AdaBoost Machine Learning Algorithms

- Both bagging & boosting use Decision Tree as base/ weak learners.

Decision Trees

Ensemble Techniques

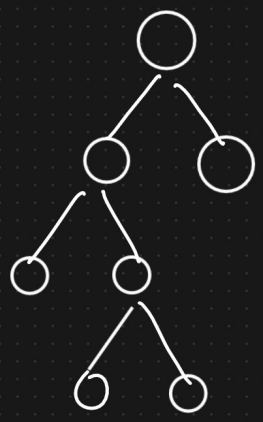
{Base Learners} Bagging

{Weak Learners} Boosting

① Random Forest Classifier
② Random Forest Regressor

① Adaboost
② Gradient Boosting
③ xgboost

Decision Tree :



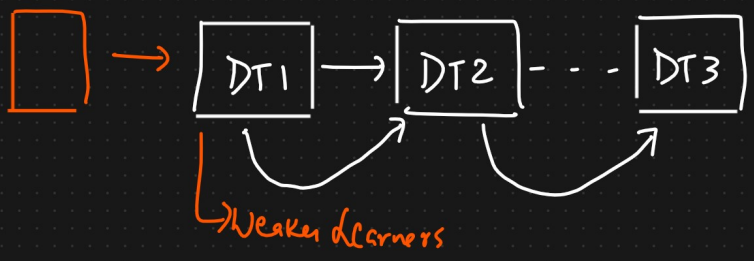
Overfitting : Train Acc ↑
Test Acc ↓

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 → Assignment weights to the weak learner

$M_1 \dots M_n$ → Decision Tree Stumps

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

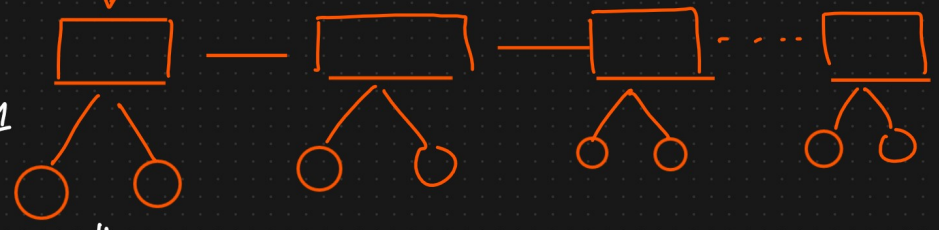
→ CLASSIFICATION
→ REGRESSION

$\{\alpha_1, \alpha_2, \alpha_3 \dots \alpha_n\}$ ⇒ weights Assigned to weak learner

Decision Tree

Stump

Depth=1



Weak Learner

↓ Leads to

↳ Underfitting

↳ Train Acc ↓ 40%

↳ Test Acc ↑ 45%

Decision Tree Stump

↓

{ High Bias, Low Variance } ⇒ { Low Bias, High Variance }

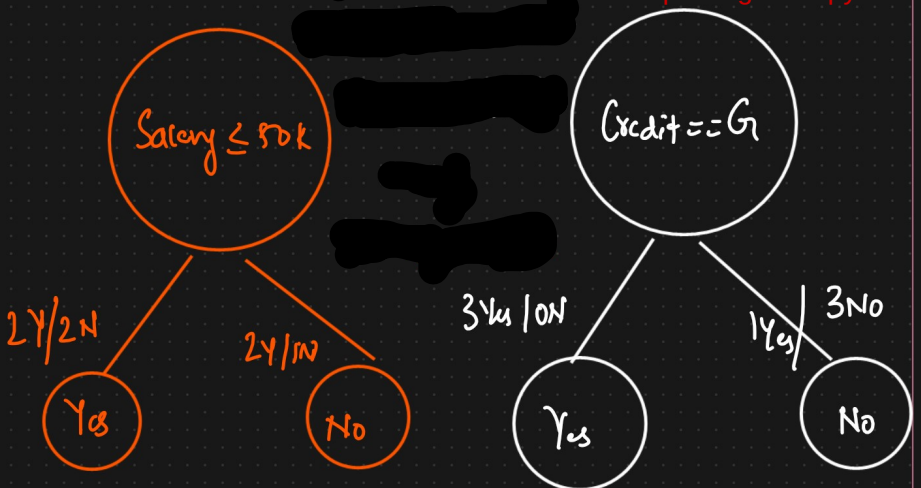
When we combine all weak learner

- Bais : -
  - Related to training data
  - Learning the wrong thing from the training data.
- Variance :
  - Related test data
  - Learning too much from the training data and not generalizing well.

Adaboost Classifier Maths Indepth Intuition

1. We create Decision Tree stump and we select the best DT stump using Entropy & GI.

Salary	Credit	Approval
<=50K	B	No
<=50K	G	Yes
<=80K	G	Yes
>50K	B	No
>50K	G	Yes
>50K	N	Yes
<=50K	N	No



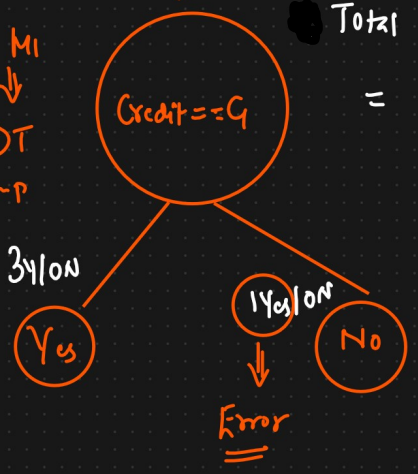
Entropy or Gini Impurity

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

(2) Sum of the Total Errors And Performance of Stump

Salary	Credit	Approval	Sample Weights
<=50K	B	No	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



① Sum of all the Total Error = 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

(3) Update the weights for correctly and Incorrectly classified points

Salary	Credit	Approval	Sample Weights	Update wts
<=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

Decrease the weight bcoz i want to increase the probability of selecting wrong record by next DT.

+ For correct classified points :-

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

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

$= 0.058$

+ For Incorrect classified :-

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

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

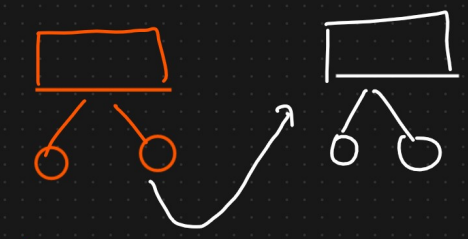
= 0.349

(4) Normalized Weights Computation And Assigning Bins

Salary	Credit	Approval	Update wts	Normalized weights	Bins Assignment
<=50K	B	No	0.058/0.697 →	0.08	0 - 0.08 (Simply adding)
<=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 - 0.70 → Maximum bin size.
<=50K	N	No.	0.058 "	0.08	0.50 - 0.58
			0.697 ←	≈ 1	

- Why do we need normalized weight?
  - → Sum of all updated weight != 1 that's why we need normalized weight which sum==1

I will try to find out a mechanism wherein I will try to send this records more frequently to DT Stump-2



$\alpha_1 = 0.896$ Prepare datapoints

- How do I say and make sure that my machine learning algorithm sends only incorrectly classified records to next DT
  - → for that particular case, we will be assigning some bins

(5) Select data points to send to Next Stump

① Iterative process selecting random value between 0 and 1

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 - 0.70
<=50K	N	No.	0.50 - 0.58

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

Incorrectly classified point getting selected again& again.

(6) Then records will be sent to next DT Stump.

S	Credit	Approval	Sample weight
>50K	N	Yes	1/7
<=50K	G	Yes	1/7
>50K	N	Yes	1/7
>50K	N	Yes	1/7
<=50K	G	Yes	1/7
>50K	B	No	1/7
>50K	N	Yes	1/7

TE

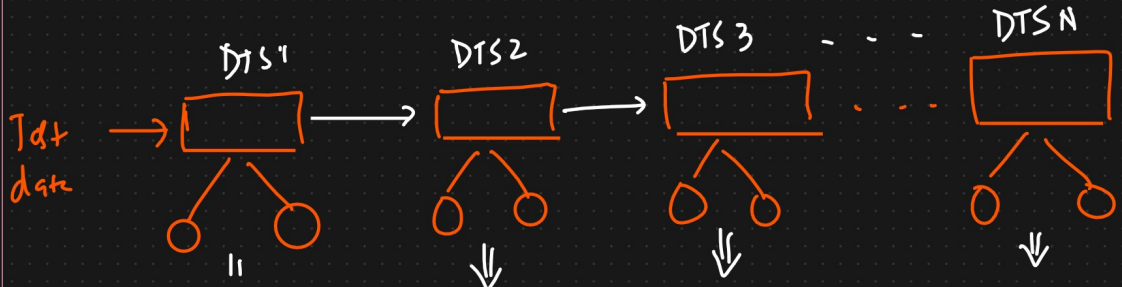
Performance Stump ⇒ 0.65

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

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

(7) Final Prediction

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



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

$f = \alpha_1(M_1) + \alpha_2(M_2) + \alpha_3(m_3) + \cdots \alpha_4(M_4)$

$= 0.896(Yes) + (0.650)(No) + 0.24(Yes) - 0.30(No)$

1 0.896
0.240
1.136

$= 1.136\,(Y_{es}) + 0.350\,(N_o)$ ⇒ O/p : Yes ✓

↑↑ ↓↓ ↑↑

→ Performance of say (Yes) = 1.136 >

Performance of say (No) = 0.350