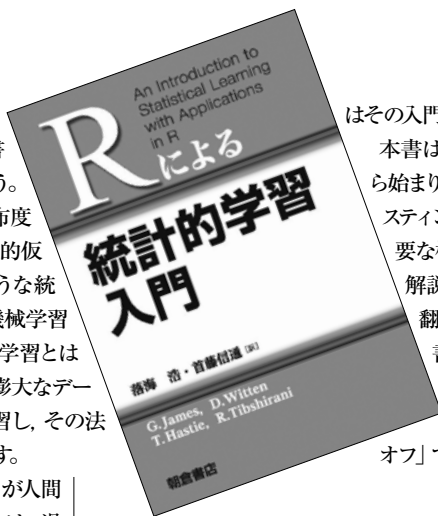


本書の書名を見て、統計学の入門的な書籍と思った読者も多いだろう。しかし、本書は代表値や散布度から始まり、相関係数や統計的仮説検定に関して解説するような統計学入門の書籍ではなく、機械学習に関する書籍である。機械学習とはAI(人工知能)に内包され、膨大なデータから何らかの法則性を学習し、その法則を予測に活かす技術を指す。

囲碁や将棋の世界ではAIが人間の力を既に凌駕した。これは、過去の膨大な対局データから、勝つための戦略を学習した強力なAIが登場したことによる。また、より身近なところでは、0から9までの手書きの文字を正しく判定する技術や、車の自動運転技術、スパムメールの判定技術などに使われている。

したがって、機械学習は過去のデータを使った予測(過去のメールデータから、今届いたメールがスパムであるかを予測・判定する)のための方法であり、社会科学の研究で使われることはこれまで余り多くなかった。これは社会科学の研究の多くは現象の説明を目指すものであり、予測とは馴染まないからと考えられる。しかし一方では、消費者行動の予測や、監視カメラを利用した行動予測など、社会科学やそれに近い分野で機械学習を用いた研究が進められている。社会調査に目を向けると、機械学習による職業コーディングの研究(高橋ら, 2004)や拙著で恐縮であるがWeb調査の不適切回答者の検知(尾崎・鈴木, 2019)などの研究がある。このように、機械学習は社会科学的研究において適用範囲を広げつつあり、今後思いもよらない有益な適用事例が見つかる可能性を秘めている。また、機械学習・AIによって変化する社会を考察するような社会科学的研究を行う際に、機械学習の内容を知っておくことも有益だろう。このような観点からすると、本書



Rによる統計的 学習入門

G. James, D. Witten,
T. Hastie, R. Tibshirani 著

落海 浩・首藤信通 訳

朝倉書店
2018年
A5判, 424頁
6,800円+税

はその入門書として非常に優れたものと言える。本書は線形回帰やロジスティック回帰等から始まり、スパース推定・ランダムフォレスト・ブースティング・サポートベクターマシンなどの重要な機械学習手法についてまんべんなく解説されている。元は洋書ではあるが、翻訳は非常に分かりやすく、日本語の書籍と同じような感覚で読むことができる。本書で特に強調されているのは、「バイアスと分散のトレードオフ」である。これは予測の観点から良い

モデルとは何かを学ぶ際に理解すべき概念である。「バイアスと分散のトレードオフ」は第2章で解説され、その後も必要に応じて触れられている。

各章は手法の解説から始まり、後半には統計ソフトウェアRによる実習がある。実習は実際のデータをRのパッケージを使って分析した事例であり、Rのスクリプトとその解説も掲載されているため、自分で実行しながら学ぶことができる。このスクリプトさえあれば、おおよそ必要な分析は実行可能である。各章末の演習問題は理論編と応用編に分かれている。理論編はやや数理的な側面を扱ったもので

あるが、そこまで難しい数学は要求されない。本書の理論的側面の理解の確認に役立つだろう。応用編では実際のデータ分析が要求されるが、そのためのデータは専用のWebページから入手可能である。

Webページは非常に充実しており、データ以外にも、実習で使われているRスクリプト、書籍の内容のスライド資料、YouTubeの動画教材などが用意されている。これらはすべて無料で利用することができる。

本書評の著者は、大学院のゼミにおいて、本書を扱い、章末の演習問題にも取り組んでもらった。演習問題には解答が用意されていないが、むしろそのことが学生間での議論を活発にし、理解を深めることにつながった。