XGBoost

## Phase I

1. An algorithm
2. A gradient boosted decision tree
3. Stands for eXtreme Gradient Boosting.
4. It is a library.
   1. Library is written in C++
   2. There is a model in scikit-learn
5. The implementation of the model supports features of scikit-learn.
6. There are three main forms of gradient boosting.
   1. Gradient boosting
   2. Stochastic Gradient boosting
      1. Stochastic means randomly determined (each input results in a different output)
      2. Regularized Gradient Boosting.
7. Algorithm features
   1. Sparse Aware: automatic handling of missing data values
   2. Block Structure: support the parallelization of tree construction.
   3. Continued training: further boost an already fitted model on new data.
8. XGBoost is a free open source software.
9. Why use XGBoost?
   1. Execution speed and model performance
10. It is a popular library for boosting.

**Phase I reflections**

1. All I know is that it is an algorithm and that it is memory efficient and of high performance.
2. It is used for boosting which I still do not know what that means.
3. XGBoost stands for extreme gradient boosting and there are three main kinds of gradient boosting. I have to find out what gradient boosting means.

## Phase 2

1. XGBoost dominates structured datasets on classification and regression problems. Evidence being that it is the go-to algorithm for competition on Kaggle.
2. Gradient boosting decision tree algorithm.
   1. Boosting:
      1. **Ensemble algorithm** used for reducing bias and also variance.
      2. An algorithm that converts weak learners to strong ones.
      3. “Can a set of weak learners create a single strong learner?
      4. It uses a number of weak classifiers to create a strong classifier
      5. You add each model sequentially to reduce errors from the prior model and this goes on until the training set is predicted perfectly or there are no more models left to add.
      6. Adaboost was the first successful boosting algorithm developed for binary classification.
   2. Gradient Boosting
      1. Gradient boosting is still a boosting technique meaning that it uses models in a sequential manner to reduce the error rate of a problem.
      2. But the ‘Gradient’ part comes in where by the Gradient Descent Algorithm is used to minimize the error rate for these models.
      3. It’s the algorithm use for reduce the loss that gives it its name.
   3. A technique for regression and classification problems.
   4. Decision tree:
      1. One of the predictive modelling approaches used in statistics.
      2. In the tree structures the leaves represent the class labels and the branches represent conjunctions of features that lead to those class labels.
         1. Each node represents a condition or a test of an attribute
         2. A branch represents the outcome of the test usually a yes or no.
         3. And each leaf label represents a class label.
   5. The goal of a decision tree model is to predict the value of a target variable based on several input variables.
3. Boosting is an ensemble technique:
   1. Sequential
      1. Meaning that new models are added to correct the errors made by existing models.
      2. The models are used sequentially under no further improvements can be made.
   2. How it works:
      1. The new models used predict the residuals / errors of the prior models
   3. Bagging
      1. One
         1. B: Bootstrap
         2. Agg: Aggregating
      2. Two
         1. A bagging model combines bootstrapping and aggregation to form an ensemble method.
         2. What bootstrapping does is to take many sub samples from the original sample of data for an objective.
         3. The results of these sub samples are aggregated using an algorithm to give the
      3. Three
         1. Bootstrapping is a statistical technique for estimating a population statistic from a sample.
            1. A statistic is something like the mean or the standard deviation. It’s a number that tells you something as a whole about a dataset.
            2. Bootstrapping is sampling with replacement.
            3. Assume we have 100 values and we want to find the mean of these values to estimate the population’s mean.
            4. You can take 1000 samples of 10 from these values each time putting back the 10 values after recording their values and then aggregate the mean of all these 1000 samples as an estimate for the population mean.

**Presentation conclusions**

1. XGBoost is an ensemble algorithm.
   1. What this means is that it is a group of algorithms just like an ensemble can also be a group of musicians.
   2. The behavior of this ensemble [ahn-sahm-buhl] is sequential (not like bagging).
      1. Meaning that one model at a time is used.
      2. The prediction error from the first model is passed on to the second model for reduction.
      3. Any remaining error is passed on to the third model and this continues on.

**Personal reflections**

1. Think of data science as a tool box filled with algorithms to use. That’s it. There is a structure to it like a sport. You can keep playing until you are familiar with the game. Think of it this way. It makes things much easier.
2. Simply repeat what other people have said or written. Just know it conceptually. You just need to know how to apply and explain conceptually.
3. Layman: It’s simply the combination of models to make the prediction more accurate.
   1. The combination is in a sequential manner where the 1st model is used followed by the second and the third and each subsequent model aims to predict what the previous models are not able to predict.

Links

1. https://machinelearningmastery.com/gentle-introduction-xgboost-applied-machine-learning/
2. <https://en.wikipedia.org/wiki/Gradient_boosting>
3. [What is bagging with reference to decision trees](4.%09https:/towardsdatascience.com/ensemble-methods-in-machine-learning-what-are-they-and-why-use-them-68ec3f9fef5f)

## Phase 3

1. Neural networks are popular but if you have limited training data and want little training time, and have little expertise for parameter tuning, boosting algorithms are useful.
2. XGBoost has regularization parameters which allows it to control over fitting. This is what allows it to have a better performance.
3. Boosting is a ensemble method which means that a collection of algorithms are used.
   1. In the case of boosting a collection of weak classifiers are used.
   2. Each model is used sequentially to correct the errors of the previous models.
   3. Bagging on the other hand does not work sequentially it takes the results of all the models and aggregates the results.
4. XGboost are good at tree-based models.
   1. XGboost models require less training data than neural nets.
   2. For image recognition , speech recognition you would use deep learning models.
   3. XGboost shows better results when there is a mixture of categorical features and numeric features and where the number of features is less than the training examples.
   4. Besides
5. Others
   1. A weak learner is an algorithm that predicts slightly better than chance

P-RFL

1. Phew! I feel so much better now that my memory and knowledge are coming back and that I remember that I do not have to be afraid.
2. I do not need to know the actual math. I should be a conceptual data scientist. I must be more confident Zhiwei. In tackling the questions. I do know all these.
3. Just know the use-cases of when to use these tools. Don’t get too technical. You are not in a period of your life where you want to do that. Just know when to use which and when and how it is when compared to others.
4. I’m starting to see the structure in all machine learning problems. Don’t get too technical. Learn what you need to learn most. Let the core idea stick. Don’t trouble yourself unnecessarily. Train test set. Cross validation. Loss function. Accuracy. Precision. And so on. Just follow what the community is saying. They are powerful enough. Don’t stress yourself unnecessarily ☺. Base your work on demand. Solve it when given. Don’t keep trying to pre-empt. It’s stressful. Just like your interview.
5. But XG boost is a good model to learn to add to my tool box. Probably learn it and test it on some dataset or against another model.
6. Same thing. Playing around with the model and its parameters, playing around with another model and then seeing the results.

**Link**

1. <https://towardsdatascience.com/catboost-vs-light-gbm-vs-xgboost-5f93620723db>
2. <https://www.kdnuggets.com/2017/10/xgboost-top-machine-learning-method-kaggle-explained.html>
3. <https://www.quora.com/Why-is-xgboost-given-so-much-less-attention-than-deep-learning-despite-its-ubiquity-in-winning-Kaggle-solutions>
4. https://www.quora.com/Why-is-xgboost-given-so-much-less-attention-than-deep-learning-despite-its-ubiquity-in-winning-Kaggle-solutions