Decision Tree

What is a Decision Tree?

Decision Trees (DTs) are a non-parametric supervised learning method used for classification and regression. The goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. A tree can be seen as a piecewise constant approximation.

Gini Impurity or Entropy

- Gini Index: Used in classification tasks, Measures the impurity of a node, Measures the likelihood of an incorrect classification of a new instance if it was randomly classified. A node with a Gini index of 0 is pure, meaning all instances belong to the same class. where pi, is the probability of an element being classified to a particular class.
- Entropy (Information Gain): Another criterion used for [classification]. It measures the amount of information needed to classify a sample. where pi, is the probability of an element being classified to a particular class.
- Information Gain (Kullback-Leibler divergence) is the reduction in entropy
- Variance Reduction (for Regression Trees): Measures the reduction in variance after a split, aiming to minimize the variance within each node.

Explain the Working of Decision Tree.

Below is the detailed breakdown of how the decision tree algorithm works:

- With all the data at its starting point, the process starts at the root node. In order to effectively divide the data into discrete classes or values, the algorithm chooses a [feature] together with a [threshold].
- Depending on the job (classification or regression), the feature and threshold are selected to [maximize information gain] or [minimize impurity].
- Depending on the outcome of the feature test, the data is separated into subgroups.
- When a characteristic like “Age” is used with a threshold of 30, for instance, the data is divided into two subsets: records with Age less than or equal to 30, and records with Age more than 30.
- For every subgroup, the splitting procedure is repeated, resulting in child nodes. Up until a given condition is satisfied, this recursive process keeps going.
- A node turns into a leaf node when a stopping requirement is satisfied. The final judgment or forecast is represented by the leaf nodes. In regression, the target variable’s mean or median value within the subset is usually found in the leaf node.

Advantages of Decision Trees:

- The Boolean logic and visual representations of decision trees make them easier to understand and consume. The hierarchical nature of a decision tree also makes it easy to see which attributes are most important, which isn't always clear with other algorithms, like neural networks.
- Decision trees have a number of characteristics, which make it more flexible than other classifiers. It can handle various data types—i.e. discrete or continuous values, and continuous values can be converted into categorical values through the use of thresholds.
- Additionally, it can also handle values with missing values, which can be problematic for other classifiers, like Naïve Bayes.
- Decision trees can be leveraged for both classification and regression tasks, making it more flexible than some other algorithms.
- It's also insensitive to underlying relationships between attributes; this means that if two variables are highly correlated, the algorithm will only choose one of the features to split on.

Disadvantages of Decision Trees:

- Complex decision trees tend to overfit and do not generalize well to new data. This scenario can be avoided through the processes of pre-pruning or post-pruning. Pre-pruning halts tree growth when there is insufficient data while post-pruning removes subtrees with inadequate data after tree construction.
- Small variations within data can produce a very different decision tree. Bagging, or the averaging of estimates, can be a method of reducing variance of decision trees. However, this approach is limited as it can lead to highly correlated predictors.
- Given that decision trees take a greedy search approach during construction, they can be more expensive to train compared to other algorithms.