**Decision Tree Regression Intuition**

**Decision Tree** is the most powerful and popular tool for classification and prediction. A Decision tree is a flowchart-like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test, and each leaf node (terminal node) holds a class label.



*A decision tree for the concept PlayTennis.*

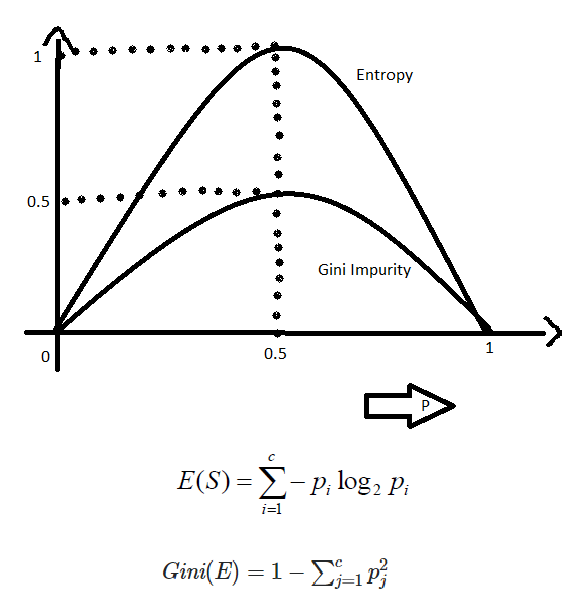
**Construction of Decision Tree:** A tree can be *“learned”* by splitting the source set into subsets based on an attribute value test. This process is repeated on each derived subset in a recursive manner called*recursive partitioning*. The recursion is completed when the subset at a node all has the same value of the target variable, or when splitting no longer adds value to the predictions. The construction of a decision tree classifier does not require any domain knowledge or parameter setting, and therefore is appropriate for exploratory knowledge discovery. Decision trees can handle high-dimensional data. In general decision tree classifier has good accuracy. Decision tree induction is a typical inductive approach to learn knowledge on classification.

**Entropy:**  
As discussed above entropy helps us to build an appropriate decision tree for selecting the best splitter. Entropy can be defined as a measure of the purity of the sub split. Entropy always lies between 0 to 1. The entropy of any split can be calculated by this formula.

The algorithm calculates the entropy of each feature after every split and as the splitting continues on, it selects the best feature and starts splitting according to it. For a detailed calculation of entropy with an example, you can refer to [this article](https://www.geeksforgeeks.org/decision-tree-introduction-example/).

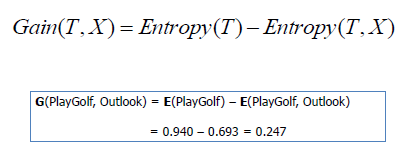
**Gini Impurity:**  
The internal working of Gini impurity is also somewhat similar to the working of entropy in the Decision Tree. In the Decision Tree algorithm, both are used for building the tree by splitting as per the appropriate features but there is quite a difference in the computation of both the methods. Gini Impurity of features after splitting can be calculated by using this formula.

For the detailed computation of the Gini Impurity with examples, you can refer to [this article](https://www.geeksforgeeks.org/decision-tree-introduction-example/). By using the above formula gini Impurity of feature/split is being calculated.

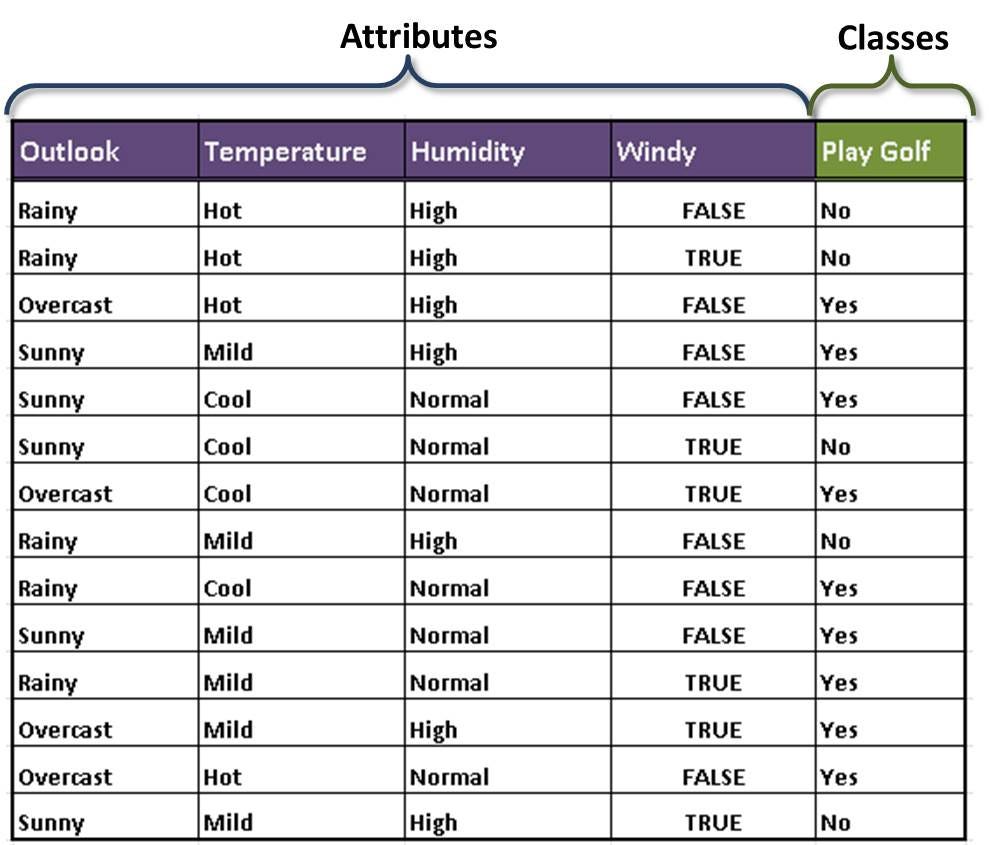


**Information gain:**

Information gain is the entropy difference of target and information 1 with respect to target.



**Lets Take an example to understand the split in decision tree**



In above dataset of golf we have **Outlook,Temperature,Humidity,Wind as** Input parameters and playing golf is output parameter, to spit the data we must have the knowledge of entropy of parent node.

Now that we know these parameters, we can start the construction of the decision tree. First, we need to determine the root node of the decision tree. As the dataset is split into two subtypes — Attributes and Class, we calculate the entropy for both, and the following Entropies are obtained.

E(Play Golf)

E(Play Golf, Outlook)

E(Play Golf, Temperature)

E(Play Golf, Humidity)

E(Play Golf, Windy)

After the calculation of the Entropies, we calculate the Information Gain.

Gain(PlayGolf, Outlook) = Entropy(PlayGolf) — Entropy(PlayGolf, Outlook)

Gain(PlayGolf, Temperature) = Entropy(PlayGolf) — Entropy(PlayGolf, Temparature)

Gain(PlayGolf, Humidity) = Entropy(PlayGolf) — Entropy(PlayGolf, Humidity)

Gain(PlayGolf, Windy) = Entropy(PlayGolf) — Entropy(PlayGolf, Windy)

Now that we have all the necessary values, we can start the splitting. The first split i.e the root node is decided on the attribute which gives us the highest information gain. In this case, it is the Outlook attribute. The further splits will be decided based on which attribute gives us the homogeneous groups. The complete decision tree is shown below.

