AdaBoost Machine Learning Algorithms

Decision Trees

Ensemble Techniques

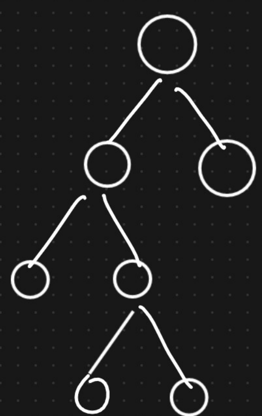
{Base Learners} **Bagging**

1. Random Forest Classifier
2. Random Forest Regressor

Boosting {weak Learners}

1. Adaboost
2. Gradient Boosting
3. Xgboost

Decision Tree :



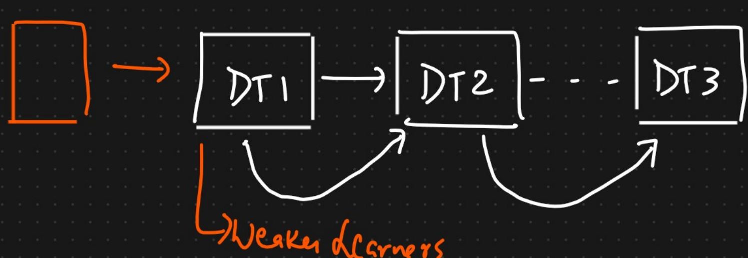
Overfitting : Train Acc ↑
Test Acc ↓

Low Bias
High Variance

Random Forest {Bagging}

Low Bias
Low Variance

Boosting {Sequentially connected}



⇒ Strong Learner

Weak Learner → Haven't Learnt Much from the Training Dataset

Random Forest → Majority Voting classifier
Average of (O/P)

AdaBoost $\longrightarrow$ Assignment weights to the weak learner

$M_1 \cdots M_n \rightarrow$ Decision Tree Stumps

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

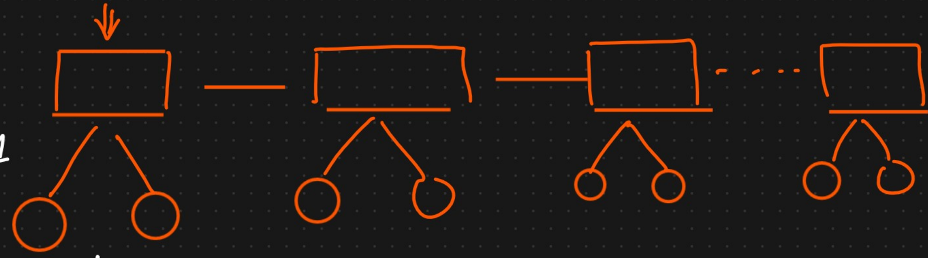
→ CLASSIFICATION

→ REGRESSION

$\{\alpha_1, \alpha_2, \alpha_3 \cdots \alpha_n\} \Rightarrow$ weights

Decision Stump

Depth=1



Weak Learner?

↳ Underfitting

↳ Train Acc ↓ 40%

↳ Test Acc ↑ 45%

Decision Tree Stump

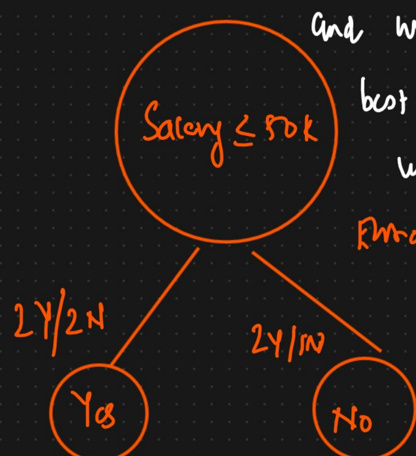
⇓

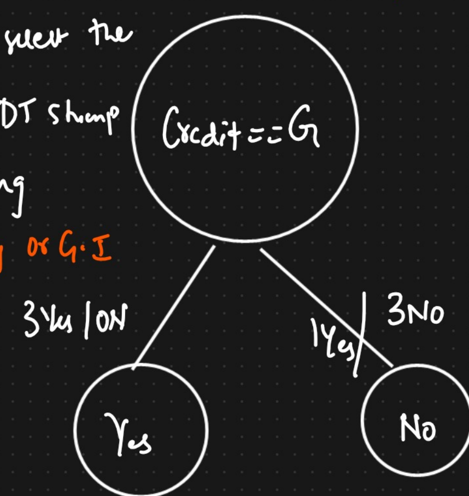
$\left\{ \begin{array}{l} \text{High Bias} \\ \text{Low Variance} \end{array} \right\} \Rightarrow \left\{ \begin{array}{l} \text{Low Bias} \\ \text{High Variance} \end{array} \right\}$

Adaboost Classifier Maths Indepth Intuition

① We create Decision Tree Stump and we select the best DT Stump using Entropy or G.I

Salary	Credit	Approval
<=50K	B	No
<=50K	G	Yes
<=50K	G	Yes
>50K	B	No
>50K	G	Yes
>50K	N	Yes
<=50K	N	No





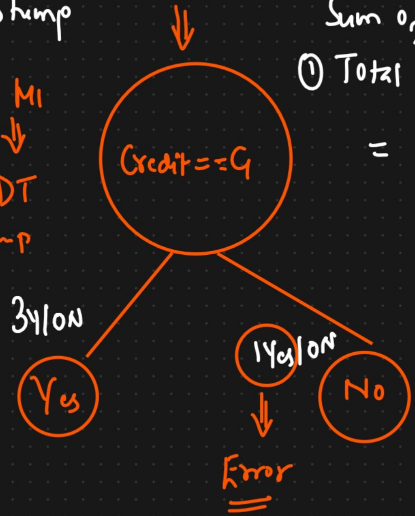
Entropy or Gini Impurity

$$H(S) = -p_+ \log_2 p_+ - p_- \log_2 p_-$$

② Sum of the Total Errors And Performance of Stump

Salary	Credit	Approval	Sample Weights
<=50K	B	No	$\frac{1}{7}$
<=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

M1 ↓ or DT Stump



Sum of all the ① Total Error $= \frac{1}{7}$

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

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

$\boxed{\alpha_1 = 0.896}$ ⇒ weight

③ Update the weights for correctly and Incorrectly classified points

Salary	Credit	Approval	Sample Weights	Update wts
<=50K	B	No	$\frac{1}{7}$ ↓ →	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 correct classified points

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

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

$$= 0.058$$

For Incorrect classified

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

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

= 0.349

④ Normalized Weights Computation And Assigning Bins

Salary	Credit	Approval	Update wts	Normalized Weights	Bins Assignment
<=50K	B	No	0.058	0.08	0 - 0.08
<=50K	G	Yes	0.058	0.08	0.08 - 0.16
<=50K	G	Yes	0.058	0.08	0.16 - 0.24
>50K	B	No	0.058	0.08	0.24 - 0.32
>50K	G	Yes	0.058	0.08	0.32 - 0.40
>50K	N	Yes	0.349	0.50	0.40 - 0.90
<=50K	N	No	0.058	0.08	0.50 - 0.58
			0.697	≈ 1	

$\alpha_1 = 0.896$

Prepare datapoints

⑤ Select data points to send to Next Stump

Salary	Credit	Approval	Bins Assignment
<=50K	B	No	0 - 0.08
<=50K	G	Yes	0.08 - 0.16
<=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	0.50 - 0.58

① Iteration process selecting random value 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

(6) These records will be sent to Next DT Stump.

S	Credit	Approval	Sample weight
>50K	N	Yes	$\frac{1}{6}$
<=50K	G	Yes	$\frac{1}{6}$
>50K	N	Yes	$1/6$
>50K	N	Yes	$1/6$
<=50K	G	Yes	$1/6$
>50K	B	No	$1/6$
>50K	N	Yes	$1/6$

TE

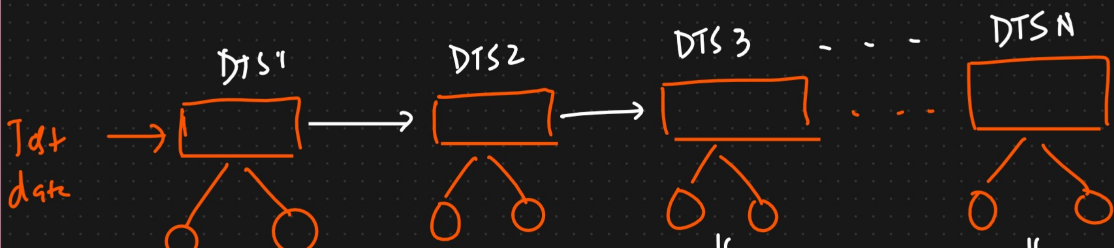
Performance Stump ⇒ 0.65

$f_1 = \alpha_1(M_1) + \alpha_2(M_2)$

$\alpha_1 = 0.896$ $\alpha_2 = 0.65$

(7) Final Prediction

Test data $(\leq 50K, G)$



↓	⇓	⇓	⇓
Yes	No	Yes	No
$\alpha_1 = 0.896$	$\alpha_2 = 0.650$	$\alpha_3 = 0.24$	$\alpha_n = -0.30$

$f = \alpha_1(M_1) + \alpha_2(M_2) + \alpha_3(M_3) + \cdots \alpha_4(M_4)$

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

$$\begin{array}{r} 0.896 \\ 0.240 \\ \hline 1.136 \end{array}$$

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

Performance of say (Yes) = 1.136 >

Performance of say (NO) = 0.350