AdaBoost and Decision Trees

Random Forests

3 concepts

Create forest and how it use to make classifciations

////////////////////

Randomforest better that decision trees.

Need a dataset,

Dataset – instances and features  
/////////////////////////

Decision Trees

Questions-classifies(categories or numeric) based on answers(conditions)

Root Node (Root) > Internal Nodes (Nodes) > Leaf Nodes (leaf)

------------------------

Choosing Root, check leaf for purity.

Check leaf impurity, “Gini”

Lower Gini = lower impurity

Gini = 1 – prob\_of\_yes^2 - prob\_of\_no^2

Where, prob\_of\_yes = yes/total\_elements\_of\_leaf

Total Gini for root =

( total\_elements\_of\_leaf\_1 / total\_elements\_of\_both\_leaf\_nodes ) \*Gini\_leaf\_1

+ ( total\_elements\_of\_leaf\_2 / total\_elements\_of\_both\_leaf\_nodes ) \*Gini\_leaf\_2

If Gini Impurity is low make it a leaf

Raw data to a decision tree

Categories based on yes/no data

-----------------------------------------------

Now Numerics

Sort data, find averages of adjacent data, find Gini impurity for the averages:target for lower than or higher than the averages. Choose lowest impurity for root

Ranked data

/////////////////