Microsoft Azure Machine Learning: 日本語版アルゴリズムチートシート

CLUSTERING（クラスタリング）

スタート

未来のデータを予測しますか？

いいえ → K-means

はい → 分類または値を予測しますか？

REGRESSION（回帰）

分類または値を予測しますか？ → 値

データをランク順のカテゴリーで予測しますか？

はい → Ordinal regression

いいえ → イベントの回数を予測しますか？

はい → Poisson regression

いいえ → 一つの値または分布を予測しますか？

分布 → Fast forest quantile regression

一つの値 → 線型近似を使用して予測できますか？

おそらく → Linear regression

いいえ → 説明できる級の境界が好ましいですか？

はい → Decision forest regression

特徴が重複する場合は → Boosted decision tree regression

いいえ → Neural network regression

データポイントが統計学的に独立している場合は → Bayesian linear regression

ANOMALY DETECTION（異常値検知）

分類 → 10万以下のデータもしくは100以上の特徴を持ちますか？

はい → One-class SVM

いいえ → PCA-based anomaly

MULTI-CLASS CLASSIFICATION（多項分類）

2つの分類もしくは複数の分類ですか？

複数 → 複数の二項分類を使用して分類しますか

いいえ → 説明できる級の境界が好ましいですか？

はい → いずれかを選択してください：
Multiclass decision forest
Muilticlass decision jungle

いいえ → Multiclass logistic regression

すべての特徴が値で、学習時間よりもパフォーマンスを優先される場合は → Multiclass neural network

2つ → 1つの分類は希薄な分類ですか？

はい → One-v-All multiclass

No → Choose a two-class classifier

TWO-CLASS CLASSIFICATION（二項分類）

10万以下のデータポイントもしくは、100以上の特徴を持ちますか？

はい → Two-class SVM

いいえ → 今の精度は高いけれども、より早く処理したい場合は → Locally Deep SVM

処理スピードまたは精度のどちらを重要視しますか？

処理スピード → Two-class averaged perception

精度 → 説明できる級の境界が好ましいですか？

いいえ → Two-class logistic regression

すべての特徴が値で、学習時間よりもパフォーマンスを優先される場合は → Two-class neural network

はい → データポイントが統計学的に独立している場合は → Two-class Bayes point machine

いずれかを選択してください：
Two-class decision forest
Two-class decision jungle

特徴が重複する場合は → Two-class boosted decision tree

このチートシートは、Azure Machine Learning Studioでアルゴリズムを選択する際に最適なものを選択するために使用します。選択するアルゴリズムは、データの種類や、解決しようとしている問題によって左右されます。

© 2015 Microsoft Corporation. All rights reserved.
Created by the Azure Machine Learning Team
Email: AzurePoster@microsoft.com
和訳元：http://aka.ms/MLCheatSheet
非公式日本語訳 by 吉田の備忘録
Microsoft