Decision Trees

Decision trees are classifiers on a target attribute (or class) in the form of a tree structure. The observations (or items) to classify are composed of attributes and their target value. The nodes of the tree can be:

- decision nodes, in these nodes a single attribute-value is tested to determine to which branch of the subtree applies. Or

- leaf nodes which indicate the value of the target attribute.

Algorithms

Hunts (TDIDT) algorithm

In Hunt's algorithm which is also named as *Top-Down Induction of Decision Tree,* a decision tree is grown in a recursive fashion by partitioning the training records into successively purer subsets. Let D_t be the set of training records that are associated with node t and $y = \{y_1, y_2, \cdots, y_c\}$ be the class labels.

1. If all reccords in D_t belong to the same class y_t, then t is a leaf node labeled as y_t which has exactly one incoming edge and no outgoing edges.

2. If D_t is an empty set, then t is a leaf node labeled by the default class, y_d

3. If D_t contains records that belong to more than one class, *an attribute test condition* is selected to partition the records into smaller subsets recursively.

Test conditions

- Depends on attribute types

  - Nominal
  - Ordinal
  - Continuous

- Depends on number of ways to split

  - 2-way split
  - 2-way split

Node Impurity

Nodes with homogeneous class distribution are preferred

- Gini Index

- Entropy

- Misclassification error