AdaBoost – Intuition ASAP

Getting Started with ML – Raunak Joshi

AdaBoost

- It is an Ensemble Learning Method.
- It is a Boosting Procedure.
- It is also known Adaptative Boosting.

AdaBoost Algorithm Steps.

- Initializing the Sample Weights.
- Decision Tree Stumps Building.
- Significance Evaluation.
- Updating Weights.
- Reforming Data.
- Forest Prediction.

Step 1 – Initializing Sample Weights.

Sleek Design	Good Suspension	Good Mileage	Good Car	Sample Weights
No	No	No	False	1/8
Yes	Yes	No	False	1/8
No	Yes	Yes	True	1/8
Yes	Yes	Yes	True	1/8
No	Yes	No	False	1/8
Yes	Yes	Yes	True	1/8
Yes	Yes	No	True	1/8
Yes	No	Yes	True	1/8

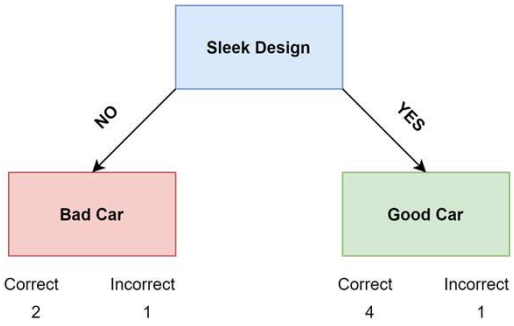
$$\boldsymbol{Initial\ Weights} = \frac{\mathbf{1}}{\boldsymbol{n}}$$
$$\boldsymbol{where\ n\ is\ total\ number\ of\ samples}$$

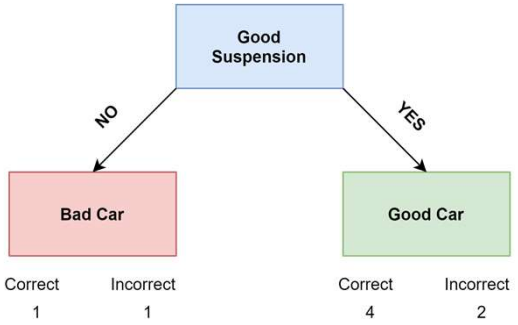
Step 2 - Decision Tree Stumps Building

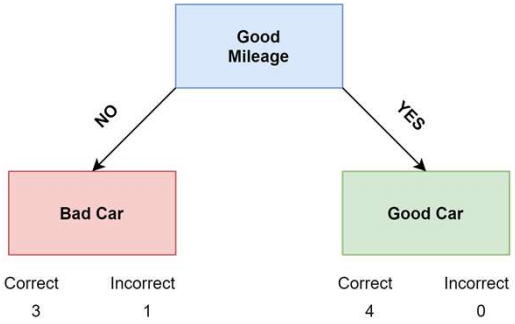
- Build a set of Weak Classifiers.
- Decision Tree with depth 1 is an example of Weak Classifier.
- This type of Decision Tree is known as **STUMP**.

Step 2 - Decision Tree Stumps Building

Sleek Design	Good Suspension	Good Mileage	Good Car	Sample Weights
No	No	No	False	1/8
Yes	Yes	No	False	1/8
No	Yes	Yes	True	1/8
Yes	Yes	Yes	True	1/8
No	Yes	No	False	1/8
Yes	Yes	Yes	True	1/8
Yes	Yes	No	True	1/8
Yes	No	Yes	True	1/8

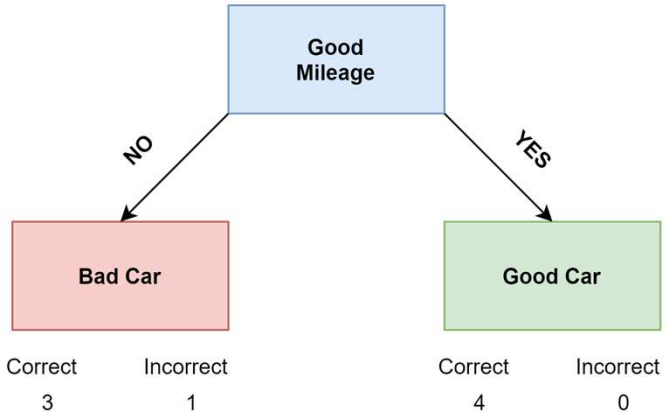






Calculate the Gini/Entropy for all Trees and select Tree with Least Impurity

Step 3 - Significance Evaluation



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

$$Total\ Error = Sum\ of\ Weights\ for\ Incorrect\ Samples$$

$$Total\ Error = \frac{1}{8}$$

$$Amount\ of\ Say = \frac{1}{2}\log\left(\frac{1 - \frac{1}{8}}{\frac{1}{8}}\right) = 0.973$$

NOTE : Log value is always base to the *e*

Step 4 - Updating Weights

$$Updated\ Weight = Sample\ Weight * e^{Amount\ of\ Say}$$

$$Updated\ Weight = \frac{1}{8} * e^{0.973}$$

$$Updated\ Weight = 0.33$$

Step 4 - Updating Weights

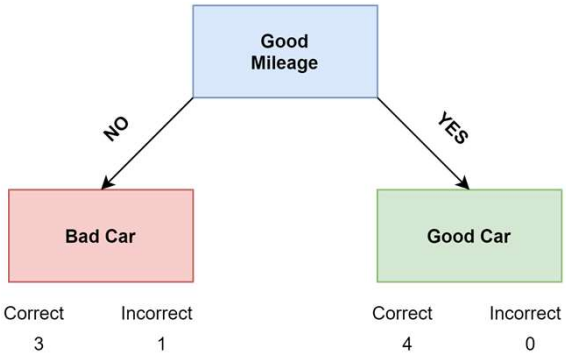
$$Updated\ Weight\ of\ other\ samples = Sample\ Weight * e^{-Amou \quad of\ Say}$$

$$Updated\ Weight = \frac{1}{8} * e^{-0.973}$$

$$Updated\ Weight = 0.05$$

Step 4 - Updating Weights

Sleek Design	Good Suspension	Good Mileage	Good Car	Sample Weights	Updated Weights
No	No	No	False	1/8	0.05
Yes	Yes	No	False	1/8	0.05
No	Yes	Yes	True	1/8	0.05
Yes	Yes	Yes	True	1/8	0.05
No	Yes	No	False	1/8	0.05
Yes	Yes	Yes	True	1/8	0.05
Yes	Yes	No	True	1/8	0.33
Yes	No	Yes	True	1/8	0.05



$$Adding\ all\ Updated\ Weights\ gives = 0.68$$

Step 4 - Updating Weights

Sleek Design	Good Suspension	Good Mileage	Good Car	Sample Weights	Updated Weights	Normalized Weights
No	No	No	False	1/8	0.05	0.07
Yes	Yes	No	False	1/8	0.05	0.07
No	Yes	Yes	True	1/8	0.05	0.07
Yes	Yes	Yes	True	1/8	0.05	0.07
No	Yes	No	False	1/8	0.05	0.07
Yes	Yes	Yes	True	1/8	0.05	0.07
Yes	Yes	No	True	1/8	0.33	0.49
Yes	No	Yes	True	1/8	0.05	0.07

$$Normalized\ Weights = \frac{Updated\ Weights}{Addition\ of\ Updated\ Weights}$$

Step 5 - Reforming Data

Sleek Design	Good Suspension	Good Mileage	Good Car	Sample Weights	Updated Weights	Normalized Weights	Cumulative Weights
No	No	No	False	1/8	0.05	0.07	0.07
Yes	Yes	No	False	1/8	0.05	0.07	0.14
No	Yes	Yes	True	1/8	0.05	0.07	0.21
Yes	Yes	Yes	True	1/8	0.05	0.07	0.28
No	Yes	No	False	1/8	0.05	0.07	0.35
Yes	Yes	Yes	True	1/8	0.05	0.07	0.42
Yes	Yes	No	True	1/8	0.33	0.49	0.91
Yes	No	Yes	True	1/8	0.05	0.07	0.98

Step 5 - Reforming Data

Sleek Design	Good Suspension	Good Mileage	Good Car
No	No	No	False
Yes	Yes	No	False
No	Yes	Yes	True
Yes	Yes	Yes	True
No	Yes	No	False
Yes	Yes	Yes	True
Yes	Yes	No	True
Yes	No	Yes	True

Sample Weights	Updated Weights	Normalized Weights	Cumulative Weights
1/8	0.05	0.07	0.07
1/8	0.05	0.07	0.14
1/8	0.05	0.07	0.21
1/8	0.05	0.07	0.28
1/8	0.05	0.07	0.35
1/8	0.05	0.07	0.42
1/8	0.33	0.49	0.91
1/8	0.05	0.07	0.98

Randomly Selected value is 0.39

Sleek Design	Good Suspension	Good Mileage	Good Car
Yes	Yes	Yes	True

Step 5 - Reforming Data

Sleek Design	Good Suspension	Good Mileage	Good Car
No	No	No	False
Yes	Yes	No	False
No	Yes	Yes	True
Yes	Yes	Yes	True
No	Yes	No	False
Yes	Yes	Yes	True
Yes	Yes	No	True
Yes	No	Yes	True

Randomly Selected value is 0.81

Sample Weights	Updated Weights	Normalized Weights	Cumulative Weights
1/8	0.05	0.07	0.07
1/8	0.05	0.07	0.14
1/8	0.05	0.07	0.21
1/8	0.05	0.07	0.28
1/8	0.05	0.07	0.35
1/8	0.05	0.07	0.42
1/8	0.33	0.49	0.91
1/8	0.05	0.07	0.98

Sleek Design	Good Suspension	Good Mileage	Good Car
Yes	Yes	Yes	True
Yes	Yes	No	True

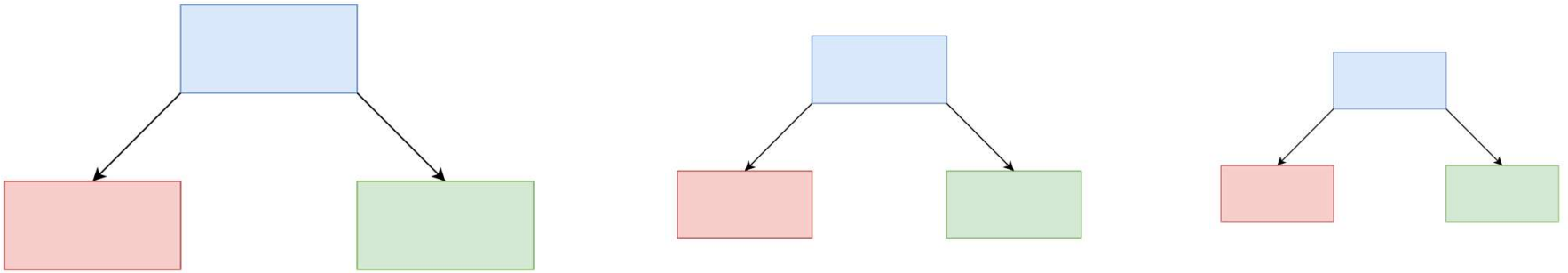
Step 5 - Reforming Data

Sleek Design	Good Suspension	Good Mileage	Good Car
No	No	No	False
Yes	Yes	No	False
No	Yes	Yes	True
Yes	Yes	Yes	True
No	Yes	No	False
Yes	Yes	Yes	True
Yes	Yes	No	True
Yes	No	Yes	True

Sleek Design	Good Suspension	Good Mileage	Good Car
Yes	Yes	Yes	True
Yes	Yes	No	True
Yes	Yes	Yes	True
Yes	Yes	No	True
Yes	Yes	No	True
No	No	No	False
Yes	Yes	No	True
Yes	No	Yes	True

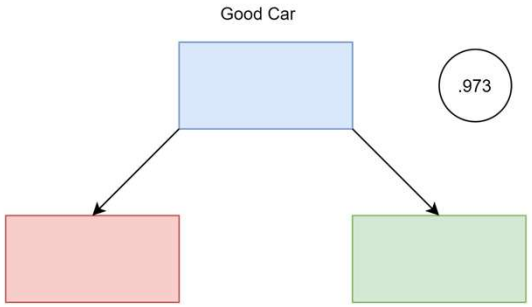
Repeat Steps 2 to 5 till you get the number of estimators in place.

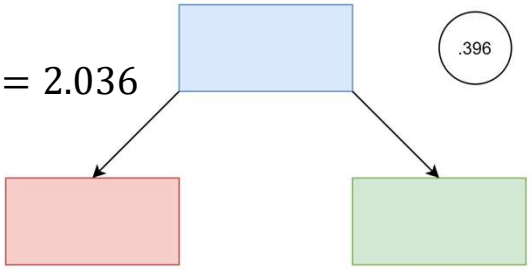
Step 5 - Reforming Data

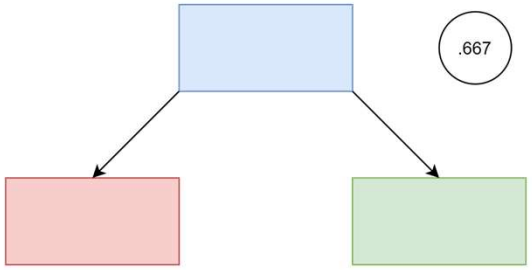


Repeat Steps 2 to 5 till you get the number of estimators in place.

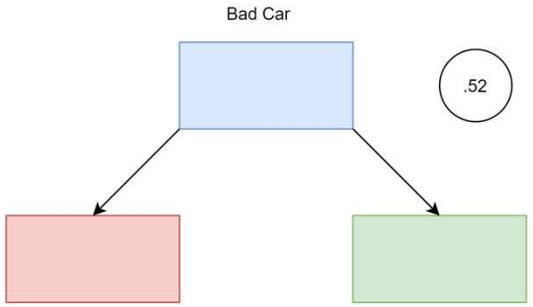
Step 6 – Forest Prediction

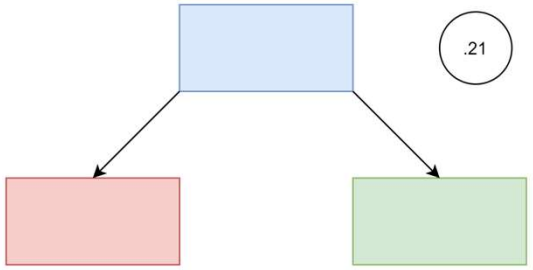


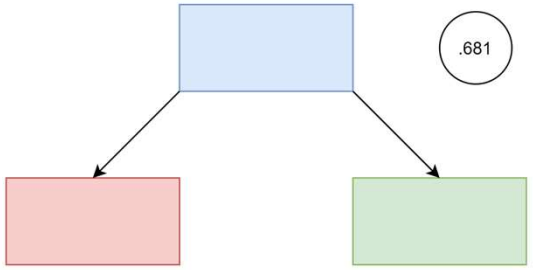




$.973 + .396 + .667 = 2.036$







$.52 + .21 + .681 = 1.411$