XGBoost implementation.

I am Alisher, PhD student, EECS International Graduate Program, National Chiao Tung University.

Introduction

In statistics and machine learning, *ensemble methods* use multiple learning algorithms to obtain better predictive performance than could be obtained from any of the constituent learning algorithms alone. Common types of ensembles: Bayes optimal classifier, Bootstrap aggregating (bagging) and Boosting. This project will be focus on Boosting algorithm. Boosting is a machine learning ensemble meta-algorithm for primarily reducing bias, and also variance in supervised learning, and a family of machine learning algorithms that convert weak learners to strong ones. One of the most popular boosting algorithm among the Kaggle community is XGBoost.

Why do I want to do this project?

In my last Kaggle competition related to an image classification problem, the victory was won by a girl from Japan who used XGBoost algorithm. Even this algorithm is very popular, the most people just try to tune hyperparameters for training the model. But very few people implement XGBoost by their own. The reason is obvious, if a powerful enough tool already exists, then why come up with something new.

However, I think that for education process is really helpful when you don't use libraries, but try to implement it by yourself. It can give more insight into how XGBoost works.

I propose a software project that implements small part of XGBoost algorithm related to regression problem.

What am I going to do in this project?

Learn and understand the math background of XGBoost algorithm.

Implement XGBoost algorithm for linear regression.

Plot all results to get better understanding of the XGBoost algorithm:

- Approximation on test data by original XGBoost and XGBoost implemented by me.
- Loss function on train and test data

Write unit tests.

Compare to the original XGBoost.

Prepare final report and presentation.

References.

https://www.kaggle.com/

https://github.com/dmlc/xgboost

https://xgboost.readthedocs.io/en/latest/

https://pypi.org/project/xgboost/

http://cinslab.com/wp-content/uploads/2019/06/Ke-Wang-XGBoost-A-Scalable-Tree-Boosting-System.pdf