
H2O.ai

Python Interface Overview

Action	Pandas or scikit-learn	H2O
Reading data	`pandas.read_csv(data_path)`	`h2o.import_file(data_path)`
Summarizing data	`pandas_frame.describe()`	`h2o_frame.describe()`
Summary statistics	`pandas_frame.mean()`	`h2o_frame.mean()`
Combining rows	`pandas.concat(list[frame1,frame2])`	`h2o_frame.rbind(h2o_frame2)`
Combining columns	`pandas.concat(list[frame1,frame2],axis = 1)`	`h2o_frame.cbind(h2o_frame2)`
Data selection	`pandas_frame[:, :]`	`h2o_frame[:, :]`
Transforming columns	`np.log(pandas_frame[x])` `np.sqrt(pandas_frame[x])`	`h2o_frame[x].log()` `h2o_frame[x].sqrt()`
Building Random Forest	`model = RandomForestClassifier(n_estimators = 100)` `model = model.fit(x_frame, y_frame)`	`model = H2ORandomForestClassifier(n_trees = 100)` `model = model.train(x, y, train_frame)`
Model Prediction	`model.predict`	`model.predict`
Model Metrics	`metrics.auc`	`metrics = model.model_performance(frame)` `metrics.auc()`

Reading Data into H2O with Python

STEP 1

Python user

```
h2o_df = h2o.import_file("../data/allyears2k.csv")
```

Python Interface Overview

Action	Pandas or scikit-learn	H2O
Reading data	`pandas.read_csv(data_path)`	`h2o.import_file(data_path)`
Summarizing data	`pandas_frame.describe()`	`h2o_frame.describe()`
Summary statistics	`pandas_frame.mean()`	`h2o_frame.mean()`
Combining rows	`pandas.concat(list[frame1,frame2])`	`h2o_frame.rbind(h2o_frame2)`
Combining columns	`pandas.concat(list[frame1,frame2],axis = 1)`	`h2o_frame.cbind(h2o_frame2)`
Data selection	`pandas_frame[:, :]`	`h2o_frame[:, :]`
Transforming columns	`np.log(pandas_frame[x])` `np.sqrt(pandas_frame[x])`	`h2o_frame[x].log()` `h2o_frame[x].sqrt()`
Building Random Forest	`model = RandomForestClassifier(n_estimators = 100)` `model = model.fit(x_frame, y_frame)`	`model = H2ORandomForestClassifier(n_trees = 100)` `model = model.train(x, y, train_frame)`
Model Prediction	`model.predict`	`model.predict`
Model Metrics	`metrics.auc`	`metrics = model.model_performance(frame)` `metrics.auc()`