**Entropy in the decision tree, how it works**

What is a Decision Tree?

Ans: A decision tree is a sort of probability tree that may be used to choose a process. For example, you may decide whether to manufacture item A or item B, or whether to invest in option 1, option 2, or option 3. Trees are a great method to cope with these sorts of complicated decisions, which typically include a lot of variables and are frequently fraught with ambiguity. Although they may be created by hand, the software is frequently utilized since trees can quickly grow complicated.

In a tree, there are three main areas:

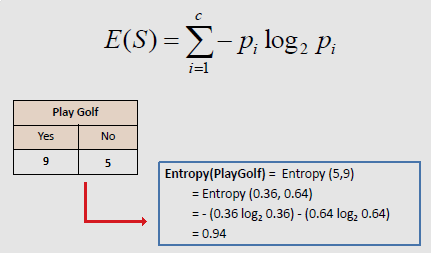
**The Decision:** a square node with two or more arcs (referred to as "decision branches") pointing to the alternatives.

**The event sequence**: represented as a circular node with two or more arcs pointing to the relevant occurrences. The circle nodes, often known as "chance nodes," can be used to represent probabilities.

**Consequences:** the costs or benefits associated with distinct decision tree paths. On a computer, the endpoint is known as a "Terminal" and is represented by a triangle or bar.

**How Entropy works in decision tree:**

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



**Information gain in the decision tree**

When we use a node in a decision tree to partition the training instances into smaller subsets the entropy changes. Information gain is a measure of this change in entropy.

Definition: Suppose S is a set of instances, A is an attribute, Sv is the subset of S with A = v, and Values (A) is the set of all possible values of A, then

Gain(S, A) = Entropy(S) - \sum\_{v \epsilon Values(A)}\frac{\left | S\_{v} \right |}{\left | S \right |}. Entropy(S\_{v})

Entropy

Entropy is the measure of uncertainty of a random variable, it characterizes the impurity of an arbitrary collection of examples. The higher the entropy more the information content.

Definition: Suppose S is a set of instances, A is an attribute, Sv is the subset of S with A = v, and Values (A) is the set of all possible values of A, then

Gain(S, A) = Entropy(S) - \sum\_{v \epsilon Values(A)}\frac{\left | S\_{v} \right |}{\left | S \right |}. Entropy(S\_{v})

Example:

For the set X = {a,a,a,b,b,b,b,b}

Total intances: 8

Instances of b: 5

Instances of a: 3

Entropy H(X) = -\left [ \left ( \frac{3}{8} \right )log\_{2}\frac{3}{8} + \left ( \frac{5}{8} \right )log\_{2}\frac{5}{8} \right ]

= -[0.375 \* (-1.415) + 0.625 \* (-0.678)]

=-(-0.53-0.424)

= 0.954

Building Decision Tree using Information Gain well as a classification tree depending upon the dependent variable.

**Information Gain use in decision tree:**

Information gain can be used as a split criterion in most modern implementations of decision trees, such as the implementation of the Classification and Regression Tree (CART) algorithm in the scikit-learn Python machine learning library in the Decision Tree Classifier class for classification.

