Microsoft Azure
Machine Learning Algorithm Cheat Sheet

This cheat sheet helps you choose the best machine learning algorithm for your predictive analytics solution. Your decision is driven by both the nature of your data and the goal you want to achieve with your data.

What do you want to do?

Extract information from text

Text Analytics

Derives high-quality information from text
Answers questions like: What info is in this text?

Extract N-Gram Features from Text	Creates a dictionary of n-grams from a column of free text
Feature Hashing	Converts text data to integer encoded features using the Vowpal Wabbit library
Preprocess Text	Performs cleaning operations on text, like removal of stop-words, case normalization
Word2Vector	Converts words to values for use in NLP tasks, like recommender, named entity recognition, machine translation

Predict between several categories

Multiclass Classification

Answers complex questions with multiple possible answers
Answers questions like: Is this A or B or C or D?

Multiclass Logistic Regression	Fast training times, linear model
Multiclass Neural Network	Accuracy, long training times
Multiclass Decision Forest	Accuracy, fast training times
One-vs-All Multiclass	Depends on the two-class classifier
Multiclass Boosted Decision Tree	Non-parametric, fast training times and scalable

Generate recommendations

Recommenders

Predicts what someone will be interested in
Answers the question: What will they be interested in?

SVD Recommender	Collaborative filtering, better performance with lower cost by reducing dimensionality

Predict between two categories

Regression

Makes forecasts by estimating the relationship between values
Answers questions like: How much or how many?

Fast Forest Quantile Regression	Predicts a distribution
Poisson Regression	Predicts event counts
Linear Regression	Fast training, linear model
Bayesian Linear Regression	Linear model, small data sets
Decision Forest Regression	Accurate, fast training times
Neural Network Regression	Accurate, long training times
Boosted Decision Tree Regression	Accurate, fast training times, large memory footprint

Predict values

Discover structure

Clustering

Separates similar data points into intuitive groups
Answers questions like: How is this organized?

K-Means	Unsupervised learning

Two-Class Classification

Answers simple two-choice questions, like yes or no, true or false
Answers questions like: Is this A or B?

Two-Class Support Vector Machine	Under 100 features, linear model
Two-Class Averaged Perceptron	Fast training, linear model
Two-Class Decision Forest	Accurate, fast training
Two-Class Logistic Regression	Fast training, linear model
Two-Class Boosted Decision Tree	Accurate, fast training, large memory footprint
Two-Class Neural Network	Accurate, long training times

Find unusual occurrences

Anomaly Detection

Identifies and predicts rare or unusual data points
Answers the question: Is this weird?

One Class SVM	Under 100 features, aggressive boundary
PCA-Based Anomaly Detection	Fast training times

Classify images

Image Classification

Classifies images with popular networks
Answers questions like: What does this image represent?

DenseNet	High accuracy, better efficiency

Share this poster: aka.ms/mlcheatsheet

Microsoft