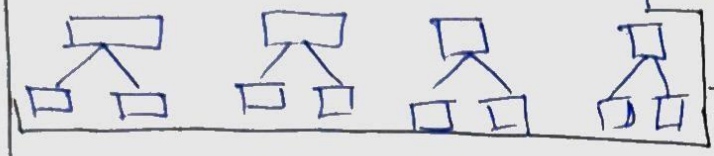
Adaboost, -using Random Forest

In Random Forest,
we grow full depth tree.
→ some trees will have more depth than other.

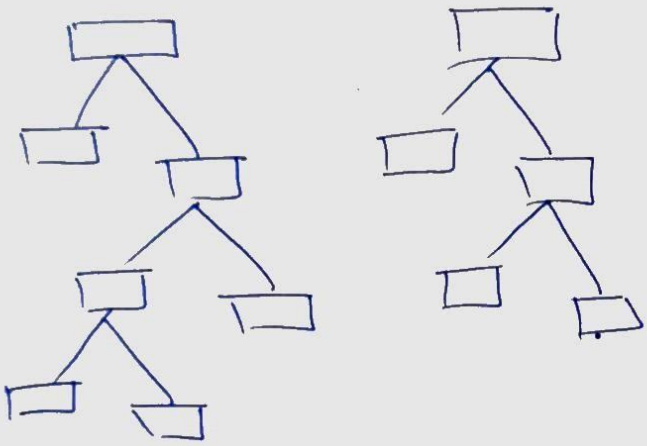
In Contrast, in forest of trees made with Adaboost, the tree are usually just a node with two leaves.



→ stumps - tree with one node and two leaves.

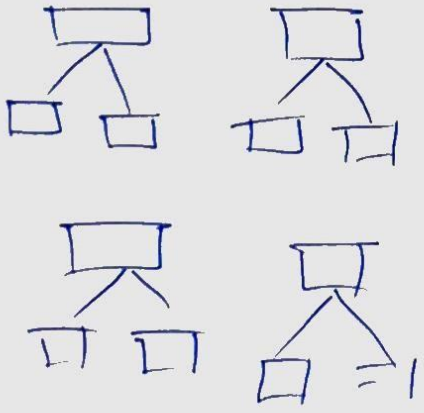
→ Forest of stumps

→ weak learners → because uses only one variable in each tree.



Random Forest

1. Each tree has equal vote.
2. Each tree was made independently.



Adaboost Forest of trees

1. Each tree has different vote in Prediction.
2. Order of trees are important
   ↓
   Error than First stump influst error made by second stump and it goes on.

Three ideas of Adaboost

1. Combines weak learners
2. Some stumps have more say in classification
3. Each stump is made by the previous 'stump's' mistakes into account.

Building - Adaboost

Step-1

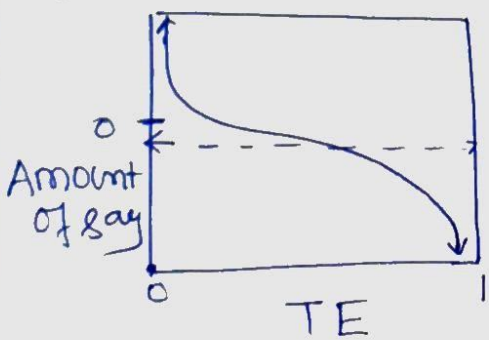
→ give equal weight to all samples in our dataset.
suppose $n=8$, so each sample will get $(1/8)$.

Step-2

Calculate gini index for all variable and choose the one having lowest value.

Step-3

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



{TE=1, ⇒ log(0) → $-\infty$}
{TE=0, ⇒ log(1)/0 → ∞}

TE=0.5 → log(0.5)/0.5 ⇒ 0/=

Total Error For a stump = Sum of weights for all misclassified samples.

Suppose there is only one misclassified sample.
⇒ (TE = 1/8) → Initially all sample weight is same.

$$\text{Amount of say} = \frac{1}{2}\log\left(\frac{1-1/8}{1/8}\right) = \frac{1}{2}\log(7)$$

$$= 0.97$$

Adaboost - 2.

Step-4: Choose second best variable to make stump. (chest_pain)

Calculate amount of say for this stump:

Suppose '3' samples are misclassified.

then $\frac{1}{8}+\frac{1}{8}+\frac{1}{8}=\frac{3}{8}=TE$

Amount of say $=\frac{1}{2}\left(\log\left(\frac{1-3/8}{1-3/8}\right)\right)=\frac{1}{2}\log(5)=0.42$

Repeat this step for all other variables..

Step-5. start at the First "stump" →

- → Take out all samples which were misclassified.
- → increase the weight of misclassified sample.
- → decrease the weight of correctly classified sample, since we will not emphasize more on correctly predictions.

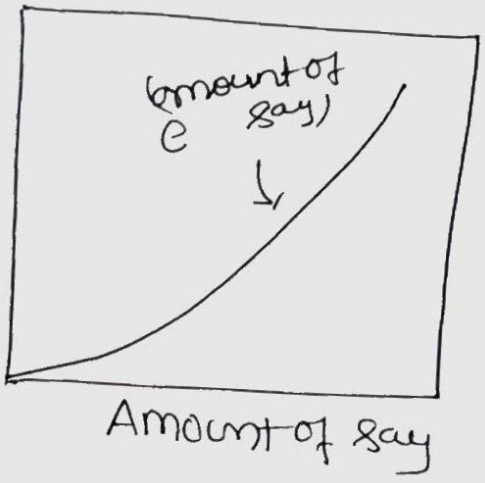
→ New Sample = Sample weight * $e^{(\text{amount of say})}$

For incorrectly classified sample ← weight

$= \frac{1}{8} * e^{(\text{amount\_of\_say})}$

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

More than Old one $(\frac{1}{8})$



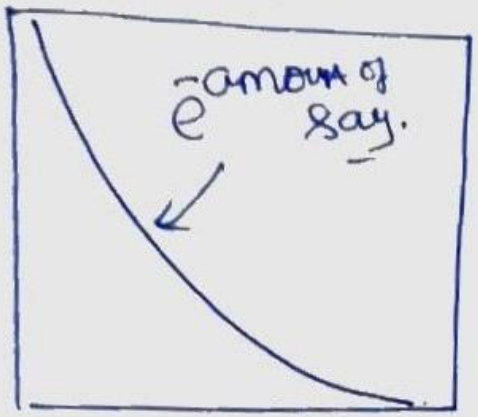
Adaboost -3

Sample weight For Correctly classified samples:-

New_sample weight = Sample_weight * $e^{-(\text{amount of say})}$

$= \frac{1}{8} * e^{-(0.97)} = \frac{1}{8} \times (0.38) = 0.005$

↓ less than old one



→ we will get new weights for each sample. Their sum may not be equals to '1'. So, we need to normalize it so that sum is 1. and pass it to next stump.