Boosting Algorithms

↳ Sequential weak learners

① Adaboost

Stump
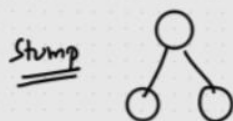
Boosting { Low Bias ← high Bias
            Low Variance ← few Variance/high
                                   Variance
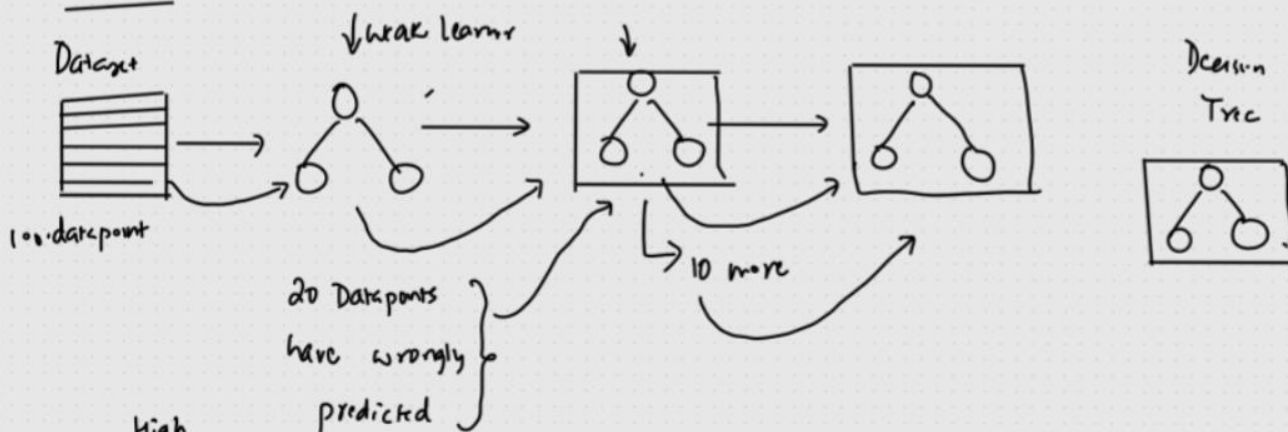
Underfitting

Training Data   Acc ↓↓ 40%
Test   Data   Acc ↑↑ 45%. }

Weak Learners →

Random Forest → Majority Voting Classifier [classification]
                Average of O/P   [Regression]

Adaboost → Weak Learners → Add the O/P of the weak learners
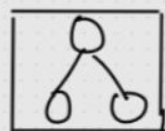   with some weights assigned to it

Adaboost

Dataset   ↓weak learner   ↓   Decision
                                                               Tree

100 datapoint

                 20 Datapoints   }
                 have wrongly
                 predicted   → 10 more

High   -ve
↓   ↓

$$f = \alpha_1(M_1) + \alpha_2(M_2) + \alpha_3(M_3) + \cdots + \alpha_n(M_n)$$

$M_1, M_2, M_3 \cdots M_n$ → Weak Learners →   Depth = 1

$\alpha_1, \alpha_2, \alpha_3 \cdots \alpha_n$ → Weights

Salary	Credit	Approval	Weights
<=50K	B	No	$\frac{1}{7}$
<=50K	G	Yes	$\frac{1}{7}$
<=80K	G	Yes	$\frac{1}{7}$
>50K	B	No	$\frac{1}{2}$
>50K	G	Yes	$\frac{1}{7}$
>50K	N	Yes	$\frac{1}{7}$
<=50K	N	No	$\frac{1}{7}$

① We create Decision Tree Stump

By selecting the best one.

Salary ≤ 50 → 2Yes/2No (Yes), 24/1N (No)

Credit==G → 3Yes/0No (Yes), 1Yes/3No (No)

$$\text{Entropy} \quad \text{or} \quad \text{Gini}$$

$$h(s) = -P_+ \log_2 P_+ - P_- \log_2 P_-$$
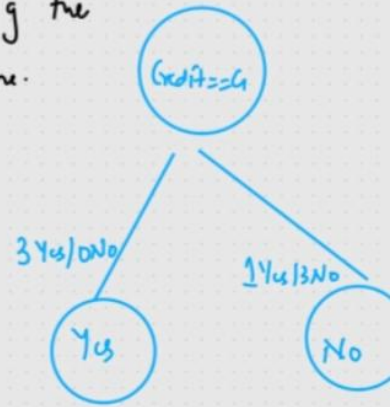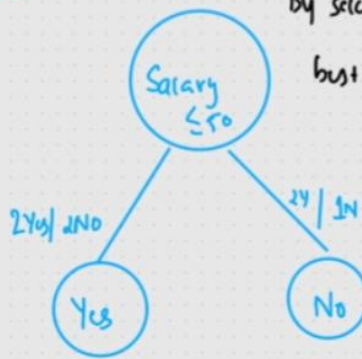
Salary	Credit	Approval	Weights
<=50K	B	No	$\frac{1}{7}$
<=50K	G	[Yes]	$\frac{1}{7}$
<=80K	G	[Yes]	$\frac{1}{7}$
>50K	B	No	$\frac{1}{7}$
>50K	G	[Yes]	$\frac{1}{7}$
>50K	N	Yes	$\frac{1}{7}$ ←
<=50K	N	No.	$\frac{1}{7}$

$$\begin{cases} B \Rightarrow Bad \\ G \Rightarrow Good \\ N \Rightarrow Normal \end{cases}$$

Credit==G → Yes: 3Yes/0No, No: 1Yes/3No
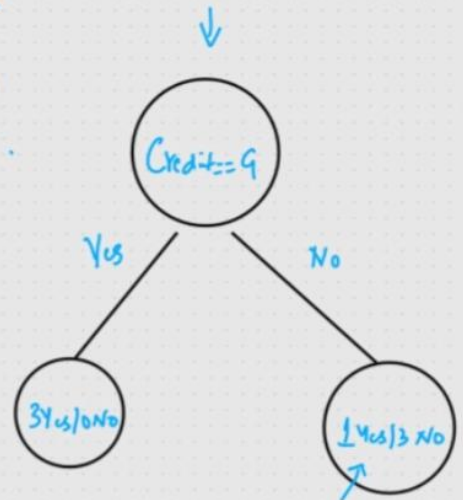
② Calculate the Total Error [Add the weights of wrong done point]

$$T.E = \frac{1}{7}$$

③ Performance of Stump $= \frac{1}{2} \ln\left[\dfrac{1-TE}{TE}\right] = \frac{1}{2} \ln[6] \approx \boxed{0.896}$

$$f = \alpha_1(M_1) + \alpha_2(M_2) + \alpha_3(M_3) + \cdots \alpha_n(M_n)$$

$f_1 = 0.896$

④ Update the weight for correctly and Incorrectly data points

Salary	Credit	Approval	Weights	Update Weight
<=50K	B	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	No	$1/7$	0.058

10K, 20K

For correctly classified point
$$= \text{Weight} * e^{-\text{Performance}}$$
$$= \frac{1}{7} * e^{-(0.896)}$$
$$= 0.058$$

for Incorrect classified point
$$= \text{Weight} * e^{\text{Performance}}$$
$$= \frac{1}{7} * e^{(0.896)}$$
$$= 0.349$$

⑤ Normalize Weights and Assign Bins $\boxed{0-1} \downarrow \rightarrow \boxed{0.12}$ $\underline{0.56}$ $\underline{0-1}$ $\underline{0.09}$

Salary	Credit	Approval	Weights	Update Weight	Normalized Weight	Bins Assignment
<=50K	B	No	$1/7$ ⁰·¹⁴ →	$0.058 \div 0.697$	0.08	$0 - 0.08$
<=50K	G	Yes	$1/7$	$0.058 \div 0.697$	0.08	$0.08 - 0.16$
<=50K	G	Yes	$1/7$	$0.058 \div 0.697$	0.08	$0.16 - 0.24$
>50K	B	No	$1/7$	0.058	0.08	$0.24 - 0.32$
>50K	G	Yes	$1/7$	0.058	0.08	$0.32 - 0.40$
>50K	N	Yes	$1/7$ <	0.349	0.50	→ $\boxed{0.40 - 0.90}$
<=50K	N	No	$1/7$	0.058	0.09	$0.90 - 1$

0.697 ≈ 1

Random (0 to 1) for Ran(0,1)
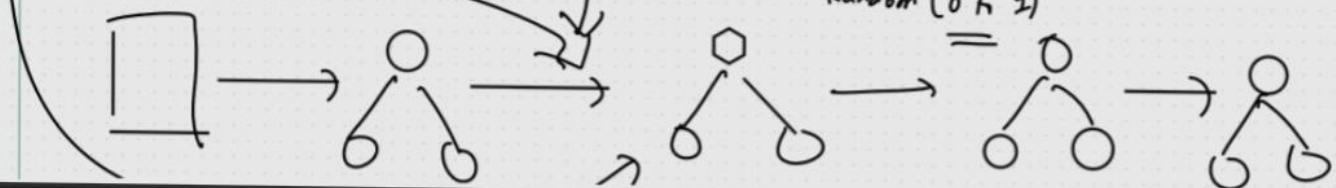
$$h_1(M_1) \quad + \quad h_2(M_2) \quad + h_3(M_3) + \cdots h_n(M_n)$$

④ Final Prediction

Test $(\leq 50x, G)$

Test $\longrightarrow$

| $d_1 = 0.846$ | $h_2 = 0.650$ | $h_3 = 0.38$ | $h_4 = 0.20$ |

Yes → No → Yes → No
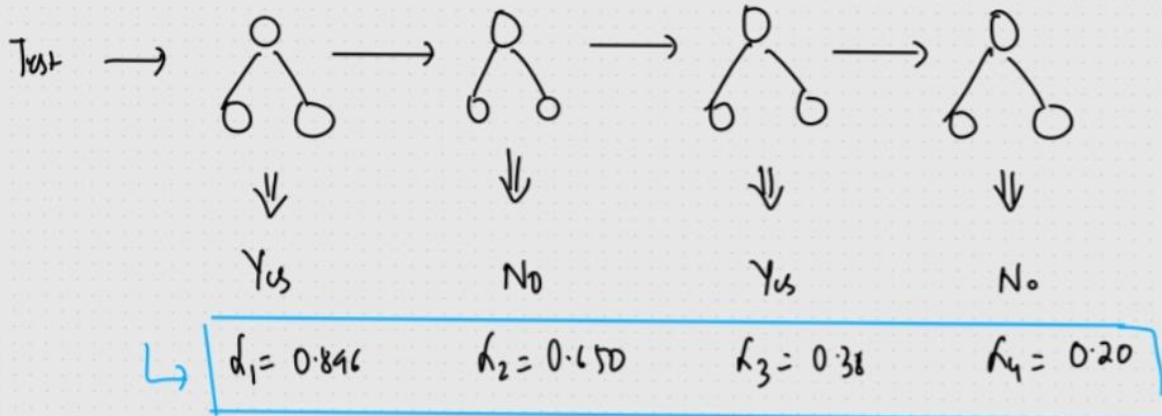
$$f = h_1(M_1) + h_2(M_2) + h_3(M_3) + h_4(M_4)$$

$$= 0.846 \, (Yes) + 0.650 \, (No) + 0.38(Yes) + 0.20 \, (No)$$

$$= 1.2(Yes) + 0.85 \, (No)$$

$$\Downarrow$$

Yes