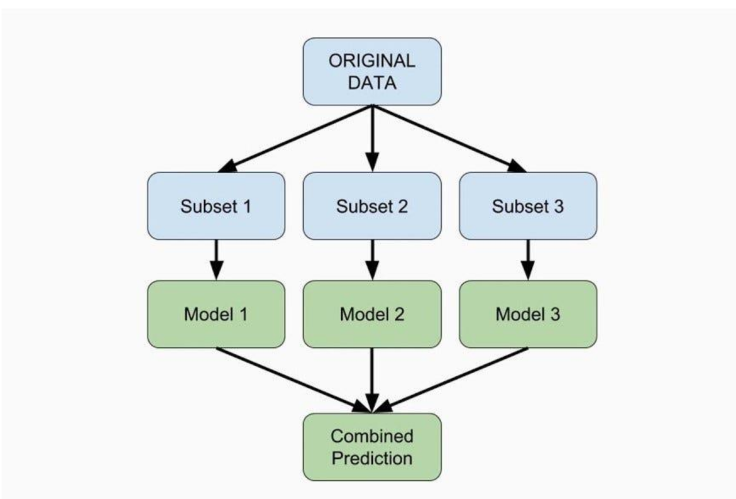
Ensemble Methods

Ensemble methods train multiple learners to solve the same problem. In contrast to ordinary learning approaches which try to construct one learner from training data, ensemble methods try to construct a set of learners and combine them. Ensemble learning is also called committee-based learning, or learning multiple classifier systems.

Figure 1.9 shows a common ensemble architecture. An ensemble contains a number of learners called base learners. Base learners are usually generated from training data by a base learning algorithm which can be decision tree, neural network or other kinds of learning algorithms. Most ensemble methods use a single base learning algorithm to produce homogeneous base learners, i.e., learners of the same type, leading to homogeneous ensembles, but there are also some methods which use multiple learning algorithms to produce heterogeneous learners, i.e., learners of different types, leading to heterogeneous ensembles. In the latter case there is no single base learning algorithm and thus, some people prefer calling the learners individual learners or component learners to base learners. The generalization ability of an ensemble is often much stronger than that of base learners. Actually, ensemble methods are appealing mainly because they are able to boost weak learners which are even just slightly better than random guess to strong learners which can make very accurate predictions. So, base learners are also referred to as weak learners.



Why Use Ensemble Techniques?

Increased Accuracy: By combining multiple models, ensemble methods can achieve higher predictive accuracy.

Reduced Overfitting: Ensemble methods tend to generalize better on unseen data.

Improved Robustness: They provide more stable and reliable predictions by averaging out biases and variances.

Types of Ensemble Techniques

Bagging (Bootstrap Aggregating)

Boosting

Stacking

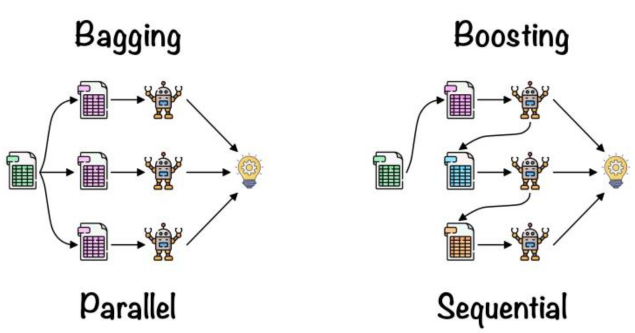
Bagging:

Bagging stands for Bootstrap Aggregating, which is a technique used in ensemble learning to reduce the variance of machine learning models. The idea behind bagging is to train multiple models on different subsets of the training data, and then combine their predictions to make the final prediction.

▶ In bagging, we randomly select a subset of the training data with replacement (bootstrap) to create a new training set for each model. This means that some data points may be present multiple times in a single subset, while others may be left out entirely.

▶ By training models on these different subsets, we can reduce the variance of the overall prediction, since each model is making its predictions based on a slightly different set of data. Bagging is particularly useful in applications where the individual models are prone to overfitting the training data, as bagging can help to reduce this overfitting by creating more diverse models.

▶ It is commonly used in applications such as classification, regression, and time series forecasting. We will discuss the detailed algorithm of Vanilla Bagging



Boosting:

Boosting is a method used in machine learning to reduce errors in predictive data analysis. Data scientists train machine learning software, called machine learning models, on labeled data to make guesses about unlabeled data. A single machine learning model might make prediction errors depending on the accuracy of the training dataset. For example, if a cat-identifying model has been trained only on images of white cats, it may occasionally misidentify a black cat. Boosting tries to overcome this issue by training multiple models sequentially to improve the accuracy of the overall system.

Boosting improves machine models' predictive accuracy and performance by converting multiple weak learners into a single strong learning model. Machine learning models can be weak learners or strong learners:

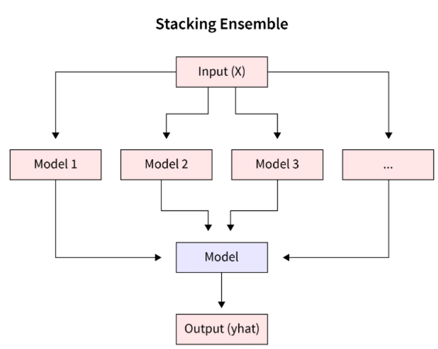
Weak learners

Weak learners have low prediction accuracy, similar to random guessing. They are prone to overfitting—that is, they can't classify data that varies too much from their original dataset. For example, if you train the model to identify cats as animals with pointed ears, it might fail to recognize a cat whose ears are curled.

Strong learners

Strong learners have higher prediction accuracy. Boosting converts a system of weak learners into a single strong learning system. For example, to identify the cat image, it combines a weak learner that guesses for pointy ears and another learner that guesses for cat-shaped eyes. After analyzing the animal image for pointy ears, the system analyzes it once again for cat-shaped eyes. This improves the system's overall accuracy.

Stacking:



Stacking in machine learning is also known as Stacking Generalisation, which is a technique where all models aggregated are utilized according to their weights for producing an output which is a new model. As a result, this model has better accuracy and is stacked with other models to be used.

We can envision it as a two-layer model where the first layer incorporates all the models, and the second one is the prediction layer which renders output.

The principle is that you can always tackle a learning issue with various models that can learn a subset of the problem but not the whole problem space. This is where Stacking is used.