← Go Back to Practical Data Science

Course Content

Bagging and Random Forest

Bagging

Before moving on to the bagging concept, let's define some key terms that are helpful in understanding the concept well.

Ensemble methods: Ensemble methods have multiple models trained on different subsets of the dataset, and the outputs of these models are somehow aggregated to get the final result.

These multiple models are also called the **base models/weak models.**

The two most common types of Ensemble methods are:

1. *Bagging*
2. *Boosting*

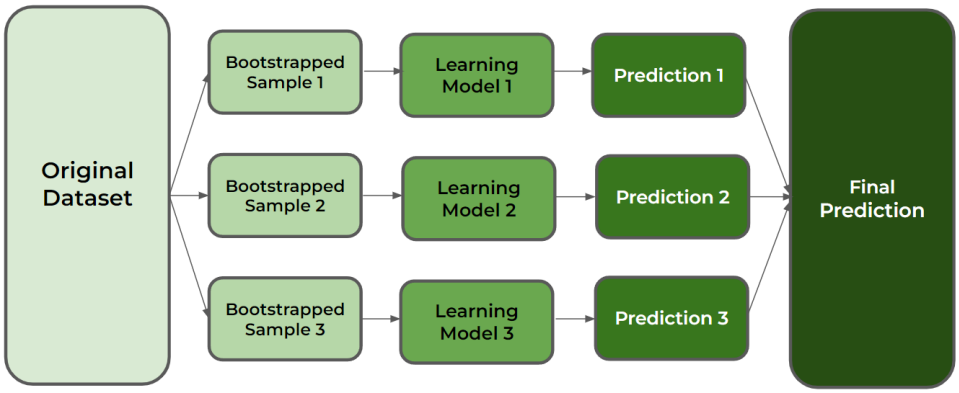
In this module, we will study Bagging.

Bagging is a combination of **Bootstrap(B) + Aggregation(Agg).**

In the previous week, we learned that bootstrapping is a statistical procedure that resamples a single dataset to create many simulated samples.

Aggregation is the process of combining the results of all the algorithms to get the final result.

A Bagging classifier is an ensemble method that trains a base model on each random subset of the original dataset. These subsets are made using bootstrapping and the classifier then aggregates their individual predictions (either by voting for classification or by averaging for regression) to form a final prediction.



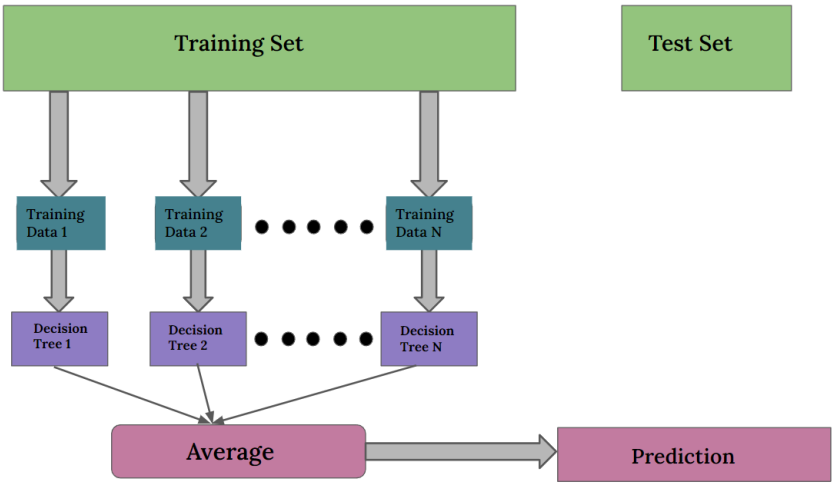
Random Forest

The Random Forest is a type of bagging algorithm where the base models are decision trees, and the algorithm uses only a subset of randomly picked independent variables (features) for each node's branching possibilities, unlike in bagging where all

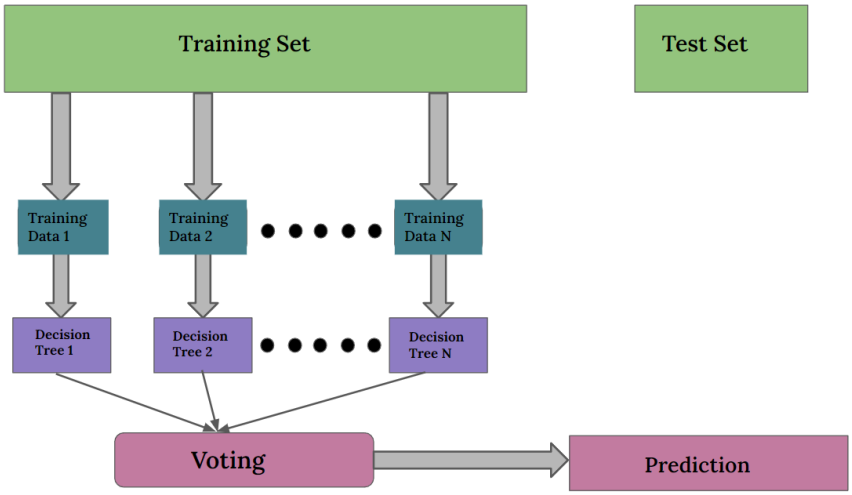
the features are considered for splitting a node.

Bootstrapped samples are taken from the original training data and on each bootstrapped training dataset, a decision tree is built by considering only a subset of features at each split. The results from all the decision trees are combined together and the final prediction is made using voting or averaging.

For regression problems, we take the average of all the predictions obtained from the different decision trees as our final prediction.



For classification problems however, we take the majority voting of all the predictions obtained from the decision trees as our final prediction.



Previous
Next