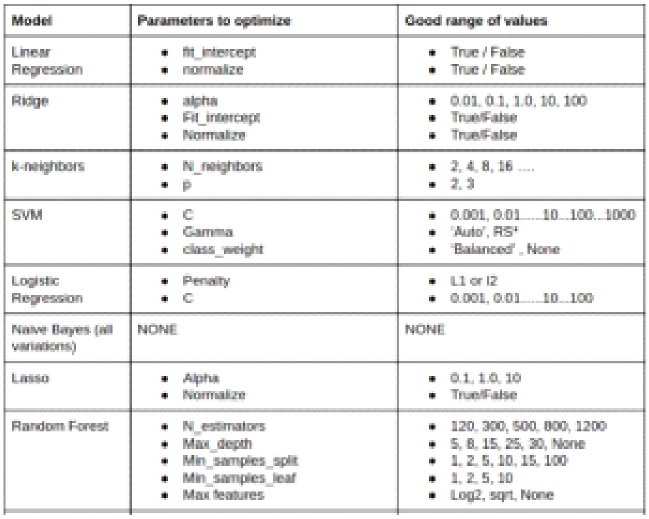
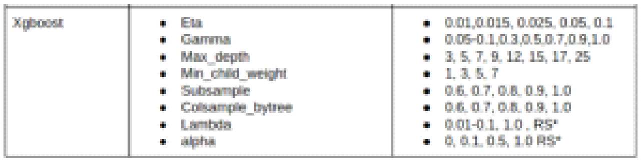
hyperparameters

Sunday, September 30, 2018 11:27 AM



Model	Parameters to optimize	Good range of values
Linear Regression	• fit_intercept • normalize	• True / False • True / False
Ridge	• alpha • Fit_intercept • Normalize	• 0.01, 0.1, 1.0, 10, 100 • True/False • True/False
k-neighbors	• N_neighbors • p	• 2, 4, 8, 16 .... • 2, 3
SVM	• C • Gamma • class_weight	• 0.001, 0.01.....10...100...1000 • 'Auto', RS* • 'Balanced' , None
Logistic Regression	• Penalty • C	• L1 or l2 • 0.001, 0.01.....10...100
Naive Bayes (all variations)	NONE	NONE
Lasso	• Alpha • Normalize	• 0.1, 1.0, 10 • True/False
Random Forest	• N_estimators • Max_depth • Min_samples_split • Min_samples_leaf • Max features	• 120, 300, 500, 800, 1200 • 5, 8, 15, 25, 30, None • 1, 2, 5, 10, 15, 100 • 1, 2, 5, 10 • Log2, sqrt, None



Xgboost	• Eta • Gamma • Max_depth • Min_child_weight • Subsample • Colsample_bytree • Lambda • alpha	• 0.01,0.015, 0.025, 0.05, 0.1 • 0.05-0.1,0.3,0.5,0.7,0.9,1.0 • 3, 5, 7, 9, 12, 15, 17, 25 • 1, 3, 5, 7 • 0.6, 0.7, 0.8, 0.9, 1.0 • 0.6, 0.7, 0.8, 0.9, 1.0 • 0.01-0.1, 1.0 , RS* • 0, 0.1, 0.5, 1.0 RS*

XG boost hyperparameters

Depth of tree:--imp as it results in complexity. 5 is better not more
Learning rate:--shrinkage parameter..learning the right weights
Max_delta_step :(xg boost handles one vs all which is beneficial, max_delta_step helps in imbalanced data sets)
No of estimators :- no of boosting rounds)
Booster--mostly tree based used
Base_score :--> not needed in parameterization
Ramdom_state ==>
Missing (specify if missing denotes some value)

Sample size
No of columns
Regularization params :-
lamda (L2 regularizn)
Alpha L1 regularizn
Gamma L0 regularizn
Gamma L0
Leads weights
Min_child weight :--bases on which a tree can be made

Random Forest
n_estiomators = number of trees in the foreset

- max_features = max number of features considered for splitting a node
- max_depth = max number of levels in each decision tree
- min_samples_split = min number of data points placed in a node before the node is split
- min_samples_leaf = min number of data points allowed in a leaf node
- bootstrap = method for sampling data points (with or without replacement)

From <https://towardsdatascience.com/hyperparameter-tuning-the-random-forest-in-python-using-scikit-learn-28d2aa77dd74>