Algorithm Cheat Sheet

BecomingHuman.AI

This cheat sheet helps you choose the best Azure Machine Learning Studio 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

CLUSTERING

Discovering structure

- K-means

ANOMALY DETECTION

Finding unusual data points

- One-class SVM — >100 features, aggressive boundary
- PCA-based anomaly detection — Fast training

REGRESSION

Predicting values

- Ordinal regression — Data in rank ordered categories
- Poisson regression — Predicting event counts
- Fast forest quantile regression — Predicting a distribution
- Linear regression — Fast training, linear model
- Bayesian linear regression — Linear model, small data sets
- Neural network regression — Accuracy, long training time
- Decision forest regression — Accuracy, fast training
- Boosted decision tree regression — Accuracy, fast training

MULTICLASS CLASSIFICATION

Predicting categories — Three or more

- Fast training, linear model — Multiclass logistic regression
- Accuracy, long training times — Multiclass neural network
- Accuracy, fast training — Multiclass decision forest
- Accuracy, small memory footprint — Multiclass decision jungle
- Depends on the two-class classifier, see notes below — One-v-all multiclass

TWO CLASS CLASSIFICATION

Two

- >100 features, linear model — Two-class SVM
- Fast training, linear model — Two-class averaged perceptron
- Fast training, linear model — Two-class logistic regression
- Fast training, linear model — Two-class Bayes point machine
- Accuracy, fast training — Two-class decision forest
- Accuracy, fast training — Two-class boosted decision tree
- Accuracy, small memory footprint — Two-class decision jungle
- >100 features — Two-class locally deep SVM
- Accuracy, long training times — Two-class neural network