

[ご購入はこちら](#)

超定番の 機械学習アルゴリズム

足立 悠

データを活用するに当たり，例えば図1のようなデータ分析の手順があります。ここで(3)データの前処理や(4)モデルの作成，を実現する手段の1つに機械学習があります。機械学習のアルゴリズムは，(1)分析問題の設定と(2)データの理解，の内容によって適したものを選択します。代表的なアルゴリズムについて理解しておく，問題を解決するために適したものを効率良く選択できるでしょう。

機械学習の全体像

機械は基本的に，人間が準備したデータの特徴量から，アルゴリズムに従って規則性(パターン)を見つけてます。この処理は，人間が試行錯誤しながら学習し，知見を獲得する様に近いと感じるでしょう。よって，一連の処理を機械が学習すると考え，機械学習と呼びます。以下に基本用語と機械学習の4つの分類を説明します。

● 基本用語

▶ 特徴量

特徴量とはデータが持つ何らかの特徴を数値化したものです。例えば，商品データの特徴量を考える場合，商品はそれぞれ，寸法や重量，色などの情報を持ちます。これらの情報が特徴量となります。特徴量はデータ項目をそのまま使うこともあれば，既存の項目を処理して新しく作成することもあります。

▶ 目的変数

目的変数とは正解データのことです。例えば，商品データにおいて，各商品の品質が良いまたは悪いを決めるとします。この品質が目的変数であると考えます。目的変数は，良いまたは悪いのように文字列値を取ることもあれば，数値を取ることもあります。

● 教師あり学習…特徴量と目的変数から状態を予測

アルゴリズムに従い過去のデータを元にパターンを学習してモデルを作成します。学習では，モデルの出



図1 データ分析の流れ

アルゴリズムの知識があると，適当なアルゴリズムをすぐを選択できて，問題解決がスピードアップ

力と正解の誤差を小さくするよう，反復処理によってモデルの最適なパラメータを見つけます。そして，そのモデルに新規データを適用して，推論(予測)結果を得ます(図2)。

● 教師なし学習…特徴量から状態を要約する

教師なし学習でも過去のデータを元に，アルゴリズムに従いパターンを学習してモデルを作成します。学習では正解データを持たないため，誤差の代わりとなる指標に基づいてモデルの最適なパラメータを見つけます。

● 深層学習(ディープ・ラーニング)…画像や時系列データなど複雑なデータにも対応できる

機械学習・教師あり学習アルゴリズムの1つであるニューラル・ネットワークが元になっています。ネットワークの中間層を多く持つことで，従来よりも複雑な問題を解ける可能性が高くなります。ディープ・ラーニングはニューラル・ネットワークが元ですが，ネットワーク構造を変化させることで，教師あり学習