Bagging and Boosting

Machine Learning Practice

Dr. Ashish Tendulkar

IIT Madras

Part 2: Boosting

There are two boosting estimators:

- AdaBoost estimator
- Gradient boosting estimator

AdaBoost estimator

Class: `sklearn.ensemble.AdaBoostClassifier`

Class: `sklearn.ensemble.AdaBoostRegressor`

Class: sklearn.ensemble.AdaBoostClassifier

base_estimator

- Default estimator is DecisionTreeClassifier with depth = 1.

n_estimators

- Maximum number of estimators where boosting is terminated. The default value is 50.

learning_rate

- Weight applied to each classifier during boosting.
- Higher value here would increase contribution of individual classifiers.
- There is a trade-off between n_estimators and learning_rate.

Class: sklearn.ensemble.AdaBoostRegressor

base_estimator

- Default estimator is DecisionTreeRegressor with depth = 3.

n_estimators

- Maximum number of estimators where boosting is terminated. The default value is 50.

learning_rate

- Weight applied to each regressor at each boosting iteration.
- Higher value here would increase contribution of individual regressor.
- There is a trade-off between n_estimators and learning_rate.

The main parameters to tune to obtain good results are

- **n_estimators** and
- Complexity of the base estimators (e.g. its depth **max_depth** or **min_samples_split**).

Attributes of AdaBoost estimators

`base_estimator_`	Base estimator of ensemble.
`estimators_`	Collection of fitted sub-estimators.
`estimator_weights_`	Weights for each estimator in ensemble.
`estimator_errors_`	Errors for each estimator in ensemble.

Gradient boosting estimators

Class: `sklearn.ensemble.GradientBoostingClassifier`

Class: `sklearn.ensemble.GradientBoostingRegressor`

There are two most important parameters of these estimators:

- `n_estimators`
- `learning_rates`

`sklearn.ensemble.GradientBoostingClassifier` supports both binary and multiclass classification.

We will directly demonstrate XGBoost through colab demonstration.