10 March 2018
20:30

Bagging, Random Forests:

Bagging, random forests, and boosting use trees as building blocks to construct more powerful prediction models.

Bootstrap or bagging, is a general-purpose procedure for reducing the bagging variance of a statistical learning method; we introduce it here because it is particularly useful and frequently used in the context of decision trees.

Hence a natural way to reduce the variance and hence increase the prediction accuracy of a statistical learning method is to take many training sets from the population, build a separate prediction model using each training set, and average the resulting predictions.

Random Forests: This method is exclusive for Decision trees only.

- Take any 5 variables and identify the minimum classification Error and consider it is Root node and start up building the decision tree.
- Again take any another 5 variables randomly with replacement and build the decision tree.
- Now, these decision trees to be connected one by one.

- Like that the process is repeated untill all the observations are exhausted. Such that we need to build one Lakh trees.
- If we want to predict for particular observation we are going to predict one lakh Predicted values from one lakh decision trees and calculated the average of all those.

- If it is the case of classification kind of problem, it will consider the number of votes which is predicted more.

So, this total process is called **Random Forests.**

So, whenever we are combining the different kinds of outputs and taking the average/No of votes and takes the decision is called **"Ensembling methods"**