



Inspire...Educate...Transform.

XGBOOST

XGBOOST Advantages

- Speed - supports parallel computing
- Input types - Accepts various inputs like dense matrix, sparse matrix, data, xgb.DMatrix
- Flexible - supports regression, classification, ranking and also user defined objective functions for model evaluation.
- Regularization - Helps in avoiding overfitting
- Missing value treatment - detects patterns in data and treats missing values.
- Uses pruning.
- Availability in Many languages



XGBOOST Approach

- Classification Problems
- Regression Problems



Hyper Parameters

1. General Hyperparameters

2. Booster Hyperparameters

- a. Hyperparameters for tree boosting
- b. Hyperparameters for linear boosting



General Parameters

1. Booster

- Sets the booster type (gmtree, gblinear etc) to use.
For classification problems, you can use gmtree. For regression, you can use any.

1. nthread

- Activates parallel computation.

1. silent

- To show the log messages on screen

Parameter for Tree Booster

- **eta** - controls learning rate. $[0,1]$. default = 0.3.
- **nrounds** - controls maximum number of iterations. (For a lower eta increase nrounds).
- **gamma** - controls the loss reduction required (regularization). More the gamma conservative the model. $[0,\text{Inf}]$ default = 0.
- **max_depth** - controls depth of tree. $[0,\text{Inf}]$ default = 6



Parameter for Tree Booster

- **min_child_weight** - minimum sum of instance weights in child node. Prevents overfitting. $[0, \text{Inf}]$ default = 1
- **subsample** - Number of samples supplied to tree. $[0, 1]$ default = 1. Typical range $[0.5-0.8]$
- **colsample_bytree** - controls number of features supplied to the tree. $[0, 1]$ default = 1, typical range $[0.5, 0.9]$
- **lambda** - controls L2 regularization
- **alpha** - controls L1 regularization



Parameter for Linear Booster

- **nrounds** - controls maximum number of iterations for gradient decent.
- **lambda** - controls L2 regularization. default = 0
- **alpha** - controls L1 regularization. default = 1



Learning Task Parameters

- **Objective**

- reg:linear - linear regression (default)
- reg:logistic - logistic regression
- multi:softmax - multiclass classification

- **eval_metric** - metric for validation of data.

- MAE - (Regression)
- Logloss - (Classification)
- AUC - (Classification)
- RMSE - (Regression)
- error - (Binary Classification)
- mlogloss - (Multiclass classification)

