ML Labs

Delivering Enterprise Machine Leaning , AI Services

Random Forest

To start learning about Random Forest, we should first know about the Ensembles.

Ensembles :

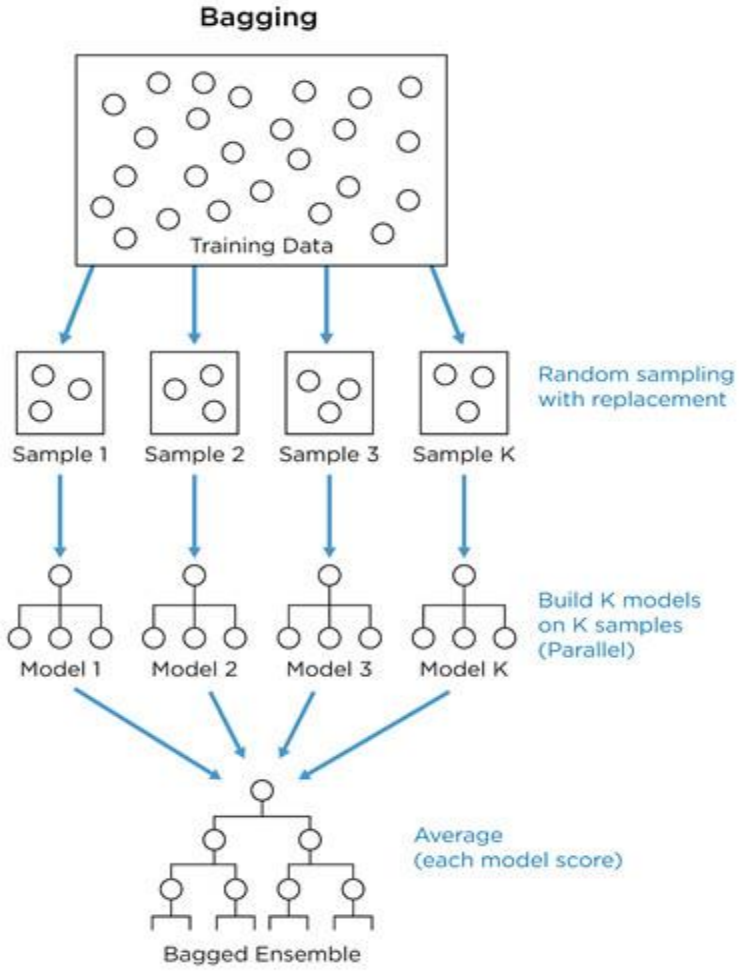
It means group of things viewed as whole rather than individually. They use a group of models to make prediction, unlike Decision Tree. Random forest is the most popular ensemble technique. Ensemble made by combinations of large number of decision trees. In Random forest Bagging technique is used which is a Ensemble method. Lets see the figure below how its work.

Creating Random Forest

Random Forest are having nearly same hyper parameters as Decision Tree. It is a collection of decision trees. Random Forest is created uses bagging (bootstrapped aggregation) as we discussed it's a ensemble technique. bagging is for choosing Random Samples from the given data and each set of samples trained in each tree in a Random forest.

The figure illustrates the total flow of Random forest:

- From the training data random bootstrap samples are taken.
- One Bootstrap sample contain 30-70% data form the dataset.
- The Decision Trees are creates on each Bootstrap sample.
- Finally we will take the Majority scores of all the predictions.



OOB(out-of-Bag) Error

OOB error is same as cross validation error. It is calculated by using each data point of training data as test point. While Bootstrap samples are taken there may be several Decision trees does not include some data points, which means that data point is unseen for that particular tree.

Lets take an example:

we have 200 data point and build Random Forest with 40 decision trees, with each tree contain 25 data points. We found that 15 trees does not have a data point “i” but still predicted the data point . Assume that 10 predicted as 0 and 5 predicted as 1. The final output will be taken as 0.

So OOB error is calculated as number of data points predicted wrongly as proportion of total number of data points.

Advantages of Random Forest

- Can Parallelize the training of forest, because each tree is constructed independently.
- Random forest is more stable than a single decision tree,
- Random forest is immune to the curse of dimensionality.
- We can calculate OOB error using training data, which gives a correct view of random forest on unseen data.

ML Labs Pvt Ltd
Marathahalli ,3rd Floor Above Khazana Jewellery Bangalore 560066
91-7338339898
91-7829396922
Connect : Bharath@pylabs.com