



ゼロからつくる Python 機械学習プログラミング入門

八谷大岳 著

講談社サイエンティフィク (2020), 3,000 円+税, 368p., ISBN: 978-4-06-520612-6



書籍の概要

機械学習を、アルゴリズムをきちんと理解した上でゼロからゴリゴリとプログラミングして実装してみたいと思っている方は少なくないに違いない。本書はその様な野心的な読者の夢を叶えてくれるかもしれない。本書は、回帰分析、分類モデル、カーネルモデル、ニューラルネットワーク、強化学習、教師なし学習など、機械学習の全般的な分析法について、原理的なことから丁寧に説明し、Pythonでの実装を示している。したがって、機械学習の概念を入門書である程度学習した初心者で、本格的な機械学習プログラミングを開始したい人に適している。あるいは、ある程度データ分析を経験している人が、機械学習全体をざっと見通したり、おさらいしたりするのに適している。最初に機械学習の概要を述べ、次に Python の導入方法や基本的な使い方の解説、線形代数や確率統計など数学の復習をした後に、回帰分析、分類モデル、カーネルモデルなど、個々の機械学習法を解説する章が続いている。Python や数学的記述を基本から丁寧に説明する構成のため、機械学習を基礎からしっかり学びたい人に向いている。残念ながら、深層学習については、畳み込みニューラルネットワークなどの最新手法があまり詳しく触れられていないので、詳細を学びたい場合はほかの書籍を当たる必要がある。

本書の特徴は、分かりやすさを優先して数式的な記述や説明を省いたりせず、Python のソースコー

ドについても丁寧に説明されていることである。機械学習の中身やプログラミングの実装に関し、ブラックボックスな部分が少ないという面では、非常にオススメの書籍である。

本書の内容

第 1 章「機械学習とは何か」では、機械学習が盛んになった歴史的経緯と、代表的な機械学習の分類である教師あり学習、教師なし学習、強化学習について概説している。

第 2 章「Python 入門」では、Python について、Anaconda のインストールから、Jupyter Notebook の導入を説明した後、Numpy の array を中心に解説している。

第 3 章「数学のおさらい」は、線形代数・最適化・確率・統計について Python で実装しながら復習する章である。

第 4 章「回帰分析」は、線形回帰とロジスティック回帰について書かれている。線形回帰は残差を反映する平均二乗誤差の最小化問題であり、最小二乗法と呼ばれるが、この辺りの説明と Python による実装がしっかり説明されている。データの標準化の効果、L2 ノルム正則化による外れ値の処理、目的変数が正規分布に従うことの意味など、線形回帰モデルの重要な項目が網羅されている。ロジスティック回帰は一般化線形モデルを用いた説明であり、交差エントロピー損失の最小化問題であるが、高速計

算を達成するために、Python による行列表現で実装している。

第5章「分類」は、教師あり学習において、目的変数がカテゴリカルデータである場合の解説である。線形判別分析 (LDA)、サポートベクトルマシン (SVM)、ナイーブベイズ (NB)、決定木 (CART) について紹介されている。LDA では、分類境界の関数の求め方について、カテゴリ間分散とカテゴリ内分散から構成される相関比の最大化問題として説明している。これを、ラグランジュ未定乗数法による制約付き最適化問題で解くが、これを Python で実装している。SVM の場合、分類境界の関数をマージン (最近傍距離) 最大化問題として解くが、これも、ラグランジュ未定乗数法による制約付き最適化問題として、Python による行列表現で実装している。また、詳細は示さないが、NB や CART についても同様に丁寧に書かれている。

以後、内容の詳細な紹介は避けるが、各話題について数式による丁寧な説明の後、Python による実装が紹介されている。

第6章「カーネルモデル」は、線形モデルを非線形に拡張するアプローチであるカーネルモデルについて取り扱っている。

第7章「ニューラルネットワーク」は、深層学習によるロジスティック回帰モデルと多クラス分類モデルを取り上げている。過学習、ドロップアウト、ミニバッチ、Adam などのトピックも取り上げられ

ている。多クラス分類モデルでは、ソフトマックス関数に基づく手書き文字の分類の例が取り上げられている。

第8章「強化学習」は、教師データを用いない機械学習で、基本的な原理と、Q 学習法の紹介である。

第9章「教師なし学習」では、代表的な主成分分析、因子分析、クラスター分析を取り上げている。

本書籍をだれに薦めるか

本書籍は、機械学習全般について、原理的なことから数式やコードを使ってきちんと理解したいと考えている学生やエンジニアにお薦めである。教養程度のリテラシーレベルを卒業して、本格的にデータ分析を始めたい人に向く書籍である。単に、AI アプリを使って AI を動かしてみたというレベルに飽きたらない人、もっとガチで AI をいじり倒して中がどうなっているのか知りたい、その上で AI を自分でコーディングできるようにになりたいという人はまず読んでみるとよい。

(2021年8月9日受付)

石井一夫 (正会員)
kishii@rs.sus.ac.jp

公立諏訪東京理科大学工学部情報応用工学科教授、久留米大学医学部内科学講座心臓・血管内科部門客員准教授。専門分野：ビッグデータ分析、計算機統計学、データマイニング、数理モデリング、機械学習、人工知能。医療ビッグデータ、気象ビッグデータ研究に従事。2015年度本会優秀教育賞受賞。日本技術士会フェロー、APEC エンジニア、IPEA 国際エンジニア。

