4.2 ENSEMBLE LEARNING

"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

The idea of ensemble learning is to employ multiple learners and combine their predictions. If we have a committee of M models with uncorrelated errors, simply by averaging them the average error of a model can be reduced by a factor of M.

• Unfortunately, the key assumption that the errors due to the individual models are uncorrelated is unrealistic in practice, the errors are typically highly correlated, so the reduction in overall error is generally small.

- Ensemble modeling is the process of running two or more related but different analytical models and then synthesizing the results into a single score or spread in order to improve the accuracy of predictive analytics and data mining applications.

•Ensembles of classifiers is a set of classifiers whose individual decisions combined in some way to classify new examples.

• Ensemble methods combine several decision trees classifiers to produce better predictive performance than a single decision tree classifier. The main principle behind the ensemble model is that a **group of weak learners come together to form a strong learner**, thus increasing the accuracy of the model.

- Why do ensemble methods work? Based on one of two basic observations

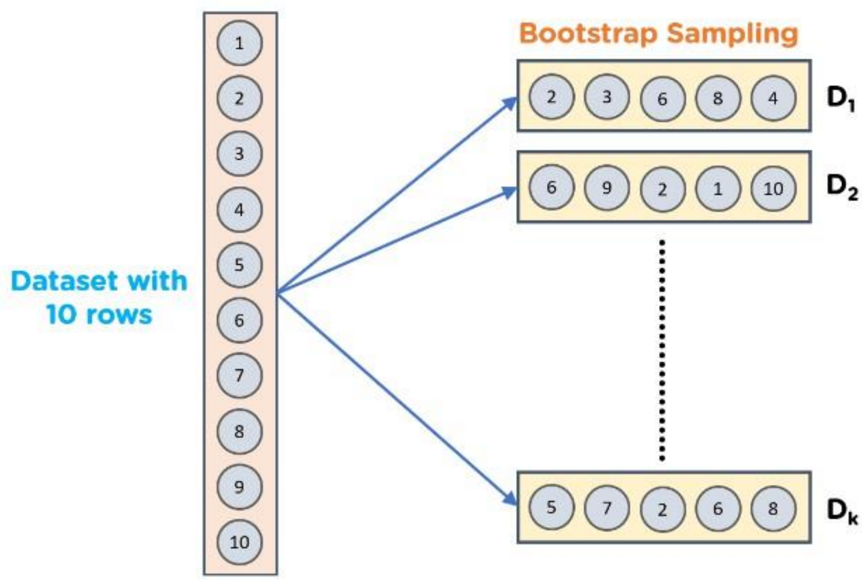
1. Variance reduction: If the training sets are completely independent, it will always helps to average an ensemble because this will reduce variance without affecting bias (e.g. bagging) and reduce sensitivity to individual data points.

2. Bias reduction: For simple models, average of models has much greater capacity than single model Averaging models can reduce bias substantially by increasing capacity and control variance by Citting one component at a time.

4.2.1 Bagging

- Bagging is also called Bootstrap Aggregating Bagging and boosting are meta algorithms that pool decisions from multiple classifiers. It creates ensembles by repeatedly randomly resampling the training data.
- Bagging is an ensemble learning technique that helps to improve the performance and accuracy of machine learning algorithms
- The meta algorithm, which is a special case of the model averaging, was originally designed for classification and is usually applied to decision tree models, but it can be used with any type of model for classification or regression.

Bootstrapping is the method of randomly creating samples of data out of a population with replacement to estimate a population parameter.



Bagging Steps:

1. Suppose there are N observations and M features in training data set. A sample from training data set is taken randomly with replacement,

2. A subset of M features is selected randomly and whichever feature gives the best split is used to split the node iteratively.

3. The tree is grown to the largest

4. Above steps are repeated in times and prediction is given based on the aggregation of predictions from n number of trees.

Advantages of Bagging

1. Reduces over-fitting of the model.

2. Handles higher dimensionality data very well.

3. Maintains accuracy for missing data.

Disadvantages of Bagging:

1. Since final prediction is based on the mean predictions from subset trees, it won't give precise values for the classification and regression model.

4.2.2 Boosting

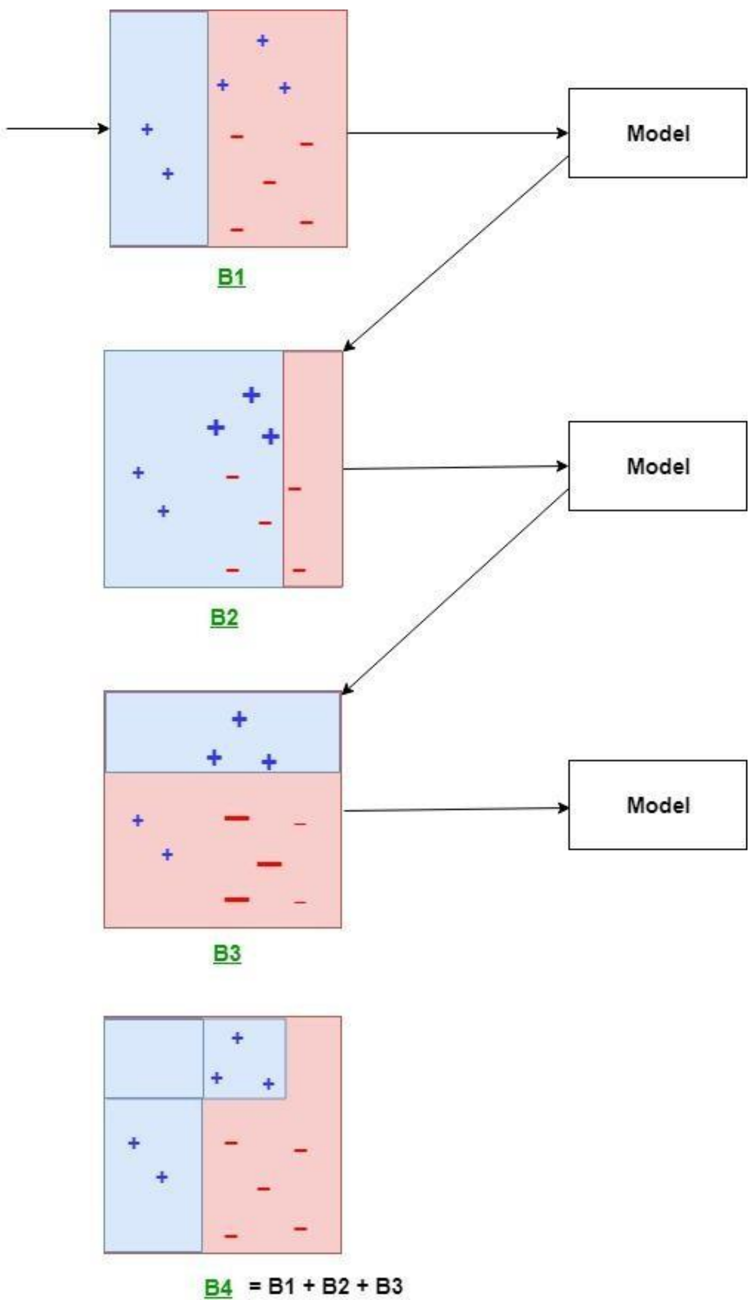
Boosting is an ensemble modeling technique that attempts to build a strong classifier from the number of weak classifiers. It is done by building a model by using weak models in series. Firstly, a model is built from the training data. Then the second model is built which tries to correct the errors present in the first model. This procedure is continued and models are added until either the complete training data set is predicted correctly or the maximum number of models are added.

Boosting is a very different method to generate multiple predictions (function estimates) and combine them linearly. Boosting refers to a general and provably effective method of producing a very accurate classifier by combining rough and moderately inaccurate rules of thumb.

Boosting is a bias reduction technique. It typically improves the performance of a single tree model

To begin, we define an algorithm for finding the rules of thumb, which we call a weak learner. The boosting algorithm repeatedly calls this weak learner, each time feeding it a different distribution over the training data. Each call generates a weak classifier and we must combine all of these into a single classifier that. hopefully, is much more accurate than any one of the rules.

Train a set of weak hypotheses: $h_{1,....,}h_T$. The combined hypothesis H is a weighted majority vote of the T weak hypotheses. During the training, focus on the examples that are misclassified.



Boosting Steps:

1. Draw a random subset of training samples d1 without replacement from the training set D to train a weak learner C1

2. Draw second random training subset d2 without replacement from the training set and add 50 percent of the samples that were previously falsely classified/misclassified to train a weak learner C2

3. Find the training samples d3 in the training set D on which C1 and C2 disagree to train a third weak learner C3

4. Combine all the weak learners via majority voting.

AdaBoost:

AdaBoost was the first really successful boosting algorithm developed for the purpose of binary classification. *AdaBoost* is short for *Adaptive Boosting* and is a very popular boosting technique that combines multiple "weak classifiers" into a single "strong classifier". It was formulated by Yoav Freund and Robert Schapire. They also won the 2003 Gödel Prize for their work.

Advantages of AdaBoost:

1. Very simple to implement

2. Fairly good generalization

3. The prior error need not be known ahead of time.

Disadvantages of AdaBoost

1. Suboptimal solution

2. Can over fit in presence of noise.

4.2.3 Stacking

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.

- Stacking, sometimes called stacked generalization, is an ensemble machine learning method that combines multiple heterogeneous base or component models via a meta-model
- The base model is trained on the complete training data, and then the meta-model is trained on the predictions of the base models. The advantage of stacking is the ability to explore the solution space with different models in the same problem.
- The stacking based model can be visualized in levels and has at least two levels of the models. The first level typically trains the two or more base learners (can be heterogeneous) and the second level might be a single meta learner that utilizes the base models predictions as input and gives the final result as output. A stacked model can have more than two such levels but increasing the levels doesn't always guarantee better performance.
- In the classification tasks, often logistic regression is used as a meta learner, while linear regression is more suitable as a meta learner for regression-based tasks.

Stacking is concerned with combining multiple classifiers generated by different learning algorithms $L_1,\ldots..L_N$ on a single dataset S, which is composed by a feature vector $S_i = (x_i, t_i)$.

- The stacking process can be broken into two phases.

1. Generate a set of base-level classifiers C1...CN where Ci=Li(S)

2. Train a meta level classifier to combine the outputs of the base level classifiers,

Fig. shows stacking frame.

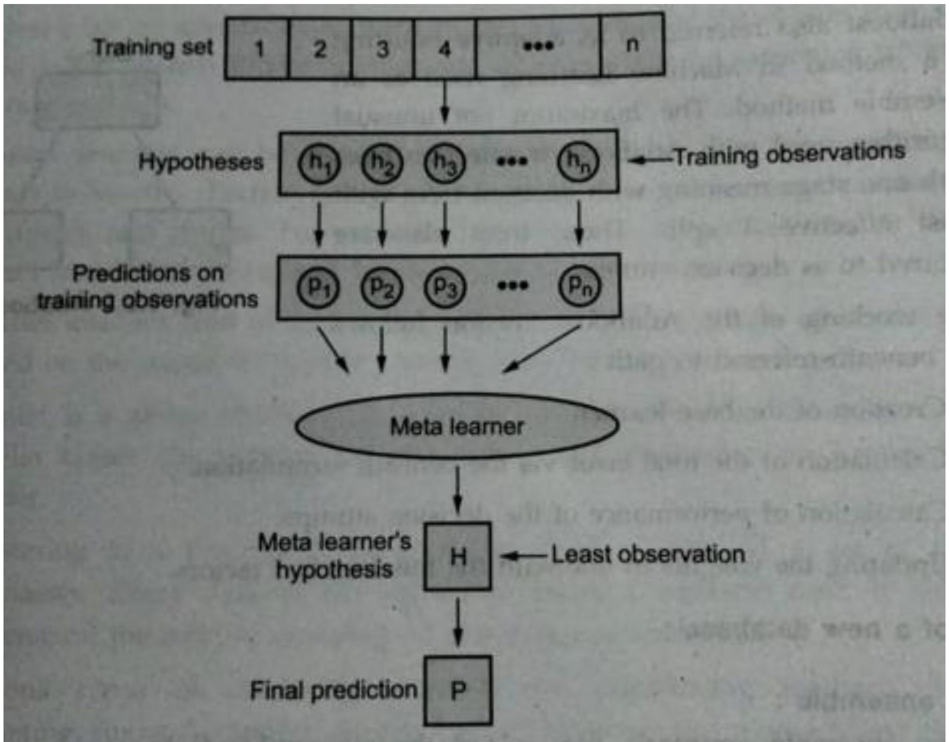


Fig. 9.2.2 Stacking frame

Difference between Bagging and Boosting

No.	Bagging	Boosting
1.	Bagging is a technique that builds multiple homogeneous models from different subsamples of the same training dataset to obtain more accurate predictions than its individual models	Boosting refers to a group of algorithms that utilize weighted averages to make weak learning algorithms stronger learning algorithms.
2.	Learns them independently from each other in parallel	Learns them sequentially in a very adaptative way
3.	It helps in reducing variance.	It helps in reducing bias and variance.
4.	Every model receives an equal weight.	Models are weighted by their performance.