

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	3
	人工知能と機械学習	3
	意思決定のプロセス	3
	モデルとアルゴリズム	4
	データと特徴量	5
	予測とラベル	5
	ラベル付きデータとラベルなしデータ	6
第 2 章	教師あり学習の概要	7
	教師あり学習	7
	回帰モデルと分類モデル	7
第 3 章	教師なし学習の概要	9
	教師なし学習	9
	教師なし学習によるデータの前処理	9
	教師なし学習の種類	10

クラスタリング	10
次元削減	10
行列分解と特異値分解	11
生成学習	11
 第 4 章 強化学習の概要	 12
強化学習	12
 第 5 章 線形回帰	 13
線形回帰	13

第 1 章


機械学習とは何か




人工知能と機械学習

人工知能（AI: artificial intelligence）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（machine learning）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**（model）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 (algorithm) 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

ref: なっとく！機械学習 p13、p19

このテーブル内の各動物は、その動物の**特徴量** (feature) によって説明される

 **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** (label) と呼ばれる特別なものがある

ref: なっとく！機械学習 p19~20

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

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 (supervised learning) 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 p22～25

- **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 (**regression model**) 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 (**classification model**) 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）

第 3 章

教師なし学習の概要



教師なし学習

教師なし学習（**unsupervised learning**）は、ラベルなしデータを扱う機械学習である

ref: なっとく！機械学習 p25～26

ラベル（予測の目的変数または正解値）がないデータから、できるだけ多くの情報を抽出することが目標となる

たとえば、ラベルが付いていない動物の画像のデータセットからは、それぞれの画像が表している動物の種類はわからないため、新しい画像がどの動物なのかを予測することはできない

しかし、2 つの画像が似ているかどうかなど、他にできることがある

つまり、教師なし学習アルゴリズムは、類似性に基づいてデータを分類できるが、それぞれのグループが何を表すのかはわからない



教師なし学習によるデータの前処理

実際には、教師なし学習はラベルが付いている場合でも利用できる

ref: なっとく！機械学習 p26

教師なし学習を使ってデータの**前処理**を行うと、教師あり学習の手法の効果を高めることができる



教師なし学習の種類

教師なし学習には、大きく分けて 3 種類の学習法がある

ref: なっとく！機械学習 p26

- **クラスタリング**: データを類似性に基づいてクラスタに分類する
- **次元削減**: データを単純化し、より少ない特徴量でデータを正確に説明する
- **生成学習**: 既存のデータに似ている新しいデータ点を生成する



クラスタリング

クラスタリング (clustering) は、データセット内の要素を類似性の高いデータ点ごとにクラスタ (グループ) に分割する

ref: なっとく！機械学習 p26~30

特徴量が 3 つを超えると、その次元を可視化できなくなるため、人間がクラスタを目で確認するのは不可能になる

コンピュータを使うことで、巨大なデータセットに対してもクラスタリングを行うことができる



次元削減

[Todo 1:]

ref: なっとく！機械学習 p30~32



行列分解と特異値分解



[Todo 2:]

ref: なっとく！機械学
習 p32~34



生成学習



[Todo 3:]

ref: なっとく！機械学
習 p34

第 4 章

強化学習の概要



強化学習



[Todo 4:]

ref: なっとく！機械学
習 p35～37

第 5 章

線形回帰



線形回帰

できる限り多くのデータ点の近くを通る直線を求めることで、その直線の式を使って新たなデータを大まかに予測することができる
このような手法を線形回帰という

ref: なっとく！機械学習 p40～

.....

Zebra Notes

Type	Number
todo	4