Ensemble Techniques and Random Forest

Prerequisite-

- Linear algebra.
- Basics of machine learning.

Objectives-

- What is ensemble learning
- Types of ensemble learning
- Understanding random forest and its working

Introduction-

In today's world where machine learning is keep getting better and better predictive model is the basic building block of machine learning and the good accuracy of the model is the ultimate goal of a machine learning engineer. This note describes one way or technique to increase the accuracy of a model known as ensemble technique. For example, when you plan to go on a vacation you will not go to some place randomly it would be highly unlikely. You will probably be going to check some websites, ask some of your friends about their favorite places and then accordingly take your decision on the basis of other people's opinions.

Ensemble technique works in the same as described in the above example. They combine multiple model to increase the overall accuracy or performance. The ensemble of models can be done in various ways.

What is Ensemble Learning-

To get better understanding Ensemble Learning, let us take an example suppose you are a cook in a restaurant and have made a new dish and now you want some feedback before you launch it in your restaurant the different tricks which you can use are:

i. You can go to one of your friends and ask him for the rating.
ii. You can ask 10 colleagues of yours to rate the food.
iii. You can ask 100 people to do the rating.

In the first case your friend might not want to break your heart and gives you a good rating even if he does not like the dish. In the second case the 10 of colleagues of yours might not have knowledge about that type of dish for example they are good at making vegetarian food but not good at making non vegetarian food. In third case out of 100 people, some of them maybe your friend, some may be your relatives and some may be complete unknown persons. The ratings, in this case would be more diversified and generalized as they have a variety of skill sets.

With this example, we can understand that diverse group of peoples have higher probability to make better decisions as compared to individual. The same analogy cab be used with creating a group of diverse set of models for predictions in comparison to just relying on one. This type of machine learning modelling method is achieved a technique known as Ensemble learning. Some simple ensemble learning techniques are:

- Averaging
- Max voting
- Weighted Average

In **averaging**, we take the prediction from each model and then use the average of these models will be your final prediction. It is used for making predictions in regression problems and in classification problems it is used for calculating probability.

Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Final rating**
5	4	5	4	4	**4.4**

In the above example each model predicts a movie rating and then the average of all the models is calculated to get an average outcome of 4.4.

Second method (**Max voting**) is mainly applied in classification problems. This method uses multiple models to make a prediction for each observation in given data where the output from each model is considered as a vote. The final prediction that we obtained is the value from most number of models.

Model 1	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Final rating**
5	4	5	4	4	**4**

In the above example each model predicts a movie rating and mode of all the prediction is considered as the final prediction which in the above case is 4.

Weighted average is extended case of averaging in this method, each model is assigned some weight based on the importance of that model. For example, if some of your model is more suitable for that model is given more weightage than the other models.

Model	Model 1	Model 2	Model 3	Model 4	Final rating
Weight	0.30	0.30	0.20	0.20	
Rating	4	4	5	4	**4.2**

In the above example model 1 and model 2 are given more importance then model 3 and model 4. Final rating is calculated using the weighted average ratings of different models.

Advance Ensemble Technique

- Bagging
- Boosting

Bagging (Bootstrap- Aggregating)

Bagging is a technique of merging the outputs of various models to get a final result. But if we go by this method it will have higher probability that these different models generate same results as they are implemented by same input data. This problem can be mitigate by the technique known as **bootstrapping.**

Bootstrapping

It is a technique of sampling by which we can create sub sample of observation with the actual dataset, with replacement. While carrying this process it should always be remembered that the size of the sub samples dataset is the same as the actual dataset.

Sampling

Sampling is a technique to handle or adjust class distribution in a dataset. There are many types of sampling method used in machine learning. These are mainly categorized as Random sampling and Stratified Sampling

Random Sampling

In random sampling there is an equal probability of getting an item from the complete dataset. Random sampling can also be done in two ways:

Sampling without replacement: In this type of sampling when we take out an object it will get removed from the dataset. For example, in a box containing 2 red and 2 blue ball the chance of obtaining a red ball is 0.5 but as soon as we take out a red ball the chance of obtaining the next red ball is 0.33.

Sampling with replacement: In this type of sampling objects are not removed from the dataset even if we take out them from the dataset. For example, if we take out a red ball from the box, we put another red ball again in the box keeping the chance of obtaining another red ball remains the same.

Stratified Sampling

Random sampling is good method to sample homogeneous data points but in case of heterogeneous data we need a different method of sampling to increase the precision of estimator. This method is known as stratified sampling. The basic idea of stratified sampling is to divide the main heterogeneous data points into smaller samples of subgroups such that the smaller samples are homogeneous with respect to the data points. These smaller samples or subgroups of data points are known as **strata**.

Bagging uses these different sampling techniques to get unbiased data distribution. Sometimes the size of subset is less than the original dataset. A base model is trained on each of the subset. The models run parallelly and generate individual predictions. These models are independent of each other. The final prediction is done by merging prediction from all the individual models. [image source : https://towardsdatascience.com/ensemble-methods-in-machine-learning-what-are-they-and-why-use-them-68ec3f9fef5f.]

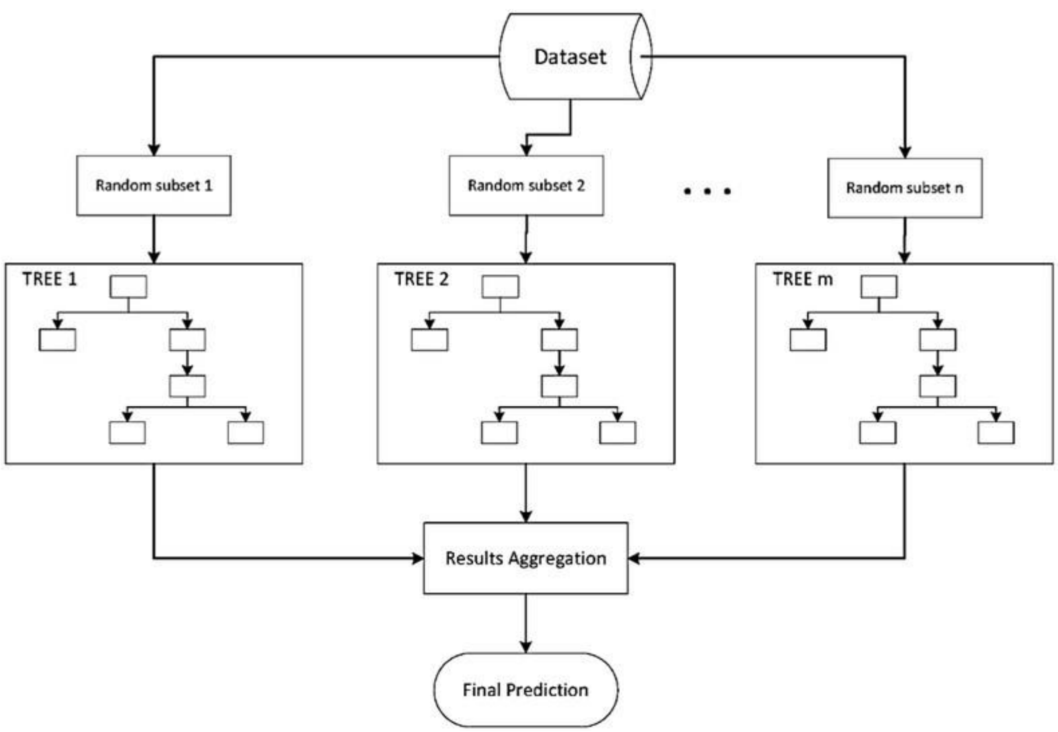

Figure 1 Bagging method

Boosting-

Suppose for example if the final prediction is wrong in two of the models and correct prediction was made in only one of the models. In this case combining the model will not give the accurate result. This situation was taken care of with the help of another method called boosting.

Boosting is a linear sequential process, where next or upcoming model tries to minimize the errors made by previous model in prediction. This method is different from bagging in the sense where each succeeding model is dependent on the previous model. Let us understand with the help of an example:

1. Various subset is created using original dataset. At first all the point is given equal weight and a first model is trained using the subset of data and now this model is used to make the prediction on the actual dataset.
2. Actual data values with predicted value are used to calculate the error in prediction. The incorrectly predicted data points are given higher weight and the second model is trained on new dataset with changed weights. The second model is also trying to reduce the error made by the previous dataset.
3. Sequentially, various models are trained and each model is minimizing the errors from the last model. The final model will be a strong learner is which is the weighted mean of all the weak learners.

Thus, boosting algorithms merge various weak learners to get one strong learner. The weak learners generally do not perform well on whole data but are efficient of subset of data. That's why when we merge them together, each model increases in efficiency and give better predictions as whole.

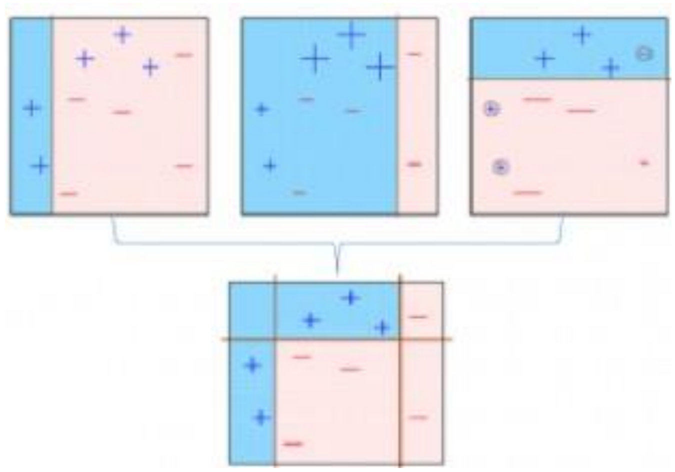

Figure 2 Boosting method

Random Forest-

Random Forest is a supervised machine learning algorithm which means it is used on a labelled dataset. It is used for both classification and regression. We can say that just like a forest which comprises many trees. Random forest algorithm comprises many decision trees.

Random forest generates decision trees on randomly selected subsample of the complete dataset and calculate prediction on each decision tree and select the best decision tree by the means of voting.

Random Forest is used in various domains such as classification of images, feature selection and recommendation engines, It lies the foundation of Boruta algorithm which is used in the selection of important features in the dataset.

Random Forest Algorithm

In Layman's term let's suppose you want to watch a movie. You can watch online, read reviews on blogs, and you can also ask some friends to help you.

Assuming you ask your friends about their favorite movies. You will get some recommendations from every friend with the help of these recommendations. You can make a list of recommended movies and ask them to vote on the list that you have made. The movie with the highest number of votes will be the final choice to watch.

In the above example you can clearly see that there are two basic steps. The first one is asking your friends about their favorite movies and choosing one recommendation out of the many movies they have watched. This part is just like a decision tree algorithm.

In the second part is the voting part that is selecting the best movie out of the list of recommended movies. This whole process of getting recommendations from friends and voting on them to select the best one is known as **Random forest algorithm.**

Random forest is basically an ensample technique which is based on philosophy of divide and conquer methodology. It generates small decision trees using random subsamples of the dataset where the collection of the generated decision tree is defined as forest. Every individual tree is created using an attribute selection indicator such as gini Index, entropy and information gain etc.

In classification problem voting is done by each tree and most voted class is considered as the final result whereas in case of regression the average method is used get the final outcome.

Basic Step Involved in the algorithm

1. Selection of random subsample of a given dataset.
2. Using attribute selection indicators create a decision tree for each subsample and record the prediction outcome from each model.
3. Apply the voting/averaging method over predicted outcomes of individual models.
4. Consider the final results as average value or most voted value.

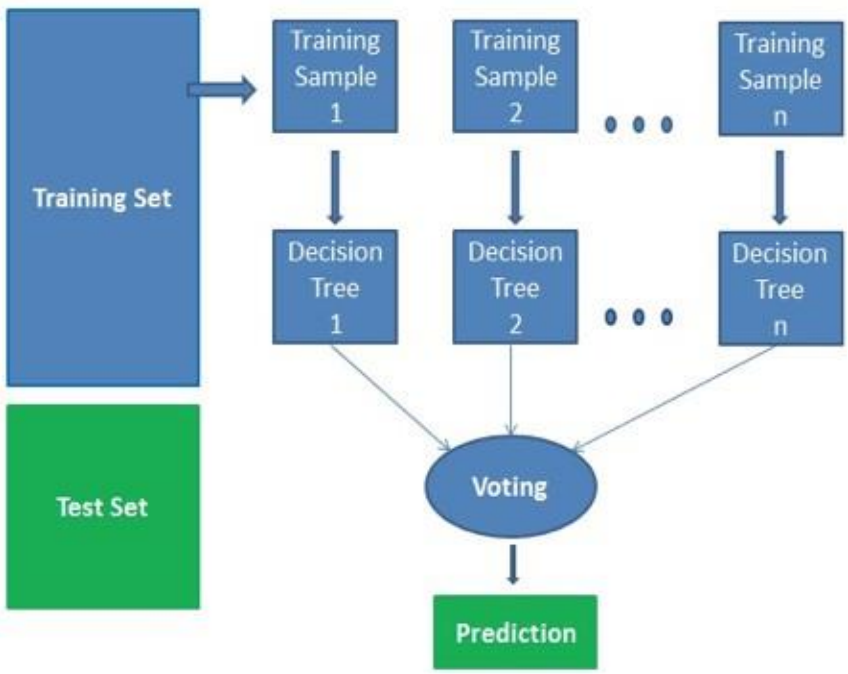

Figure 1 Random Forest

Advantages

- It can be used for both classification and regression process.
- It is one of the most accurate algorithms present out there because of the number of decision trees taking part in the process.
- Random Forest does not suffer overfitting.
- It is used to select features of relatively more importance and helps in feature selection.

Disadvantages

- Random Forest algorithm is very slow compared to others because it calculates prediction for each decision tree for every sub sample and then votes on them to select the best one which is time consuming.

- It is difficult to explain the model as compared to a decision tree where you can easily make the decision following the path of the tree.

Random Forest vs Decision Tree

- Random Forest is a collection of multiple decision trees.
- Decision trees are computationally faster than Random Forest.
- Random forest does not suffer with overfitting but decision tree suffers.
- It is easy to interpret a decision tree as compared to a random forest.

****