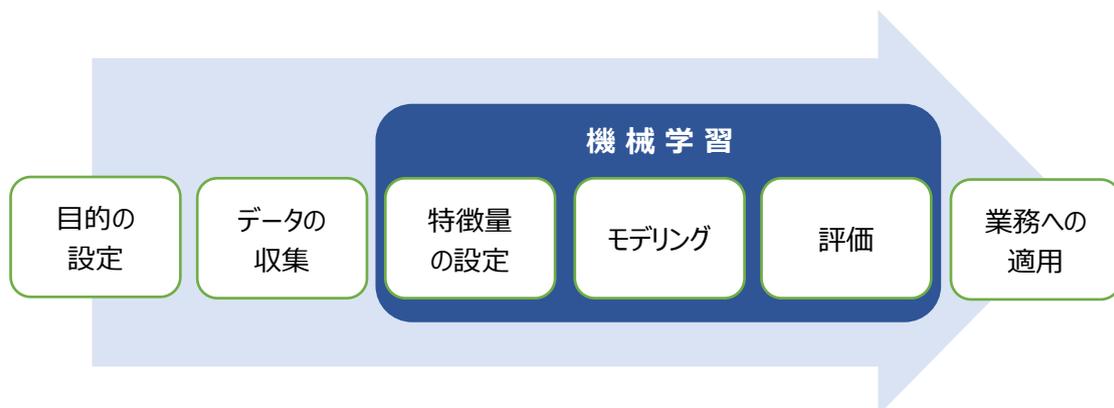


機械学習の仕組みと AI 分析

1. 機械学習を用いたデータ分析のプロセス

【プロセスの全体図】



- (1) 目的の設定 : 機械学習 (モデル) 適用の目的を設定する
- (2) データの収集 : 機械学習に必要なデータを収集する
△の受注データを収集する
- (3) 機械学習 : 機械学習を行うため「特徴量の設定」「モデリング」「モデルの評価」を行う
- (4) 業務への適用 : 機械学習の結果を自裁の業務に適用する

2. 「ぱぱっと AI スコア」における機械学習

【機械学習の仕組み】

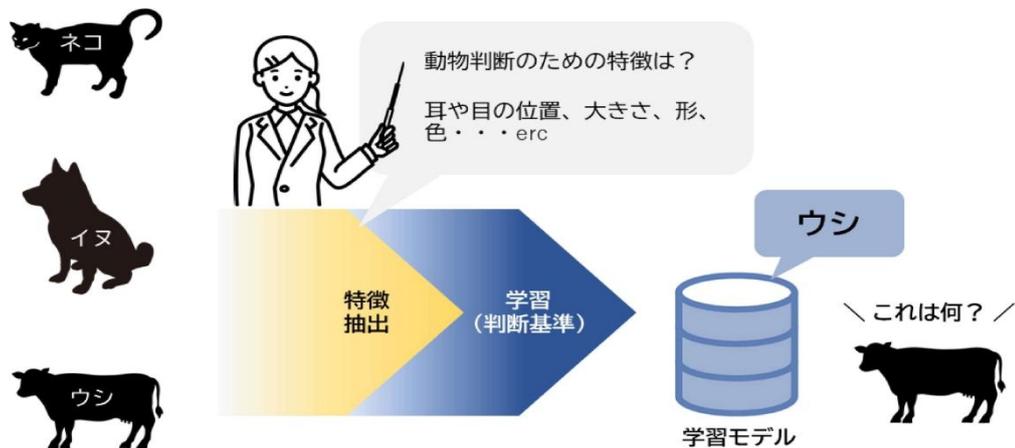
「ぱぱっと AI スコア」では、「機械学習 (教師あり学習) 」を採用しています。

「教師あり機械学習」とは、既知となった過去のデータを「機械学習アルゴリズム」に与えることで、それらを「正解データ」として学習し予測モデルを生成します。

予測モデルは、「特徴量」に基づいて生成されます。

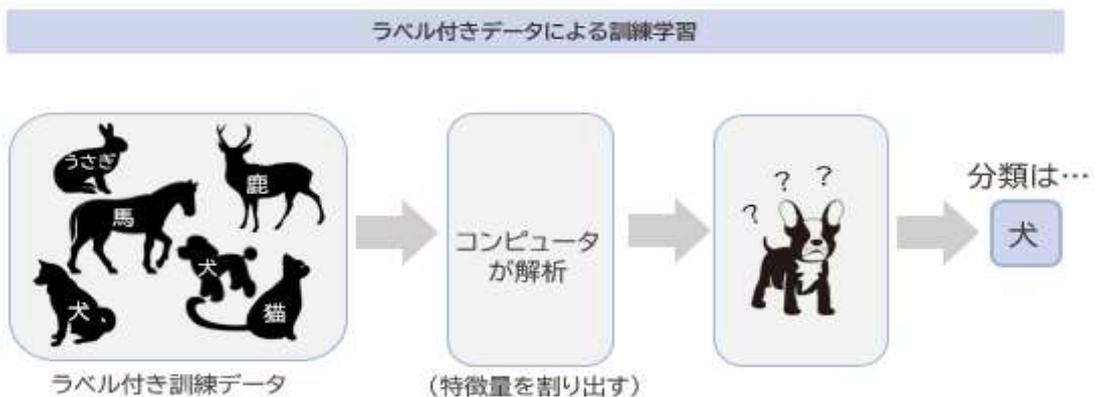
【特徴量】

「特徴量」とは、分析のためにそろえたデータの中で、「機械学習アルゴリズム」に読み込ませるために必要な「データ項目」のことです。例えば、動物の画像を判別する場合で考えると、動物の「耳や目の形や位置・大きさ」などのことを言います。特徴量の「質と量」により、「予測モデルの精度」が大きく左右されます。



【教師あり機械学習】

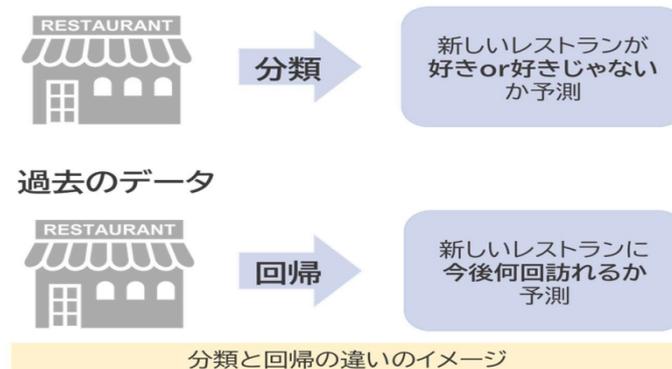
大量の動物の画像データが存在した時、「これは“ネコ”」「これは“イヌ”」・・・といったように、あらかじめ画像に「ラベリング」しておきます。十分な量の正解データを用意し、それらを教師（正解）として機械学習を行います。機械学習の結果、得られた予測モデルを新たな画像に適用することにより、“ネコ” か “イヌ”かを判定することが可能になります。



【分類と回帰】

「教師あり機械学習」は、あらかじめ定めた分類に予測データを振り分ける「分類」と、連続するデータの将来の値を予測する「回帰」のふたつに分かれます。

「ぱぱっと AI スコア」では「分類」の手法を用いて予測を行います。



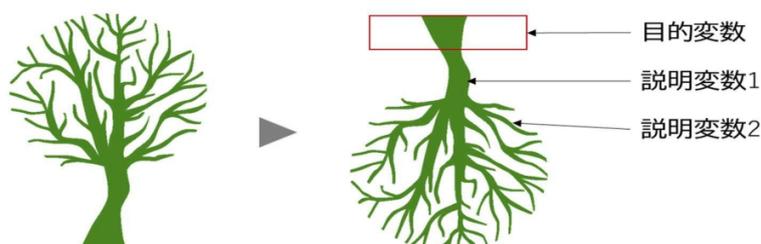
【判別分析】

「分類」の手法を用いて「機械学習」を行う場合、大切なポイントは「判別分析」です。

「判別分析」とは、二つの母集団を設定し、「あるサンプルがどちらの母集団に属するか」を推測するための手法です。母集団への所属があらかじめわかっているサンプルを「教師あり機械学習（判別分析）」することにより、予測対象のデータがどちらの母集団に所属するのかを判別することができます。判別の基準は、「サンプルデータと各母集団との距離」を算出し、その距離の小さい母集団に所属という考えをとります。

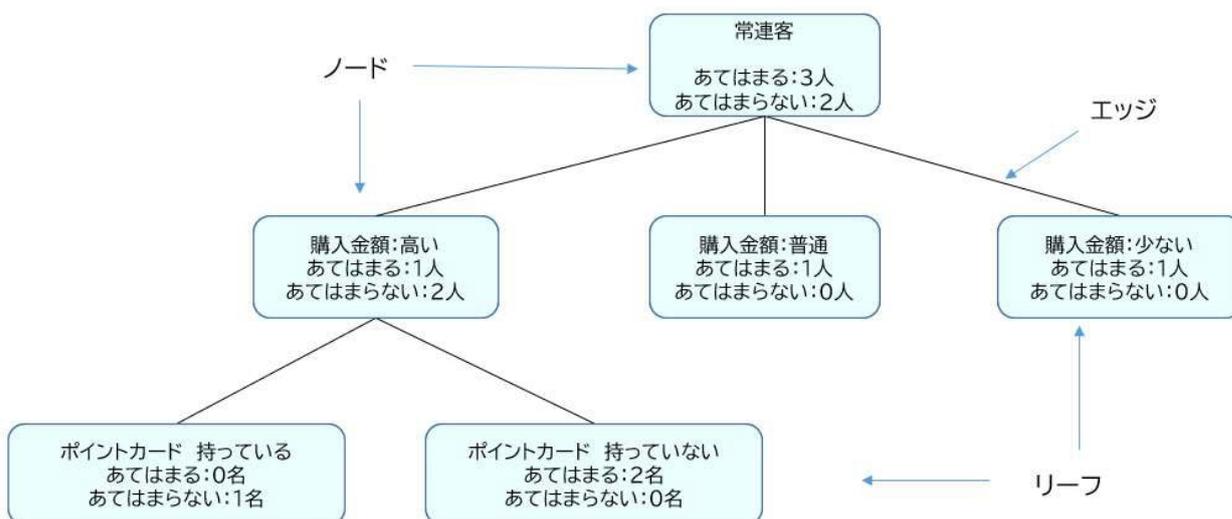
【決定木（デシジョン・ツリー）】

「分類」を多段階で実行する場合の「教師あり機械学習」では「決定木（デシジョン・ツリー）」を活用します。決定木（デシジョン・ツリー）では分類過程を階層化し樹形図で表現します。



【例】常連客の分析

顧客番号	常連客である	購入金額	ポイントカード所持
A	あてはまる	高い	持っている
B	あてはまる	普通	持っていない
C	あてはまらない	普通	持っている
D	あてはまる	高い	持っていない
E	あてはまらない	少ない	持っていない



「ぱぱっと AI スコア」では、「特徴量」ごとに「見込み客」か「見込み客ではない」かを「判別（分類）」することにより、精度の高い「予測モデル」を生成しています。

以 上

株式会社ネットビジネスサポート

出典：総務省 統計局「統計学習の指導のために」

<https://www.stat.go.jp/teacher/comp-learn-04.html>