

まえがき

人工知能時代をエンジョイするのか、人工知能のエジキになるのか

2016年に、大阪大学理学部数学科の学生を対象に（統計的）機械学習に関する講義をはじめたことが、本書を執筆するきっかけになったと思います。当時、テキストとして適当なものがなく、2018年に日本語に翻訳された『Rによる統計的学習入門』“*Introduction to Statistical Learning with Application in R*”を参考に講義を組み立てました。説明が丁寧で、わかりやすく、非常に気に入り、日本語の翻訳をしてみたいと思うぐらいになりました（その時点で他の方が権利をすでに獲得していて、翻訳は実現しませんでした）。本書の扱っている単元や、章立てもその著述と似ているのも、そうしたところによっています。

2017年に現在の所属、基礎工学部情報科学科数理科学コースに異動となりました。早速、3年生後期の計算数理Bという科目で、2016年と同様の内容の講義を試みました。数理科学コースは2年生でR言語を習熟していて、当初は違和感なく首尾よく講義を行えました。しかし、よりどころとしていた『Rによる統計的学習入門』については、2018年、2019年というように講義を重ねるごとに、問題意識を抱くようになりました。初学者の方が機械学習の概要を把握するのには、最適な書籍とは思われますが、

1. 本質的な理解というよりは、感覚的な理解だけで十分
2. プログラムのステップを見ないで、パッケージにデータを放り込めば十分

という姿勢には、どうしても合点がいきませんでした。

現在は、人工知能の時代と言われます。言うまでもありませんが、インターネットのおかげで、必要な情報を即座に得ることができ、職場の業務が効率化され、生活も豊かになりました。その一方で、人間が行ってきた業務の多くが人工知能に置き換えられるのではないかという危惧が生じています。私自身は、データサイエンスや機械学習の業務に携わる人との付き合いが多いのですが、「業務でどのような資質が求められますか」と聞くと、知識や経験というよりは、「ロジック」という答えが多く、活躍している人ほどその傾向が強いように思われます。情報の真偽を吟味する、人が見えない本質やチャンスを見る、制約にとらわれない発想などが、「ロジック」の結果として生まれているという視点です。逆に、そういう「ロジック」が欠如していて、ヤマカンに頼るというのであれば、人工知能のエジキになる可能性が高くなるように思われます。

もっとも、こうした「ロジック」が、数学やプログラミングをやらないと身につかない、というのは真実ではないかもしれません、大学教員30年間の中で、多くの学生をみてきた経験からす

ると、それらには非常に強い相関があるように思われます。本書は、機械学習に関する知識も提供しますが、それと同時に、数学的に本質を理解して、プログラミングで処理を構成して、検証するという経験を通じて、読者の方々の「ロジック」を脳裏に構築することを、目標の一つに掲げています。

『Rによる統計的学習入門』は、講義を組み立てる上で大変参考にはなりましたが、人工知能時代をエンジョイするために不可欠な「ロジック」を構築するという視点が十分ではなく、どうしても本書を執筆せざるを得なかつた、というのが私の正直な気持ちです。

また、数年前に日本語に翻訳された“*Elements of Statistical Learning*”『統計的学習の基礎』(共立出版)は、分量が多く、輪講して挫折したという話をよく聞きます。困ったときに参考にする百科事典として使う方が多いように思われます(私は頻繁に利用しています)。『統計的学習の基礎』を大学の講義半期程度に圧縮できたら、という気持ちも本書を執筆するきっかけとなりました。また、『統計的学習の基礎』は信頼できる知識を提供していますが、ソースコードをおいたり、読者が具体的にスキルを身につけるような誘導があつてもよいように思いました。

本書の100問は、2017～2019年の講義で学生に課した演習問題で、特に『Rによる統計的学習入門』を漫然と読んでいる学生にツッコミを入れるために作成したものです。数式を正しく導出し、プログラムを組んで実行結果を見るなど、手を動かして自分のものにすべきと伝えました。ただ、問題だけを与えると、優秀な学生以外は自力では解けないので、全員が課題を提出できるよう、講義で解答に限りなく近い丁寧なヒントを与えました。それが本書の本文です。問題1～100と本文で重複している記述が若干あるのは、そのためです。また、講義で提供している10～15分間の復習ビデオも提供しています。

本シリーズの特徴

本書というよりは、本シリーズの特徴を以下のようにまとめてみました。

1. 身につける：ロジックを構築する

数学で本質を把握し、プログラムで処理を構成して、データを処理していきます。読者の皆さんのが脳裏に、「ロジック」を構築していきます。機械学習の知識だけではなく、視点が身につきますので、新しい機械学習の技術が出現しても追従できます。100問を解いてから、「大変勉強になりました」と言う学生がほとんどです。

2. お話だけで終わらない：コードがあるのですぐにコード（行動）に移せる

機械学習の書籍でソースプログラムがないと、非常に不便です。また、パッケージがあっても、ソースプログラムがないとアルゴリズムの改良ができません。gitなどでソースが公開されている場合もありますが、MATLABしかなかったり、十分でない場合もあります。本書では、ほとんどの処理にプログラムのコードが書かれていて、数学がわからなくても、それが何を意味するかを理解できます。

3. 使い方だけで終わらない：大学教授が書いた学術書

パッケージの使い方、実行例ばかりなる書籍も、よく知らない人がきっかけを掴めるなど、存在価値はありますが、手順にしたがって機械学習の処理を実行できても、どのような動作をしているかを理解できないので、満足感として限界があります。本書では、機械学習の各処理の数学的原理とそれを実現するコードを提示しているので、疑問の生じる余地がありません。本書はどちらかというと、アカデミックで本格的な書籍に属します。

4. 100問を解く：学生からのフィードバックで改善を重ねた大学の演習問題

本書の演習問題は、大学の講義で使われ、学生からのフィードバックで改良を重ね、選びぬかれた最適な100問になっています。そして、各章の本文はその解説になっていて、本文を読めば、演習問題はすべて解けるようになっています。

5. 書籍内で話が閉じている (self-contained)

定理の証明などで、詳細は文献〇〇を参照してください、というように書いてあって落胆した経験はないでしょうか。よほど興味のある読者（研究者など）でない限り、その参考文献をたどって調査する人はいないと思います。本書では、外部の文献を引用するような状況を避けるように、題材の選び方を工夫しています。また、証明は平易な導出にし、難しい証明は各章末の付録においています。本書では、付録まで含めれば、すべての議論が完結しています。

6. 読者ページ：章ごとのビデオ、オンラインの質疑応答、プログラムファイル

大学の講義では、slackで24/365体制で学生からの質問に回答していますが、本書では Disqus という質疑応答のシステムを採用しています。また、講義で利用した各章10～15分のビデオを公開しています。また、本書にあるプログラムをすべて手打ち込むのは難しいので、プログラムのリストのファイルをサイトにおいています。

7. 線形代数

機械学習や統計学を学習するうえでネックになるのが、線形代数です。研究者向きのものを除くと、線形代数の知識を仮定しているものは少なく、本質に踏み込めない書籍がほとんどです。そのため、シリーズ第1号の『統計的機械学習の数理100問 with R』と第2号の『同 with Python』では、第0章として、線形代数という章を用意しています。14ページしかありませんが、例だけでなく、証明もすべて掲載しています。ご存知の方はスキップしていただいて結構ですが、自信のない方は休みの日を1日使って読まれてもよいかと思います。

本書の使い方

各章は、問題、その解説（本文）、付録（証明、プログラム）からなっています。問題を解いてみて、わからないときだけ本文を読める方（上級者）もいらっしゃるでしょうが、本文から読み始めて最後に問題をとくという形でも問題ありません。最後まで読破して、身につけることを優先してください。

講義で使われる場合、第1章を3回、第4章を1～2回、第6章を2回、それ以外の章で各1回で、合計12～13回程度の講義が組めるかと思います。線形代数の章（第0章）に2～3回かけて15回としてもよいですが、第1章～第9章をやりながら適宜戻ってくるという進め方でもよいかと思います。かなりできる学生なら、最初に100問を課してもよいでしょう。本文をじっくり読めば、回答できるようになっています。輪講（担当を決めて交代で読み続けていく）でも、同じ程度の回数で読めるかと思います。

謝辞

機械学習の数理100問の執筆をご提案いただいた共立出版の皆様、特に本書の担当編集者で、シリーズ化を受け入れ、本書の出版に関して細かい点までチェックしていただいた大谷早紀氏に感謝します。また、大阪大学大学院の学生である稻岡雄介君と新村亮介君には、提出前の原稿に目を通して、プログラムや数式の誤りを指摘してもらいました。2018～2019年の講義の学生（大阪大学基礎工学部）からのフィードバックが、本書にとって有益な情報となったことも、特記すべきと思っています。