
# Modèles Boosting Comparaison

# 5. Modèles – Boosting - Comparaison

| Function | XGBoost | CatBoost | Light GBM |
|---|---|---|---|
| Important parameters which control overfitting | 1. **learning_rate or eta** – optimal values lie between 0.01-0.2<br>2. **max_depth**<br>3. **min_child_weight:** similar to min_child leaf; default is 1 | 1. **Learning_rate**<br>2. **Depth** - value can be any integer up to 16. Recommended - [1 to 10]<br>3. No such feature like min_child_weight<br>4. **l2-leaf-reg**: L2 regularization coefficient. Used for leaf value calculation (any positive integer allowed) | 1. **learning_rate**<br>2. **max_depth**: default is 20. Important to note that tree still grows leaf-wise. Hence it is important to tune **num_leaves** (number of leaves in a tree) which should be smaller than 2^(max_depth). It is a very important parameter for LGBM<br>3. **min_data_in_leaf**: default=20, alias= min_data, min_child_samples |
| Parameters for categorical values | Not Available | 1. **cat_features**: It denotes the index of categorical features<br>2. **one_hot_max_size**: Use one-hot encoding for all features with number of different values less than or equal to the given parameter value (max – 255) | 1. **categorical_feature**: specify the categorical features we want to use for training our model |
| Parameters for controlling speed | 1. **colsample_bytree**: subsample ratio of columns<br>2. **subsample**: subsample ratio of the training instance<br>3. **n_estimators**: maximum number of decision trees; high value can lead to overfitting | 1. **rsm**: Random subspace method. The percentage of features to use at each split selection<br>2. No such parameter to subset data<br>3. **iterations**: maximum number of trees that can be built; high value can lead to overfitting | 1. **feature_fraction:** fraction of features to be taken for each iteration<br>2. **bagging_fraction:** data to be used for each iteration and is generally used to speed up the training and avoid overfitting<br>3. **num_iterations:** number of boosting iterations to be performed; default=100 |