

機械学習の整理帳

tomixy

2025 年 6 月 20 日

目次

| | | |
|-------|-----------------------------|---|
| 第 1 章 | 機械学習とは何か | 2 |
| | 人工知能と機械学習 | 2 |
| | 意思決定のプロセス | 2 |
| | モデルとアルゴリズム | 3 |
| | データと特徴量 | 4 |
| | 予測とラベル | 4 |
| | ラベル付きデータとラベルなしデータ | 5 |
| 第 2 章 | 教師あり学習の概要 | 6 |
| | 教師あり学習 | 6 |
| | 回帰モデルと分類モデル | 6 |

第 1 章


機械学習とは何か




人工知能と機械学習

人工知能は包括的な用語

ref: なっとく！機械学習 p4～6

 人工知能 コンピュータが決定を下すことができるすべてのタスクを集めたもの

機械学習は人工知能の一部

 機械学習 コンピュータが「データに基づいて」決定を下すことができるすべてのタスクを集めたもの

データとは、「経験」を表すコンピュータ用語



意思決定のプロセス

ref: なっとく！機械学習 p8～9、p15

経験に基づいて意思決定を行うために人間が用いるプロセスは**記憶・定式化・予測フレームワーク**と呼ばれ、次の 3 つのステップで構成されている

1. 記憶：過去の同じような状況を思い出す
2. 定式化：全般的なルールを定式化する
3. 予測：このルールを使って将来起こるかもしれないことを予測する

コンピュータに「記憶・定式化・予測」フレームワークを使わせることで、コンピュータに私たちと同じように考えさせることができる


1. 記憶：巨大なデータテーブルを調べる
2. 定式化：さまざまなルールや式を調べてデータに最適な**モデル**を作成する
3. 予測：モデルを使って未来（未知）のデータについて予測を行う



モデルとアルゴリズム

コンピュータはデータを使って**モデル**を構築するという方法で問題を解く

ref: なっとく！機械学習 p9、p15～16

 **モデル** データを表すルールの集まりであり、予測を行うために使うことができる

モデルは、次のようなものと考えることができる


既存のデータをできる限り厳密に模倣する一連のルールを使って現実を表すもの

そして、最適なモデルとは、次のようなものである

新しいデータに最もうまく**汎化**するもの

最適なモデルを構築するためのさまざまなアルゴリズムがある

アルゴリズムは、モデルを構築するために使ったプロセスのこと

 **アルゴリズム** 問題を解いたり計算を行ったりするために使われる手続き（一連のステップ）


データと特徴量

データがテーブルに含まれている場合、各行はデータ点である

たとえば、動物のデータセットがある場合、各行は異なる動物を表している

このテーブル内の各動物は、その動物の**特徴量**によって説明される

ref: なっとく！機械学習 p13、p19

 **特徴量** モデルが予測を行うために使うことができるデータの特性や属性

データがテーブルに含まれている場合、特徴量はテーブルの列であり、特徴量は各データを説明する

予測とラベル

特徴量の中には、**ラベル**と呼ばれる特別なものがある

一般に、特定の特徴量を他の特徴量に基づいて予測しようとしているなら、その特徴量はラベルである

ref: なっとく！機械学習 p19～20

機械学習モデルの目標は、

データに含まれているラベルを推測すること

であり、モデルが行う推測を予測と呼ぶ



ラベル付きデータとラベルなしデータ

データには、大きく分けて、

ref: なっとく！機械学習 p20～21

- ラベル付きデータ：ラベルが付いているデータ
- ラベルなしデータ：ラベルが付いていないデータ

の 2 種類がある

予測したいと思うような列を持たないデータセットは、ラベルなしデータである

ラベル付きデータとラベルなしデータは、教師あり学習と教師なし学習という 2 種類の機械学習を生み出している

第 2 章

教師あり学習の概要



教師あり学習

教師あり学習は、ラベル付きデータを扱う機械学習であり、その目標はラベルを予測すること

ref: なっとく！機械学習 p21

ラベルが付いていない新しいデータが渡された場合、教師あり学習モデルはそのデータ点のラベルを予測する

意思決定を行うためのフレームワーク「記憶・定式化・予測」は、教師あり学習の仕組みそのもの

1. データセットを記憶する
2. 特徴と考えられるものをモデル（ルール）として定式化する
3. 新しいデータが与えられたときに、そのデータのラベルを予測する



回帰モデルと分類モデル

教師あり学習モデルでは、数値と状態の 2 種類のデータが使われる

ref: なっとく！機械学習 p22～23

- **数値データ** : 数値を用いるあらゆる種類のデータ
- **カテゴリ値データ** : カテゴリ（状態）を用いるあらゆる種類のデータ

そして、この 2 種類のデータから、次の 2 種類の機械学習モデルが生まれた

- **回帰モデル** : 数値データを予測する機械学習モデル
- **分類モデル** : カテゴリ値データを予測する機械学習モデル

回帰モデルは数値のラベルを予測するモデルであり、この数値を特徴量に基づいて予測する

分類モデルは状態の有限集合に含まれている状態を予測するモデルである
(カテゴリ値データでは、各データ点に有限のカテゴリ集合が紐づけられる)