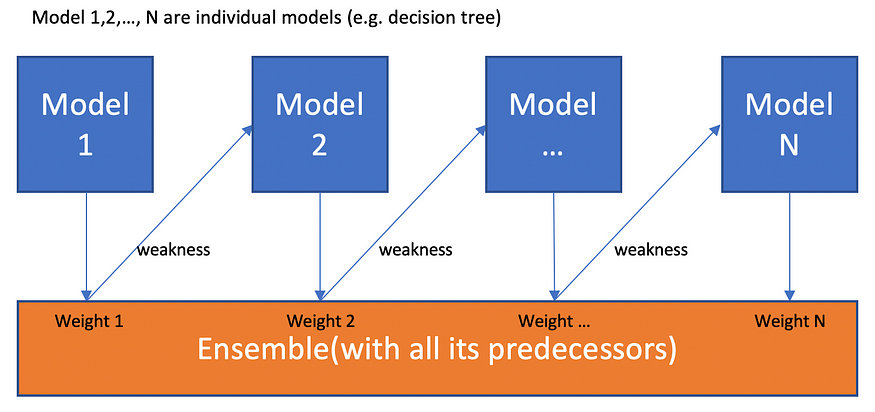
Boosting Algorithm

Boosting algorithms seek to improve the prediction power by training a sequence of weak models, each compensating the weaknesses of its predecessors.



One is weak, together is strong, learning from past is the best

To understand Boosting, it is crucial to recognize that boosting **is ageneric algorithm rather than a specific model**. Boosting needs you to specify a weak model (e.g. regression, shallow decision trees, etc) and then improves it.

With that sorted out, it is time to explore different definitions of weakness and their corresponding algorithms.

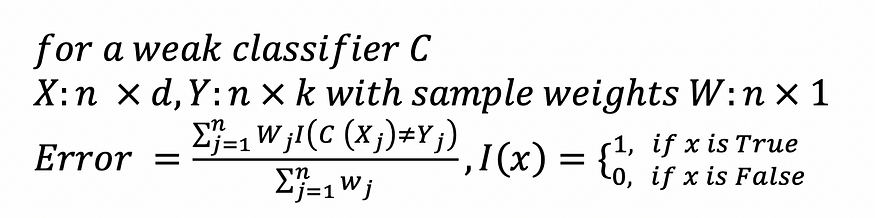
1.) Ada Boost Algorithm (Adaptive Boosting)

2.) X G Boosting Algorithm

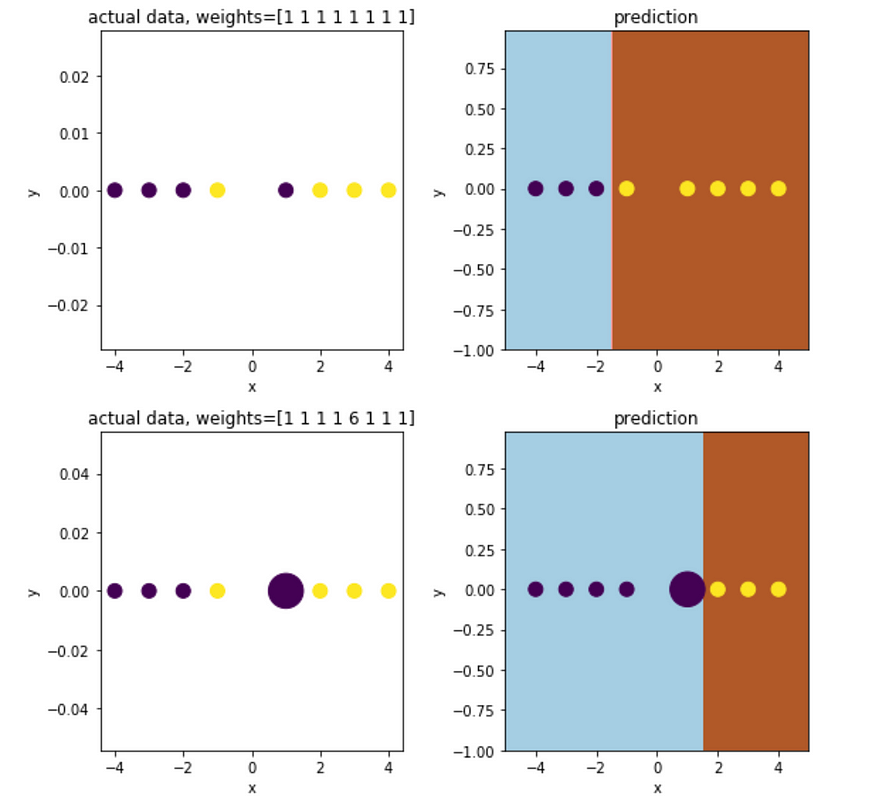
3.) L G Boosting Algorithm

1.) Ada Boost Algorithm

AdaBoost is a specific Boosting algorithm developed for classification problems (also called discrete AdaBoost). The weakness is identified by the weak estimator’s error rate:

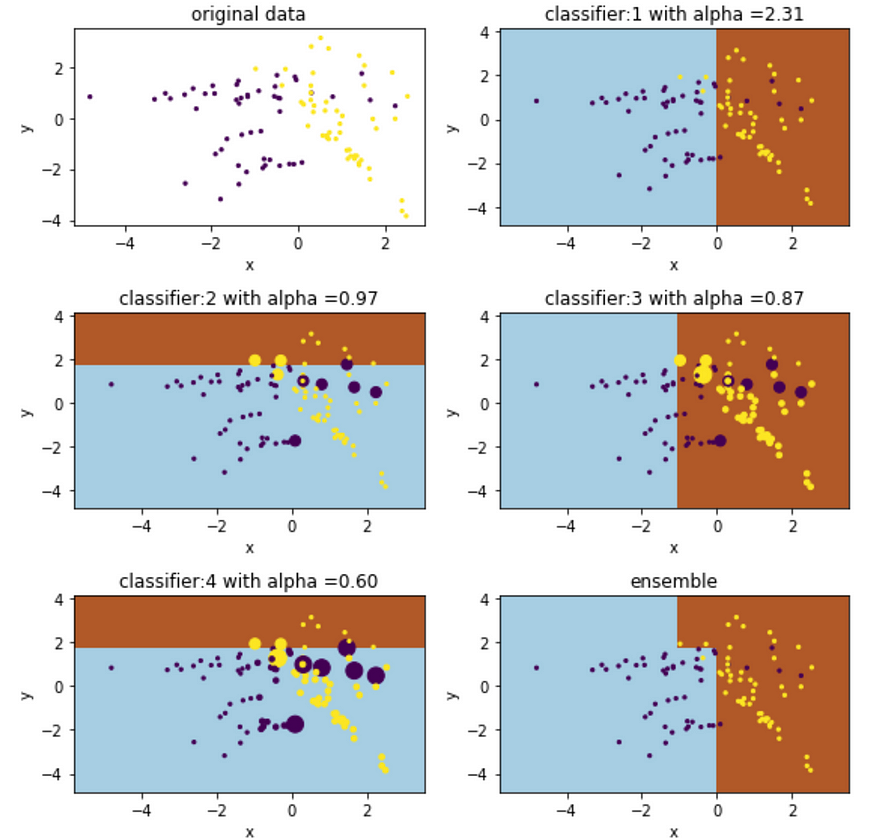


In each iteration, AdaBoost identifies miss-classified data points, increasing their weights (and decrease the weights of correct points, in a sense) so that the next classifier will pay extra attention to get them right.



How sample weights affect the decision boundary

AdaBoost trains a sequence of models with augmented sample weights, generating ‘confidence’ coefficients Alpha for individual classifiers based on errors. Low errors leads to large Alpha, which means higher importance in the voting.



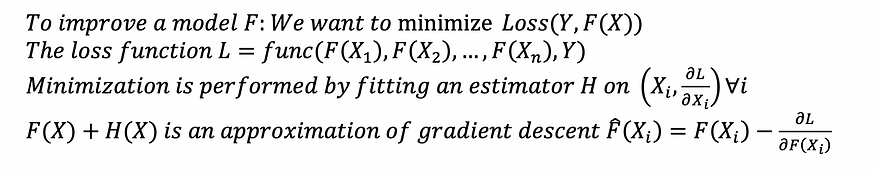
The size of dots indicates their weights

2.) X G Boosting Algorithm

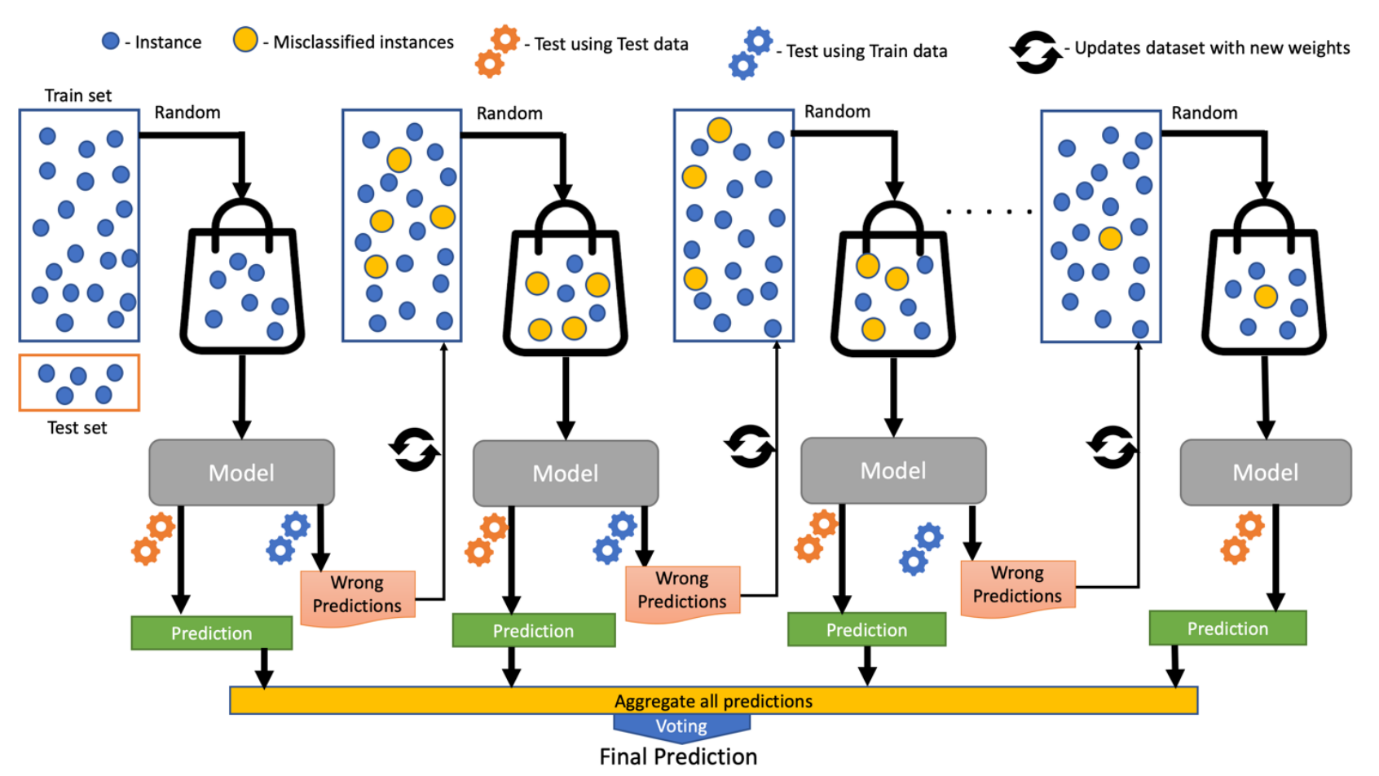
Extreme Gradient Boosting (XG Boost) is an open-source library that provides an efficient and effective implementation of the gradient boosting algorithm.

Shortly after its development and initial release, XG Boost became the go-to method and often the key component in winning solutions for a range of problems in machine learning competitions.

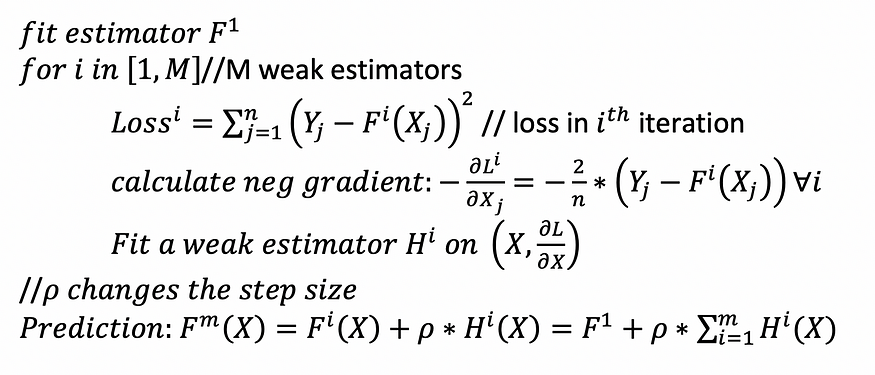
Regression predictive modelling problems involve predicting a numerical value such as a dollar amount or a height. **XG Boost** can be used directly for **regression predictive modelling .**



weakness is defined by gradient

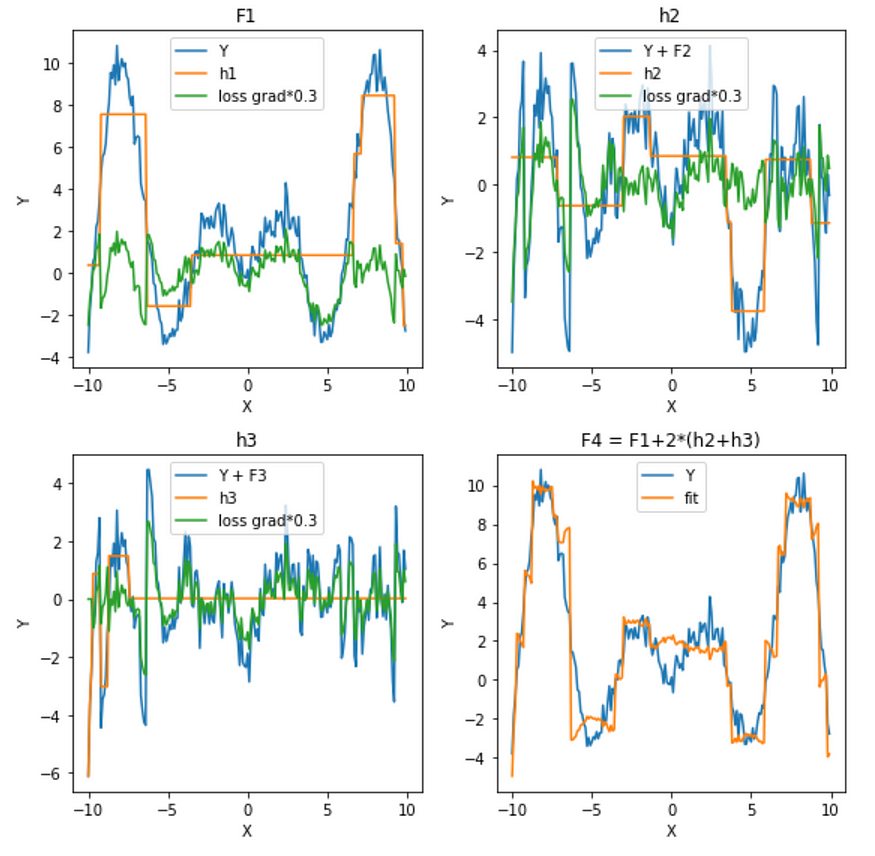


Gradient boosting requires a differential loss function and works for both regression and classifications.



Gradient Boosting with Least Square

Following is a visualization of how weak estimators H are built over time. Each time we fit a new estimator (regression tree with max \_depth =3 in this case) to the gradient of loss (LS in this case).

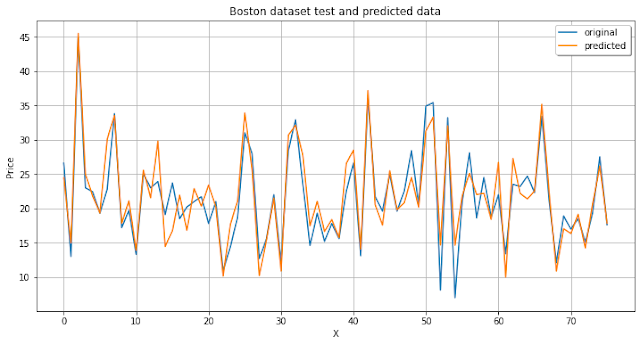


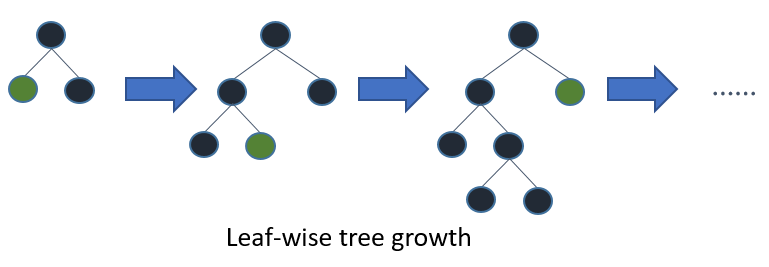
gradient is scaled down for visualization purpose

3.) L G Boosting Algorithm

Light GB is an open-source gradient boosting framework that based on tree learning algorithm and designed to process data faster and provide better accuracy. It can handle large datasets with lower memory usage and supports distributed learning.

We use Boston Housing Price dataset as a target regression data and we can easily load it from sklearn . datasets module. To keep the feature column names, I'll use pandas DataFrame type for feature data. Then, we'll splint data into train and test parts.

[](https://blogger.googleusercontent.com/img/b/R29vZ2xl/AVvXsEi0cb9OEpmUhn5_737F7OzVUt2E2bxfd7HydYoMlEnojyMYxCXf3LVqsvp3aAkjhudxQ2rO2mn3qsh_P7ZQC7nsEEShRavnOWd5VLodOBwbUGVkRa3rEPZlQQ0ZMJ9VuScblcmq8CUum3TiWa0oaEHvVBrMq8M5JEfQCzFMUvqre1y_mMOlAzvauTlp/s717/lightgbm_regression.png)



Because of this growth strategy, it isn’t straightforward to use max\_depth alone to limit the complexity of trees. The num\_leaves parameter sets the maximum number of nodes per tree. Decrease num\_leaves to reduce training time.