

訳者序文

「ビッグデータ」、「データサイエンティスト」、「AI」といった言葉がよく聞かれるようになってきた昨今、アクセスログや株価情報など、日々蓄積されるさまざまなデータの利活用が推進されている。その中で、「機械学習」は、これらの言葉と切っても切れない存在となっている。

Michael Bowles による本書の原著、*Machine Learning in Python: Essential Techniques for Predictive Analysis* は、さまざまなデータから未来を予測する“回帰問題”や“分類問題”に焦点を当て、多くの Python のコードとともに、機械学習の効率的なアルゴリズムを取り上げている。本書は、アルゴリズムがどのように振る舞うかをプログラムから理解し、結果を適切に解釈することに重点を置いており、動作を理解するためのコードと、実践で利用する最適化されたコードの 2 種類を用意している。数学的な観点よりも、この方法がなぜ、どのように役立つかという実践的な観点を重視して書かれていることも、本書の特徴である。さらに、無償で入手可能なデータセットに対してアルゴリズムを適用し、その実行結果から解釈の仕方に至るまでを丁寧に解説している。この一連の流れをなぞることで、解析に必要な考え方を一通り身につけられるだろう。

データ解析でよく遭遇する予測問題では、過学習を避けつつその精度をいかに高めていくかといった議論を、特に学術的な観点でよく耳にする。一方で、ビジネス的な観点からは、ただ精度を上げればよいといった議論ではなく、むしろ精度を多少犠牲にしても、どういった変数が予測に寄与しているのか、得られた知見をもとにどのように次につなげていけばよいのかといった議論が重要視されることもあるだろう。本書では、さまざまな種類や特性を持つデータについて、どのようにすれば重要な変数を特定できるのかについても丁寧に記述しており、まさにそういったニーズにも合致している。また、紹介されている実践的なコードを利用することで、解析のテクニックを自身の問題へとすぐに展開することができ、現場での問題解決の一助となるだろう。

データ解析の手順を料理の手順にたとえると、世の中はデータという食材を収穫する段階から、どのように美味しく調理するかを考える段階に移ってきてていると感じる。本書を手に取られた読者は、まさにその過渡期にいる方だろう。本書は、その下ごしらえから調理の仕方まで丁寧に説明しているため、データ解析の入門書として読んでいただ

くには、ぴったりの1冊である。本文を読み進めていくだけではなく、ぜひ実際に手を動かしながら理解を深めていっていただきたい。

2019年2月

訳者一同