Decision Trees

Epoch IIT Hyderabad

Himani Agrawal
MA22BTECH11008

Contents

Introduction

Decision tree learning is a method for approximating discrete-valued target functions, in which the learned function is represented by a decision tree. Learned trees can also be re-represented as sets of if-then rules to improve human readability. Decision trees classify instances by sorting them down the tree from the root to some leaf node, which provides the classification of the instance.

Attribute selection measures

- Construction of Decision Tree**:** A tree can be *"learned"* by splitting the source set into subsets based on Attribute Selection Measures. Attribute selection measure (ASM) is a criterion used in decision tree algorithms to evaluate the usefulness of different attributes for splitting a dataset. The goal of ASM is to identify the attribute that will create the most homogeneous subsets of data after the split, thereby maximizing the information gain.This process is repeated on each derived subset in a recursive manner called *recursive partitioning*.
- Entropy is the measure of the degree of randomness or uncertainty in the dataset. In the case of classifications, It measures the randomness based on the distribution of class labels in the dataset.
- Gini Impurity is a score that evaluates how accurate a split is among the classified groups. The Gini Impurity evaluates a score in the range between 0 and 1, where 0 is when all observations belong to one class, and 1 is a random distribution of the elements within classes. It is given by : $1-\sum p_i^2$ where pi is the proportion of the elements in the set that belongs to ith category

- Information gain measures the reduction in entropy or variance that results from splitting a dataset based on a specific property. It is used in decision tree algorithms to determine the usefulness of a feature by partitioning the dataset into more homogeneous subsets with respect to the class labels or target variable. The higher the information gain, the more valuable the feature is in predicting the target variable. Information gain is calculated by:

  $$\text{Information gain}(H,A)=H-\sum(|H_v|/|H|).H_v$$

  Where, A=Specific attribute,

  $|H|$=Entropy of dataset sample S

  $|H_v|$= number of instances in the subset S that have the value v for attribute A

Classification and regression tree algorithm for classification

Creating a CART model involves selecting input variables and split points on those variables until a suitable tree is constructed.The selection of which input variable to use and the specific split or cut-point is chosen using a greedy algorithm to minimize a cost function. Tree construction ends using a predefined stopping criterion, such as a minimum number of training instances assigned to each leaf node of the tree.

Greedy splitting is a numerical procedure where all the values are lined up and different split points are tried and tested using a cost function. The split with the best cost (lowest cost because we minimize cost) is selected.

The recursive binary splitting procedure described above needs to know when to stop splitting as it works its way down the tree with the training data.The most common stopping procedure is to use a minimum count on the number of training instances assigned to each leaf node. If the count is less than some minimum then the split is not accepted and the node is taken as a final leaf node.