

XGBoost

Epoch IIT Hyderabad

Arin Aggarwal
MA22BTECH11006

1 INTRODUCTION

eXtreme Gradient Boosting (XGBoost) is an ensemble learning method that combines the predictions of multiple weak models to produce a stronger prediction. It employs a few added techniques to enhance results provided by Gradient Boost Algorithm.

2 ADDED TECHNIQUES

2.1 *Regularisation*

It penalizes more complex models through regularization to prevent overfitting. Ridge and Lasso regression can be done.

2.2 *Sparsity Awareness*

XGBoost can handle sparse data that may be generated from preprocessing steps or missing values. It uses a special split finding algorithm that is incorporated into it that can handle different types of sparsity patterns.

2.3 *Weighted Quantile Sketch*

Quantile Sketches are used to find optimal split which increases efficiency.

2.4 *Parallel Learning*

For faster computing, XGBoost can make use of multiple cores on the CPU.

2.5 *Cache Awareness*

XGBoost has been designed to make optimal use of hardware. This is done by allocating internal buffers in each thread, where the gradient statistics can be stored.

2.6 *Tree Pruning*

XGBoost uses `max_depth` parameter as specified the stopping criteria for the splitting of the branch, and starts pruning trees backward. This depth-first approach improves computational performance significantly.