What is Ensemble learning in Machine Learning?

**An ensembled model is a machine learning model that combines the predictions from two or more models.**

Ensemble learning is primarily used to improve the model performance, such as classification, prediction and function approximation.

Basic idea is to learn a set of classifiers and make them to vote.

There are 3 most common ensemble learning methods in machine learning.

* Bagging
* Boosting
* Stacking

### **1. Bagging(Homogeneous)**

Bagging is a method of ensemble model, which is primarily used to solve supervised machine learning problems.

Here, Multiple models are trained independently on different subsets of the training data, and their outputs are combined by taking an average or majority vote.

Bagging increases bias but reduces variance.

It is generally completed in two steps as follows:

* **Bootstrapping:** It is a random sampling method that is used to derive sample subsets with equal tuples from the training data.
* **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.



### **2. Boosting(Homogeneous)**

Boosting is an ensemble method that enables each member to learn from the previous member's mistakes and make better predictions for the future.

In this approach, a sequence of models is trained, where each subsequent model focuses on the data that were previously misclassified by the previous models. The outputs of these models are combined in a weighted manner to produce the final prediction.

One of the most popular algorithms that uses boosting is the AdaBoost algorithm, which stands for Adaptive Boosting.

In AdaBoost, a series of weak classifiers are trained sequentially, where each subsequent classifier is trained to correct the mistakes of the previous classifiers. The output of each classifier is combined using a weighted sum to produce the final prediction.

A picture containing chart

Description automatically generated

### **3. Stacking(Heterogeneous)**

**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.**

Diagram

Description automatically generatedIn this approach, multiple models are trained and their outputs are used as input to a meta-model, which then produces the final prediction.

* **Level-0 Predictions:** Each base model is trained on some training data and provides different predictions.
* **Meta Model:**  It helps to combine the predictions of the base models and is trained on different predictions made by individual base models.

**VOTING**

Voting is a popular ensemble learning technique that involves combining the predictions of multiple models by taking the majority vote (for classification) or the average (for regression).

There are different types of voting methods, including:

**Hard Voting**: The predicted output class is a class with the highest majority of votes. Hard voting is typically used for classification problems.

Example : Suppose three classifiers predicted the output class(A, A, B), so here the majority predicted A as output. Hence A will be the final prediction.

**Soft Voting**: The output class is the prediction based on the average of probability given to that class. Soft voting is typically used for regression problems.

Diagram

Description automatically generatedExample: Suppose given some input to three models, the prediction probability for class A = (0.30, 0.47, 0.53) and B = (0.20, 0.32, 0.40). So the average for class A is 0.4333 and B is 0.3067, the winner is clearly class A because it had the highest probability averaged by each classifier.