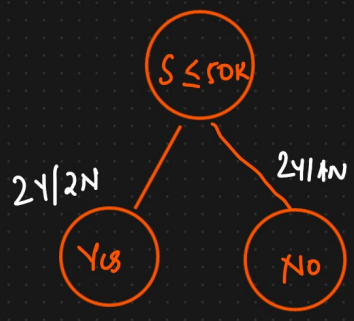
Adaboost Classifier → Boosting Technique

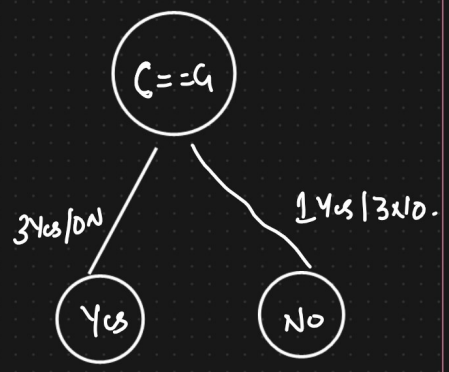
Dataset

Salary	Credit	Approval	SW
<=50K	B	No	1/7
<=50K	G	Yes	1/7
<=50K	G	Yes	1/7
>50K	B	No	1/7
>50K	G	Yes	1/7
>50K	N	Yes	1/7
<=50K	N	No	1/7

① We create Decision Tree Stump and we select the best stump



Best DT Stump.



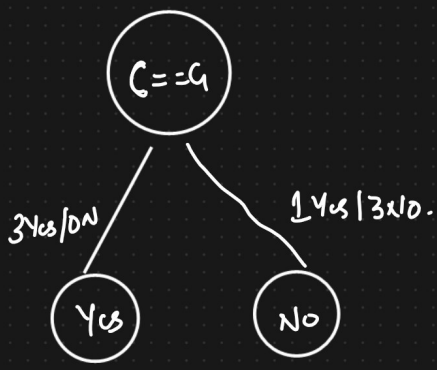
Entropy or Gini and Information gain

③ Sum of Total Errors And performance of stump

Dataset

Salary	Credit	Approval	SW
<=50K	B	No	1/7
<=50K	G	Yes	1/7
<=50K	G	Yes	1/7
>50K	B	No	1/7
>50K	G	Yes	1/7
>50K	N	Yes	1/7
<=50K	N	No	1/7

Best DT Stump.



$$\boxed{\text{Sum of TE}} = \frac{1}{7}.$$

② Performance of Stump $= \frac{1}{2} \ln\left[\frac{1-TE}{TE}\right]$

$$= \frac{1}{2} \ln\left[\frac{1-1/7}{1/7}\right]$$

$$= \frac{1}{2} \ln[6] \approx 0.896$$

Performance of Stump = 0.896

$$f = \alpha_1(M_1) + \alpha_2(M_2) + \cdots + \alpha_n(M_n)$$

$\alpha_1 = 0.896$ ⇒ Weight.

④ Update the weights for correctly and Incorrectly classified points.

Salary	Credit	Approval	SW	Updated weights
<=50K	B	No	1/7 ↓	0.058
<=50K	G	Yes	1/7 ↓	0.058
<=50K	G	Yes	1/7 ↓	0.058
>50K	B	No	1/7 ↓	0.058
>50K	G	Yes	1/7 ↓	0.058
>50K	N	Yes	1/7 ↑↑	0.349
<=50K	N	No	1/7 ↓	0.058

For correctly Classified points

$= \text{weight} * e^{-\text{Performance of stump}}$

$= \frac{1}{7} * e^{-(0.896)}$

$= 0.058$

For Incorrect Classified points

$= \text{weight} * e^{\text{Performance of stump}}$

$= \frac{1}{7} * e^{(0.896)}$

$= 0.349$

⑤ Normalized Weights Computation And Assigning Bins

0.45 0.17

Salary	Credit	Approval	SW	Updated weights	Normalized Wt	Bin Assignment
<=50K	B	No	1/7 ↓	0.058	0.083	0 - 0.08
<=50K	G	Yes	1/7 ↓	0.058	0.083	0.08 - 0.16 ← 0.095
<=50K	G	Yes	1/7 ↓	0.058	0.083	0.16 - 0.24
>50K	B	No	1/7 ↓	0.058	0.083	0.24 - 0.32
>50K	G	Yes	1/7 ↓	0.058	0.083	0.32 - 0.40
>50K	N	Yes	1/7 ↑↑	0.349	0.500	→ 0.40 - 0.90
<=50K	N	No	1/7 ↓	0.058	0.083	0.90 - 0.98
				0.697	1	

⑥ Select data points to the Next Stump

Salary	Credit	Approval	Bin Assignment
<=50K	B	No	0 - 0.08
<=50K	G	Yes	0.08 - 0.16 ← 0
<=50K	G	Yes	0.16 - 0.24
>50K	B	No	0.24 - 0.32
>50K	G	Yes	0.32 - 0.40
>50K	N	Yes	0.40 - 0.90
<=50K	N	No	0.90 - 0.98

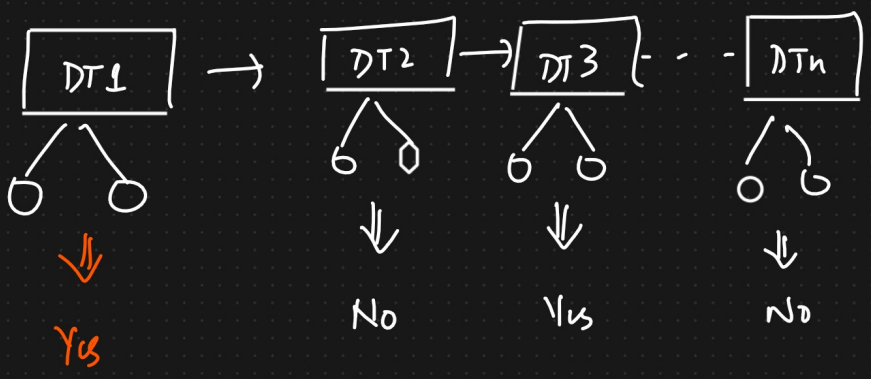
⑤ Iteration process (selecting random values between 0 and 1)

S	Credit	Approval	Random
>50K	N	Yes	0.50
<=50K	G	Yes	0.10
>50K	N	Yes	0.60
>50K	N	Yes	0.75
<=50K	G	Yes	0.24
>50K	B	No	0.32
>50K	N	Yes	0.87

⑦ Final Prediction

Test data (≤50K, G) ←

f = Test data →



$\alpha_1 = 0.896$ $\alpha_2 = 0.650$ $\alpha_3 = 0.24$ $\alpha_n = -0.30$

$$f = 0.896\,(\text{Yes}) + 0.650\,(\text{No}) + 0.24\,(\text{Yes}) - 0.30\,(\text{No})$$

$$= 1.136\,(\text{Yes}) + 0.350\,(\text{No}) \Rightarrow \text{O/P} \Rightarrow \text{Yes}$$

Performance of say(Yes) = 1.136 >>

Performance of say(No) = 0.350