

Boosting Algorithm: Gradient Boosting

Gradient boosting is a technique for regression and classification problems. The prediction model produced in the form of an ensemble of weak prediction models.

The accuracy of a predictive model can be boosted in two ways:

- a. Either by using feature engineering or
- b. By applying boosting algorithms.

There are many boosting algorithms like

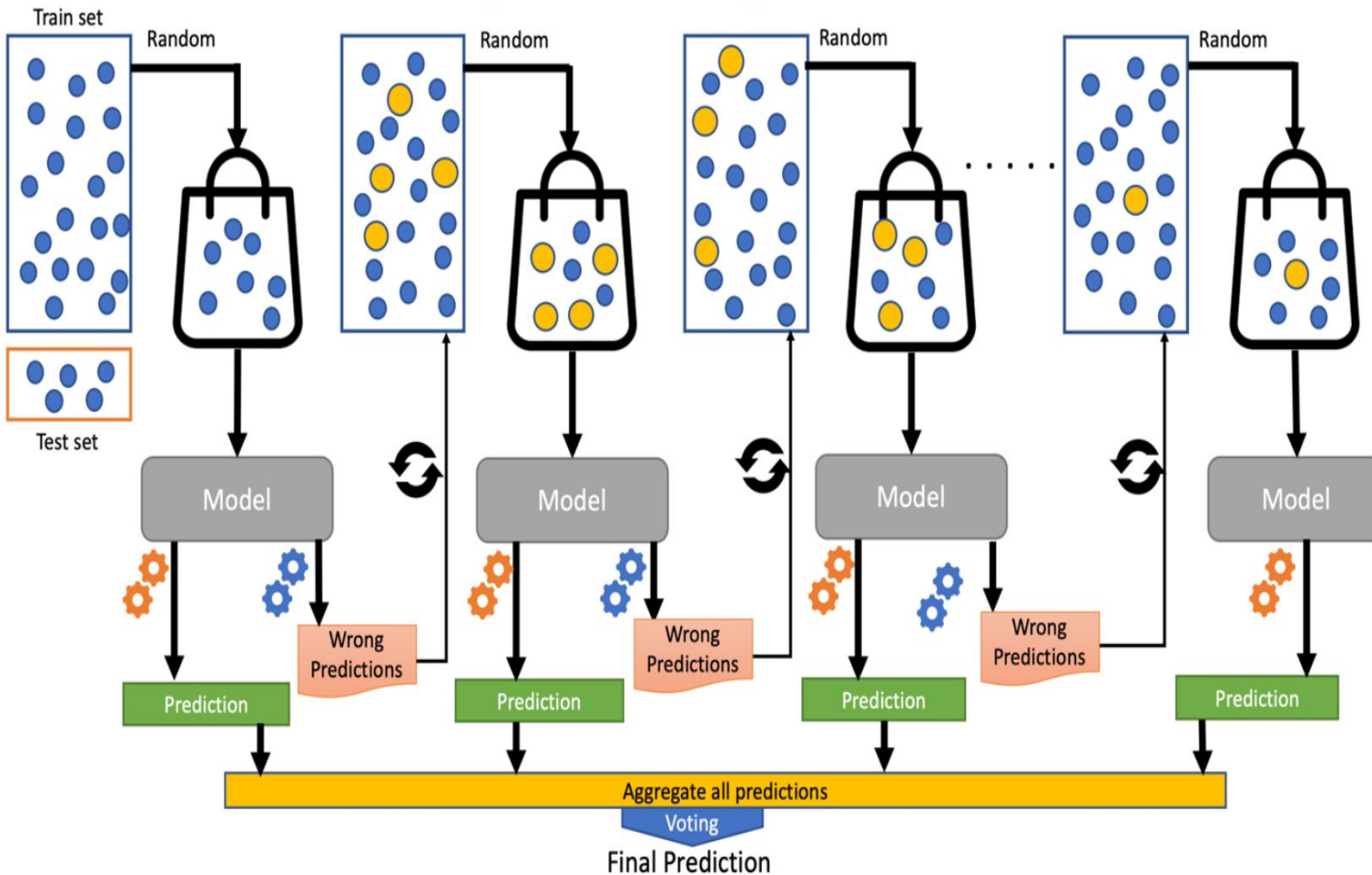
- Gradient Boosting
- XGBoost
- AdaBoost

Internal working of boosting algorithm



INTERNSHIPSTUDIO

● - Instance ● - Misclassified instances ⚙ - Test using Test data ⚙ - Test using Train data ↻ - Updates dataset with new weights



Gradient Boosting

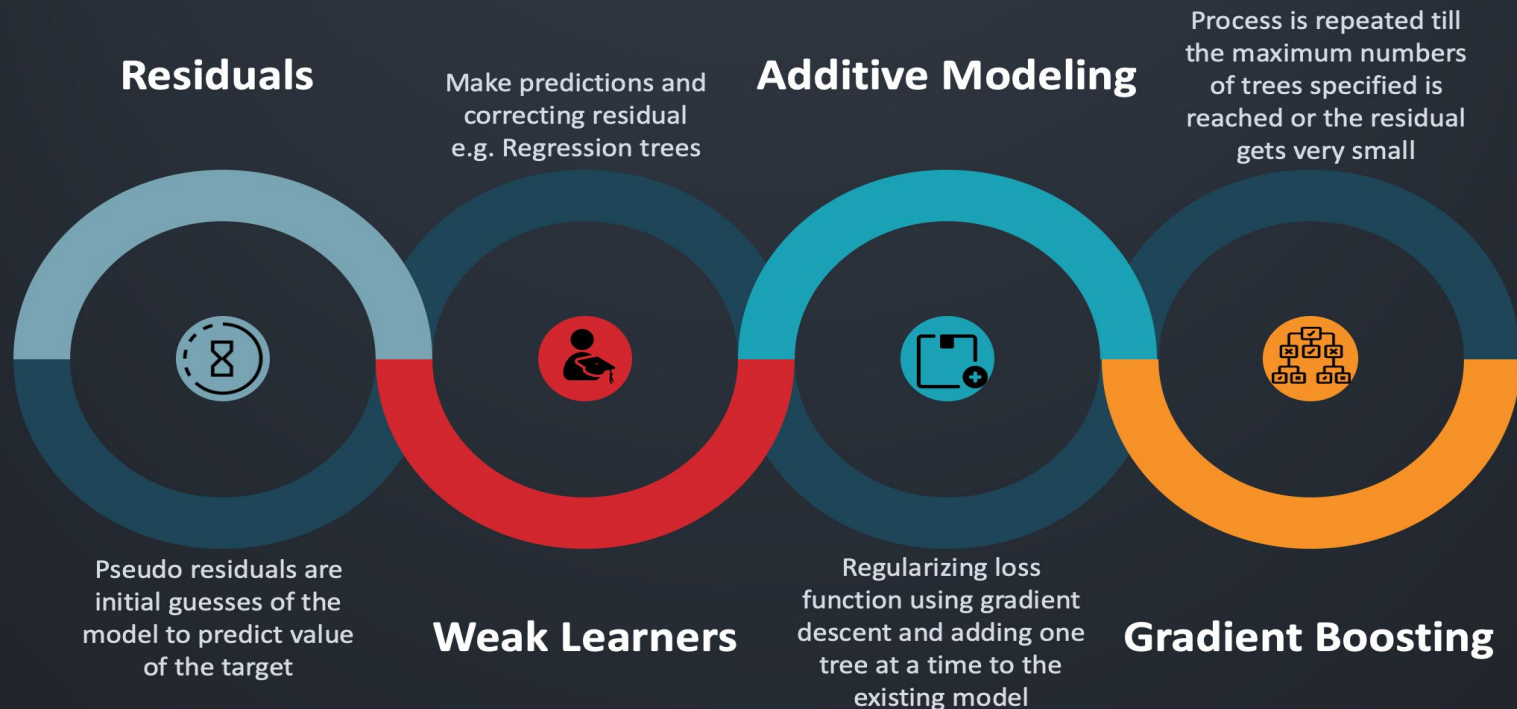
Gradient boosting Algorithm involves three elements:

- A loss function to be optimized.
- Weak learner to make predictions.
- An additive model to add weak learners to minimize the loss function.



INTERNSHIPSTUDIO

Gradient Boosting Process Diagram



Extreme Gradient Boosting (XGBoost)



INTERNSHIPSTUDIO

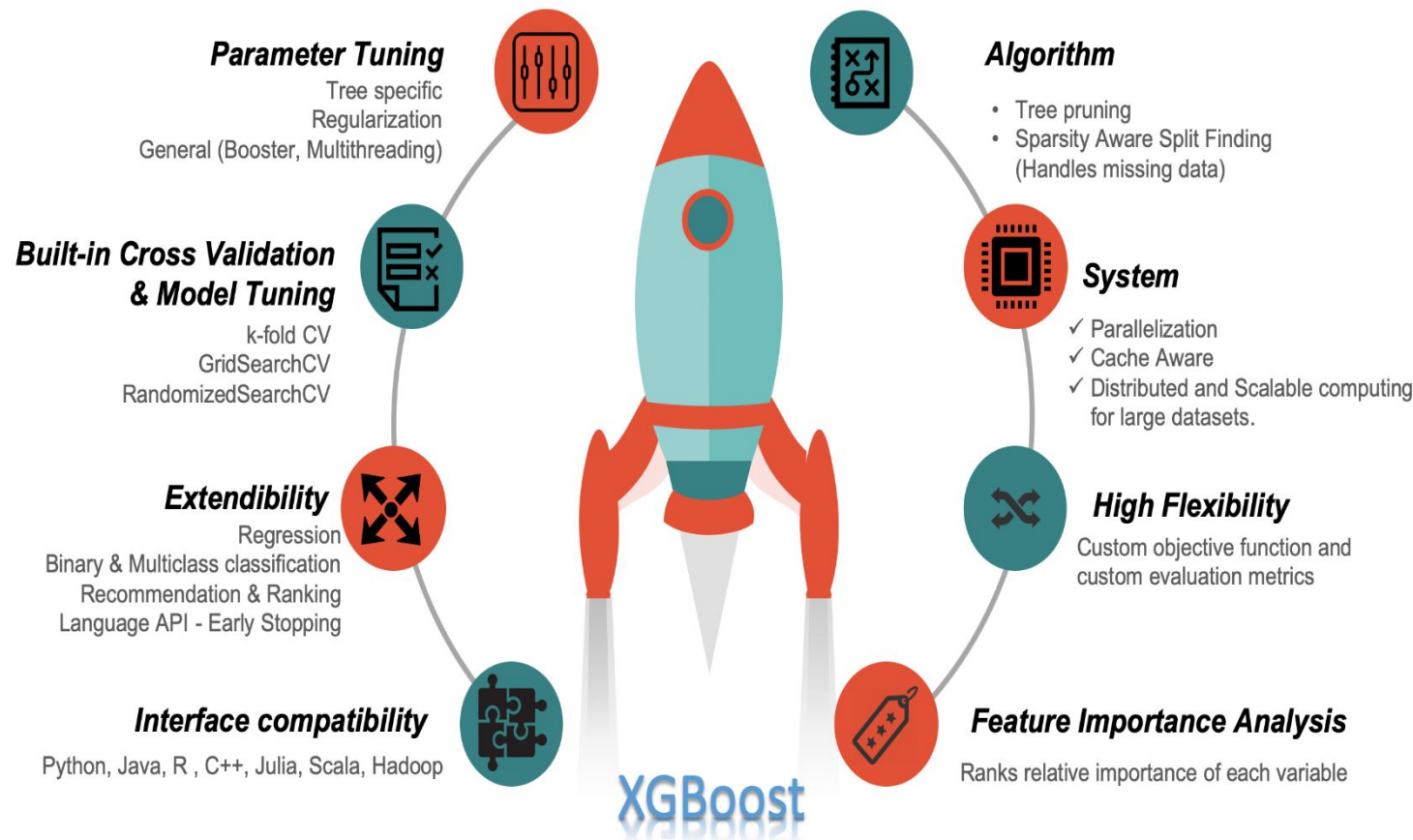
- XGBoost Algorithm is an implementation of gradient boosted decision trees, designed for speed and performance.
- Basically, it is a type of software library. It can be used for supervised learning tasks such as Regression, Classification, and Ranking.
- It is built on the principles of gradient boosting framework and designed to “push the extreme of the computation limits of machines to provide a scalable, portable and accurate library.”



System Feature- XGBoost

For use of a range of computing environments this library provides:

- Parallelization of tree construction using all of your CPU cores during training.
- Distributed Computing for training very large models using a cluster of machines & Out-of-Core Computing for very large datasets that don't fit into memory.



INTERNSHIPSTUDIO

Comparison- XGBoosting

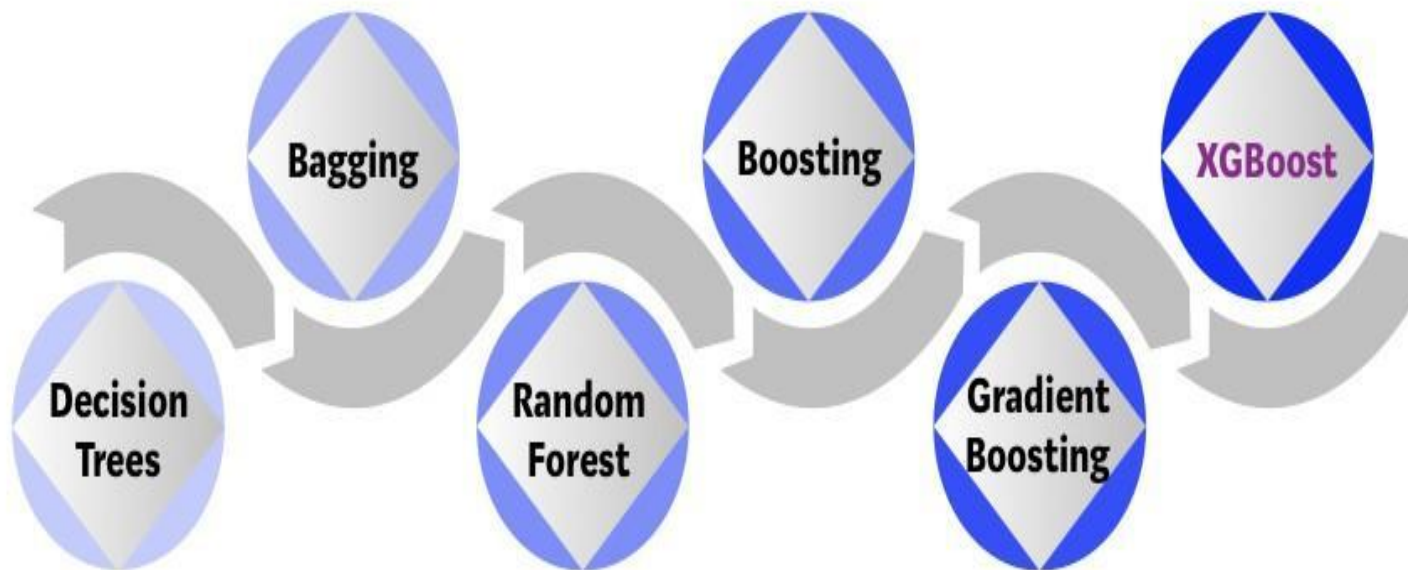


INTERNSHIPSTUDIO

Bootstrap aggregating or Bagging is a ensemble meta-algorithm combining predictions from multiple-decision trees through a majority voting mechanism

Models are built sequentially by minimizing the errors from previous models while increasing (or boosting) influence of high-performing models

Optimized Gradient Boosting algorithm through parallel processing, tree-pruning, handling missing values and regularization to avoid overfitting/bias



**Decision
Trees**

Bagging

**Random
Forest**

Boosting

**Gradient
Boosting**

XGBoost

A graphical representation of possible solutions to a decision based on certain conditions

Bagging-based algorithm where only a subset of features are selected at random to build a forest or collection of decision trees

Gradient Boosting employs gradient descent algorithm to minimize errors in sequential models

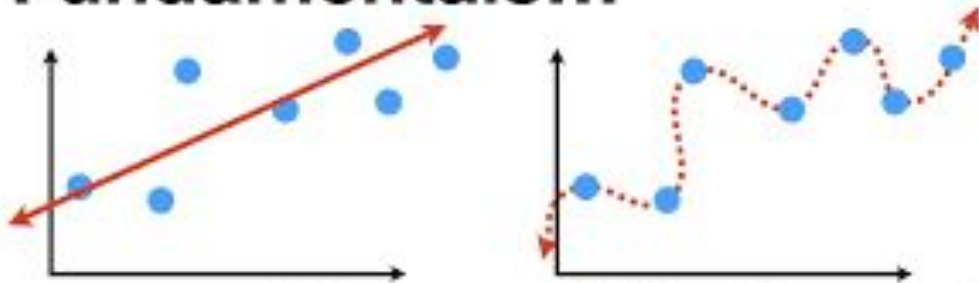
What is Bias?



INTERNSHIPSTUDIO

- Bias is how far are the predicted values from the actual values. If the average predicted values are far off from the actual values then the bias is high.
- High bias causes algorithm to miss relevant relationship between input and output variable. When a model has a high bias then it implies that the model is too simple and does not capture the complexity of data thus underfitting the data.

Machine Learning Fundamentals...

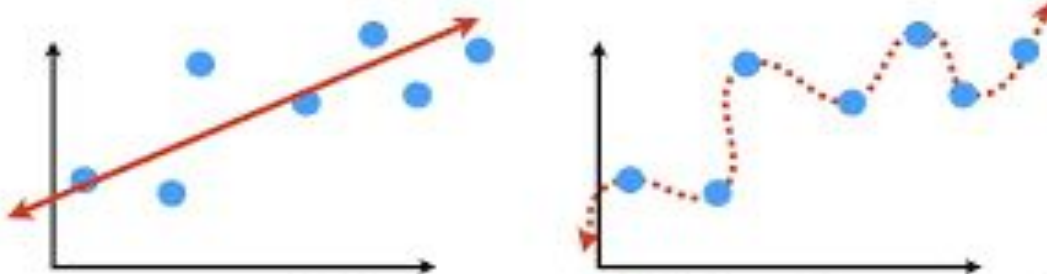


...Bias and Variance!!!

What is Variance ?

- Variance occurs when the model performs good on the trained dataset but does not do well on a dataset that it is not trained on, like a test dataset or validation dataset. Variance tells us how scattered are the predicted value from the actual value.
- High variance causes overfitting that implies that the algorithm models random noise present in the training data.

Machine Learning Fundamentals...



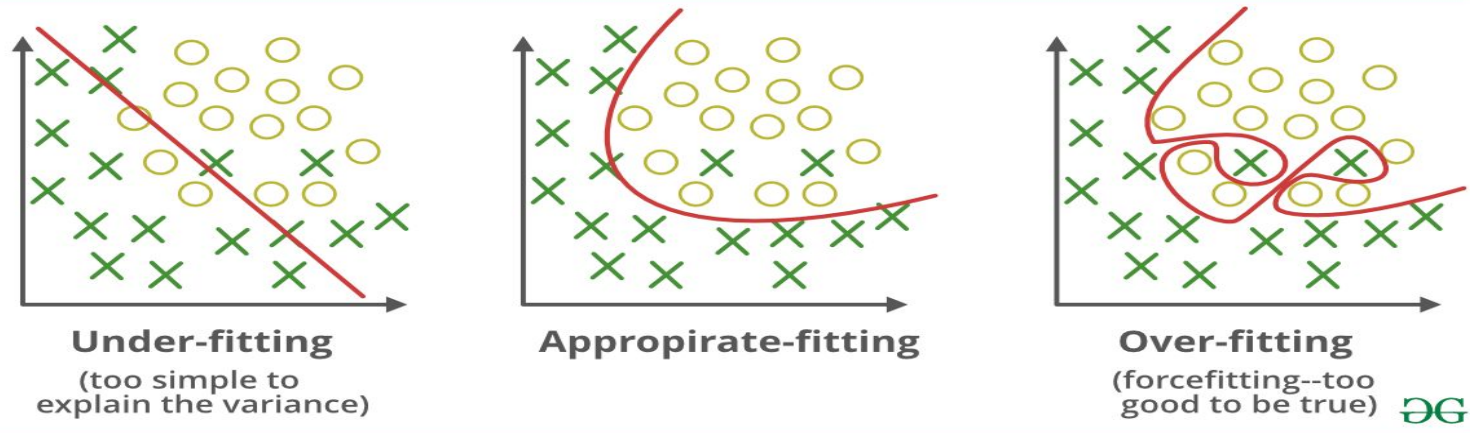
...Bias and Variance!!!



INTERNSHIPSTUDIO

What is Underfitting?

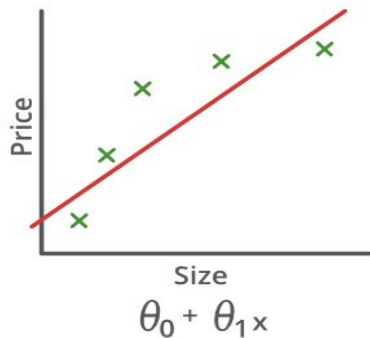
- A statistical model or a algorithm is said to have underfitting when it cannot capture the underlying trend of the data.
- Underfitting destroys the accuracy of our machine learning model.
- Its occurrence simply means that our model or the algorithm does not fit the data well enough.
- It usually happens when we have less data to build an accurate model and also when we try to build a linear model with a non-linear data.



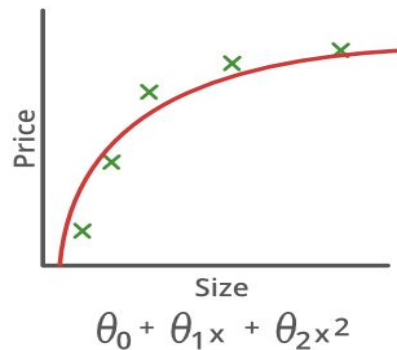
Underfitting

- Underfitting can be avoided by using more data and also reducing the features by feature selection.

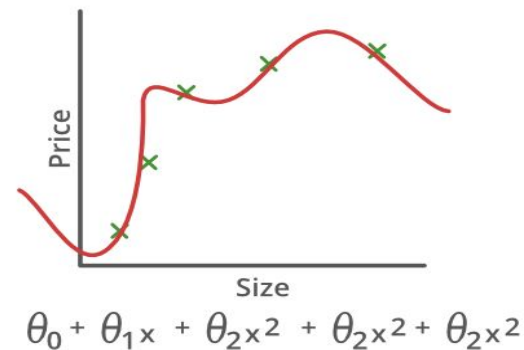
Underfitting – High bias and low variance



High bias (underfit)



High bias (underfit)

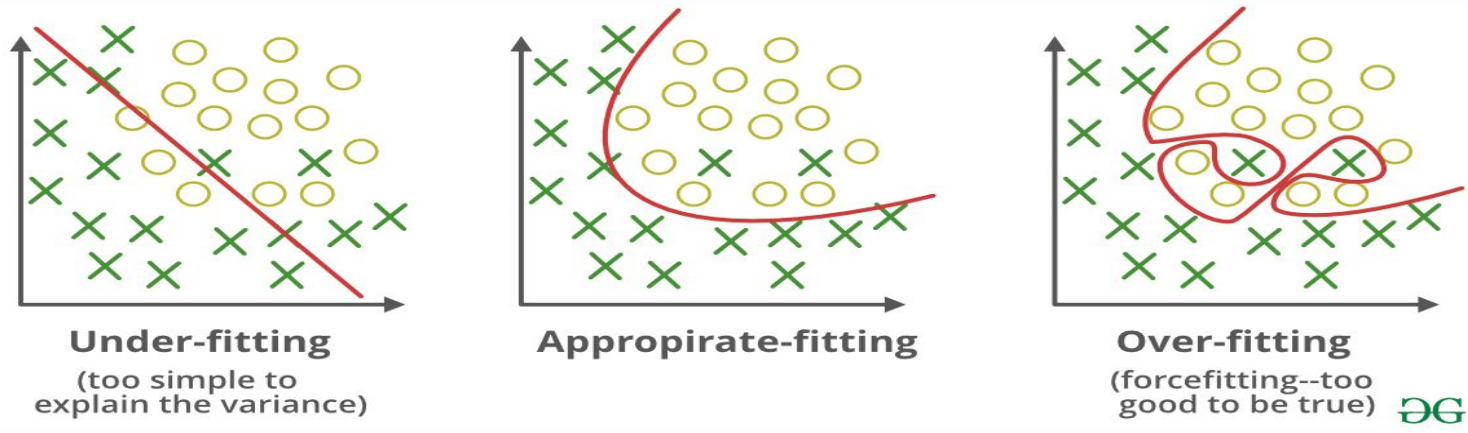


High variance
(overfit)



What is Overfitting?

- Overfitting refers to a model that models the training data too well.
- Overfitting happens when a model 'learns' the detail and noise in the training data to the extent that it try to 'cheat' predictions on new data.
- This means that the noise or random fluctuations in the training data is picked up and learned as concepts by the model.



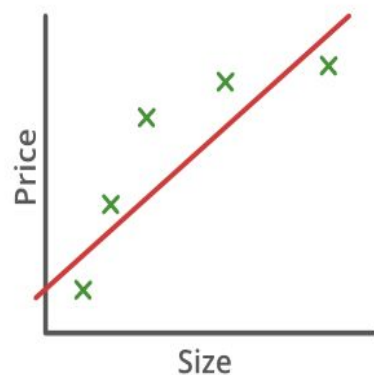
Overfitting



INTERNSHIPSTUDIO

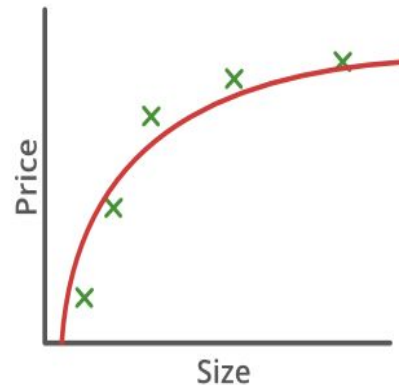
- When a model gets trained with so much of data, it starts learning from the noise and inaccurate data entries in our data set.

Overfitting – High variance and low bias



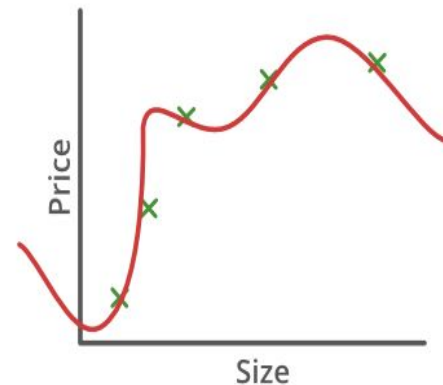
$$\theta_0 + \theta_1 x$$

High bias (underfit)



$$\theta_0 + \theta_1 x + \theta_2 x^2$$

High bias (underfit)



$$\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$$

High variance
(overfit)



How to reduce Overfitting?

Techniques to reduce overfitting :

1. Increase training data.
2. Reduce model complexity.
3. Early stopping during the training phase (have an eye over the loss over the training period as soon as loss begins to increase stop training).
4. Use dropout for neural networks to tackle overfitting.

How to reduce Underfitting?

Techniques to reduce underfitting

1. Increase model complexity
2. Increase number of features, performing feature engineering
3. Remove noise from the data.
4. Increase the number of epochs or increase the duration of training to get better results.

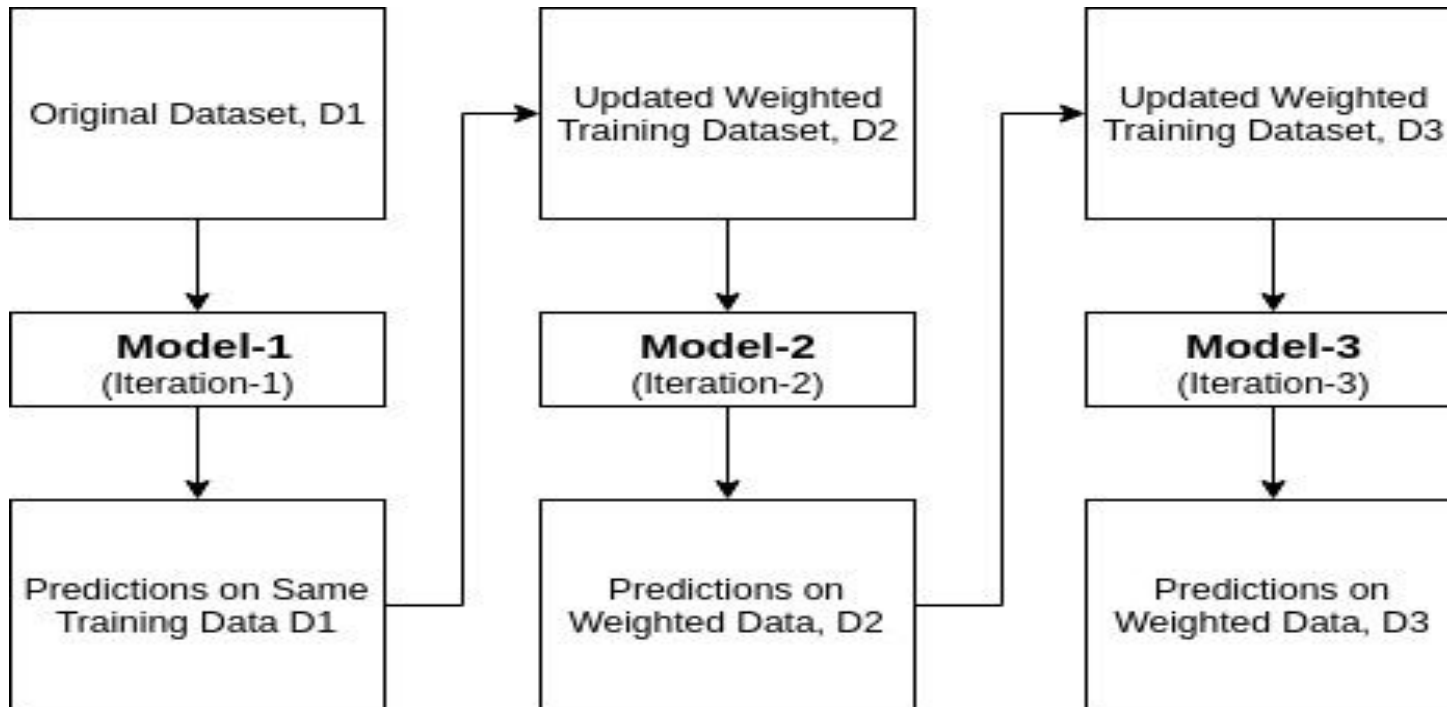
**What is
Ensemble Learning ?**

AdaBoost



INTERNSHIPSTUDIO

- AdaBoost is short for Adaptive Boosting.
- It combines multiple classifiers to increase the accuracy of classifiers.
- AdaBoost is an iterative ensemble method.
- AdaBoost classifier builds a strong classifier by combining multiple poorly performing classifiers so that you will get high accuracy strong classifier.



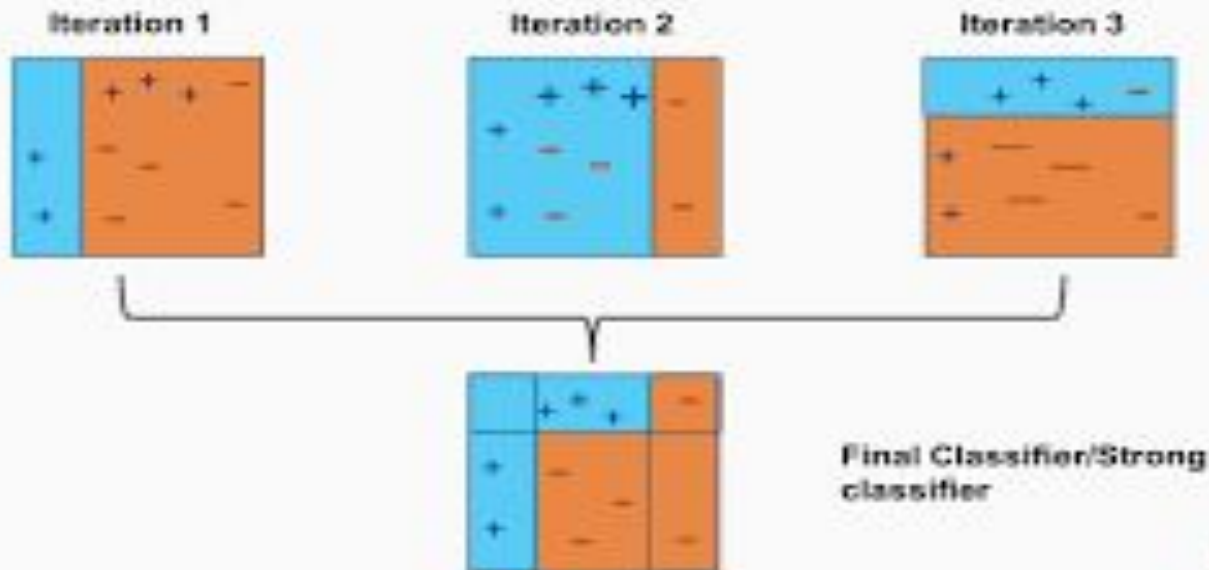
AdaBoosting



INTERNSHIPSTUDIO

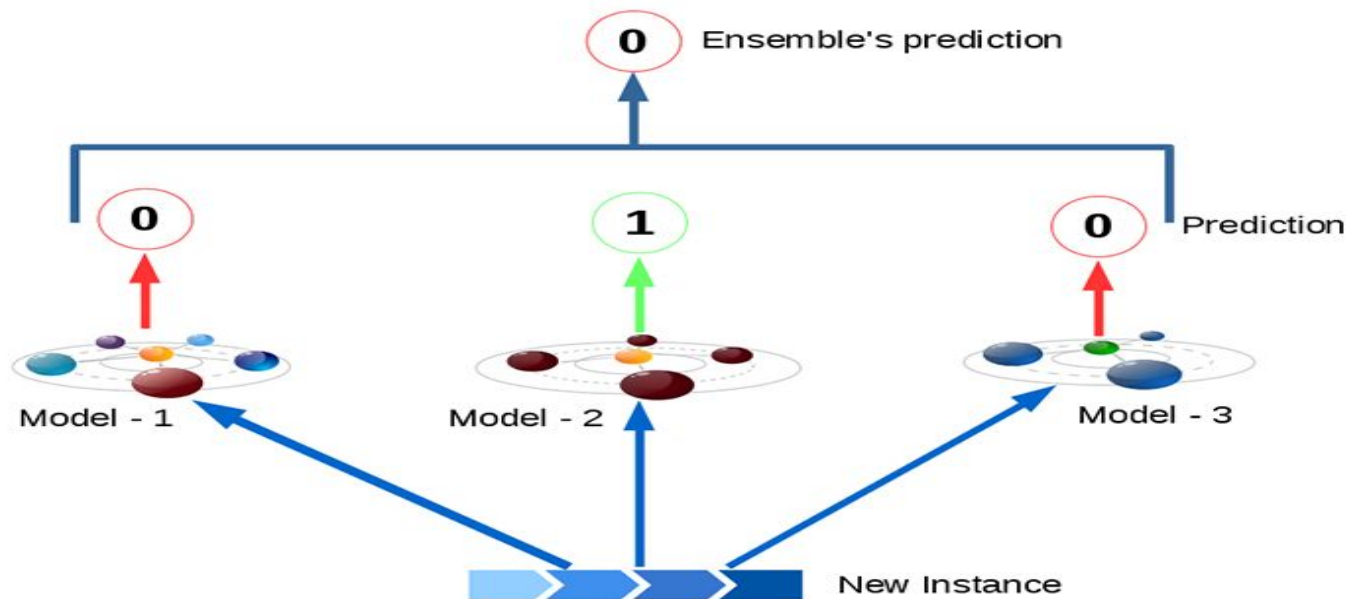
- The weak learners in AdaBoost are decision trees with a single split, called decision stumps.
- AdaBoost works by putting more weight on difficult to classify instances and less on those already handled well.
- AdaBoost algorithms can be used for both classification and regression problem.

Boosting



Voting

- Voting is one of the simplest ways of combining the predictions from multiple machine learning algorithms.
- It works by first creating two or more standalone models from your training dataset. A Voting Classifier can then be used to wrap your models and average the predictions of the sub-models when asked to make predictions for new data.
- You can create a voting ensemble model for classification using the [VotingClassifier](#) class.





Q.1 What is Gradient Boosting?

Q.2 What is XGBoosting?

Q.3 What is Overfitting?

Q.4 What is Underfitting?

Q.5 How to reduce Overfitting?



INTERNSHIPSTUDIO

Thank You