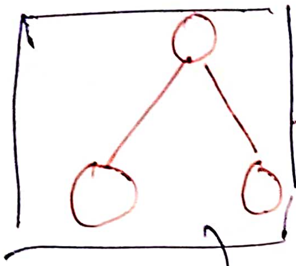
Boosting Algorithms

↓

Sequential weak learners

① Adaboost



Stump

Decision tree with one depth (1) → leads to overfitting

Train acc ↓↓ 40%

Test acc ↑↑ / ↓↓ 45% / 35%

↓

Low bias

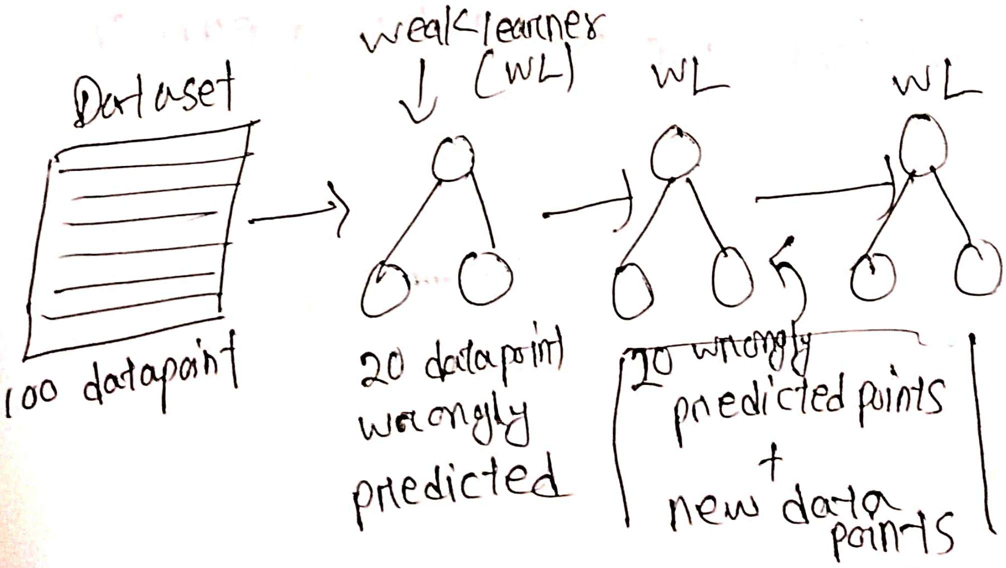
high variance / Low var

So we use Adaboost to make it

→ Low bias and low variance.

weak learner → not learn much from data.
~~(WL)~~ (WL)

Adaboost → weak learners + Add the o/p of the weak learners with some weight asigned to it.



$$f = \alpha_1(M_1) + \alpha_2(M_2) + \alpha_3(M_3) + \alpha_4(M_4) + \dots + \alpha_n(M_n)$$

↓
weights

$\alpha \uparrow$ +ve → importance of that ~~feature~~ model is more. (+ve)

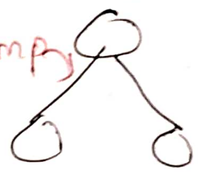
$\alpha \downarrow$ -ve → importance of that model is less. (-ve)

α = It is the weight ~~associ~~ asigned to the models. ($\alpha_1, \alpha_2, \alpha_3 \dots \alpha_n$)

$M_1, M_2, M_3 \dots M_n$ → weak learners

↓

Stump

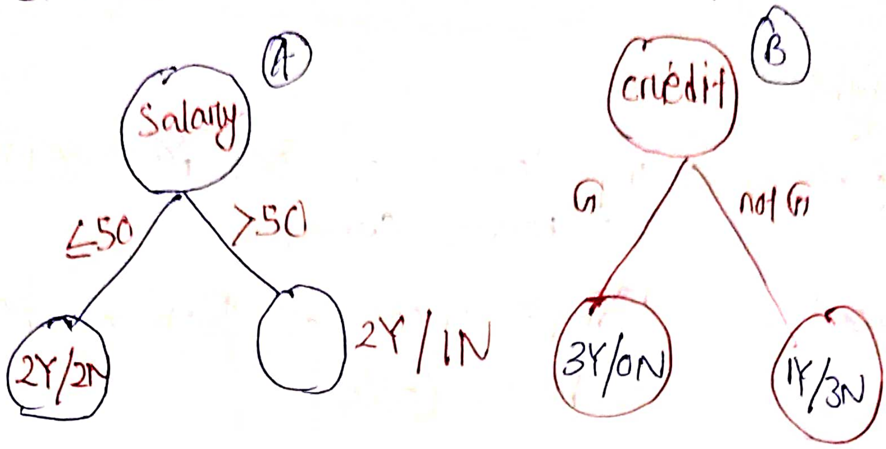


(depth = 1)

Salary	Credit (Good (G) bad (B) Normal (N))	Approval
<= 50K	B	No
<= 50K	G	Yes
<= 50K	G	Yes
> 50K	B	No
> 50K	G	Yes
> 50K	N	Yes
<= 50K	N	NO

Step-1

we create a decision tree stump

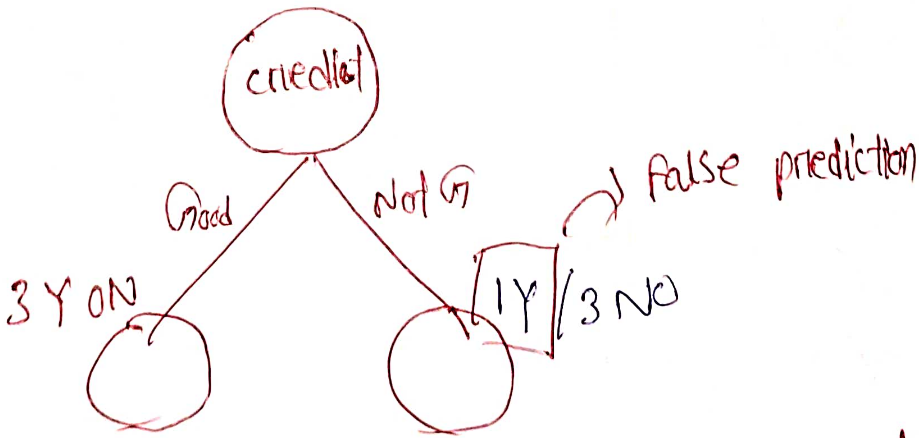


→ Entropy or ~~info~~ Gini impurity.

Step 2

Salary	Credit	Approval	weights
<=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

B will be selected as it's ~~less~~ less entropy.



Total error = sum of of all error weight

= 1/7

step-3

Salary	credit	Approval	weight	updated weight
<=50K	B	~~Yes~~ No	1/7 (0.14)	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	Yes	1/7	0.058
<=50K	N	NO	1/7	

performance of stump

$$\frac{1}{2}\ln\left[\frac{1-TE}{TE}\right]$$

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

$$= \frac{1}{2}\ln\left[\frac{6/7}{1/7}\right] = \frac{1}{2}\ln(6) \approx 0.896.$$

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

$$\alpha_1 = 0.896$$

step-4 update weights for correctly and incorrectly datapoints.

$$\boxed{\text{for correct points} = \text{weight} * e^{-\text{performance}}}$$

$$= 1/7 \times e^{-0.896} = 0.058$$

$$\boxed{\text{for incorrect points} = \text{weight} \times e^{\text{performance}}}$$

$$= \frac{1}{7} \times e^{0.896} = 0.349$$

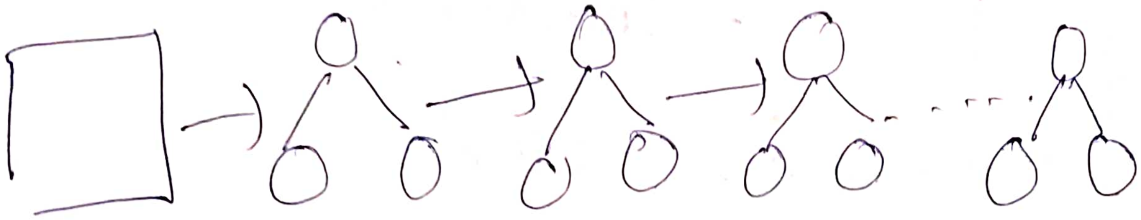
Step-5 Normalize weights and assign Bins

$U/0.697$

Salary	Credit	Approved	weight	Uweight	Normalized	Bin assigned
<=50K	B	No	1/7	0.58	0.08	0 — 0.08
<=50K	G	Yes	1/7	0.58	0.08	0.08 — 0.16
<=50K	G	Yes	1/7	0.58	0.08	0.16 — 0.24
>50K	B	No	1/7	0.58	0.08	0.24 — 0.32
>50K	G	Yes	1/7	0.58	0.08	0.32 — 0.40
>50K	N	Yes	1/7	0.349	0.50	0.40 — 0.90
<=50K	N	No	1/7	0.58	0.08	0.90 — 0.98
			Sum = 1	0.697		

→ chance of taking wrong data points will be more (0.40 — 0.90)

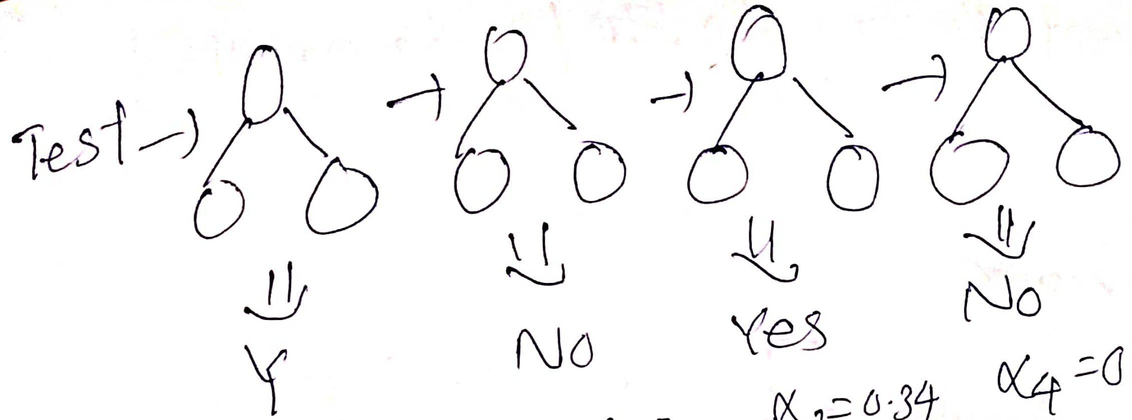
→ random value of 0-1 will be selected for the bin and data points will be selected. for next DT stump.



$$\alpha(M_1) + \alpha_2(M_2) + \alpha_3(M_3) \dots + \alpha_n(M_n)$$

Final prediction

Test (<=50K, G)



Let $\alpha_1 = 0.896$ $\alpha_2 = 0.650$ $\alpha_3 = 0.34$ $\alpha_4 = 0.20$

$$\text{Final pred} = \alpha_1(M_1) + \alpha_2(M_2) + \alpha_3(M_3) + \alpha_4(M_4)$$

$$= 0.896(\text{Yes}) + 0.65(\text{No}) + 0.34(\text{Y}) + 0.2(\text{N})$$

$$= 1.2(\text{Yes}) + 0.85(\text{No})$$

final prediction

Yes (more weight)

Regression

In case of regression we use MSE instead of entropy. Other things will be same.