Microsoft Azure Machine Learning: Algorithm Cheat Sheet

START

CLUSTERING

K-means

No ← Predict future data points?

Yes ↓

Categories →

Predict categories or values?

ANOMALY DETECTION

Select one:
One-class SVM
PCA-based anomaly detection

Yes ←

Values ↓

REGRESSION

Ordinal regression ← Yes

Data in rank-ordered categories?

No ↓

Poisson regression ← Yes

Predict event counts?

No ↓

Fast forest quantile regression ← Distribution

Predict single values or a distribution?

Single values ↓

Linear regression ← Close enough

Linear approximation okay?

If your data points are statistically independent, try:

Bayesian linear regression

If you prefer performance over training time, and all features are numerical, try:

No → **Decision forest regression**

If you have overlapping features, try:

Boosted decision tree regression

Neural network regression

MULTI-CLASS CLASSIFICATION

>2 →

Two categories or more than two?

No → Prefer classifier built from >1 two-class classifiers?

No → Classes can be separated by a hyperplane?

No → Select one:
Multiclass decision forest
Muilticlass decision jungle

If you prefer performance over training time, and all features are numerical, try:

Multiclass neural network

2 ↓

One of the categories rare?

Yes ↓

One-v-All multiclass

Yes ↓

Multiclass logistic regression

No ↓

Choose a two-class classifier

TWO-CLASS CLASSIFICATION

Speed →

Prefer speed or accuracy?

Classes can be separated by a hyperplane?

Yes → **Two-class SVM**

No → **Two-class averaged perceptron**

If the accuracy is good but you want it faster, try:

Locally Deep SVM

Accuracy ↓

Classes can be separated by a hyperplane?

Yes → Data points statistically independent?

Yes → **Two-class Bayes point machine**

No → **Two-class logistic regression**

Not necessarily ↓

Select one:
Two-class boosted decision tree
Two-class decision forest
Two-class decision jungle

If you prefer performance over training time, and all features are numerical, try:

Two-class neural network

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.

Microsoft