

XGBoost Classifier

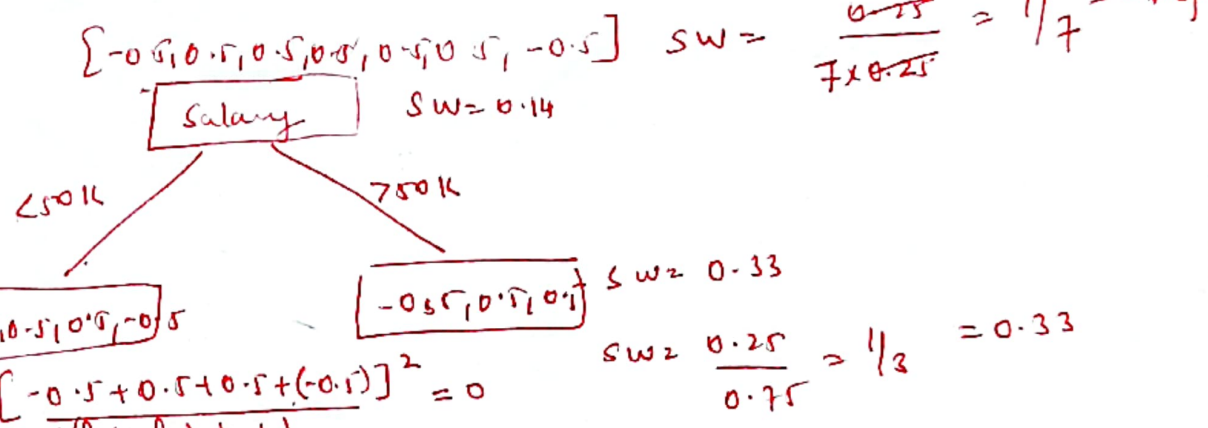
→ updated $\mu = 0.6$

	Credit	Approval	Res	Pr	Res	Base model
Salary						$Pr = 0.5$
$\leq 50K$	B	0	-0.5	1	1	
$\leq 50K$	G	1	0.5	1	1	
$\leq 50K$	G	1	0.5	1	1	
$\leq 50K$	G	0	-0.5	1	1	
$> 50K$	B	1	0.5	1	1	
$> 50K$	G	1	0.5	1	1	
$> 50K$	N	1	0.5	1	1	
$> 50K$	N	0	-0.5	1	1	

$\sigma(0 + (0.1)(1))$ \xrightarrow{SW} $\sigma(0.1) = \frac{1}{1 + e^{-0.1}} = 0.6$

$Pr = \sigma[\alpha(T_1) + \alpha(T_2) + \dots + \alpha(T_n)]$
 Construct a DT

Learning Rate



① Construct tree with Root \rightarrow gamma

② Calculate $SW = \frac{\sum (Residual)^2}{\sum (Pr(1-Pr) + 1)}$

$\log(Odds) = \log\left(\frac{p}{1-p}\right)$

$\sum (Pr(1-Pr) + 1)$ \rightarrow hyperparameter

③ Calculate gain

$gain = (SW)_{left} + (SW)_{right} - (SW)_{root}$
 $= 0.21$

We compare gain of different features. and choose one for which we get most of the gain

Post pruning in XGB classifier = $\frac{Pr(1-Pr)}{\text{cover value}} > \frac{\text{gain}}{\text{Post pruning}} \rightarrow \text{cut the branch}$

$$Pr = \sigma \left[\log \left(\frac{p}{1-p} \right) + \alpha (fW_1) \right] + \alpha (fW_2) + \dots$$

$$Pr = \sigma \left[\log \left(\frac{0.5}{0.4} \right) + \alpha (1) \right]$$

$$= \sigma \left[0 + 0.1(1) \right] = \frac{1}{e^{0.1} + 1} = 0.6$$

XG Boost CART

Imbalanced dataset

Balanced and Imbalanced

Imbalanced dataset may lead to provide a biased output for majority of the cases.

① Down Sampling

Take all minority points and sample equal majority points.

Disadvantage:

① Losing some critical information.

② Upsampling

Repeat minority points.

Disadvantage:- overfitting

③ ~~Another type~~ → create artificial / synthetic data points → extrapolation

↓
Another type → class weight - assign weight in even ratio