

第2部

実践編①：機械学習

第1部を一通り学ぶことで、実際のビジネス現場でデータ分析をどのように始めるのかについて、イメージがつかめるようになったのではないのでしょうか。データを読み込み、「綺麗」でないデータの処理を行いながら、全体を可視化して傾向を掴むという一通りの流れを理解できれば、ほとんどのデータ分析業務を「はじめる」力は十分に身についたと言えます。

ここからは、いよいよ、本格的にデータ分析業務で「結果を出す」ための技術についての学びを深めていきます。データ数が急激に増加している昨今、機械（コンピュータ）の力を借りながら対話的に分析を進めていく手段である機械学習は必須スキルとなりつつあります。

こうした背景から、ここでは、機械学習による分析、そして、将来予測に関する技術を学びます。これらの技術を理解し、現場で適切に活用していくことができれば、ほとんどのデータ分析業務は怖くありません。現場で「結果を出す」ための力を身につけることができると言えます。

第3章では、統計を使って、機械学習に向けた事前分析を行います。機械学習といってもさまざまな手法があり、どのような手法を利用すべきかを判断するために、事前分析をうまく行うことが肝になります。第4章では、機械学習の中でも「教師なし学習」と呼ばれるクラスタリングと、将来予測を可能にする回帰分析を行います。最後に、第5章では、「教師あり学習」を呼ばれる決定木という手法をもちいたデータの分類方法を学んでいきます。

第2部で取り扱うPythonライブラリ

データ加工：Pandas

可視化：Matplotlib

機械学習：scikit-learn