XG Boost (Extreme Gradient Boosting)

→ XGBoost Classifier
→ " Regressor.

Step-1 → Create a base model.

(0.5)

XG Boost classifier

Salary	Credit Score	Approval (y)	Base model	R_1
<=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

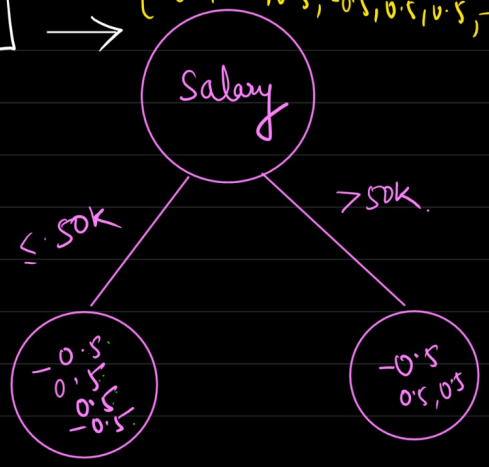
Step-2 Construct a decision tree with root.

Step-3 Calculate Similarity weight.

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

$\lambda = 0$

0.5 → [-0.5, 0.5, 0.5, -0.5, 0.5, 0.5, -0.5)



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

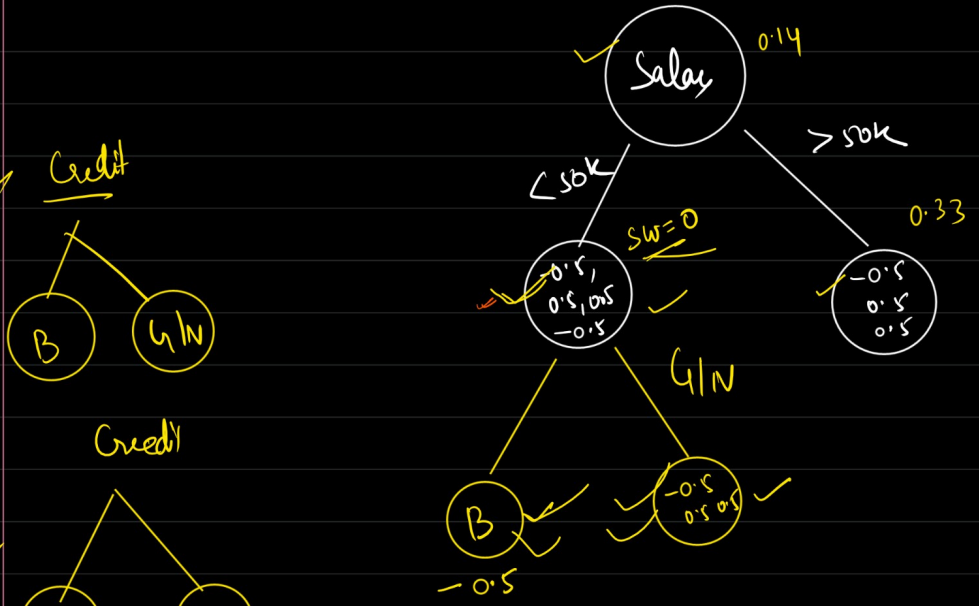
$$SW_{root} = \frac{(0.5)^2}{0.25 \times 7} = \frac{0.25}{1.75} = 0.14$$

$$SW_{LC} = \frac{(-0.5+0.5+0.5-0.5)^2}{(\quad)} = 0$$

$$SW_{RC} = \frac{(-0.5+0.5+0.5)^2}{0.25 \times 3} = \frac{0.25}{0.75} = 0.33$$

$$SW_{gain} = (SW_{LC} + SW_{RC}) - Root$$
$$= (0 + 0.33) - 0.14$$
$$= 0.19.$$

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



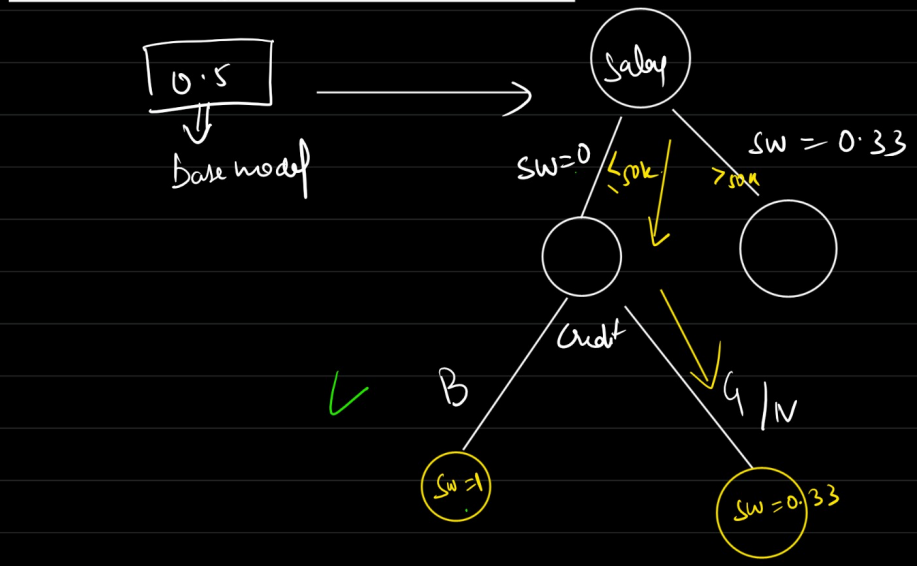
$$SW(LC) \rightarrow \frac{(-0.5)^2}{0.5(1-0.5)} = \frac{0.25}{0.25} = 1 \text{ (Pure node)}$$

$$SW(RC) = \frac{(-0.5+0.5+0.5)^2}{0.25 \times 3} = \frac{0.25}{0.75} = 0.33$$

$$SW(Root) = \frac{(-0.5+0.5+0.5-0.5)^2}{(\quad)} = 0$$

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

* How prediction is made



< 50k, B

$$o/p = \sigma(b_{made} + \alpha(sw))$$
$$= \sigma\left(\log\left(\frac{0.5}{1-0.5}\right) + \alpha(0+1)\right)$$
$$= \sigma(\log(1) + 0.1 \times 1)$$
$$= \frac{1}{1+e^{-0.1}} = 0.52$$

2nd row <= 50k, G

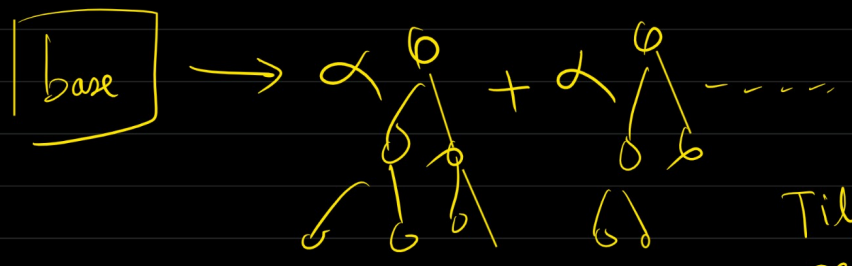
$$o/p = \sigma(b_{mod} + \alpha(sw))$$
$$= \sigma(0 + 0.1(0+0.33))$$
$$= \frac{1}{1+e^{-0.033}} = 0.501$$

XGBclassifier $\Rightarrow$ $o/p = \sigma(\text{base model} + \alpha(sw))$

σ: $\frac{1}{1+e^{-x}}$; base model: $\log\frac{p}{1-p}$

Salary	Credit Score	Approval	Base model	R_1	$\hat{y}_{pred}$
<=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-3 Again construct dT with all features $\hat{y}$ as target values



Till required no of Trees are achieved

$$O/P = \sigma\left(Base + \alpha_1 DT_1 + \alpha_2 DT_2 \dots \alpha_n(DT_n)\right)$$