

***Faraz Ahmed***

**Artificial Intelligence using Machine Learning and Deep Learning**

***Assignment # 5***

***Faraz Ahmed***

***(Group B)***

**Bagging & Boosting:**

Bagging is a method of merging the same type of predictions. Boosting is a method of merging different types of predictions. Bagging decreases variance, not bias, and solves over-fitting issues in a model. Boosting decreases bias, not variance. Bagging and boosting are two ensemble-learning techniques in machine learning used to improve the performance and robustness of models, especially for decision trees and other weak learners.

Bagging (Bootstrap Aggregating)**:**

Bagging is an ensemble method where multiple instances of a base model are trained on different subsets of the training data. The key idea behind bagging is to reduce the variance of the model. Here is how it works:

1. **Bootstrap Sampling:** Multiple random subsets (with replacement) of the training data are created. Each subset is called a "bootstrap sample."
2. **Model Training:** A base model (e.g., a decision tree) is trained on each bootstrap sample. Since each sample is slightly different, it results in slightly different models.
3. **Prediction:** When making predictions, each model predicts an outcome, and majority voting (for classification) or averaging (for regression) the results from all models often determines the final prediction.

Bagging is typically used to reduce overfitting, improve model stability, and increase accuracy. A well-known bagging algorithm is Random Forest, which applies bagging to decision trees.

**Reducing Variance with Bagging**

We use bagging for combining weak learners of high variance. Bagging aims to produce a model with lower variance than the individual weak models. These weak learners are homogenous, meaning they are of the same type.

Bagging is also known as Bootstrap aggregating. It consists of two steps: bootstrapping and aggregation.

**Bootstrapping**

Involves resampling subsets of data with replacement from an initial dataset. In other words, subsets of data are taken from the initial dataset. These subsets of data are called bootstrapped datasets or, simply, bootstraps. Resampled ‘with replacement’ means an individual data point can be sampled multiple times. Each bootstrap dataset is used to train a weak learner.

**Aggregating**

The individual weak learners are trained independently from each other. Each learner makes independent predictions. The results of those predictions are aggregated at the end to get the overall prediction. The predictions are aggregated using either max voting or averaging.

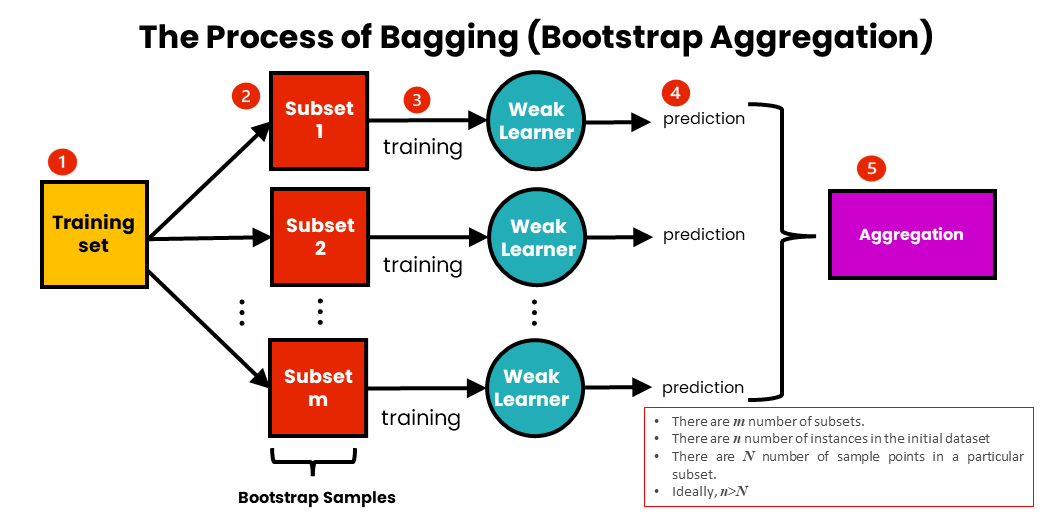
**Max Voting**

It is a commonly used for classification problems that consists of taking the mode of the predictions (the most occurring prediction). It is called voting because like in election voting, the premise is that ‘the majority rules’. Each model makes a prediction. A prediction from each model counts as a single ‘vote’. The most occurring ‘vote’ is chosen as the representative for the combined model.

**Averaging**

It is generally used for regression problems. It involves taking the average of the predictions. The resulting average is used as the overall prediction for the combined model.

**Steps of Bagging:**



**The steps of bagging are as follows:**

1. We have an initial training dataset containing n-number of instances.
2. We create an m-number of subsets of data from the training set.  We take a subset of N sample points from the initial dataset for each subset. Each subset is taken with replacement. This means that a specific data point can be sampled more than once.
3. For each subset of data, we train the corresponding weak learners independently. These models are homogeneous, meaning that they are of the same type.
4. Each model makes a prediction.
5. The predictions are aggregated into a single prediction. For this, either max voting or averaging is used.

**Boosting:**

Boosting is another ensemble method that combines multiple base models to create a strong model. Unlike bagging, boosting focuses on reducing both bias and variance, making it particularly suitable for improving accuracy. Here is how boosting works:

1. **Weighted Data:** Each data point in the training set is assigned a weight. Initially, all weights are equal.
2. **Model Training:** A base model (often a weak learner) is trained on the weighted training data. The model is evaluated, and its errors are measured.
3. **Adjusting Weights:** The weights of the training data points are adjusted based on the errors made by the model. Misclassified points are given higher weights, while correctly classified points are given lower weights.
4. **Repeat:** Steps 2 and 3 are repeated for a fixed number of iterations or until a certain level of accuracy is reached.
5. **Aggregating Models:** The base models are combined, with each model having a weight based on its performance. In boosting, models that perform well are given more influence in the final prediction.

Popular boosting algorithms include AdaBoost (Adaptive Boosting), Gradient Boosting, and XGBoost.

In summary, both bagging and boosting are ensemble techniques used to enhance the performance and robustness of machine learning models. Bagging aims to reduce variance, while boosting aims to reduce both bias and variance. The choice between them depends on the specific problem and the characteristics of the dataset.

**Reducing Bias by Boosting**

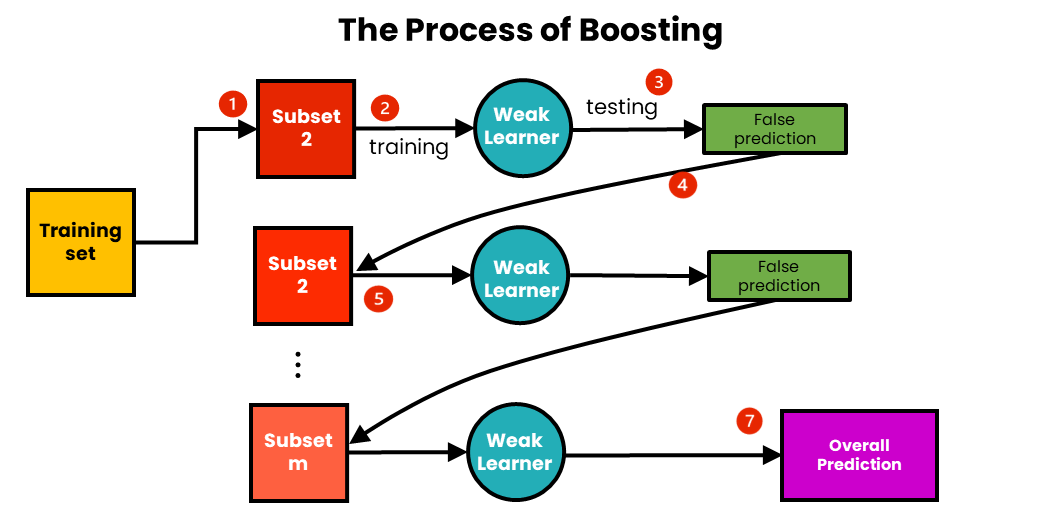
We use boosting for combining weak learners with high bias. Boosting aims to produce a model with a lower bias than that of the individual models. Like in bagging, the weak learners are homogeneous.

Boosting involves sequentially training weak learners. Here, each subsequent learner improves the errors of previous learners in the sequence. A sample of data is first taken from the initial dataset. This sample is used to train the first model, and the model makes its prediction. The samples can be either correctly or incorrectly predicted. The samples that are wrongly predicted are reused for training the next model. In this way, subsequent models can improve on the errors of previous models.

Unlike bagging, which aggregates prediction results, boosting aggregates the results at each step. They are aggregated using weighted averaging.

**Weighted averaging** involves giving all models different weights depending on their predictive power. In other words, it gives more weight to the model with the highest predictive power. This is because the learner with the highest predictive power is considered the most important.

**Steps of Boosting:**



**Boosting works with the following steps:**

1. We sample m-number of subsets from an initial training dataset.
2. Using the first subset, we train the first weak learner.
3. We test the trained weak learner using the training data. Because of the testing, some data points will be incorrectly predicted.
4. Each data point with the wrong prediction is sent into the second subset of data, and this subset is updated.
5. Using this updated subset, we train and test the second weak learner.
6. We continue with the following subset until the total number of subsets is reached.
7. We now have the total prediction. The overall prediction has already been aggregated at each step, so there is no need to calculate it.

**Improving Model Accuracy with Stacking**

We use stacking to improve the prediction accuracy of strong learners. Stacking aims to create a single robust model from multiple heterogeneous strong learners.

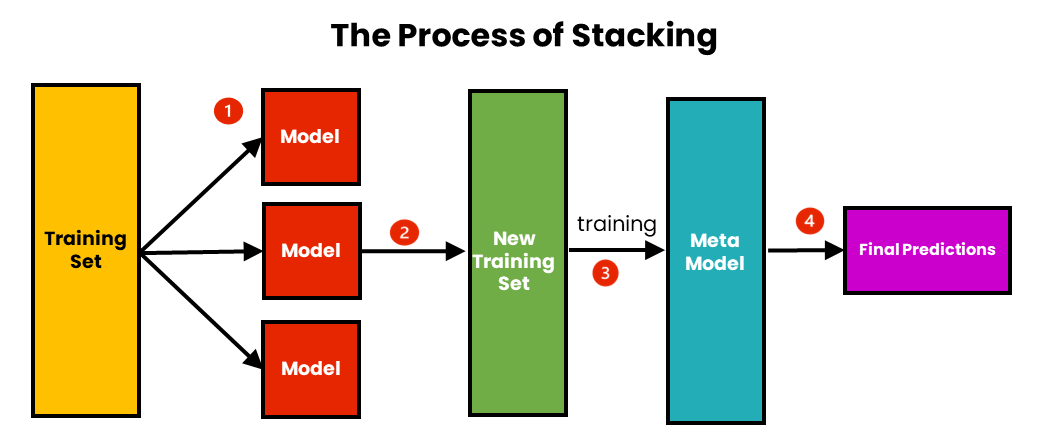
Stacking differs from bagging and boosting in that:

* It combines strong learners
* It combines heterogeneous models
* It consists of creating a Metamodeling. A metamodeling is a model created using a new dataset.

Individual heterogeneous models are trained using an initial dataset. These models make predictions and form a single new dataset using those predictions. This new data set is used to train the metamodeling, which makes the final prediction. The prediction is combined using weighted averaging.

Because stacking combines strong learners, it can combine bagged or boosted models.

**Steps of Stacking:**



**The steps of Stacking are as follows:**

1. We use initial training data to train m-number of algorithms.
2. Using the output of each algorithm, we create a new training set.
3. Using the new training set, we create a meta-model algorithm.
4. Using the results of the meta-model, we make the final prediction. The results are combined using weighted averaging.

**When to use Bagging vs Boosting vs Stacking?**



If you want to reduce the overfitting or variance of your model, you use bagging and if you are looking to reduce underfitting or bias, you use boosting. However, if you want to increase predictive accuracy, use stacking.

Bagging and boosting both works with homogeneous weak learners. Stacking works using heterogeneous solid learners.

All three of these methods can work with either classification or regression problems.

One disadvantage of boosting is that it is prone to variance or overfitting. It is thus not advisable to use boosting for reducing variance. Boosting will do a worse job in reducing variance as compared to bagging.

On the other hand, the converse is true. It is not advisable to use bagging to reduce bias or underfitting. This is because bagging is more prone to bias and does not help reduce bias.

Stacked models have the advantage of better prediction accuracy than bagging or boosting. However, because they combine bagged or boosted models, they have the disadvantage of needing much more time and computational power.   If you are looking for faster results, it is advisable not to use stacking. However, stacking is the way to go if you are looking for high accuracy.

**Conclusion**

Bagging, boosting and stacking are important for ensuring the accuracy of models. They can help prevent undesirable consequences caused by inaccurate models. Below are some of the key takeaways from the article:

* Ensemble learning combines multiple machine learning models into a single model. The aim is to increase the performance of the model.
* Bagging aims to decrease variance, boosting aims to decrease bias, and stacking aims to improve prediction accuracy.
* Bagging and boosting combine homogenous weak learners. Stacking combines heterogeneous solid learners.
* Bagging trains models in parallel and boosting trains the models sequentially. Stacking creates a meta-model.