Adaboost Algorithm

1. Adaboost is an Ensemble technique
2. Ensemble means a group producing a single effect
3. Example : concerted music
4. Boosting means improve something
5. AdaBoost - Adaptive Boosting
6. AdaBoost can be used in joining with many other types or learning algorithms to improve performance

Adaboost Algorithm

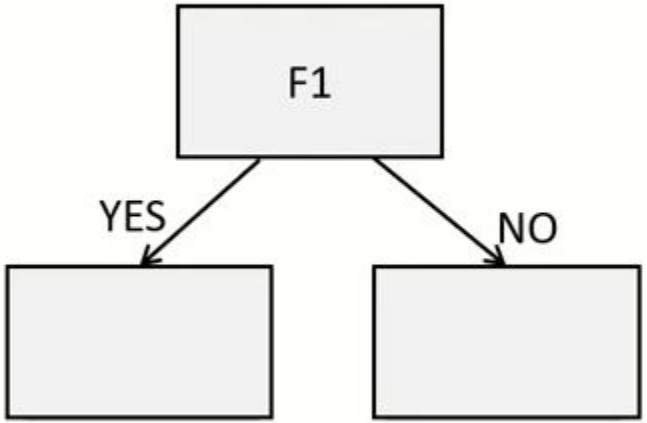
- STEP1: SAMPLE WEIGHT CREATION
- STEP2: STUMP CREATION
- STEP3: STUMP SELECTION
- STEP4: CALCULATE TOTAL ERROR
- STEP5: CALCULATE AMOUNT OF SAY (OR) PERFORMANCE SAY
- STEP6: UPDATE WEIGHTS
- STEP7: NORMALIZE THE WEIGHTS
- STEP8: NEW SAMPLE FORMATION

Step 1: Weight Creation

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

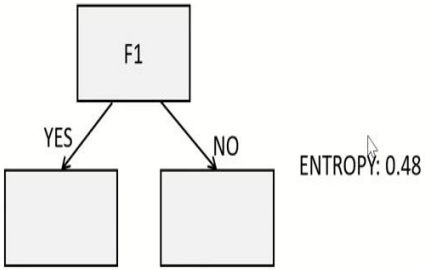
1) Sample weight creation
2) Is 1/n - n is total number of records
3) All the weights combine together has to be 1
4) Initially all of them has the same weight (⅛)

Step 2:Stumps Creation

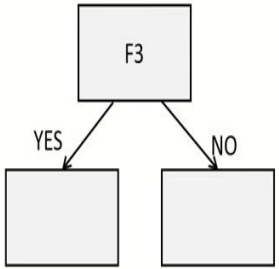


1) For each feature it will have only 2 leaf nodes. (only 1 depth)
2) It is called has stumps

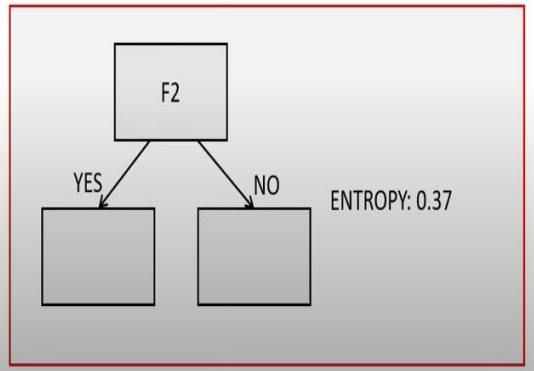
Step 3:Stump Selection



ENTROPY: 0.48



ENTROPY: 0.50



ENTROPY: 0.37

1) **Entropy means measure of quality of energy**
2) **Lower the entropy higher the quality**
3) Each stump entropy or Gini index will be calculated
4) Stump with less in entropy value will be selected.

Step 4:Calculate total error

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

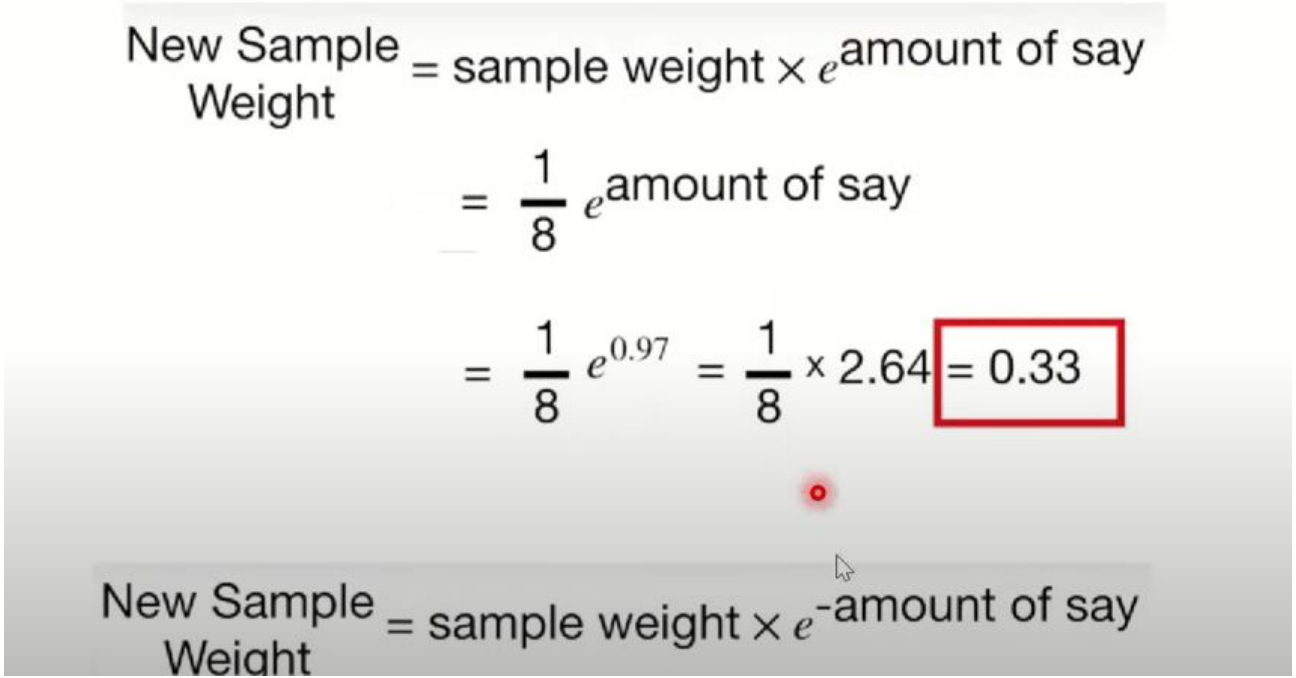
1) Total Error will be between 0 and 1
2) TE is the sum of all the errors in the classified record for the same weights
3) TE is 0 means is a perfect stump
4) TE is 1 mean is a weak stump
5) In that image TE is 1/8

Step 5:Calculate Amount of Say or Performance Say

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

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

Step 6:Update Weights



1) Updated weight calculated and will be updated in table.
2) For all other weight also will be calculated and is updated

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

Step 7: Normalize 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

Step 8: New Sample Formation

F1	F2	F3	O/P	SAMPLE WEIGHT
12	3	23	YES	0.07
23	5	45	YES	0.07
34	3	43	NO	0.07
21	4	65	YES	0.49
45	5	34	NO	0.07
12	2	23	NO	0.07
34	5	43	YES	0.07
16	6	45	YES	0.07

BUCKETS
0 - 0.07
0.07 - 0.14
0.14 - 0.21
0.21 - 0.70
0.70 - 0.77
0.77 - 0.84
0.84 - 0.93
0.93 - 1

F1	F2	F3	O/P	SAMPLE WEIGHT
34	3	43	NO	0.07
23	5	45	YES	0.07
12	3	23	YES	0.07
34	5	43	YES	0.07
16	6	45	YES	0.07
12	2	23	NO	0.07
45	5	34	NO	0.07
21	4	65	YES	0.49

Finally

- 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.