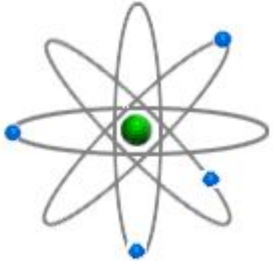
Class 9

H Academy

March 15th, 2021 - By Nathan Landman

Class Agenda

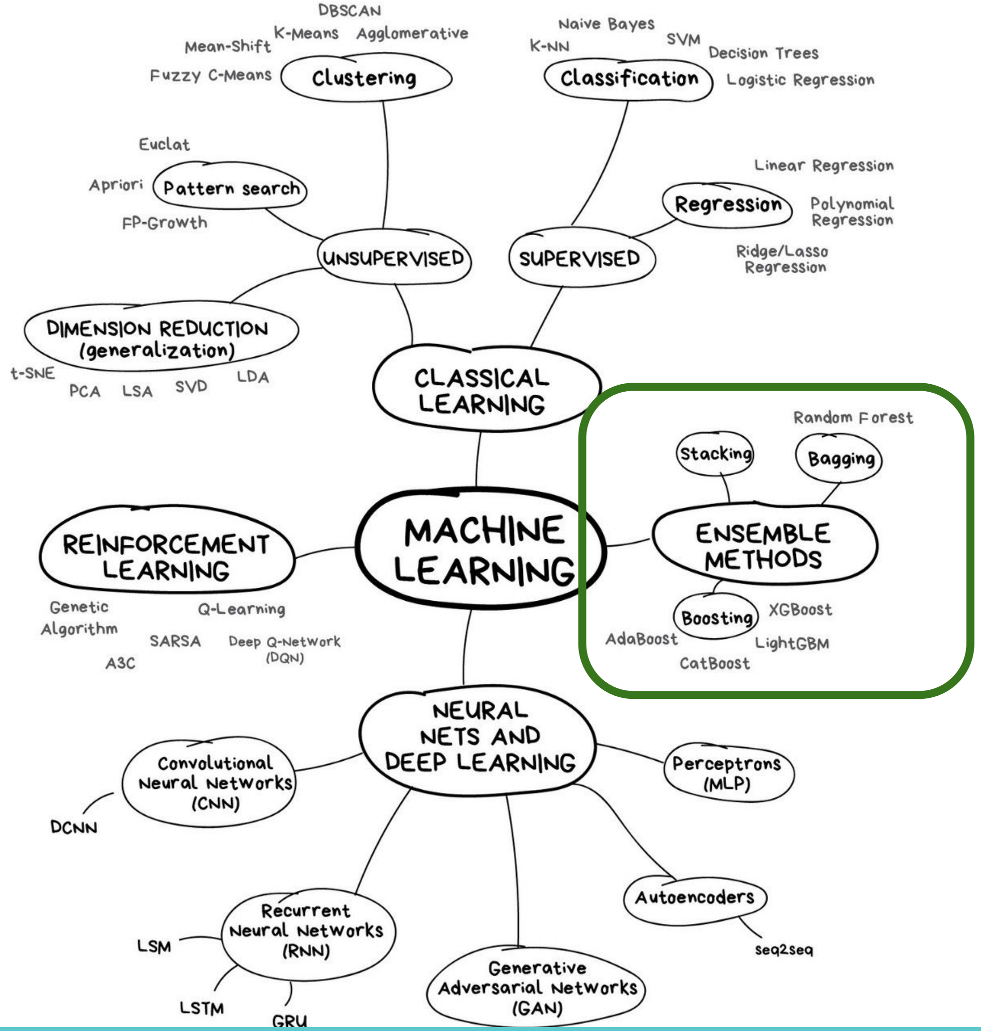


1. Ensemble Methods

2. Bagging

3. Boosting

4. Stacking



Next class….

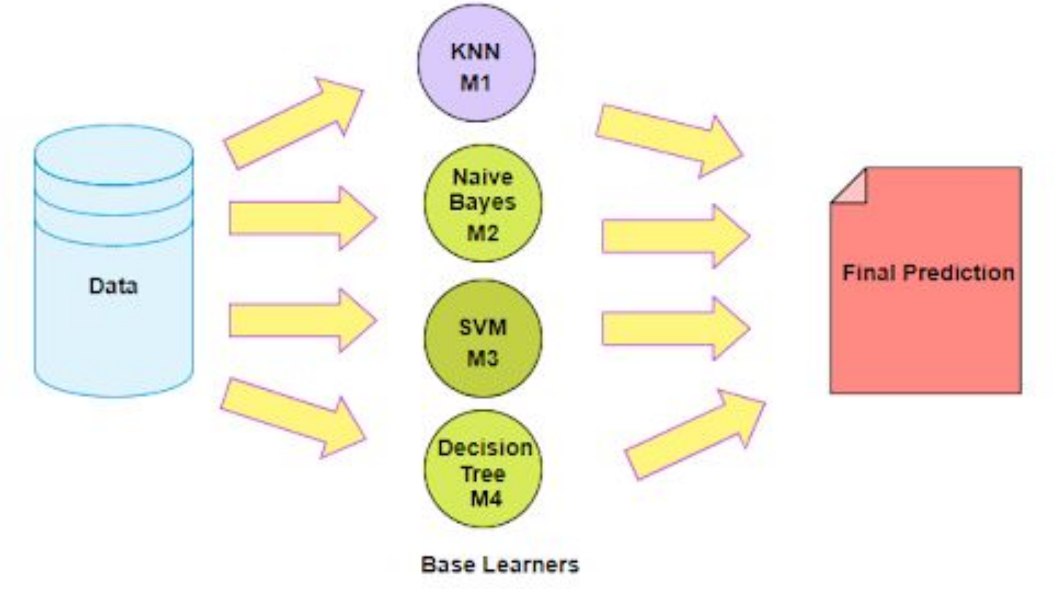
Machine Learning Landscape

1. Ensemble Methods

Ensemble methods is a machine learning technique that combines several base models in order to produce one optimal predictive model.

We'll talk about three types of ensemble methods:

1. **Bagging** - where modeling happens in parallel.
2. **Boosting** - where modeling happens in sequence
3. **Stacking** - Where models are averaged together.

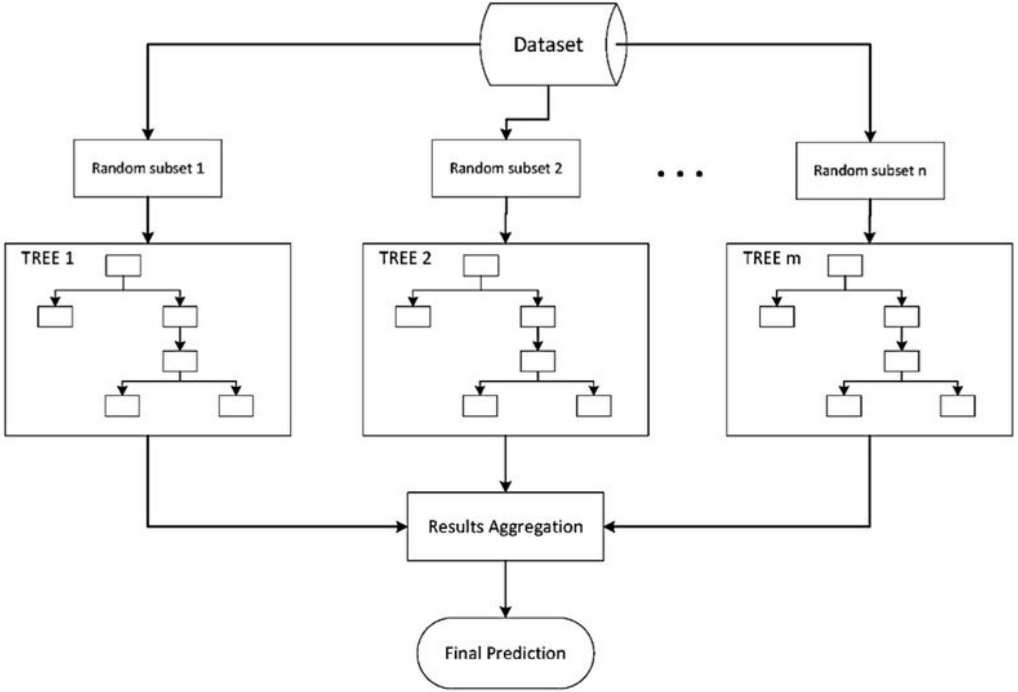


2. **BAGG**ing - **B**ootstrap **AGG**regation

Given a sample of data, multiple bootstrapped subsamples are pulled.

A Decision Tree is formed on each of the bootstrapped subsamples. After each subsample Decision Tree has been formed, an algorithm is used to aggregate over the Decision Trees to form the most efficient predictor.

Bagging models reduce the **variance** of a model.

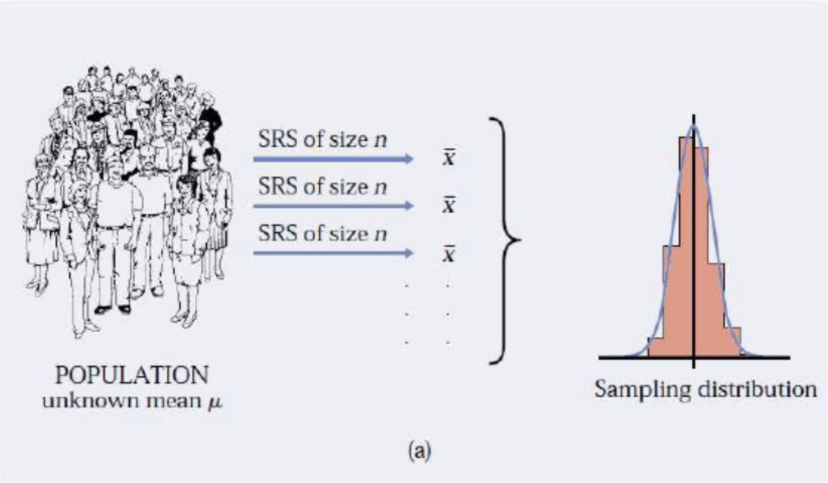


2. **BAGG**ing - **B**ootstrap **AGG**regation

Bootstra-what?

Bootstrapping is a statistical procedure that resamples a single dataset to create many simulated samples.

This process allows you to calculate standard errors, construct confidence intervals, and perform hypothesis testing for numerous types of sample statistics.



The idea of the sampling distribution of the sample mean *x-bar*: take very many samples, collect the *x*-values from each, and look at the distribution of these values

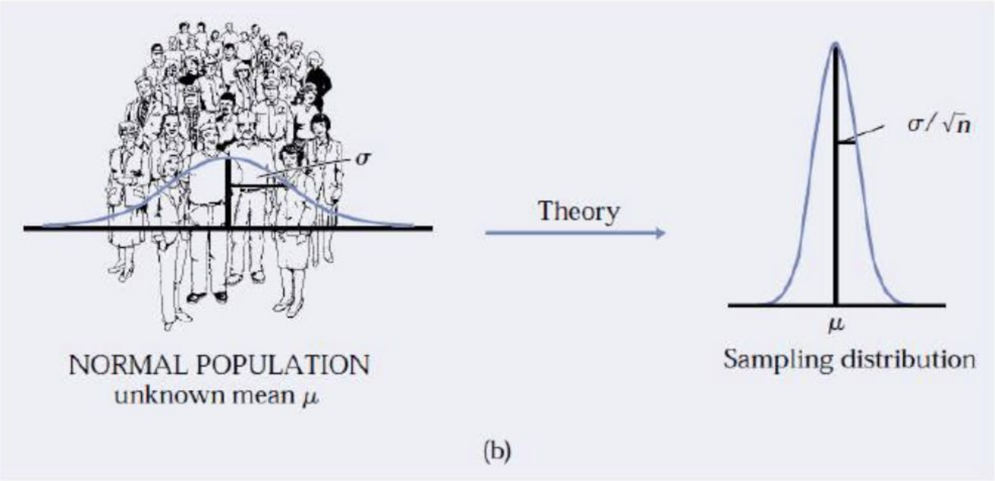
From Hesterberg et al. (2003)

2. **BAGG**ing - **B**ootstrap **AGG**regation

Bootstra-what?

Bootstrapping is a statistical procedure that resamples a single dataset to create many simulated samples.

This process allows you to calculate standard errors, construct confidence intervals, and perform hypothesis testing for numerous types of sample statistics.



The theory shortcut: if we know that the population values follow a normal distribution, theory tells us that the sampling distribution of *x-bar* is also normal.

From Hesterberg et al. (2003)

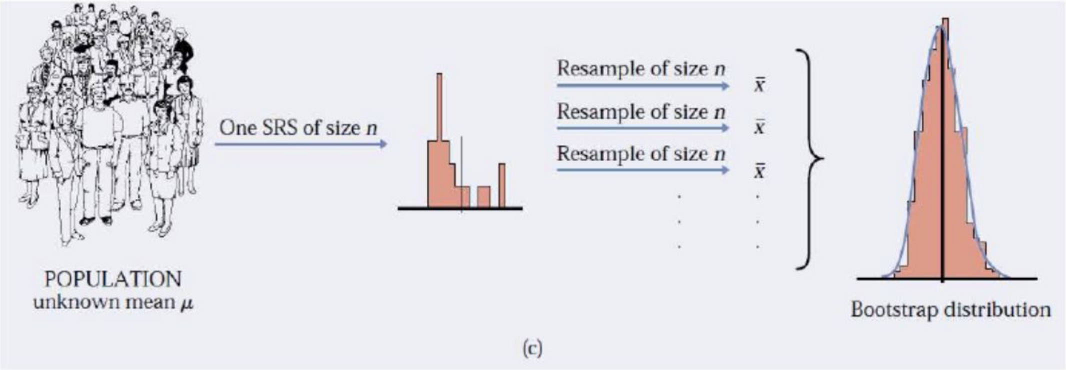
This is known as the central limit theorem

2. **BAGG**ing - **B**ootstrap **AGG**regation

Bootstra-what?

Bootstrapping is a statistical procedure that resamples a single dataset to create many simulated samples.

This process allows you to calculate standard errors, construct confidence intervals, and perform hypothesis testing for numerous types of sample statistics.



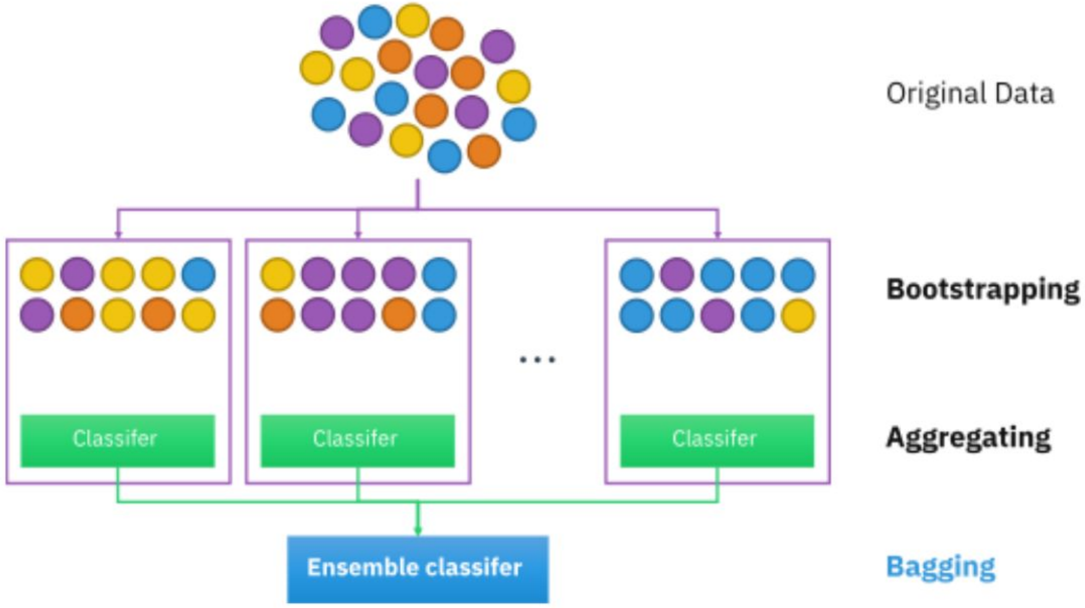
The bootstrap idea: when theory fails and we can afford only one sample, that sample stands in for the population, and
the distribution of *x* in many resamples stands in for the sampling distribution

2. **BAGG**ing - **B**ootstrap **AGG**regation

Given a sample of data, multiple bootstrapped subsamples are pulled.

A Decision Tree is formed on each of the bootstrapped subsamples. After each subsample Decision Tree has been formed, an algorithm is used to aggregate over the Decision Trees to form the most efficient predictor.

Bagging models reduce the **variance** of a model.



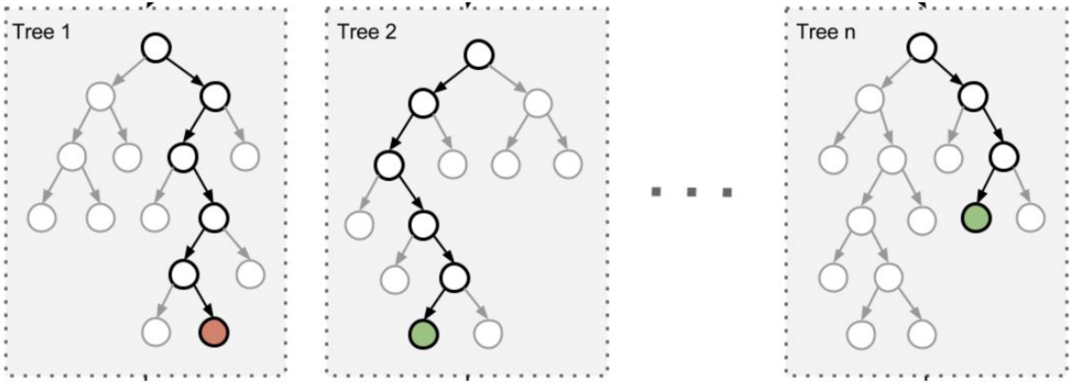
$$f(x) = 1/M \sum_{m=1}^{M} f_m(x)$$

2. **BAGG**ing - Random Forests

In **random forests**, each tree in the ensemble is built from a **bootstrapped** sample drawn with replacement.

Instead of using all the features, **a random subset of features is selected**, further randomizing the tree.

As a result, the **bias** of the forest **increases** slightly, but due to the averaging of less correlated trees, its variance decreases, resulting in an overall better model.

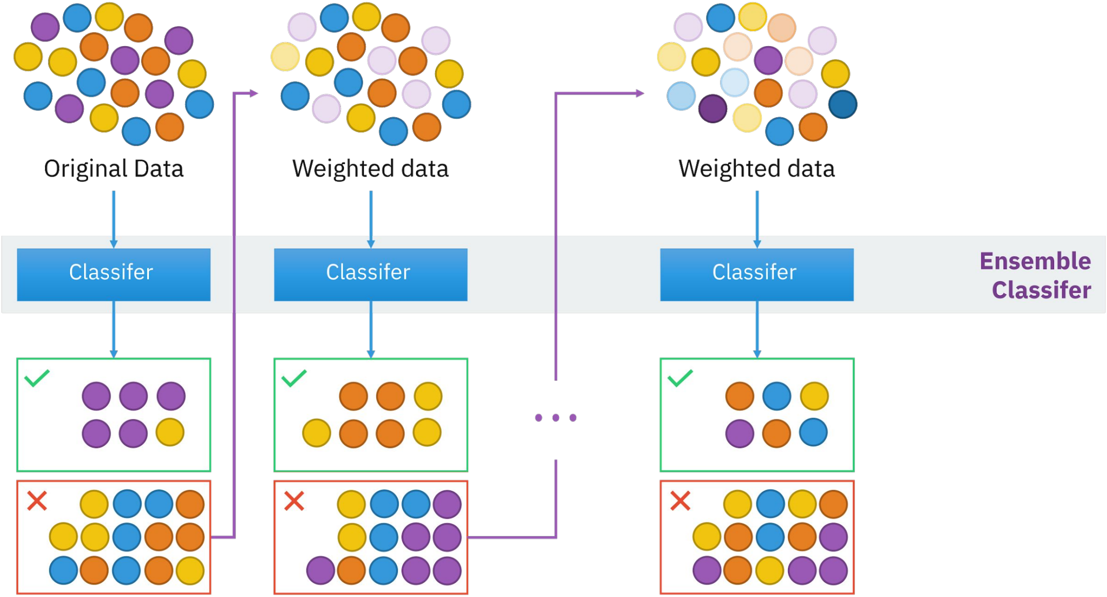


This shows a decision tree, but any classifier/regressor could be used.

3. Boosting

In **boosting**, models are fit on weighted versions of the dataset; more weight is given to examples that were misclassified by earlier rounds.

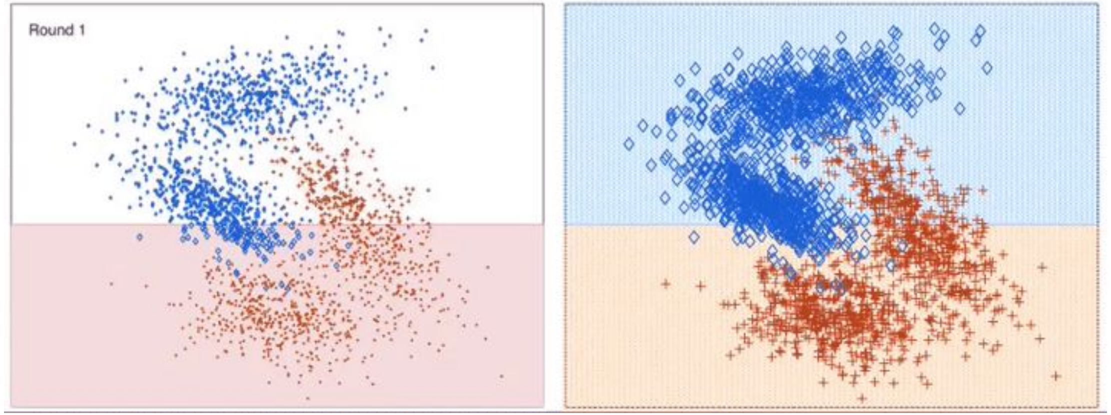
The predictions are then combined through a weighted majority vote (classification) or a weighted sum (regression) to produce the final prediction.



3. Boosting - *AdaBoost* - Adaptive Boosting

AdaBoost works by weighting the observations, putting more weight on difficult to classify instances and less on those already handled well. New weak learners are added sequentially that focus their training on the more difficult patterns.

Predictions are made by majority vote of the weak learners' predictions, weighted by their individual accuracy.

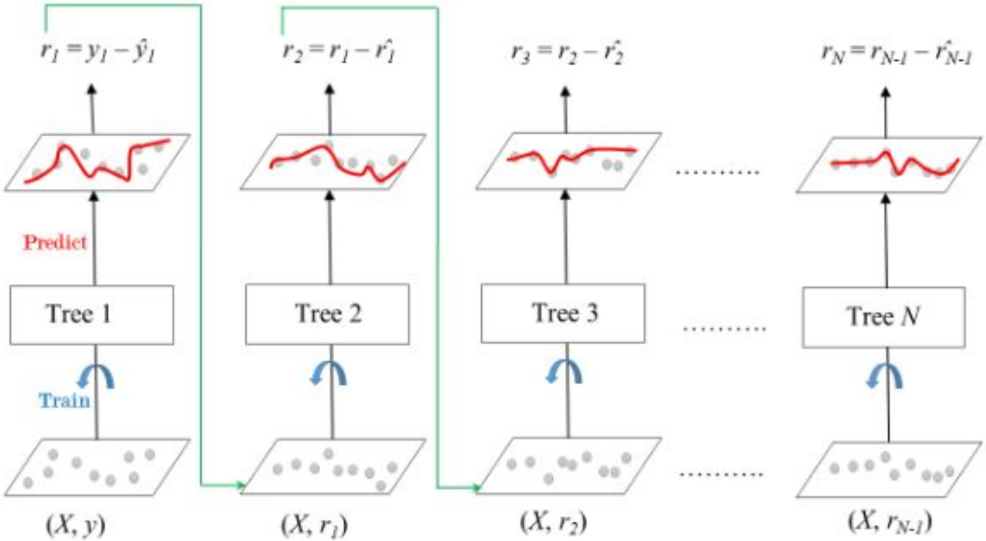


The weak learners in AdaBoost are decision trees with a single split, called decision stumps.

3. Boosting - Gradient Boosting Tree

The main difference between a GBT and AdaBoost is that GBT defines a **loss function** to be optimized.

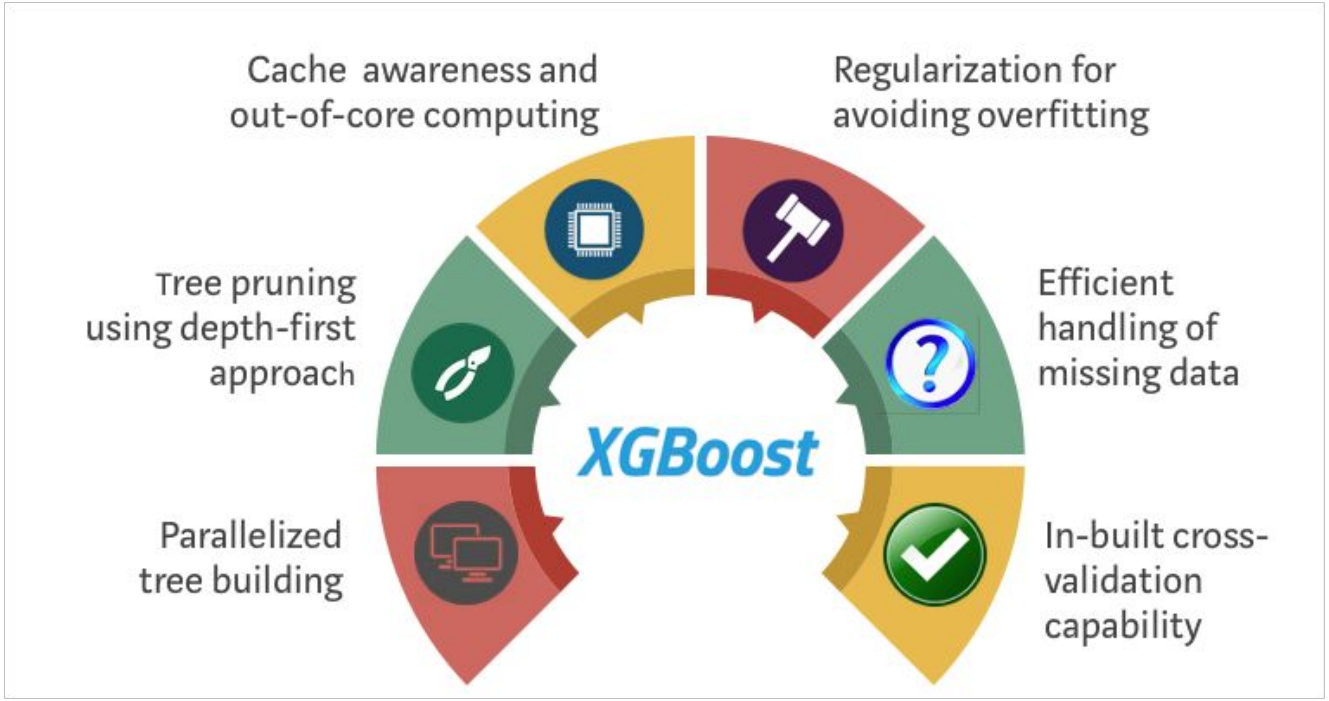
After calculating the loss, to perform the gradient descent procedure, we must add a tree to the model that reduces the loss (i.e. follow the gradient). We do this by parameterizing the tree, then modify the parameters of the tree and move in the right direction by (reducing the residual loss.



Read this blog for more details.

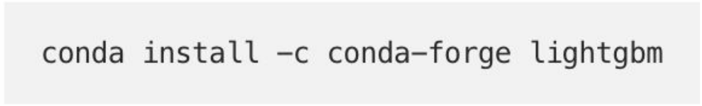
3. Boosting - XGBoost - eXtreme GBT!

What's the difference between a GBT and XGBoost? Performance!

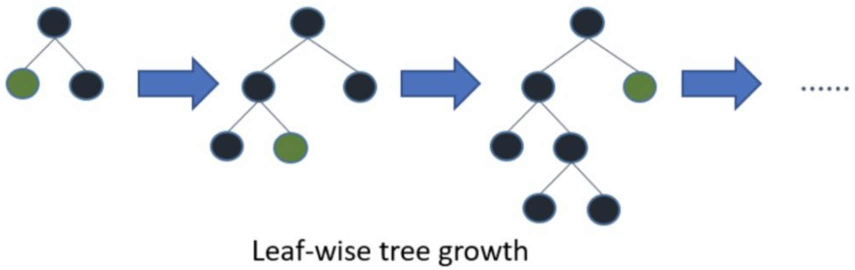

Cache awareness and out-of-core computing
Regularization for avoiding overfitting
Tree pruning using depth-first approach
Efficient handling of missing data
XGBoost
Parallelized tree building
In-built cross-validation capability

3. Boosting - LightGBM

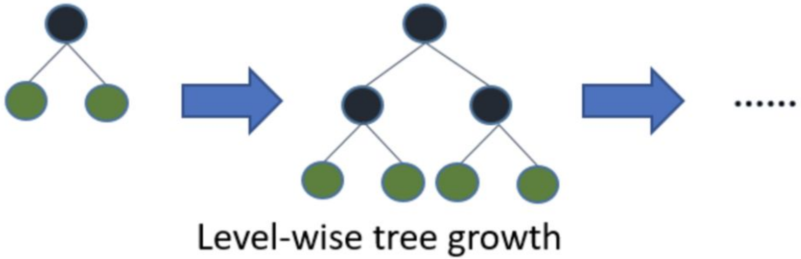
LightGBM is a gradient boosting framework that uses tree based learning. It grows vertically while other algorithms grow horizontally.



```
conda install -c conda-forge lightgbm
```

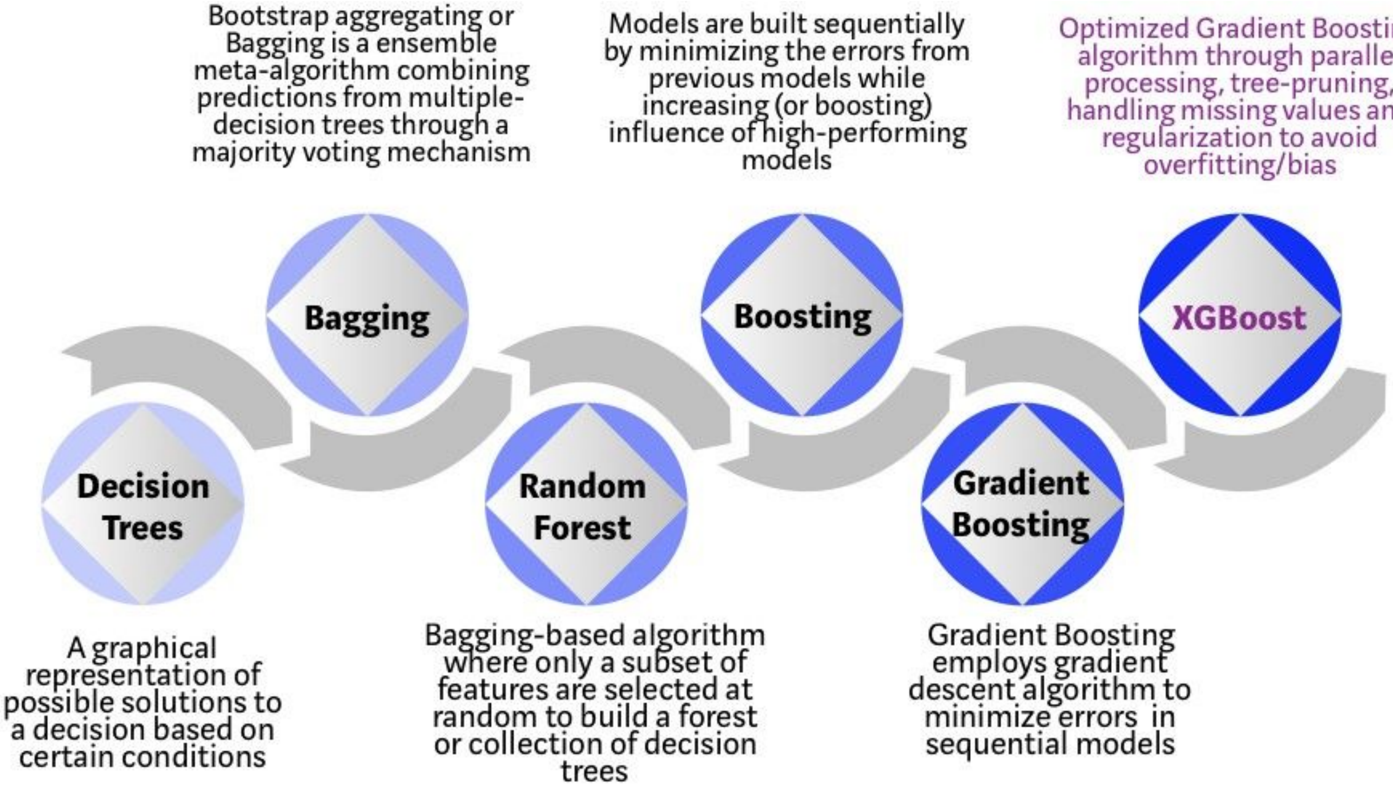


Explains how LightGBM works



How other boosting algorithm works

Summary of Bagging and Boosting



Decision Trees

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

Bagging

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

Random Forest

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

Boosting

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

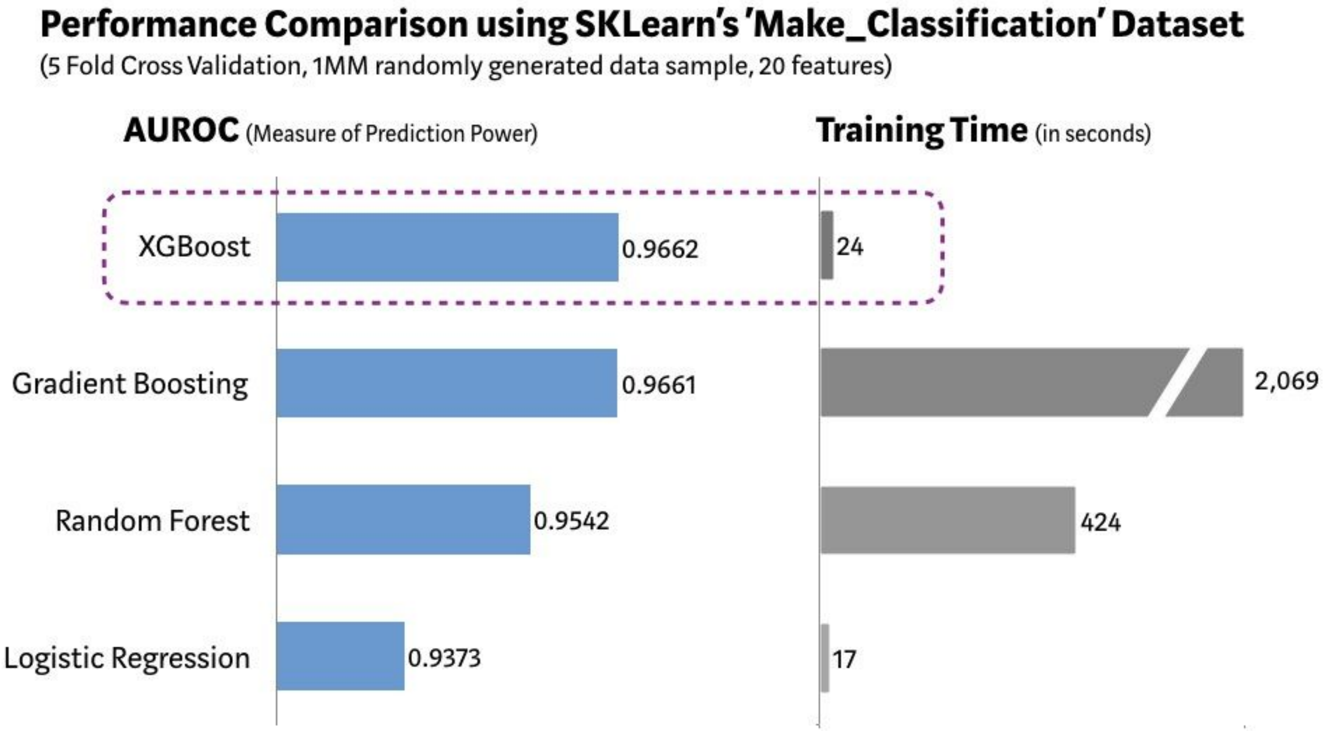
Gradient Boosting

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

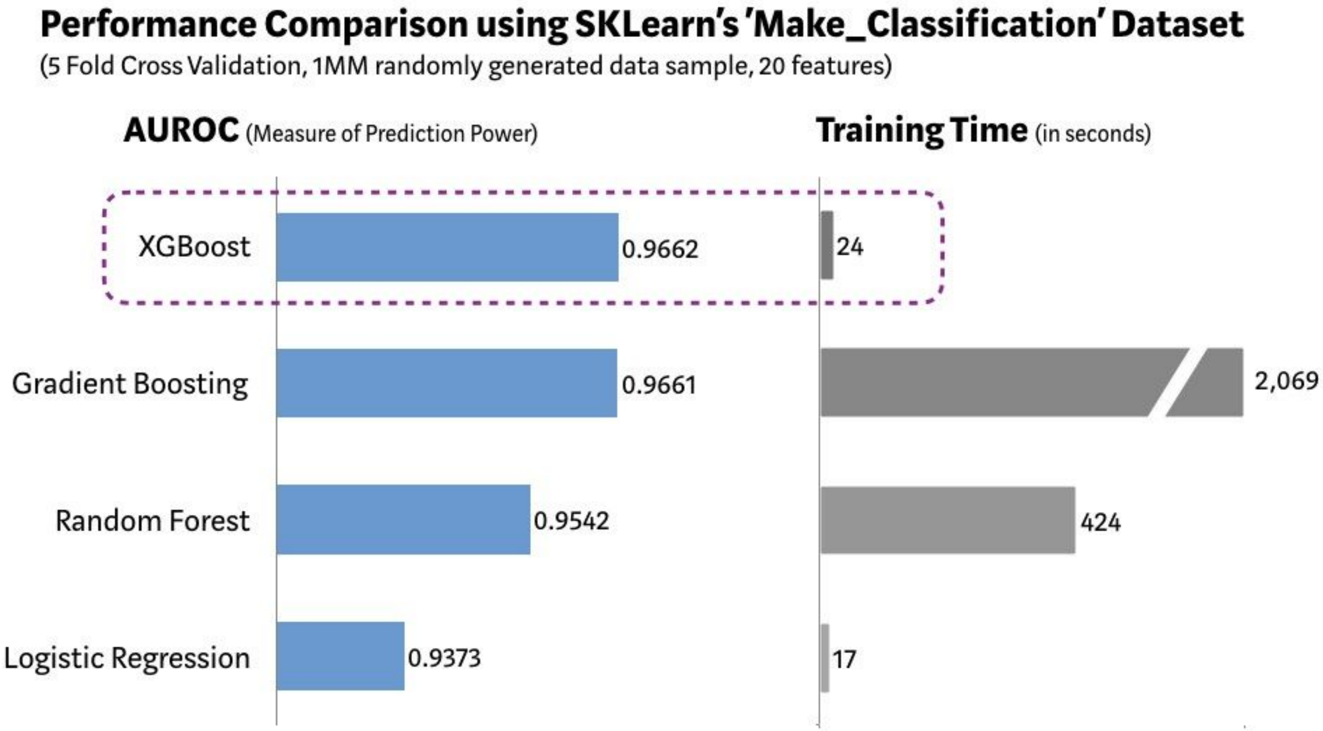
XGBoost

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

Summary of Bagging and Boosting



Summary of Bagging and Boosting - Can we do better?

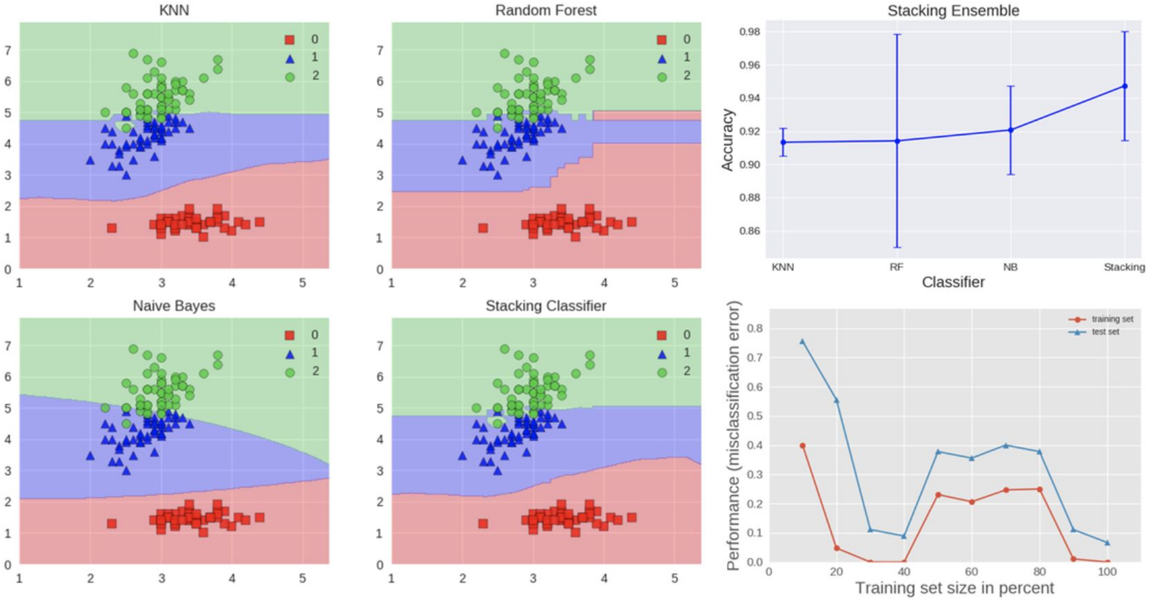


4. Stacking

Stacked Generalization or **Stacking** involves combining the predictions from multiple machine learning models on the same dataset.

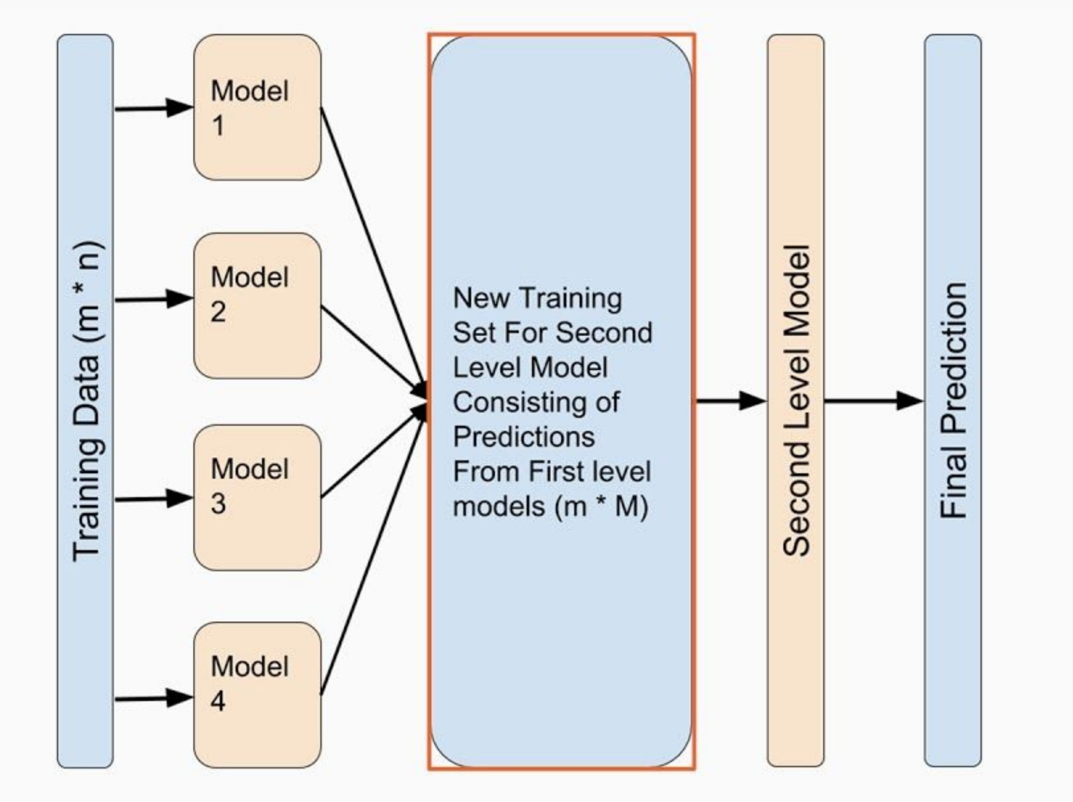
It contains level-0 models (base-models) and level-1 models (meta-models).

The meta model is trained on the predictions made by the base models.



In this case, the model consists of kNN, Random Forest, and Naive Bayes base classifiers whose predictions are combined by Logistic Regression as a meta-classifier.

4. Stacking


Training Data (m * n)
Model 1
Model 2
Model 3
Model 4
New Training Set For Second Level Model Consisting of Predictions From First level models (m * M)
Second Level Model
Final Prediction

5. Exercise

Your task is to use the Titanic dataset and get the highest accuracy possible from now until the end of class.

Note the time limits!

1. Load the dataset.
2. Establish a benchmark (Classify all as *alive* and see accuracy).
3. Use a logistic regressor and see if you can beat the benchmark.
4. Continue with more sophisticated models until you achieve the best score possible!

For inspiration, look at this extensive tutorial.