

AdaBoost

Dr. Parimala, SCORE, VIT

Ensemble Boosting



STEP1: SAMPLE WEIGHT CREATION

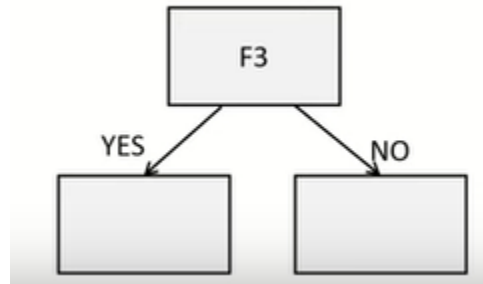
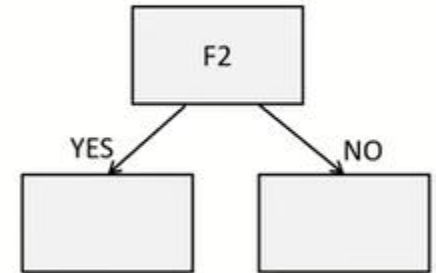
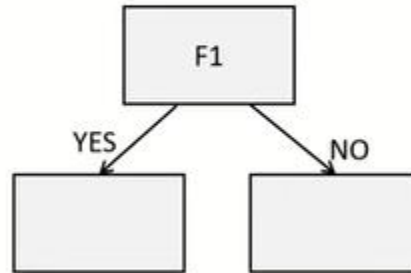
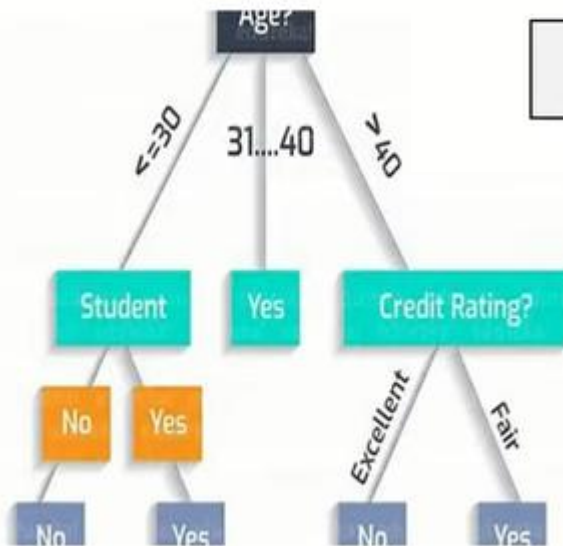
F1	F2	F3	O/P
12	3	23	YES
23	5	45	YES
34	3	43	NO
21	4	65	YES
45	5	34	NO
12	2	23	NO
34	5	43	YES
16	6	45	YES

SAMPLE WEIGHT
1/8
1/8
1/8
1/8
1/8
1/8
1/8
1/8

Sample weights should be from 0 to 1.
Initially all the weights are assigned
equally ($w=1/n$)

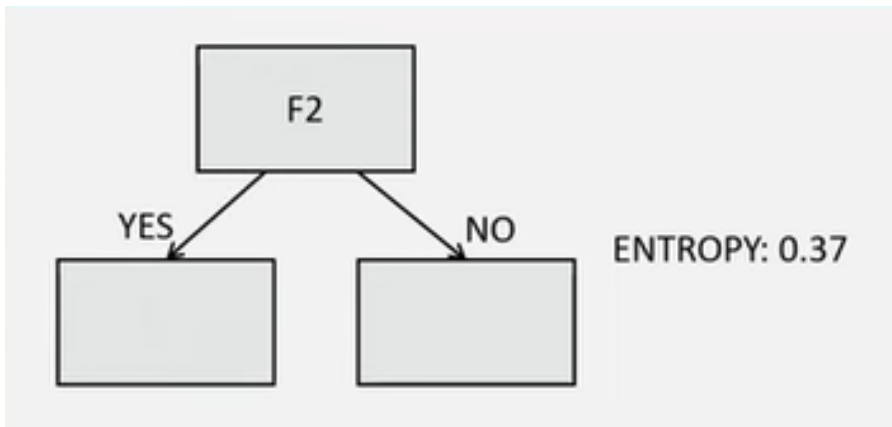
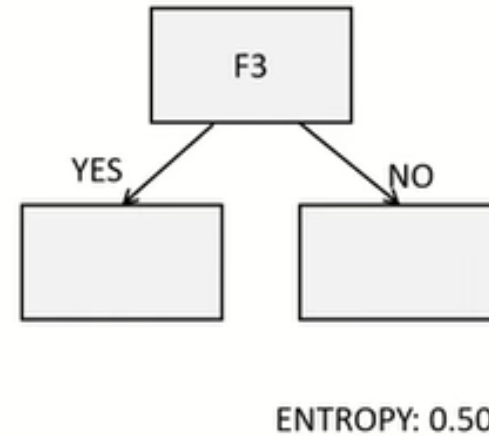
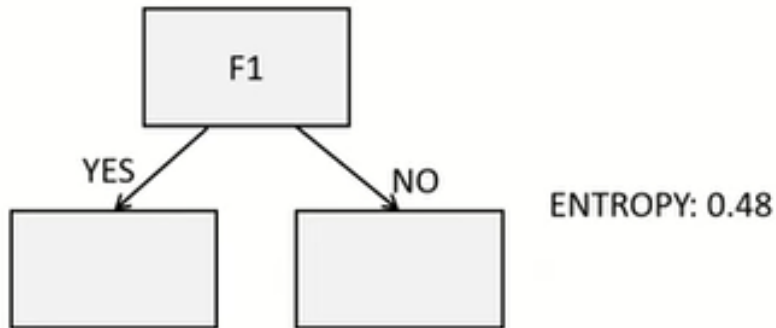
STEP2: STUMP CREATION

STUMPS



Tree with two leaf nodes and one depth is called as stumps
Stumps are created for every feature

STEP3: STUMP SELECTION



Entropy or gini index is calculated for each stumps
Select the stump which has minimum entropy

STEP4: CALCULATE TOTAL ERROR

- Between 0 and 1
- **Total error in Adaboost is the sum of weights of misclassified records.** The total error is the sum of all the errors in the classified record for sample weights. The total error will always be between 0 and 1, with 0 representing perfect stump (correct classification) and 1 representing weak stump (misclassification).

F1	F2	F3	O/P	SAMPLE WEIGHT
12	3	23	YES	1/8
23	5	45	YES	1/8
34	3	43	NO	1/8
21	4	65	YES	1/8
45	5	34	NO	1/8
12	2	23	NO	1/8
34	5	43	YES	1/8
16	6	45	YES	1/8

$$TE = 1/8$$

Total error is the sum of all weights which are not classified correctly

Assume that only one sample is misclassified here, so the $TE=1/8$

Incase if two samples are misclassified, then $TE=1/8+1/8$

STEP5: CALCULATE AMOUNT OF SAY (OR) PERFORMANCE SAY

$$\text{Amount of Say} = \frac{1}{2} \log\left(\frac{1 - \text{Total Error}}{\text{Total Error}}\right)$$

$$\text{Amount of Say} = \frac{1}{2} \log(7) = 0.97$$

STEP6: UPDATE WEIGHTS

New Sample Weight = sample weight $\times e^{\text{amount of say}}$

$$= \frac{1}{8} e^{\text{amount of say}}$$

$$= \frac{1}{8} e^{0.97} = \frac{1}{8} \times 2.64 = 0.33$$

Formula to update weight for misclassified samples

New Sample Weight = sample weight $\times e^{-\text{amount of say}}$

Formula to update weights for correctly classified samples

STEP6: UPDATE WEIGHTS

Sum of all sample weight is 1 but sum of all updated weights is 0.68. So we have to normalize the updated weights.

F1	F2	F3	O/P	SAMPLE WEIGHT	UPDATED WEIGHTS
12	3	23	YES	1/8	0.05
23	5	45	YES	1/8	0.05
34	3	43	NO	1/8	0.05
21	4	65	YES	1/8	0.33
45	5	34	NO	1/8	0.05
12	2	23	NO	1/8	0.05
34	5	43	YES	1/8	0.05
16	6	45	YES	1/8	0.05

STEP7: NORMALIZE THE WEIGHTS

F1	F2	F3	O/P	SAMPLE WEIGHT	UPDATED WEIGHTS	NORMALIZED WTS
12	3	23	YES	1/8	0.05	0.07
23	5	45	YES	1/8	0.05	0.07
34	3	43	NO	1/8	0.05	0.07
21	4	65	YES	1/8	0.33	0.49
45	5	34	NO	1/8	0.05	0.07
12	2	23	NO	1/8	0.05	0.07
34	5	43	YES	1/8	0.05	0.07
16	6	45	YES	1/8	0.05	0.07

Normalized weights= Updated /sum of
updated weight
For eg., $0.05/0.68=0.07$

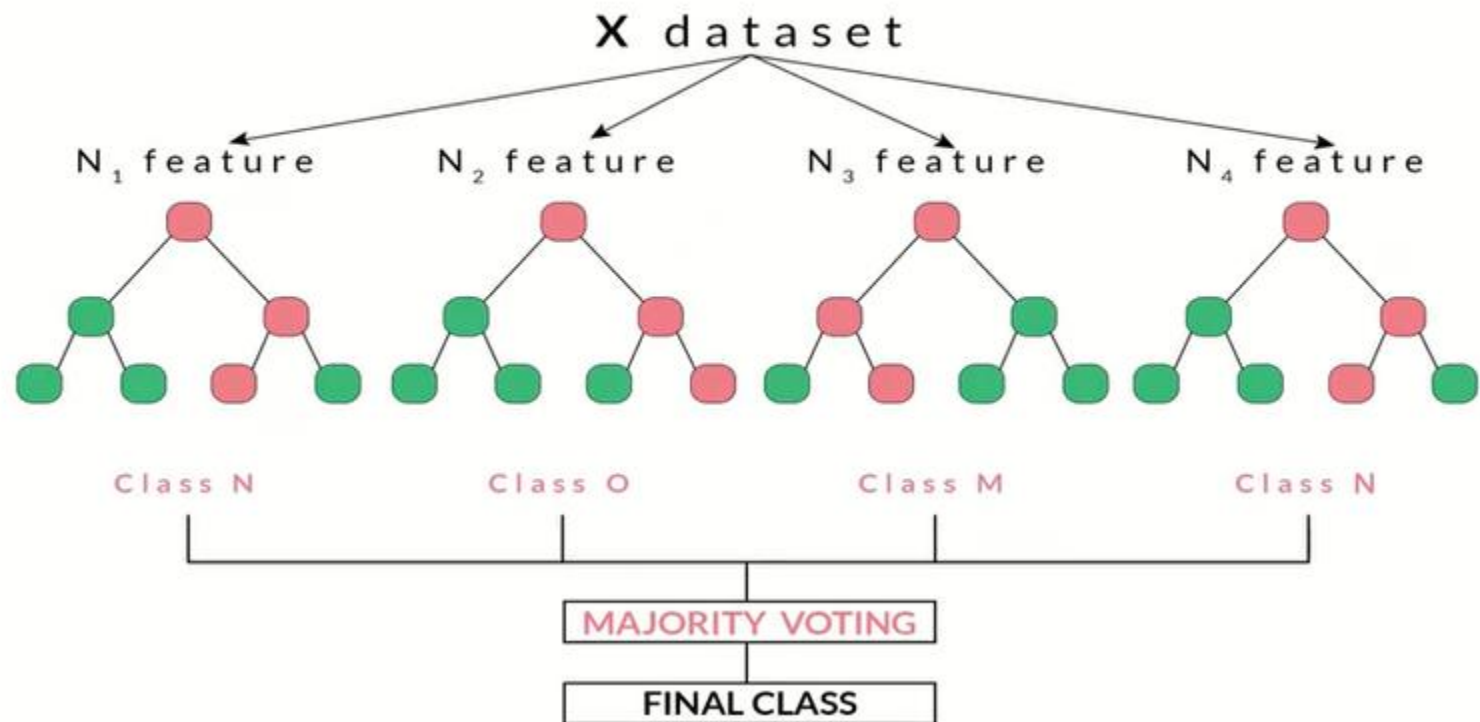
STEP8: NEW SAMPLE FORMATION

F1	F2	F3	O/P	SAMPLE WEIGHT	BUCKETS	F1	F2	F3	O/P	SAMPLE WEIGHT
12	3	23	YES	0.07	0 - 0.07	34	3	43	NO	0.07
23	5	45	YES	0.07	0.07 - 0.14	23	5	45	YES	0.07
34	3	43	NO	0.07	0.14 - 0.21	12	3	23	YES	0.07
21	4	65	YES	0.49	0.21 - 0.70	34	5	43	YES	0.07
45	5	34	NO	0.07	0.70 - 0.77	16	6	45	YES	0.07
12	2	23	NO	0.07	0.77 - 0.84	12	2	23	NO	0.07
34	5	43	YES	0.07	0.84 - 0.93	45	5	34	NO	0.07
16	6	45	YES	0.07	0.93 - 1	21	4	65	YES	0.49

BACK TO STEP 2

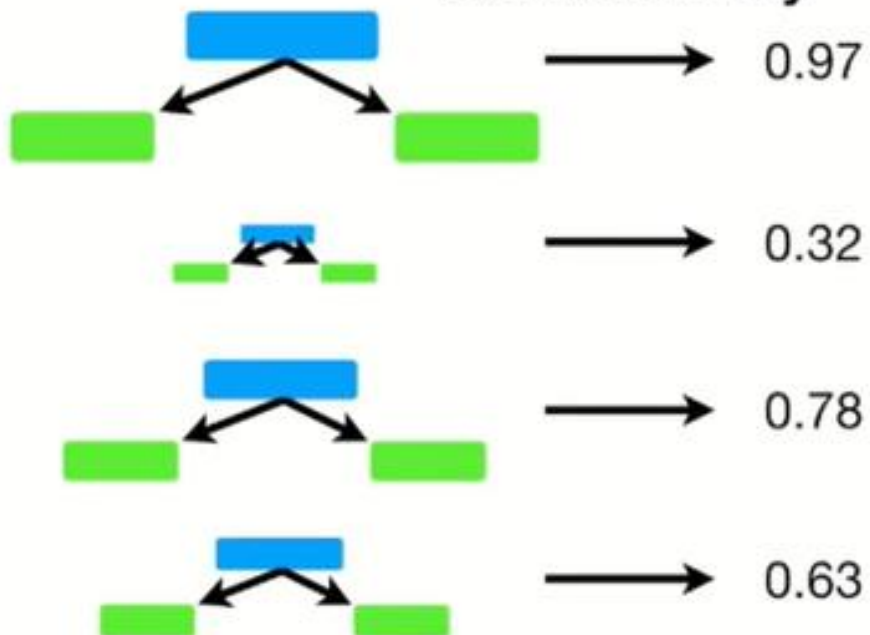
- STEP1: SAMPLE WEIGHT CREATION
- STEP2: STUMP CREATION
- STEP3: STUMP SELECTION
- STEP4: CALCULATE TOTAL ERROR
- STEP5: CALCULATE AMOUNT OF SAY (OR) PERFORMANCE SAY
- STEP6: UPDATE WEIGHTS
- STEP7: NORMALIZE THE WEIGHTS
- STEP8: NEW SAMPLE FORMATION

WHAT DOES RANDOM FOREST DO?



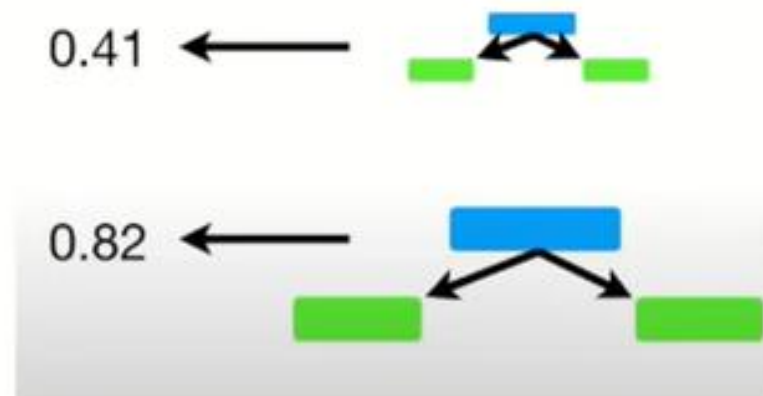
YES

Amount of Say



NO

Amount of Say



IMPORTANT POINTS ABOUT ADABOOST

- ADABOOST combines lot of “weak learners” to make classifications.
- Some Stumps get more say in the classification than others.
- Each stump is made by taking the previous stumps mistakes.