Xgboost ML Algorithm (Classification)

$\hat{y} = 0.5$

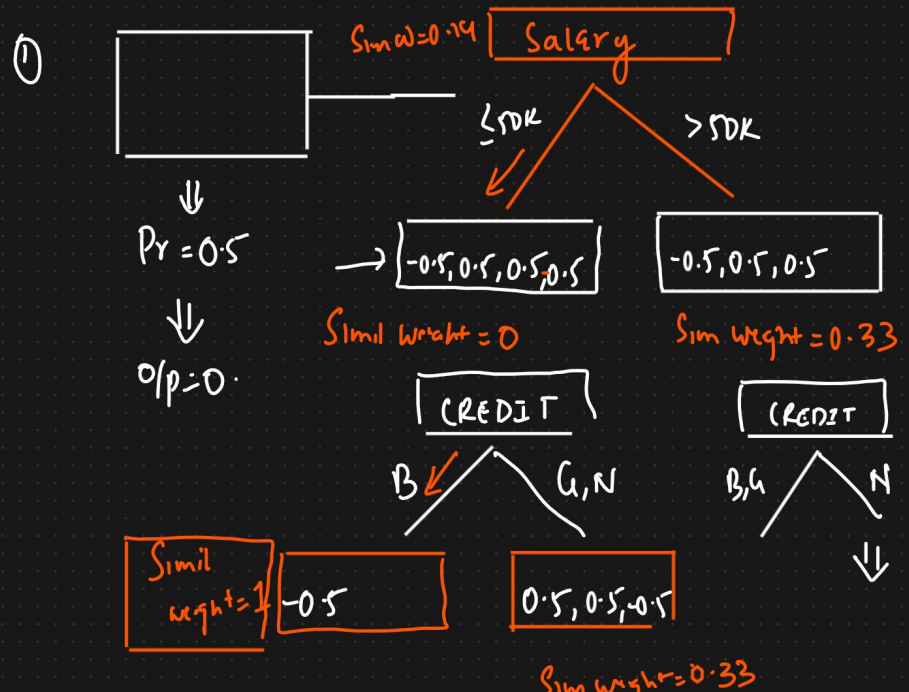
Dataset

Salary	Credit	y Approval	R1	$\hat{y}$	R2
<=50K	B	0	-0.5	0.52	-0.48
<=50K	G	1	0.5	0.58	0.42
<=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.
3. Calculate Similarity Weight
   $= \sum (\text{Residual})^2$
   Cover value $\leftarrow \sum Pr(1-Pr) + \lambda$ (hyperparameter)
4. Calculate Gain

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



Test data

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

$$\log(\text{odds}) = \log\left(\frac{0.5}{0.5}\right)$$

$$= 0$$

$$\text{Similarity (Root)} = \frac{0.25}{0.25} = 1$$

$$\text{Similarity (Rest)} = \frac{0.25}{0.75} = 0.33$$

$$\text{Gain} = 1 + 0.33 - 0 = \boxed{1.33}$$

Similarity weight $(\lambda) = \frac{\Sigma (\text{Residual})^2}{\Sigma P_r(1-P_r)}$

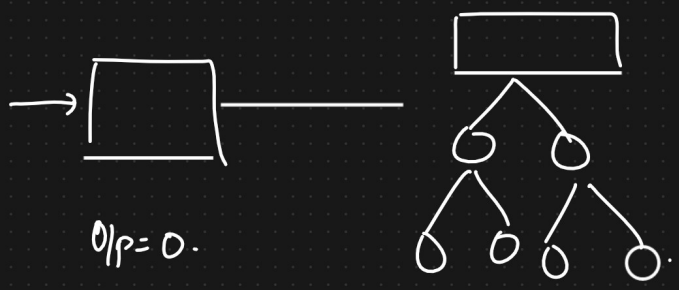
Similarity (RC) $= \frac{(-0.5+0.5+0.5)^2}{0.75} = \frac{0.25}{0.75} = \frac{1}{3}$

$= [(-0.5+0.5+0.5-0.5)^2] = 0$

$1 \Leftarrow [0.5(1-.5)+0.5(1-.5)+ .5(1-.5)+0.5(1-.5)]$

Gain $= 0 + 0.33 - .14 = .21$

Final O/p



New Data $= \sigma\left(0 + \lambda(1)\right)$ ← (Similarity weight), ↓

$\lambda = 0.1$

Logistic function ⇐ Sigmoid Activation $= \sigma\left(0 + (0.1)(1)\right)$

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

Second Record O/p $= \sigma\left(0 + \lambda(0.33)\right)$

$= \sigma\left(0 + 0.1(0.33)\right)$

$$= \frac{1}{1+e^{-0.033}} = 0.508$$

Xgboost classifier

$$O/P = \sigma\left(\text{Base learner} + \alpha_1(DT1) + \alpha_2(DT2) + \alpha_3(DT3)\right).$$