

イラストで学ぶ 機械学習 最小二乗法による識別モデル学習を中心に

杉山 将 著

講談社 (2013 年)

A5 判 230 ページ 定価 (本体 2,800 円+税)

ISBN : 978-4-06-153821-4



さまざまなデータのもつ意味をコンピュータに理解させることを目的とした技術の総称は機械学習と呼ばれている。近年のコンピュータの性能向上やインターネットの発展、ビッグデータの登場といった背景もあり、非常に注目される研究領域の一つといえる。計測・制御の場面においても、ニューラルネットワークなどのコンピュータサイエンスの知見を基礎とした制御手法が古くから用いられているように、機械学習はデータのもつ意味を適切に処理することで多くの応用が期待されている。

機械学習は高度な数学の理論から実用的な応用まで幅広くカバーしており、次々と新たな学習アルゴリズムが開発されている。このため、この分野を新たに学ぶ人にとって最先端の研究動向を知ることは非常に難しい。そのうえ、機械学習のコンテンツを幅広くカバーした書籍は少ないのが現状である。教科書的な存在として知られる本もあるが、入門書としては難しい内容となっている。

本書は機械学習の入門書として執筆されており、機械学習の基礎から近年の発展的課題まで簡潔に幅広くカバーした内容となっている。本書は全 6 部 20 章構成となっており、基礎から順を追って理解できる。タイトルにもあるように随所にはイラストが挿入されているが、マンガ形式で理解できるという訳ではなく、基本的には式を追って展開されている。また、本書の最大の特徴はイラストではなく、付属されている Matlab のサンプルコードであるように感じる。紹介されている学習アルゴリズムの多くは自分で実装して実際に動作を確認できる。このサンプルコードを実際に試すだけでも機械学習の概観を把握できるだろう。また、アルゴリズムは導出だけではなく、導出された式が数学的にどのような意味をもつのかについても解説されており、深い理解に役立っている。

第 I 部の第 1~2 章は機械学習の基礎を紹介している。第 1 章で機械学習の全体像と基本概念、第 2 章で学習に用いられるモデルについて初心者向けに説明している。

続く第 II 部の第 3~6 章では教師付き回帰について解説している。第 3 章では本書の基礎となる最小二乗学習についてカーネルの概念も含めて述べられている。つぎの第 4

章では制約付き最小二乗学習の説明となっており、雑音への過適合をどのようにして軽減するかを理解できる。第 5 章と第 6 章は、L1 ノルム正則化に基づくスパース学習および損失最小化に基づくロバスト学習についての説明であり、高速な計算方法や異常値を含むデータを用いた学習についてわかりやすく解説されている。

第 III 部の第 7~11 章は教師付き分類に関する解説となっている。第 7 章では回帰問題を発展させることでデータ分類に利用できることを示している。また、第 8 章はクラス分類で最も一般的に使用されているサポートベクトルマシンについて、理論の基礎となるカーネルトリックや KKT 条件も踏まえて解説されている。第 9 章から第 11 章ではさらなる発展的な機械学習の枠組みとして、ブースティングに代表されるアンサンブル学習やロジスティック回帰に基づく確率的分類、条件付き確率場を用いた系列データの分類について、最小二乗学習との関連も含めて紹介している。

第 IV 部の第 12~14 章は教師なし学習について解説されている。第 12 章は異常検出について、著者である杉山先生の専門の一つである密度比に関連した手法も紹介されており、とても興味深い内容となっている。第 13 章と第 14 章では、線形空間だけでなく非線形空間への拡張も含めたデータの次元削減法とクラスタリング法について、結果を比較しながら説明されており非常に理解しやすい。

第 V 部の第 15~19 章は発展的課題として、機械学習の先進的なアルゴリズムを著者の杉山先生自身の研究も数多く含めて解説している。第 15 章はビッグデータと相性がいいことから注目されているオンライン学習のアルゴリズムとして Passive-Aggressive と AROW が紹介されている。また、第 16 章ではラベルは得られないが大量に入手できるデータを有効活用できる半教師付き学習について簡単に述べられている。第 17 章は教師付き次元削減について局所フィッシャー判別分析を紹介している。第 18 章と第 19 章でも同様に機械学習における学習データを有効活用する方法として、転移学習とマルチタスク学習についてそれぞれ解説している。

そして最後に、第 VI 部 20 章では本書のまとめと今後の機械学習の展望について述べられている。

本書は数少ない機械学習の入門書として、主に正則化付きの識別モデル学習を中心として、初心者や専門分野外の人であっても式の展開を把握しながら読み進めることができる良書であるといえる。リファレンスとしての利用には向かないが、本書の読了後には機械学習についての基礎は身につけているであろう。とくに、これから機械学習を新たに学ぶ学生に薦めたい一冊である。

(慶應義塾大学 高橋 巧一)