## XGBoost

Extreme Gradient Boosting (or) XGBoost is a supervised Machine-learning algorithm used to predict a target variable ‘y’ given a set of features – Xi. It combines several weak learners into a strong learner to provide a more accurate & generalizable ML model.

XGBoost falls in the same class of ML algorithms as the GBM (Gradient Boosting model). Both these models use an iterative technique known as boosting that builds a number of decision trees one after the other while focusing on correctly predicting those data points that were predicted wrongly in the previous tree.

While both these ML models follow the same principle, XGBoost is better than traditional GBM for three main reasons –

* Regularization: XGB uses regularization while training the model thereby controlling over-fitting of the ML model, which could lead to incorrect predictions on unseen data.
* Performance: XGB is faster because it allows multi-core processing (parallel processing)
* Handling sparse data sets: XGB better handles missing values in the data
* Cache Optimization: Data structures & Algorithms are cache optimized to best use system’s hardware resources.

### Explained

The goal of a supervised learning algorithm is to predict accurately a label ‘y’ based on pattern in the other features - Xi - of the data set. XGBoost constructs several decision trees iteratively for this purpose. Initially, a decision tree is constructed and the label is predicted. The next decision tree will focus on correctly predicting the data points that were predicted incorrectly by the first tree. This process continues until the user-specified number of trees is reached.

The focus on wrong predictions is maintained by assigning weights to those data points to lay more emphasis on a correct prediction in the next tree. Such a model is also called an ensemble model.

### Advantages –

* Produces highly accurate models as a result of multiple decision trees & regularization
* Very good model training performance
* Users can specify custom optimization objectives and evaluation criteria

### Disadvantages –

* Outliers in the data set can affect model quality
* More training time since trees are built iteratively.

XGB:

Extreme Gradient Boosting (xgboost) is similar to [gradient boosting](https://www.analyticsvidhya.com/blog/2015/09/complete-guide-boosting-methods/)  but more efficient. It has both linear model solver and tree learning algorithms. So, what makes it fast is its capacity to do parallel computation on a single machine.

This makes xgboost at least **10 times faster** than existing gradient boosting implementations. It supports various objective functions, including regression, classification and ranking.