Gradient Boosting Algorithm

Regression & Classification

Dataset

$(y-\hat{y}) \rightarrow$ Step 2

x_1	x_2	y	
Exp	Degree	Salary	R_1
2	BE	50k	-25K
3	Masters	70k	-5K
5	Masters	80k	5K
6	P.HD	100k	25K

Avg salary = 75 K

Step : 1 Create a base model

I/P → $\boxed{}$ → O/P

Avg sal = 75K = $\hat{y}$

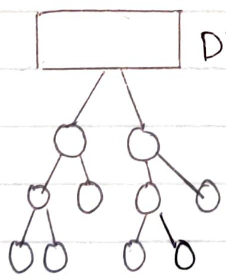
Step : 2 Compute residuals, Error

↳ $R_1 = y - \hat{y}$

Step : 3 Construct a decision tree, consider inputs x_i and outputs as R_1

Base model

$\boxed{= .75}$



$DT_1 \{x_1, R_1\}$

Decision tree 1 gives Predicted R_2

Exp	Deg	Sal	R_1	Predicted R_2	$\hat{y}$
2	BE	50	-25	-23	74.77
3	Master	70	-5	-3	74.97
5	Mas	80	5	3	
6	P.hD	100	25	23	

final output is calculated by combining prediction of base model and result of decision tree

Predicted output calculation:

$75 + (-23) = 75 - 23 = 52$ {over fitting}

Because 52 is much closer to actual o/p 50

thus,

$= 75 + \alpha(-23)$

$= 75 + -0.23$

$= 74.77$

α = learning rate = {0 to 1} = 0.

= 0.01

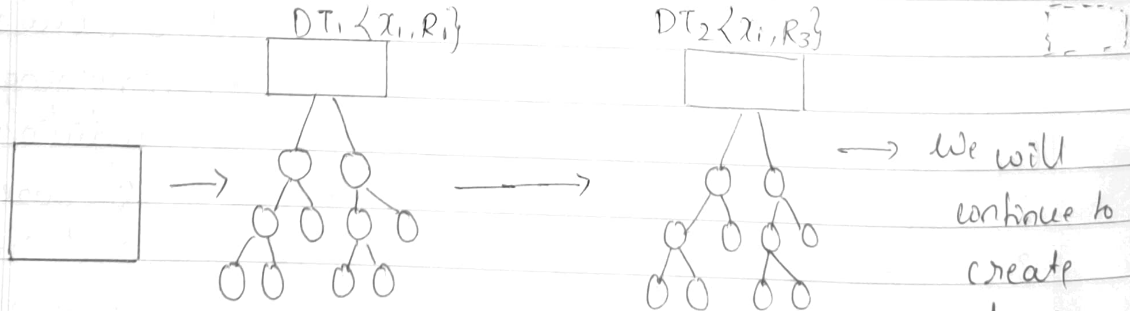
record-2 $= 75 + \alpha(-3)$

$= 75 + -0.03$

$= 74.97$

Exp	Degree	Salary	R_1	Predicted R_2	$\hat{y}$	R_3
2	B.E	50	-25	-23	74.77	-24.77
3	Master	70	-5	-3	74.97	-4.97
5	Mas	80	5	3		
6	P.hD	100	25	20		

Next decision tree



We will continue to create decision tree

Mathematical representation $\boxed{\alpha_0 = 1}$

$$F(x) = \alpha_0 h_0(x) + \alpha_1 h_1(x) + \alpha_2 h_2(x) + \cdots + \alpha_n h_n(x)$$

$h_0(x)$ = Base model

$h_1(x)$ = Decision tree

$\{\alpha_0, \alpha_1, \ldots, \alpha_n\}$ - learning rate

$\alpha = [0 \text{ to } 1]$

Xgboost Classification Algorithm

Dataset

x_1 Salary	x_2 Credit	y Approval	Error $(y - 0.5)$ R_1 (step 1)
<=50K	B	0	-0.5
<=50K	G	1	0.5
<=50K	G	1	0.5
>50K	B	0	-0.5
>50K	G	1	0.5
>50K	N	1	0.5
<=50K	N	0	-0.5

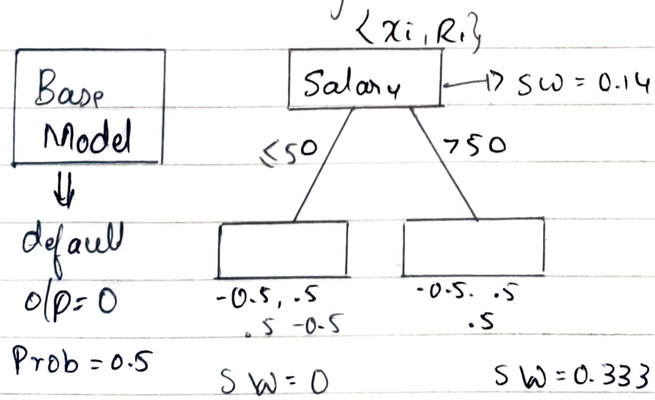
Steps

1. Construct a base model
2. Construct a decision tree with root node
3. Calculate similarity weight

$$S.W = \frac{(\sum \text{Residual})^2}{\sum \text{Prob}(1 - \text{Prob})}$$

4. Calculate gain

Step : 1 - Constructing base model



Similarity weight (left child)

$$S.W(LC) = \frac{\left(\sum \text{residual}\right)^2}{\sum Pr(1-Pr)} = \frac{[-0.5 + 0.5 + 0.5 + (-0.5)]}{0.5(0.5) + 0.5(0.5) + 0.5(0.5) + 0.5(0.5)} = 0$$

S.W of Right child

$$SW(RC) = \frac{\left(\sum \text{residual}\right)^2}{\sum Pr(1-Pr)} = \frac{[-0.5 + 0.5 + 0.5]^2}{0.5(0.5) + 0.5(0.5) + 0.5(0.5)}$$

$$= \frac{0.25}{0.75} = 0.333$$

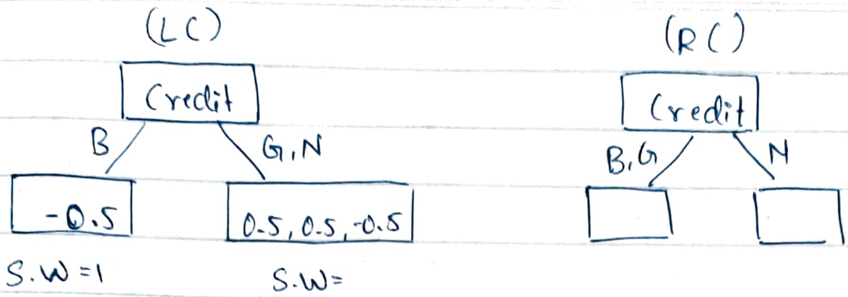
Similarity weight of root

$$S.W(\text{root}) = \frac{0.25}{1.75} = 0.14$$

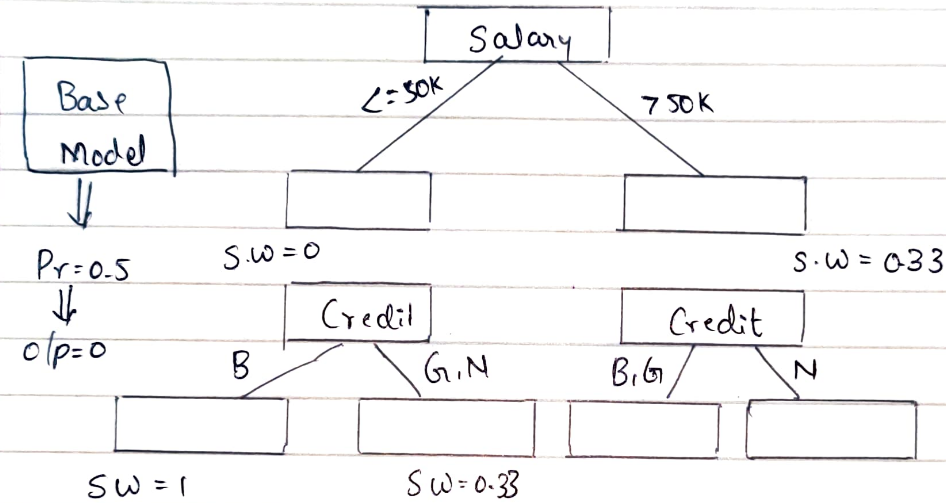
Calculating gain

$$\text{gain} = 0 + 0.33 - 0.14 = \cancel{0.21}$$

$$= 0.19$$



Final output for classification problem



log odds formula

$$\log(\text{odds}) = \log\left(\frac{P}{1-P}\right) = \log\left(\frac{0.5}{0.5}\right) = 0$$

$\log(1) = 0$

$\alpha \rightarrow$ learning rate

$\alpha = [0 \text{ to } 1]$

Base model $\rightarrow$ Test data o/p $= 0 + \alpha(1)$

$\sigma(0 + \alpha(1)) \longrightarrow$ Sigmoid activation function

$\sigma(0 + 0.1)$

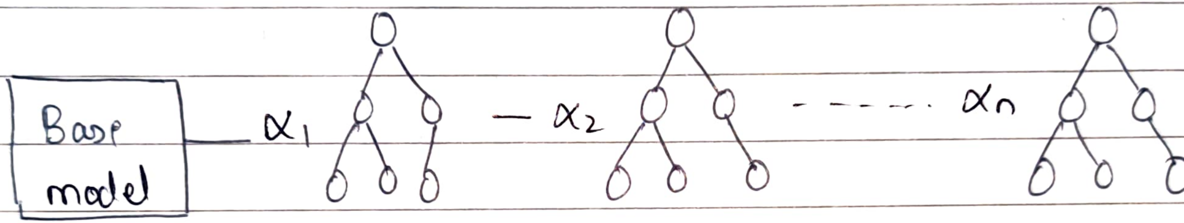
$\sigma(0.1)$ $\quad\quad \sigma = \dfrac{1}{1 + e^{-z}}$

o/p = 0.52 $\Longrightarrow$ Setting threshold = 0.6

threshold is set by Domain expert

$\boxed{0.52 < 0.6} \Longrightarrow 0$

Xg boost Summary



$$o/p = \sigma\left(\text{Base learner} + \alpha_1 (DT_1) + \alpha_2 (DT_2) + \cdots + \alpha_n (DT_n)\right)$$