XGBoost (Extreme Gradient Boosting)

→ XGBoost Classifier
→ XG Boost Regressor

* XGBoost classifier:

Salary	Credit Score	Approval (Y_{act})	Base model (Y_{baseP})	R_1 $(Y_{act} - Y_{baseP})$
<= 50k	B	0	0.5	-0.5
<= 50k	G	1	0.5	0.5
<= 50k	G	1	0.5	0.5
> 50k	B	0	0.5	-0.5
> 50k	G	1	0.5	0.5
> 50k	N	1	0.5	0.5
<= 50k	N	0	0.5	-0.5

Step-1 → Create a base model

* Since base model should be unbiased, take 0.5 here

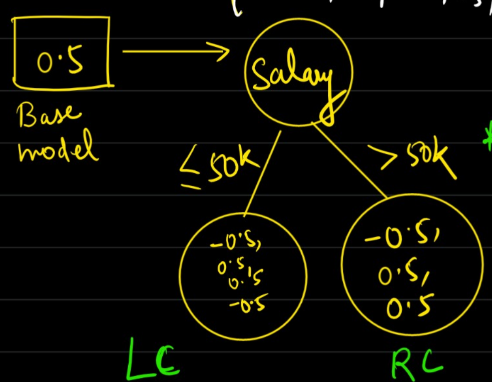
Step 2.- Construct a decision tree with root

Step-3 Calculate similarity weights

$$\text{Similarity weight} = \frac{\sum (\text{Residual})^2}{\sum Pr(1-Pr) + \lambda}, \quad \lambda = 0$$

Step 4: Highest Similarity weight gain

[-0.5, 0.5, 0.5, -0.5, 0.5, 0.5, -0.5]



$$* \; SW_{root} = \frac{[\cancel{-0.5}+\cancel{0.5}+\cancel{0.5}-\cancel{0.5}+0.5+\cancel{0.5}-\cancel{0.5}]^2}{0.5(1-0.5)+0.5(1-0.5)+0.5(1-0.5)+0.5(1-0.5)+0.5(1-0.5)+0.5(1-0.5)+0.5(1-0.5)}$$

$$= \frac{(0.5)^2}{0.25+0.25+0.25+0.25+0.25+0.25+0.25}$$

$$= \frac{0.25}{1.75} = 0.14$$

$$* \; SW.\text{Left child} = \frac{(\cancel{-0.5}+\cancel{0.5}+\cancel{0.5}-\cancel{0.5})}{0.5(1-0.5)+0.5(1-0.5)+0.5(1-0.5)} = 0$$

$$* \; SW\text{ Right child} = \frac{(\cancel{-0.5}+0.5+\cancel{0.5})^2}{0.5(1-0.5)+0.5(1-0.5)+0.5(1-0.5)}$$

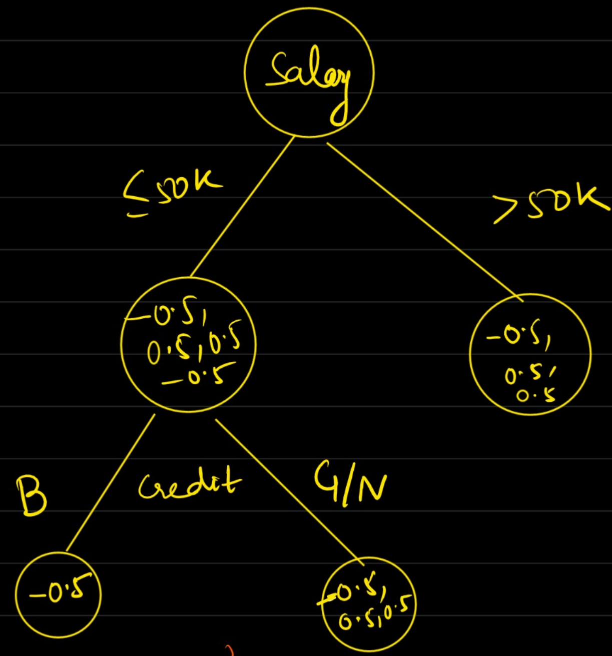
$$= \frac{0.25}{0.75} = \frac{1}{3} = 0.33.$$

$$SW_{gain} = (\text{Leftchild} + \text{Rightchild}) - \text{Root}$$

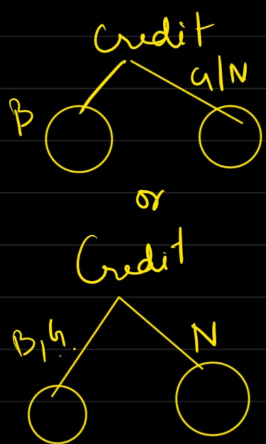
$$= 0 + 0.33 - 0.14 = 0.19$$

* Whichever feature has the highest similarity weight gain, the split will happen on that feature.

* Here assuming Salary column is giving similarity weight gain



Two possibilities



⇓

based on similarity gain score

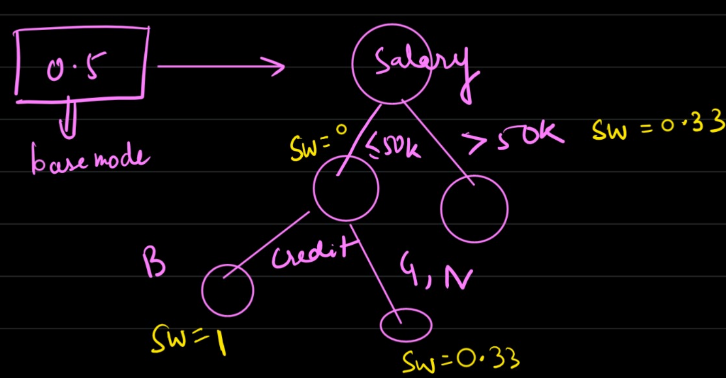
$SW(LC) = \frac{(-0.5)^2}{0.5(1-0.5)} = \frac{0.25}{0.25} = 1$ (Pure node)

$SW(RC) = \frac{(\cancel{-0.5} + \cancel{0.5} + 0.5)^2}{0.5(1-0.5)+0.5(1-0.5)+0.5(1-0.5)} \Rightarrow 0.25/0.75 = 0.33$

$SW(Root) = \frac{(\cancel{-0.5} + \cancel{0.5} + \cancel{0.5} \cancel{-0.5})}{(0.25+0.25+0.25+0.25)} = 0$

$SW_{gain} = 1 + 0.33 - 0) = \underline{1.33}$

* How to make prediction ?



⇛ In Gradient Boosting

$f(x) = \text{base model} + \alpha_1 M_1 + \alpha_2 M_2 + \alpha_3 M_3 + \cdots \alpha_n M_n$

XGB classifier

$O/P = \sigma(\text{Base model} + \alpha(SW))$

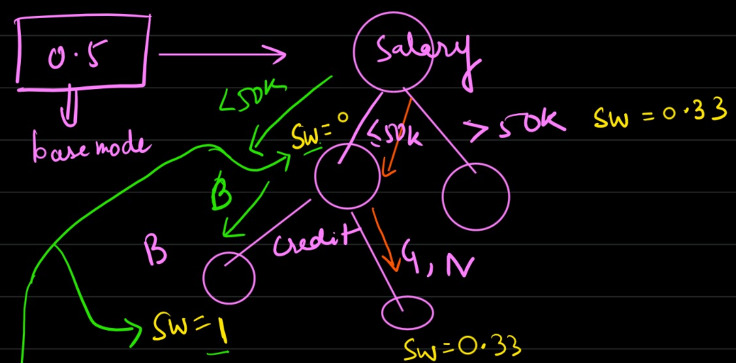
$\sigma \rightarrow \frac{1}{1+e^{-x}}$

Since its a classification problem, we need to Sigmoid fn.

⇓

gives the value b/w 0 & 1

* first row prediction $(\leq 50K, B)$



range of sigmoid is 0 to 1

* Base model $= \log\left(\frac{P}{1-P}\right)$ (log odds)

$$O/P = \sigma(\text{Basemodel} + \alpha(SW))$$

$$= \sigma\left(\log\left(\frac{0.5}{1-0.5}\right) + \alpha(0+1)\right)$$

say - 0.1

$$= \sigma(0 + 0.1 \times 1)$$

$$= \frac{1}{1+e^{-0.1}} = \underline{\underline{0.52}}$$

2nd row

$(\leq 50K, G)$

$$O/P = \sigma(0 + \alpha(0.33))$$

$\alpha = 0.1$

$$= \frac{1}{1+e^{-0.33}} = 0.508$$

Step 4 * Now do this for all rows calculate $\hat{y}$

	Credit Score	Approv	Base model	(R_1) $(y_{act} - y_{base\,P})$	$\hat{y}$
<=50k	B	0	0.5	-0.5	0.52
<=50k	G	1	0.5	0.5	0.508
<=50k	G	1	0.5	0.5	-
>50k	B	0	0.5	-0.5	-
>50k	G	1	0.5	0.5	-
>50k	N	1	0.5	0.5	-
<=50k	N	0	0.5	-0.5	-

Step 5 * Again Construct DT will all and features and $\hat{y}$ as target variable

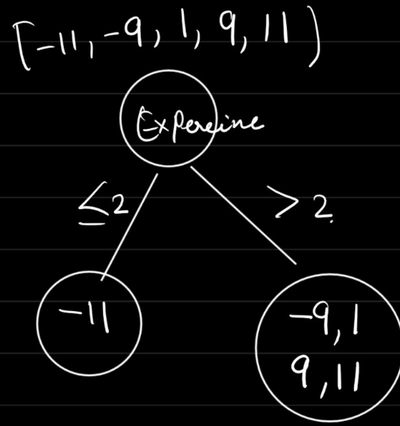
Till required

of tree is achieved.

$$O/P = \sigma\left(\text{Base learner} + \alpha_1(DT_1) + \alpha_2(DT_2) \ldots \alpha_n(DT_n)\right)$$

* XGBoost Regressor

Exp	Gap	Salary	Base Prediction	Res 1
→ 2	Yes	40	51	-11
→ 2.5	Yes	42	51	-9
3	No	52	51	1
4	No	60	51	9
4.5	Yes	62	51	11



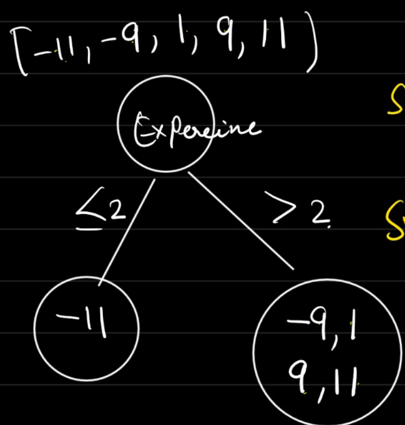
$$\text{Base Prediction} = \frac{(40+42+52+60+62)}{5} = 51.2 \approx \underline{\underline{51}}$$

* SW in case of classification

$$= \frac{(\Sigma \text{Residual})^2}{\Sigma Pr(1-Pr) + \lambda}$$

* SW in case of regression

$$= \frac{(\Sigma \text{Residual})^2}{\text{No of residuals} + \lambda}$$



$$SW_{Root} = \frac{(-\cancel{11} + -\cancel{9} + 1 + \cancel{9} + \cancel{11})^2}{5+1} = \frac{1}{6}$$

$$SW_{LC} = \frac{(-11)^2}{1+1} = \frac{121}{2} = 60.5$$

$$SW_{RC} = \frac{(-\cancel{9} + 1 + \cancel{9} + 11)^2}{4+1} \quad \left(\text{Here taking } \underline{\underline{\lambda = 1}}\right)$$

$$= \frac{144}{5} = \underline{\underline{28.8}}$$

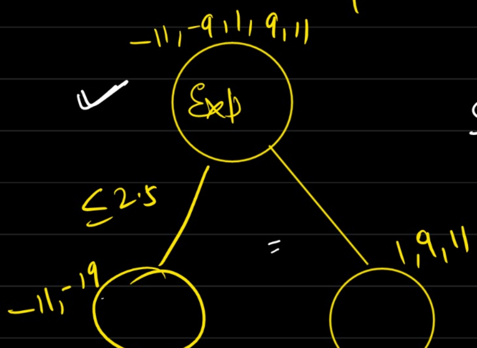
$$SW\ Gain = (60.5 + 28.8) - 0.166$$

$$= \underline{\underline{89.134}}$$

* Split will be based on highes SW gain

*

✓



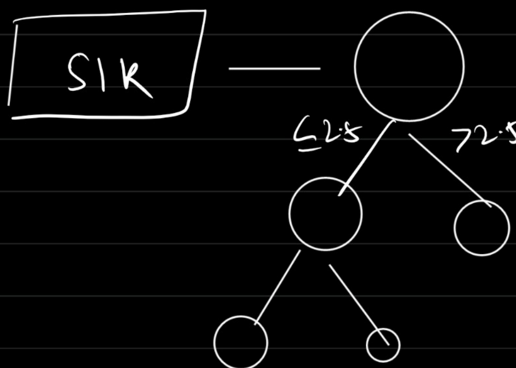
$$SW_{LC} = \frac{(-11-9)^2}{2+1} = \frac{400}{3} = 133.3$$

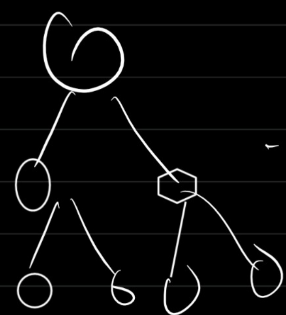
$$SW_{RC} = \frac{(1+9+11)^2}{3+1} = \frac{(21)^2}{4} = 110.23$$

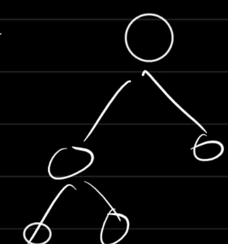
$$SW_{root} = 0.166$$

$$SW_{gain} = 133.3 + 110.23 - 0.166$$

$\simeq 244 \Rightarrow$ Highest gain, So will split on this row





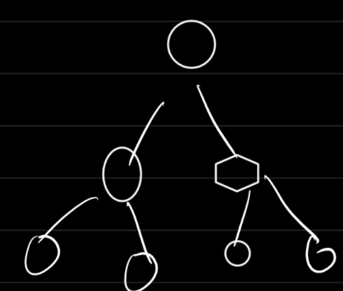


until the required no of dt is made.

$$\text{Prediction} = 51 + \alpha_1(SW) + \alpha_2(\ ) + \dots \alpha_n(DT_n)$$

* $\underline{SW} = \dfrac{(\Sigma \text{ Residual})^2}{\Sigma Pr(1-Pr) + \lambda}$ $\rightarrow$ Hyperparameter.

One use of λ $\rightarrow$



$\underbrace{\lambda \uparrow \ SW \downarrow}$
$\Downarrow$
Chances are there similarity weight gain can be negative $\longrightarrow$ Stop splitting

* $\Sigma Pr(1-Pr) \Rightarrow$ lower value.

$\Uparrow$

any SW less than $\Sigma Pr(1-Pr)$ we stop splitting.

Say $0.5(1-0.5) = 0.25$ if whole SW is less 0.25, stop splitting.