XGBoost

EXTREME GRADIENT BOOSTING

INNOMATICS RESEARCH LABS

Contents

- What is XGBoost?
- Key Highlights
- XGBoost Working
- XGBoost Tuning Parameters

INNOMATICS RESEARCH LABS

What is XGBoost?

- XGBoost (Extreme Gradient Boosting) is an optimized distributed gradient boosting library.
- It uses Gradient Boosting (GBM) framework at core.
- Gradient boosting is a supervised learning algorithm, which attempts to accurately predict a target variable by combining the estimates of a set of simpler, weaker models.

INNOMATICS RESEARCH LABS

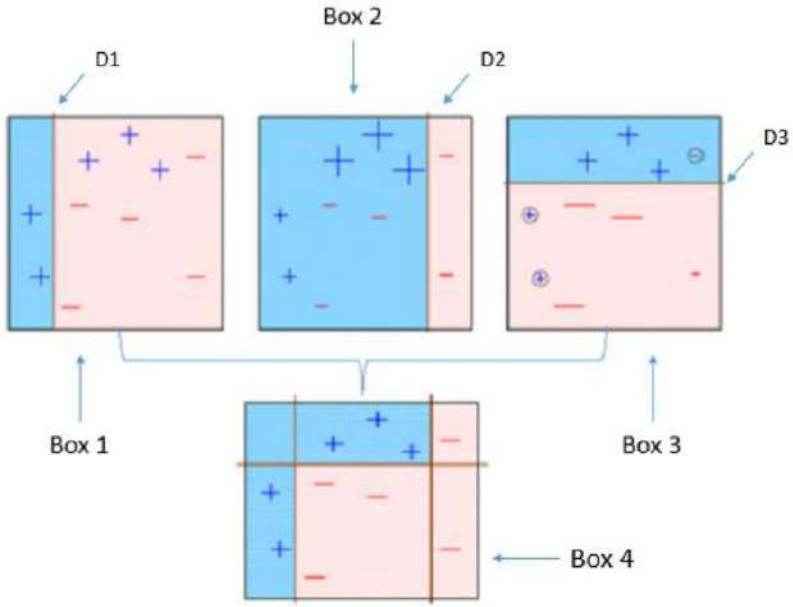
Why is it so good ?

- **Parallel Computing**: It is enabled with parallel processing (using OpenMP); i.e., when you run xgboost, by default, it would use all the cores of your laptop/machine.
- **Regularization**: Regularization is a technique used to avoid overfitting in linear and tree-based models.
- **Enabled Cross Validation**: XGboost is enabled with internal CV function.
- **Missing Values**: XGBoost is designed to handle missing values internally.

INNOMATICS RESEARCH LABS

How does XGBoost work ?

- Boosting is a sequential process; i.e., trees are grown using the information from a previously grown tree one after the other. This process slowly learns from data and tries to improve its prediction in subsequent iterations.



INNOMATICS RESEARCH LABS

XGBoost Tuning Parameters

- Choose a relatively high learning rate to start with.
- Tune tree-specific parameters (max_depth, min_child_weight, gamma, subsample, colsample_bytree) for decided learning rate and number of trees
- Tune regularization parameters (lambda, alpha) for xgboost which can help reduce model complexity and enhance performance.
- Lower the learning rate and decide the optimal parameters