

書評

「データ科学入門Ⅳ」

写像とグラフによる統一的体系化
—変数／定数の観測可能性の視点から



松嶋敏泰 監修
早稲田大学データ科学教育
チーム 著
サイエンス社 (2025 年)
A5 判, 176 ページ,
定価 2,090 円
ISBN : 978-4-7819-1629-3

データサイエンスの学習において、回帰、分類、クラスタリングと個別の手法を学んできたものの、それらの関係性や背後にある思想が見えず、知識が断片化している感覚を抱く学習者は少なくないであろう。本書『データ科学入門Ⅳ』は、シリーズを通じて「データからの科学的方法による意思決定の科学」という一貫した視点を提示し、まさにその知的な停滞感を打破してくれる。

本書が扱うのは、いわゆる「教師なし学習」のクラスタリングと次元圧縮である。多くの成書がアルゴリズムと使い方の羅列に陥りがちな中、本書はあえて k -平均法と主成分分析のみに解説を絞り、その分、一つの手法を多様な視点から繰り返し解釈し直すというユニークなアプローチを採っている。その核心は、個別の手法を「データ生成観測メカニズム」の視点から捉え直す試みにある。多くの入門書が個別の手法のアルゴリズムと使い方の解説に終始するのに対し、本書は一歩引いた立ち位置から、「主成分分析とは、結局のところデータの背後にあるどのような構造を仮定し、どのような写像を構築しようとしているのか」、「クラスタリングにおけるクラスターとは、観測不可能な潜在変数としてどう位置づけられるのか」といった根源的な問いに答えていく。これにより学習者は、単に手法を適用するだけでなく、「なぜこの手法が有効なのか」、「その限界はどこにあるのか」を深く洞察する力を養うことができる。また、随所に配置されたコラムでは、本文で触れきれない関連手法や概念などが紹介されており、読者の知的好奇心を刺激し、より広い視野を与えてくれるであろう。理論的側面を補強する EM アルゴリズムや固有値、固有ベクトル

の解説が付録に含まれている点も、本書のコンセプトを支えている。

第5章では、シリーズⅠ～Ⅳ巻の内容を「意思決定写像」や「データ生成観測メカニズムグラフ」によって総括している。シリーズを通して読んできた読者にとっては、学んできた知識が昇華され、データ科学の全体像が統一的な視点からクリアになるはずだ。Ⅰ巻で意思決定の基本、Ⅱ巻で回帰や分類、Ⅲ巻でモデル選択を学んだ上で本書に至ることで、データサイエンスの主要な柱が有機的に結びつく構成となっている。初めて本シリーズに触れる方は、これら既刊のうち、特に回帰や分類を扱うⅡ巻も併せて読まれることをお勧めする。

本書は、学部上級生から大学院生、そして理論的基礎に立ち返りたい実務家にとって、断片的な知識を構造化するための得難い羅針盤となるであろう。特に実務家にとっては、日々利用するアルゴリズムをブラックボックスとしてではなく、その仮定や限界を理解した上で使いこなすための、強固な理論的支柱となるはずだ。専門的な書名から、クラスタリングや次元圧縮を扱っているとは分かりにくいかもしれないが、そのタイトルに臆することなく手に取ってほしい。そこには、単なる道具の寄せ集めではない、体系的な学問としてのデータサイエンスの探求が待っている。

本書は、手っ取り早く特定の手法を学びたい、あるいはプログラミングコードをコピー＆ペーストして結果を出したい、というタイプの読者には向かないかもしれない。しかし、データサイエンスを単なる道具の寄せ集めとしてではなく、一つの体系的な学問として真摯に学びたいと願う全ての探求者にとって、本書は知的刺激に満ちた一冊となるであろう。手法の森で道に迷いそうになったとき、あるいは自身の知識の現在地を確認したくなったときに、何度も立ち返る価値のある良書であると考える。

参考文献

[1] 水田正弘：書評「データ科学入門Ⅲ」, 横幹, 第19巻1号, pp. 31-32 (2025)

(東京農工大学 山下 善之)