

CHAPTER 23

Ensemble Methods

Ensemble learning is a technique that combines the output of multiple classifiers also called weak learners to build a more robust prediction model. Ensemble methods work by combining a group of classifiers (or models) to get an enhanced prediction accuracy. The idea behind an “ensemble” is that the performance from the average of a group of classifiers will be better than each classifier on its own. So each classifier is called a “weak” learner.

Ensemble learners are usually high-performing algorithms for both classification and regression tasks and are mostly competition-winning algorithms. Examples of ensemble learning algorithms are Random Forest (RF) and Stochastic Gradient Boosting (SGB). We will motivate our discussion of ensemble methods by first discussing decision trees because ensemble classifiers such as RF and SGB are built by combining several decision tree classifiers.

Decision Trees

Decision trees, more popularly known as classification and regression trees (CART), can be visualized as a graph or flowchart of decisions. A branch connects the nodes in the graph, the last node of the graph is called a terminal node, and the topmost node is called the root. As seen in Figure 23-1, when constructing a decision tree, the root is at the top, while the branches connect nodes at lower layers until the terminal node.

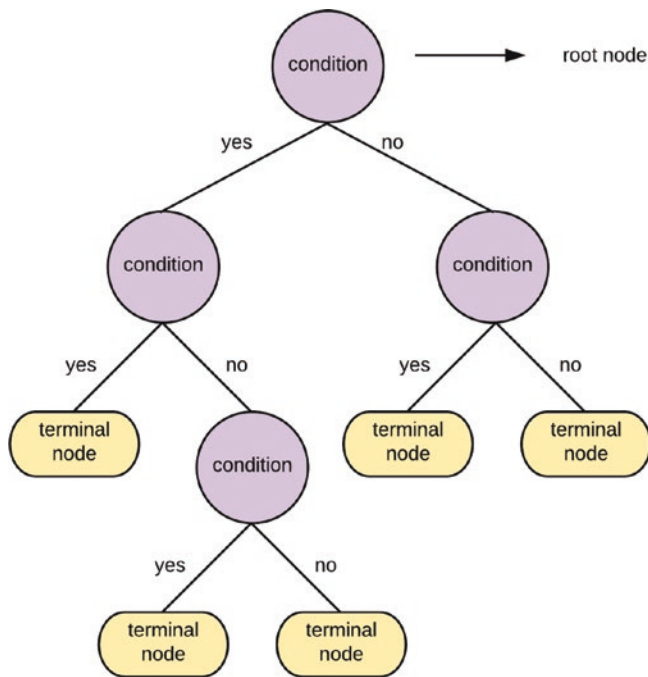


Figure 23-1. *Illustration of a decision tree*

On Regression and Classification with CART

A classification or regression tree is built by randomly splitting the set of attributes of the given dataset into distinct regions. The data points that fall within a particular region are used to form the predictor from the means of the targets in the regression case and the highest occurring class in the classification setting.

Thus, if an unseen observation or test data falls within a region, the mean or modal class is used to predict the output for regression and classification problems, respectively. In regression trees, the output variable is continuous, whereas in classification trees, the output variable is categorical. The terminal node of a regression tree takes the average of the samples in that region, while the terminal node of a classification tree is the highest occurring class in that area.

The process of splitting the features of the dataset into regions is by a greedy algorithm called recursive binary splitting. This strategy works by continuously dividing the feature space into two new branches or regions until a stopping criterion is reached.