Decision Tree Learning

Decision tree learning is one of the most common and practical methods used for inductive inference and successfully applied in many subjects from health to finance. Decision trees can be used for discrete and real-value data sets.

A decision tree is a tree where each node represents a feature, each link represents a decision and each leaf represents an outcome(categorical or continuous value). Most algorithms developed for learning in decision trees consist of variations of greedy search. This approach, which will explain the focus of our discussions with the ID3 algorithm.

ID3 Algorithm

The ID3 algorithm starts with the question of which feature should be tested at the root of the tree. The answer to this question is possible by statistically testing how well each feature classifies the training set. Once the best feature is selected, a root child is created for each value of the feature. The training set samples are then sorted by the appropriate child node. After this stage, the appropriate child for each feature is repeated for the knot. The reason why the algorithm is considered a greedy search is that it doesn't take a step back to rethink any choice.

Entropy and Information Gain

The crucial question at this stage is how to choose the best feature.While the ID3 algorithm finds the best feature for each node, it takes advantage of a statistical feature called information gain. Before defining knowledge gain, we need to define the concept of entropy as follows, we measure the purity of an arbitrary set with entropy.

$$Entropy(S) = \sum_{i=1}^{c} -p_i log_2 p_i$$

The target attribute can take on c different values, then the entropy of S is defined as where

p_i is the proportion of S belonging to class i.