MicrosoftML: Algorithm Cheat Sheet for R / Python

This cheat sheet helps you choose the best MicrosoftML algorithm for your predictive analytics solution. Your decision is driven by both the nature of your data and the question you're trying to answer.

START

Predicting categories

TWO-CLASS CLASSIFICATION (type = "binary")

rxFastTrees() / rx_fast_trees()	Accuracy, fast training, large memory footprint
rxFastForest() / rx_fast_forest()	Fast training, large memory footprint
rxFastLinear() /rx_fast_linear()	Linear model, fast training large data sets
rxNeuralNet() / rx_neural_net()	Accuracy, long training time, large data sets
rxLogisticRegression() / rx_logistic_regression()	Linear model, fast training large data sets
rxOneClassSvm() / rx_one_class_svm()	Highly unbalanced classes

Two

Predicting values

REGRESSION (type = "regression")

Accuracy, fast training, large memory footprint	**rxFastTrees() / rx_fast_trees()**
Fast training, large memory footprint	**rxFastForest() / rx_fast_forest()**
Linear model, fast training, large data sets	**rxFastLinear() /rx_fast_linear()**
Accuracy, long training time, large data sets	**rxNeuralNet() / rx_neural_net()**

Finding unusual data points

MULTI-CLASS CLASSIFICATION (type="multiClass")

rxNeuralNet() / rx_neural_net()	Accuracy, long training time, large data sets
rxLogisticRegression() / rx_logistic_regression()	Linear model, fast training, large data sets

Three or more

ANOMALY DETECTION

>100 features, aggressive boundary	**rxOneClassSvm() / rx_one_class_svm()**

Microsoft