Instructor: Rahul Tiwari
Email: support@rahultiwari.co.in
LinkedIn: rahultiwari1986 | Mob: 9643393302

# Boosting Technique

- Weak Learners
- Weight
- Dependency

**Decision Stump we create as we need High Bias Models**

In the random forest all the trees are having the same weightage

In the boosting we will have different weightage based on the performance of the model

$$F = \alpha_1 H_1 + \alpha_2 H_2 + \alpha_3 H_3 + \cdots \alpha_n H_n$$

The models which we are creating is a High Bias model and we will make it low bias

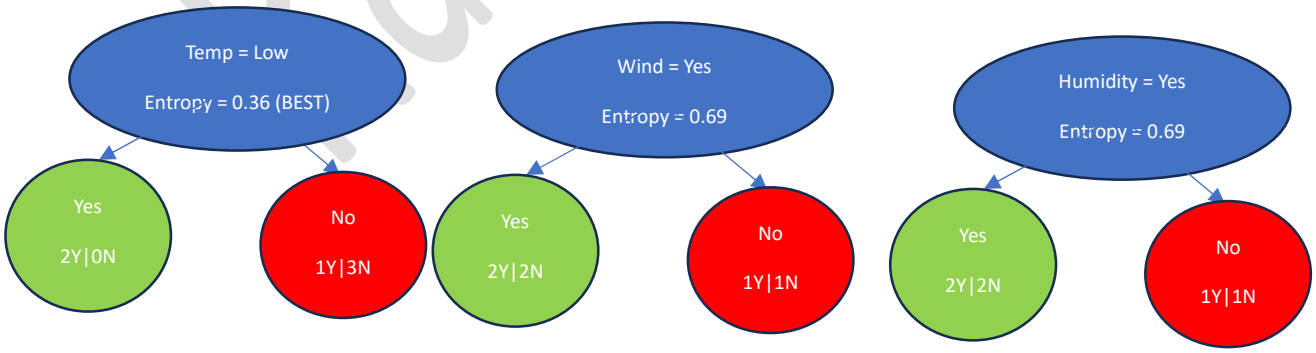
Input of one model goes to the other model

**Steps:**

1) Assign Weights
2) Create Decision Stumps
3) Decide which one is the best decision stump (Entropy)
4) Find $\alpha_1$
5) Increase the weight for the incorrectly classified points
6) Decrease the weight for the correctly classified points

**Data**

| Temp | Wind | Humidity | Play | Sample Weight |
|---|---|---|---|---|
| High | No | Yes | No | 1/6 |
| Low | Yes | No | Yes | 1/6 |
| High | Yes | Yes | No | 1/6 |
| High | No | Yes | Yes | 1/6 |
| Low | Yes | Yes | Yes | 1/6 |
| High | Yes | No | No | 1/6 |

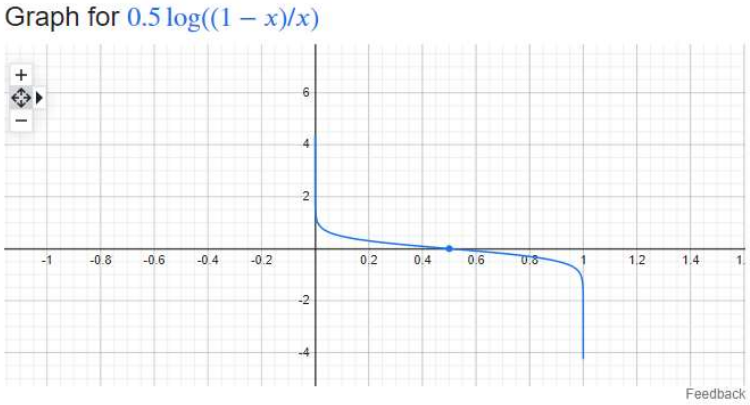


Instructor: Rahul Tiwari
Email: support@rahultiwari.co.in
LinkedIn: rahultiwari1986 | Mob: 9643393302

# Calculating the weights

**Consider a scenario!**

- Model 1 – 100% Prediction correct – More positive Alpha value
- Model 2 – 100% Prediction Incorrect – More negative Alpha value
- Model 3 – 50% Prediction correct and 50% incorrect – 0 Alpha value




- This function will not work well when we have the pure data
- Where X = Error
- The value of X will vary from 0 to 1

As in the above example only one row incorrectly classified the value of $\alpha_1 = \frac{1}{2}\log\left(\frac{1-\frac{1}{6}}{\frac{1}{6}}\right) = 0.80$

Increase the weight for the incorrect data and decrease the weight for the correct data classification

| Temp | Wind | Humidity | Play | Sample Weight | Adjusted W | Normalized W |
|---|---|---|---|---|---|---|
| High | No | Yes | No | 1/6 | 0.07 | 0.09 |
| Low | Yes | No | Yes | 1/6 | 0.07 | 0.09 |
| High | Yes | Yes | No | 1/6 | 0.07 | 0.09 |
| High | No | Yes | Yes | 1/6 | 0.37 | 0.54 |
| Low | Yes | Yes | Yes | 1/6 | 0.07 | 0.09 |
| High | Yes | No | No | 1/6 | 0.07 | 0.09 |

To calculate the adjusted weight the formula is

For correct prediction: $\frac{1}{6} * e^{-0.8} = 0.07$

For incorrect prediction: $\frac{1}{6} * e^{0.8} = 0.37$

In the above weights the sum is not 1 so we will normalize the weights so we will sum the weights = Total Sum = 0.07+0.07+0.07+0.37+0.07+0.07 = 0.72

We will divide all the individual numbers with 0.72

Instructor: Rahul Tiwari
Email: support@rahultiwari.co.in
LinkedIn: rahultiwari1986 | Mob: 9643393302

| Temp | Wind | Humidity | Play | Sample Weight | Adjusted W | Normalized W | Bin |
|---|---|---|---|---|---|---|---|
| High | No | Yes | No | 1/6 | 0.07 | 0.09 | 0-0.09 |
| Low | Yes | No | Yes | 1/6 | 0.07 | 0.09 | 0.09-0.18 |
| High | Yes | Yes | No | 1/6 | 0.07 | 0.09 | 0.18-0.27 |
| High | No | Yes | Yes | 1/6 | 0.37 | 0.54 | 0.27-0.81 |
| Low | Yes | Yes | Yes | 1/6 | 0.07 | 0.09 | 0.81-0.90 |
| High | Yes | No | No | 1/6 | 0.07 | 0.09 | 0.90-1 |

Now we will generate new records based on the importance of weights as the record highlighted was predicted wrong, we are giving more weight to that record. So in the new data we will have more number of entries for that record

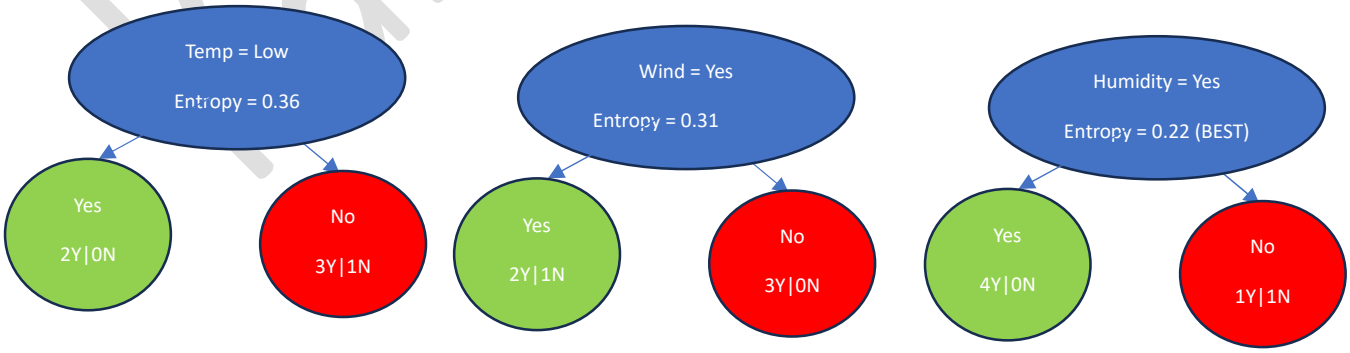
We will randomly generate the number from 0-1

For example:

- 0.93
- **0.64**
- **0.39**
- 0.13
- **0.5**
- 0.87

So the new records will be and you can see there will be more number of instances for the wrongly predicted data in the new data

| Temp | Wind | Humidity | Play | Sample Weight |
|---|---|---|---|---|
| High | Yes | No | No | 1/6 |
| High | No | Yes | Yes | 1/6 |
| High | No | Yes | Yes | 1/6 |
| Low | Yes | No | Yes | 1/6 |
| High | No | Yes | Yes | 1/6 |
| Low | Yes | Yes | Yes | 1/6 |



Instructor: Rahul Tiwari
Email: support@rahultiwari.co.in
LinkedIn: rahultiwari1986 | Mob: 9643393302

$\alpha_1 = 0.80$ (Temperature)

$\alpha_2 = 0.80$ (Humidity)

$\alpha_3 = 0.62$ (Wind)

$$F = 0.80H_1 + 0.80H_2 + 0.62H_3$$

New Data point:

Temperature = High

Wind = Yes

Humidity = Yes

0.8*N + 0.8*Y + 0.62*Y

So there is a more weightage to Y so the prediction will be Y