$ + M2 $f_1(x)$

log(odds) probability

$$\gamma = \frac{\sum \text{Residual}}{\sum[(\text{Previous Prob} * (1 - \text{Previous Prob})]}$$

Find for node#3

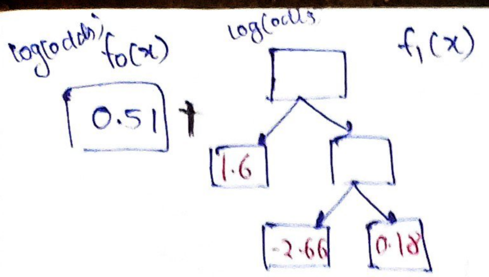
- In that sample is 2, two points

$$= \frac{-0.625 - 0.625}{0.625(1-0.625) + 0.625(1-0.625)}$$

$$= \frac{-2 \times 0.625}{0.625 \times 0.375 + 0.625 \times 0.375}$$

$$= \frac{-\cancel{2 \times 0.625}}{\cancel{2 \times 0.625} \times 0.575}$$

$$= \frac{-1}{0.375} = -2.66$$



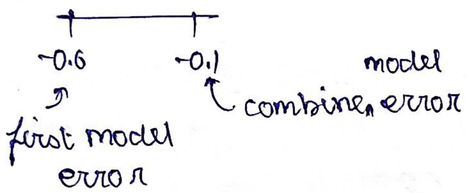
Then check each row of where that fall in leaf after that

$f_0(x) + f_1(x)$

$0.51 + (-2.66) = -2.15$

Just bec. we calculate probability for residuals

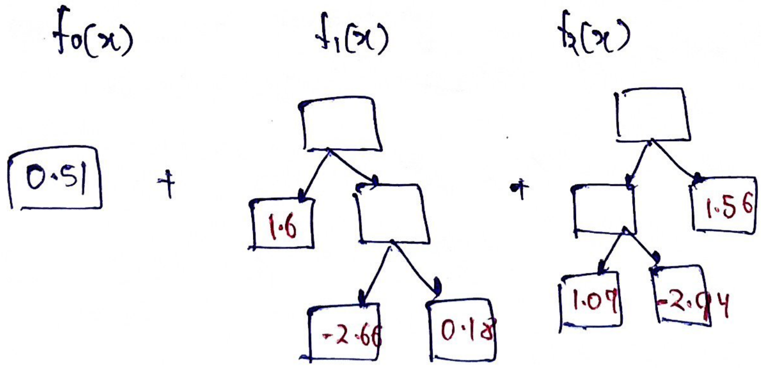
$p = \frac{1}{1+e^{-2.159}} = 0.10$



- Again find the leaf nodes apply

$$\gamma = \frac{\sum \text{Residual}}{\sum[(\text{Pre.Prob} * (1-\text{Pre.Prob})]}$$

Now combine log(odds)

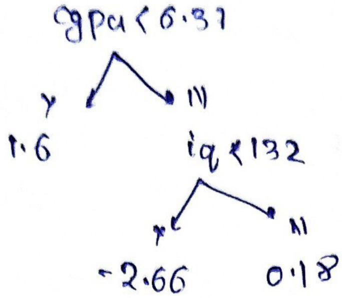


The o/p we not provide in log so we convert into probability

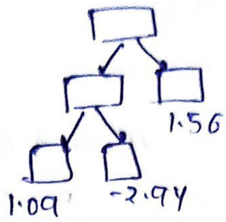
Prediction

$x_q = \{7.2, 100\}$
cgpa

$\boxed{0.51}$ +



+



$= 0.51 + (-2.66) + 1.56$

$= -0.59$

this is combine log(odds)

$$P = \frac{1}{1+e^{0.59}} = 0.35$$

if threshold is 0.5

Then $= x_q = 0$

No placement