Decision Tree

decision trees are constructed via an algorithmic approach that identifies ways to split a data set based on different conditions. Decision Trees are a non-parametric supervised learning method used for both classification and regression tasks.

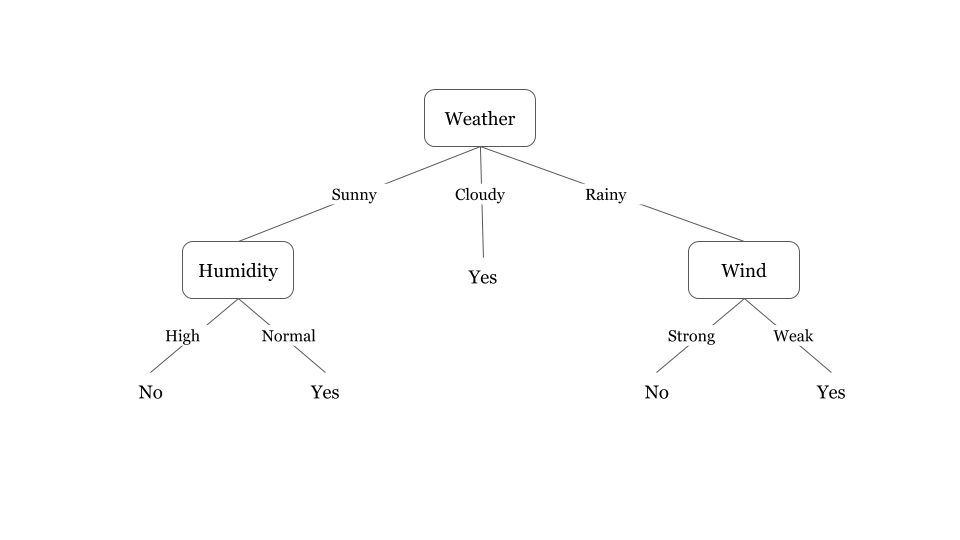
It is a tree-structured classifier, where**internal nodes represent the features of a dataset, branches represent the decision rules** and **each leaf node represents the outcome.**

## Structure

A decision tree is drawn upside down with its root at the top., the bold text in black represents a condition/**internal node**, based on which **the tree splits** into branches/ **edges**. The end of the branch that doesn’t split anymore is the decision/**leaf**, Decision Tree algorithms are referred to as CART or Classification and Regression Trees.

Nonparametric methods, or distribution-free methods, are statistical methods that do not rely on assumptions that the data are drawn from a given probability distribution.

## Example



Illustrates a learned decision tree. We can see that each node represents an attribute or feature and the branch from each node represents the outcome of that node. Finally, its the leaves of the tree where the final decision is made.

The best split is one which separates two different labels into two sets.

## Expressiveness of decision trees

Decision trees can represent any boolean function of the input attributes. Let’s use decision trees to perform the function of three boolean gates AND, OR and XOR.

|  |
| --- |
| https://www.hackerearth.com/practice/machine-learning/machine-learning-algorithms/ml-decision-tree/tutorial/ |

# Attribute selection measure or ASM

## we can easily select the best attribute for the nodes of the tree with ASM. There are two popular techniques for ASM,

* **Information Gain**
* **Gini Index**

## how to choose the best attribute?

The best attribute in terms of which attribute has the most **information gain**, a measure that expresses **how well an attribute splits that data into groups** based on classification. “Information gain” which tells us how much the parent entropy has decreased after splitting it with some feature.

## information gain

Information gain is the measurement of **changes in entropy** after the segmentation of a dataset based on an attribute. It calculates how **much information a feature provides** us about a class. According to the value of information gain, we split the node and build the decision tree and a node/attribute having the **highest information gain** is split first.

|  |  |
| --- | --- |
| https://s3-ap-southeast-1.amazonaws.com/he-public-data/high%20information%20gaine8d3940.png |  |

* **S= Total number of samples**
* **P(yes)= probability of yes**
* **P(no)= probability of no**

|  |
| --- |
| 1. Information Gain= Entropy(S)- [(Weighted Avg) \*Entropy(each feature) |

## Entropy

To define information, gain precisely, we need to measure **entropy**that measures the level of impurity in a group of examples.  It **specifies randomness** in data. Entropy is nothing but the uncertainty in our dataset. In order to make a decision tree, we need to calculate the impurity of each split, and when the purity is 100%, we make it as a leaf node.

|  |
| --- |
| **Entropy(s)= -P(yes)log2\*P(yes)- P(no)\*log2 P(no)** |

To check the impurity of feature 2 and feature 3 we will take the help for Entropy formula.

|  |  |
| --- | --- |
| feature 2 | entropy calculation |

## Gini Index

Gini index is a measure of **impurity** or purity used while creating a decision tree in the CART (Classification and Regression Tree) algorithm. An attribute with **the low Gini index** should be preferred as compared to the high Gini index. It only creates **binary splits**, and the CART algorithm uses the Gini index to create binary splits

|  |
| --- |
| **Gini Index= 1- ∑jPj2** |

## Pruning

Pruning is a process of deleting the unnecessary nodes from a tree in order to get the optimal decision tree.

## Confusion Matrix

Confusion Matrix is used to understand the trained classifier behavior over the test dataset or validate dataset.

## Accuracy score

Accuracy score is used to calculate the accuracy of the trained classifier.

# Disadvantages of the Decision Tree

* The decision tree contains lots of layers, which makes it complex.
* It may have an overfitting issue, which can be resolved using the **Random Forest algorithm.**
* For more class labels, the computational complexity of the decision tree may increase.