R Interface Overview

Action	R	H2O
Reading data	`read_csv(data_path)`	`h2o.importFile(data_path)`
Summarizing data	`summary(data_frame)`	`h2o.summary(h2o_frame)`
Summary statistics	`mean(data_frame[["x"]])`	`h2o.mean(h2o_frame)`
Combining rows	`rbind(data_frame1, data_frame2)`	`h2o.rbind(h2o_frame1, h2o_frame2)`
Combining columns	`cbind(data_frame1, data_frame2)`	`h2o.cbind(h2o_frame1, h2o_frame2)`
Data selection	`data_frame[,]`	`h2o_frame[,]`
Transforming columns	`log(data_frame[,"x"])` `sqrt(data_frame[,"x"])`	`log(h2o_frame[,"x"])` `sqrt(h2o_frame[,"x"])`
Building Random Forest	`model = randomForest(y ~ x, data_frame)`	`model = h2o.randomForest(x, y, train_frame)`
Model Prediction	`predict(model, data_frame)`	`h2o.predict(model, h2o_frame)`
Model Metrics	`performance(model)` `auc(model)`	`metrics = model.model_performance(frame)` `h2o.auc(model)`

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()`

H₂O.ai

R Interface Overview

Action	R	H2O
Reading data	`read_csv(data_path)`	`h2o.importFile(data_path)`
Summarizing data	`summary(data_frame)`	`h2o.summary(h2o_frame)`
Summary statistics	`mean(data_frame[["x"]])`	`h2o.mean(h2o_frame)`
Combining rows	`rbind(data_frame1, data_frame2)`	`h2o.rbind(h2o_frame1, h2o_frame2)`
Combining columns	`cbind(data_frame1, data_frame2)`	`h2o.cbind(h2o_frame1, h2o_frame2)`
Data selection	`data_frame[,]`	`h2o_frame[,]`
Transforming columns	`log(data_frame[,"x"])` `sqrt(data_frame[,"x"])`	`log(h2o_frame[,"x"])` `sqrt(h2o_frame[,"x"])`
Building Random Forest	`model = randomForest(y ~ x, data_frame)`	`model = h2o.randomForest(x, y, train_frame)`
Model Prediction	`predict(model, data_frame)`	`h2o.predict(model, h2o_frame)`
Model Metrics	`performance(model)` `auc(model)`	`metrics = model.model_performance(frame)` `h2o.auc(model)`

H₂O.ai