

Estimator Configuration

ML Advanced Configuration



Hayley Bresina
One AI Client Enablement



03 Jul 2024

Topics Covered

- Overview of estimators
- Estimators available in One AI
- The default estimator configuration settings in One AI
- How to configure estimators for One AI models

Learning Outcomes

You will:

- Understand the role & types of estimators in machine learning tasks, enabling you to select the most suitable estimators for each model
- Gain the ability to configure estimators in One AI, including how to enable & adjust default & optional estimators for specific models
- Adjust estimator parameters & hyperparameters to fine-tune & optimize model performance, understanding the impact of these adjustments on the accuracy & reliability of predictions



Overview of Estimators



Estimators

- Model or tool that learns from data to make predictions or transform data
 - Set of rules that takes input data (OM data) & produces outputs (predictions)
- Related to, but different than an algorithm
 - **Estimator:** like a predictive model that learns from data to make predictions
 - **Algorithm:** set of rules or steps that the estimator follows to learn from the data; method used to train the estimator
- Term originates from statistics
- Estimators can be used for various tasks:
 - Classifications or regressions
 - Clustering
 - Data preprocessing



Estimators Available in One AI



Classification Estimators

- **AdaBoostClassifier** - ensemble method that combines multiple weak classifiers
- **DecisionTreeClassifier** - splits data into branches at each node based on feature values
- **KNeighborsClassifier** - classifies new cases based on a similarity measure
- **LightGBMClassifier** - builds an ensemble of decision trees sequentially where each new tree corrects errors of previous trees
- **LogisticRegression** - linear model that estimates the probability of a binary outcome using a logistic function
- **RandomForestClassifier** - ensemble learning method that constructs multiple decision trees during training and merges their results
- **Support Vector Classifier (SVC)** - finds the decision boundary that best separates classes

Regression Estimators

- **AdaBoostRegressor, DecisionTreeRegressor, LightGBMRegressor, RandomForestRegressor & Support Vector Regressor (SVR)**
 - Function essentially the same as their classifier counterparts
- **Lasso (L1)** - selects the simpler model to only include the most important predictors
- **Ridge (L2)** - keeps all predictors in the model & effective in handling multicollinearity
- **ElasticNet** - combines the properties of Lasso and Ridge to enhance accuracy & interpretability
- **GaussianProcessRegressor** - provides predictions along with a measure of uncertainty
- **HuberRegressor** - gives less weight to outliers, reducing their impact on the model
- **LinearRegression** - fits a linear equation to the data; simple technique
- **StochasticGradientDescentRegressor (SGD)** - uses the entire dataset to calculate the gradients & updates the model parameters using one training example at a time



Default Settings in One AI



Default Classification Estimators in One AI

Default classifiers:

- AdaBoostClassifier
- LightGBMClassifier
- LogisticRegression
- RandomForestClassifier

Optional classifiers:

- DecisionTreeClassifier
- KNeighborsClassifier
- SVC

Default Regression Estimator in One AI

Default regressors:

- ElasticNet
- Lasso
- LightGBMRegressor
- LinearRegression
- RandomForestRegressor

Optional regressors:

- AdaBoostRegression
- DecisionTreeRegressor
- GaussianProcessRegressor
- HuberRegressor
- Ridge
- SGDRegressor
- SVR

Check out the "One AI Machine Learning Algorithms and Settings" help article for each estimator's default parameter & hyperparameter configuration



Estimator Configuration in One AI





Thanks for watching!

