Ensemble Methods

Stacking in Machine Learning

- There are many ways to ensemble models in machine learning, such as Bagging, Boosting, and stacking.
- Stacking is one of the most popular ensemble machine learning techniques used to predict multiple nodes to build a new model and improve model performance.
- Stacking enables us to train multiple models to solve similar problems, and based on their combined output, it builds a new model with improved performance.
- In this topic, "Stacking in Machine Learning", we will discuss a few important concepts related to stacking, the general architecture of stacking, important key points to implement stacking, and how stacking differs from bagging and boosting in machine learning.
- Before starting this topic, first, understand the concepts of the ensemble in machine learning. So, let's start with the definition of ensemble learning in machine learning.

What is Ensemble learning in Machine Learning?

- Ensemble learning is one of the most powerful machine learning techniques that use the combined output of two or more models/weak learners and solve a particular computational intelligence problem.
- E.g., a Random Forest algorithm is an ensemble of various decision trees combined.
- Ensemble learning is primarily used to improve the model performance, such as classification, prediction, function approximation, etc.
- In simple words, we can summarise the ensemble learning as follows:
- An ensembled model is a machine learning model that combines the predictions from two or more models.”

There are 3 most common ensemble learning methods in machine learning. These are as follows:

- Bagging
- Boosting
- Stacking

However, we will mainly discuss Stacking on this topic.

1. Bagging:

Bagging is a method of ensemble modeling, which is primarily used to solve supervised machine learning problems. It is generally completed in two steps as follows:

- **Bootstrapping**: It is a random sampling method that is used to derive samples from the data using the replacement procedure. In this method, first, random data samples are fed to the primary model, and then a base learning algorithm is run on the samples to complete the learning process.
- **Aggregation:** This is a step that involves the process of combining the output of all base models and, based on their output, predicting an aggregate result with greater accuracy and reduced variance.

Example: In the Random Forest method, predictions from multiple decision trees are ensembled parallelly. Further, in regression problems, we use an average of these predictions to get the final output, whereas, in classification problems, the model is selected as the predicted class.

Bagging

- Bagging is used when our objective is to reduce the variance of a decision tree.
- Here the concept is to create a few subsets of data from the training sample, which is chosen randomly with replacement.
- Now each collection of subset data is used to prepare their decision trees thus, we end up with an ensemble of various models.

- The average of all the assumptions from numerous tress is used, which is more powerful than a single decision tree.
- Random Forest is an expansion over bagging.
- It takes one additional step to predict a random subset of data.
- It also makes the random selection of features rather than using all features to develop trees. When we have numerous random trees, it is called the Random Forest.

These are the following steps which are taken to implement a Random forest:

- Let us consider X observations Y features in the training data set. First, a model from the training data set is taken randomly with substitution.
- The tree is developed to the largest.
- The given steps are repeated, and prediction is given, which is based on the collection of predictions from n number of trees.

Advantages of using Random Forest technique:

- It manages a higher dimension data set very well.
- It manages missing quantities and keeps accuracy for missing data.

Disadvantages of using Random Forest technique:

Since the last prediction depends on the mean predictions from subset trees, it won't give precise value for the regression model.

2. Boosting

- Boosting is an ensemble method that enables each member to learn from the preceding member's mistakes and make better predictions for the future.
- Unlike the bagging method, in boosting, all base learners (weak) are arranged in a sequential format so that they can learn from the mistakes of their preceding learner.
- Hence, in this way, all weak learners get turned into strong learners and make a better predictive model with significantly improved performance.

- We have a basic understanding of ensemble techniques in machine learning and their two common methods, i.e., bagging and boosting.
- Now, let's discuss a different paradigm of ensemble learning, i.e., Stacking.

Boosting:

- Boosting is another ensemble procedure to make a collection of predictors.
- In other words, we fit consecutive trees, usually random samples, and at each step, the objective is to solve net error from the prior trees.
- If a given input is misclassified by theory, then its weight is increased so that the upcoming hypothesis is more likely to classify it correctly by consolidating the entire set at last converts weak learners into better performing models.
- Gradient Boosting is an expansion of the boosting procedure.

1. Gradient Boosting = Gradient Descent + Boosting

It utilizes a gradient descent algorithm that can optimize any differentiable loss function. An ensemble of trees is constructed individually, and individual trees are summed successively. The next tree tries to restore the loss (It is the difference between actual and predicted values).

Advantages of using Gradient Boosting methods:

- It supports different loss functions.
- It works well with interactions.

Disadvantages of using a Gradient Boosting methods:

It requires cautious tuning of different hyper-parameters.

Bagging Vs Boosting

- We all use the Decision Tree Technique on day to day life to make the decision.
- Organizations use these supervised machine learning techniques like Decision trees to make a better decision and to generate more surplus and profit.
- Ensemble methods combine different decision trees to deliver better predictive results, afterward utilizing a single decision tree.
- The primary principle behind the ensemble model is that a group of weak learners come together to form an active learner.
- There are two techniques given below that are used to perform ensemble decision tree.

Bagging	Boosting
Various training data subsets are randomly drawn with replacement from the whole training dataset.	Each new subset contains the components that were misclassified by previous models.
Bagging attempts to tackle the over-fitting issue.	Boosting tries to reduce bias.
If the classifier is unstable (high variance), then we need to apply bagging.	If the classifier is steady and straightforward (high bias), then we need to apply boosting.
Every model receives an equal weight.	Models are weighted by their performance.
Objective to decrease variance, not bias.	Objective to decrease bias, not variance.
It is the easiest way of connecting predictions that belong to the same type.	It is a way of connecting predictions that belong to the different types.
Every model is constructed independently.	New models are affected by the performance of the previously developed model.