

1章 機械学習とは

機械学習とは

機械学習とは、機械を使って大量のデータから特性や傾向を見つけ、計算式の塊を作って予測などを行っていく処理のことです。

この計算式の塊を モデル と呼びます。

機械学習の種類

1. 教師あり学習
2. 教師なし学習

1. 教師あり学習

機械学習を行う際に 正解となるラベルデータが存在する 場合に用いられる学習方式です。

この場合のデータは、目的変数（正解ラベル）と 説明変数（目的変数以外のデータ）で構成されています。

教師あり学習の目標は、目的変数を上手く予測できるようなモデルを作り運用していくことです。

教師あり学習には 回帰 と 分類 の二種類があります。

回帰は連続値を予測し、分類はカテゴリを予測します。

例えば、住宅価格や気温を予測するのが回帰で、動物の種類を予測するのが分類です。

（教師あり学習の代表的な手法）

- 線形回帰（回帰）
- 決定木分析（回帰/分類）
- ランダムフォレスト（回帰/分類）
- サポートベクタマシン（回帰/分類）

2. 教師なし学習

正解となるラベルデータが存在しない場合に用いられる学習方式です。

教師なし学習の目標は、データ間の特徴や関係性を調べることです。

クラスタリングや次元削減といった分析を行います。

クラスタリングはデータをグループ分けしてクラスタを作り、クラスタ内の相関などを確認することができます。

次元削減は、データの情報量を保ったまま、説明変数の数を減らすことです。

（教師なし学習の代表的な手法）

- k近傍法（クラスタリング）
- 主成分分析（次元削減）

機械学習の処理の手順

1. データ入手
2. データの前処理
3. データを詳しく見る
4. アルゴリズムを選ぶ
5. データを学習させる（モデルを作る）
6. 精度を確かめる
7. 試験運用（新たなデータ）
8. 実務に応用