● 免責

本書に記載された内容は，情報の提供だけを目的としています。したがって，本書を用いた運用は，必ずお客様自身の責任と判断によって行ってください。これらの情報の運用の結果について，技術評論社および著者はいかなる責任も負いません。

本書記載の情報は，2015年9月現在のものを掲載していますので，ご利用時には，変更されている場合もあります。

以上の注意事項をご承諾いただいた上で，本書をご利用願います。これらの注意事項をお読みいただかずに，お問い合わせいただいても，技術評論社および著者は対処しかねます。あらかじめ，ご承知おきください。

● 商標，登録商標について

- 本書に登場する製品名などは，一般に各社の登録商標または商標です。なお，本文中に ™，® などのマークは省略しているものもあります。

はじめに

「機械学習」にかかわるITエンジニアが、予想以上に増えているのかもしれない——そんな疑問を抱いたのは1年ほど前の出来事でした。「データサイエンス」や「ディープラーニング」、果ては「人工知能」まで、メディア好みのバズワードが溢れる中、データ分析を専門としない一般のITエンジニアに対しても、機械学習の活用が期待される時代がやって来ました。世の中では、「専門知識がなくても使える」と宣伝する機械学習サービスすら提供されています。

しかしながら、そこには大きな落とし穴があります。さまざまな機械学習のツールやライブラリーがオープンソースして提供されるようになり、機械学習の計算処理は誰でもできるようになりました。データを投入してプログラムを実行すれば、何らかの結果が出てきます。しかしながら、その結果にはどのような「意味」があるのでしょうか？　機械学習の結果をビジネスに活用するには、その背後にあるアルゴリズムを理解して、その結果が持つ意味を正しくとらえる必要があります。

本書では、機械学習のビジネス活用を念頭に置き、機械学習の基礎となるアルゴリズムを根本から解説しています。具体的な例題を用いて、「どのような考え方で、何を計算しているのか」という点をごまかさずに説明します。それによって、機械学習、さらにはデータサイエンスの本質を理解していただくことが本書の目標です。機械学習には、さまざまなアルゴリズムがありますが、その根底には「データのモデル化とパラメーターの最適化」という共通した考え方があります。本書では、このような「考え方」に重点をおき、それぞれの数式の意味をできる限り平易に説明するように心がけました。この点が理解できれば、ディープラーニングやニューラルネットワークなど、本書の範囲を超えた、さらに高度なアルゴリズムも恐れることはないでしょう。

「機械学習のビジネス活用企画を頼まれて困っている」「販売分析アプリケーションの開発プロジェクトに、突然、参加が決まった」——知人のITエンジニアからのこのような声が、冒頭の疑問の出どころでした。これからの時代、機械学習を理解して使いこなすことは、ITエンジニアとして新たな人生を切り開くチャンスになるのは間違いなさそうです。そして何よりも、機械学習には、ITエンジニアの知的好奇心、技術への探究心を存分に満たしてくれる面白さがあります。一人でも多くの方に、本書を足がかりとして、機械学習の世界への第一歩を踏み出していただけることを期待しています。

2015年　初秋　中井 悦司

謝 辞

本書の執筆、出版にあたり、お世話になった方々にお礼を申し上げます。

本書の構想は、技術評論社の池本公平氏の発案からスタートしました。機械学習に関する筆者の知識を体系的に整理して、ITエンジニアに向けた書籍としてまとめあげる、またとない機会を提供いただいたことを感謝します。

また、本書の原稿の査読を快く引き受けていただき、Mac OS X/Windows版のインストール手順を提供いただいた織学(おりまなぶ)氏にも感謝します。

本書の内容をまとめるにあたっては、国立情報学研究所「トップエスイー」の有志による勉強会で行った議論から、数多くのヒントをいただきました。同勉強会に参加いただいた方々にも改めて感謝します。

そして、本書の執筆作業の大部分は、今年から小学校に通い始めた愛娘の歩実を駅まで送り届けた後、早朝のスターバックスの店内で進められました。「早寝早起き朝ごはん」をモットーに、こんな健康的な生活を支えてくれる妻の真理にもあらためて感謝の気持ちを伝えたいと思います。「いつもありがとう！」

本書が対象とする読者

本書は、機械学習のアルゴリズムについて、その背後にある理論を理解して、ビジネスに役立てたいと考えるITエンジニアを対象としています。機械学習にはさまざまな利用目的がありますが、本書では、「データ分析結果をビジネス判断に役立てる」という観点から、各種のアルゴリズムの解説を進めます。機械学習のツールやライブラリーの使い方を説明した書籍ではありませんので、その点はご注意ください。

また、本書で取り上げる例題の多くは、機械学習の世界ではバイブルとも言える、次の書籍から引用しています。

『パターン認識と機械学習・上／下』C.M. ビショップ（著）、元田 浩、栗田 多喜夫、樋口 知之、松本 裕治、村田 昇（監訳）、丸善出版（2012年）

機械学習を学ぼうとこの書籍に挑戦したものの、高度で理解しきれなかったという方も多いかもしれません。ITエンジニアに限らず、この「バイブル」を読破するための入門書として、本書を活用していただくこともできるでしょう。

本書の読み方

本書は、第1章から順に読み進めることで、機械学習のさまざまなアルゴリズムを体系的に理解できるように構成されています。第1章では、「機械学習のビジネス活用」という観点を明確にするために、より大きなデータサイエンスの枠組みから機械学習をとらえます。その後、第2章〜第8章では、第1章で紹介する代表的な例題について、具体的なアルゴリズムを適用していきます。同じ問題に対して、複数のアルゴリズムを適用することで、それぞれのアルゴリズムの特徴や共通する考え方が理解できるようになります。

また、本書では、それぞれのアルゴリズムを実装した、Pythonによるサンプルコードを提供しています。サンプルコードを実行して得られる、具体的な回答を観察することで、数式だけではわからない、アルゴリズムの本質を捉えることができるでしょう。

そして、機械学習のアルゴリズムを理解する上では、一定レベルの数学の知識が必要となります。本書では、「その数式は何を計算しているのか」という点をできるだけ平易に解説していますが、大学初等程度の数学の知識があれば、よりスムーズに理解を進めることができます。機械学習の前提となる数学を学びたい方は、参考文献の「数学の基礎」に示した書籍を参考にしてください。

出版後に発見された修正点や補足情報については、技術評論社のWebサイトで公開していきます。

・http://gihyo.jp/book/2015/978-4-7741-7698-7

なお、「もう10年近く数学はやっていない……」という方のために、本書で使用する主な数学記号と基本公式を下記にまとめておきます。必要に応じて参照してください。

主な数学記号と基本公式

和の記号

記号 Σ（シグマ）は和を表します。次は、x_1 から x_N までの足し算になります。

$$\sum_{n=1}^{N} x_n = x_1 + x_2 + \cdots + x_N \tag{1}$$

積の記号

記号 Π（パイ）は積を表します。次は、x_1 から x_N までのかけ算になります。

$$\prod_{n=1}^{N} x_n = x_1 \times x_2 \times \cdots \times x_N \tag{2}$$

指数関数

記号 exp（エクスポネンシャル）は、自然対数の底 $e \fallingdotseq 2.718$ を用いた指数関数を表します。次は、e の x 乗を表わす関数になります。

$$\exp x = e^x \tag{3}$$

指数関数の積は、引数の和に変換されます。

$$\prod_{n=1}^{N} e^{x_n} = e^{x_1} \times \cdots \times e^{x_N} = e^{x_1 + \cdots + x_N} = \exp\left\{\sum_{n=1}^{N} x_n\right\} \tag{4}$$

指数関数 e^x は、微分しても関数が変化しません。

$$\frac{d}{dx} e^x = e^x \tag{5}$$

対数関数

記号 ln（ログナチュラル）は、自然対数の底 $e \fallingdotseq 2.718$ を用いた対数関数を表します。

$$\ln x = \log_e x \tag{6}$$

$x = e$ を代入すると、ちょうど1になります。

$$\ln e = 1 \tag{7}$$

対数関数は、次の対数法則を満たします。

$$\ln \frac{ab}{c} = \ln a + \ln b - \ln c \tag{8}$$

$$\ln a^b = b \ln a \tag{9}$$

これらより、(4) の形式の指数関数を対数関数に代入すると、式が簡単になります。

$$\ln\left(\exp \sum_{n=1}^{N} x_n\right) = \sum_{n=1}^{N} x_n \times \ln e = \sum_{n=1}^{N} x_n \tag{10}$$

これは、対数関数 $\ln x$ は、指数関数 e^x の逆関数である事を意味しています。

対数関数の微分は、次のとおりです。

$$\frac{d}{dx} \ln x = \frac{1}{x} \tag{11}$$

■ 偏微分

複数の変数を持つ関数について、特定の変数で微分することを偏微分と呼びます（記号∂は、「デル」、「ラウンドディー」などと読みます）。

$$\frac{\partial f(x,y)}{\partial x} \text{：} y \text{ を固定して } x \text{ で微分する}$$

$$\frac{\partial f(x,y)}{\partial y} \text{：} x \text{ を固定して } y \text{ で微分する}$$

偏微分についても合成関数の微分の公式が成り立ちます。

$$\frac{\partial f(g(x,y))}{\partial x} = f'(g(x,y)) \times \frac{\partial g(x,y)}{\partial x} \tag{12}$$

$f'(x)$は、一階の微分係数を表します。

$$f'(x) = \frac{df(x)}{dx} \tag{13}$$

■ ベクトルの内積と外積

ボールドフォントの変数は、ベクトル、および、行列を表します。ベクトルについては、成分を縦にならべた「縦ベクトル」を基本とします。

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \tag{14}$$

表記の都合上、横ベクトルで記載するときは、転置記号Tを用いて、縦ベクトルであることを示します。

$$\mathbf{x} = (x_1, x_2, x_3)^{\mathrm{T}} \tag{15}$$

逆に、縦ベクトルを転地すると、横ベクトルになります。

$$\mathbf{x}^{\mathrm{T}} = (x_1, x_2, x_3) \tag{16}$$

「横ベクトル」×「縦ベクトル」は内積を表します。

$$\mathbf{w}^{\mathrm{T}}\mathbf{x} = (w_1, w_2, w_3)\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \sum_{i=1}^{3} w_i x_i \tag{17}$$

「縦ベクトル」×「横ベクトル」は外積を表します。

$$\mathbf{w}\mathbf{x}^{\mathrm{T}} = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix}(x_1,\ x_2,\ x_3) = \begin{pmatrix} w_1x_1 & w_1x_2 & w_1x_3 \\ w_2x_1 & w_2x_2 & w_2x_3 \\ w_3x_1 & w_3x_2 & w_3x_3 \end{pmatrix} \tag{18}$$

(12)を用いると、ベクトルの内積を代入した関数について、特定の成分で偏微分することができます。

$$\frac{f(\mathbf{w}^{\mathrm{T}}\mathbf{x})}{\partial w_i} = f'(\mathbf{w}^{\mathrm{T}}\mathbf{x})\frac{\partial(\mathbf{w}^{\mathrm{T}}\mathbf{x})}{\partial w_i} = f'(\mathbf{w}^{\mathrm{T}}\mathbf{x})x_i \tag{19}$$

ベクトルの大きさは、次の記号で表します。

$$\|\mathbf{x}\| = \sqrt{\mathbf{x}^{\mathrm{T}}\mathbf{x}} = \sqrt{x_1^2 + x_2^2 + x_3^2} \tag{20}$$

■ 確率変数の期待値と分散

確率的にさまざまな値をとる変数Xを確率変数と呼び、$X=x$という値をとる確率（Probability）を$P(x)$で表します。確率変数の期待値（Expected value）Eと分散（Variance）Vは次式で定義されます。

$$E[X]=\sum_x xP(x) \tag{21}$$

$$V[X]=E[\{X-E(X)\}^2] \tag{22}$$

(21)の和$\sum_x$は、すべての場合のxについて合計します。

平均と分散について、次の公式が成り立ちます。

$$E[aX+b]=aE[X]+b \tag{23}$$

$$V[aX]=a^2V[X] \tag{24}$$

$$V[X]=E[X^2]-(E[X])^2 \tag{25}$$

(23)より、$\overline{x}=E[X]$として、次が成り立ちます。

$$E[X-\overline{x}]=E[X]-\overline{x}=0 \tag{26}$$

2つの確率変数XとYが「独立である」というのは、$X=x$かつ$Y=y$となる確率（同時確率）$P(x,y)$がそれぞれの確率の積で表されることを示します。

$$P(x,y)=P_X(x)\times P_Y(y) \tag{27}$$

たとえば、2個のサイコロを振った時に、1のゾロ目がでる確率は、それぞれのサイコロの目が1になる確率1/6の積で計算できます。これは、それぞれのサイコロの目の確率が独立であることを示します。

確率変数XとYが独立な場合、$\overline{x}=E[X]$、$\overline{y}=E[Y]$として、次が成り立ちます[注1]。

$$\begin{aligned}E[(X-\overline{x})(Y-\overline{y})]&=\sum_{x,y}(X-\overline{x})(Y-\overline{y})P(x,\ y)\\&=\sum_x(X-\overline{x})P_X(x)\sum_y(Y-\overline{y})P_Y(y)\\&=E[X-\overline{x}]E[Y-\overline{y}]=0\end{aligned} \tag{28}$$

注1　この関係は、「3.3.1　標本平均／標本分散の一致性と不偏性の証明」で用いています。

各章概要

■第１章　データサイエンスと機械学習

機械学習のアルゴリズムを学ぶ準備として、より大きなデータサイエンスの枠組みから機械学習をとらえます。「データサイエンスにおける機械学習の役割」を理解することで、機械学習のビジネス活用という観点から、アルゴリズムの特性をより明確に理解することができるようになります。また、第2章～第8章で使用する例題を先に解説した上で、サンプルコードの実行環境を準備する手順を説明します。

■第２章　最小二乗法：機械学習理論の第一歩

機械学習の基礎となる「回帰分析」の中で、もっとも基本的な「最小二乗法」のアルゴリズムを解説します。計算そのものはそれほど難しくありませんが、この手続きを通して、機械学習の理論的基礎となる「統計モデル」の考え方を理解していきます。また、機械学習の結果をビジネスに適用する上でのポイントとなる、「オーバーフィッティング」の検出について説明します。

■第３章　最尤推定法：確率を用いた推定理論

確率を利用した統計モデルの基礎となる「最尤推定法」の手続きを解説します。第2章と同じ例題を扱いながら、最小二乗法との類似点／相違点を整理した上で、機械学習における、確率を用いたモデルの役割を理解していきます。少し高度な話題として、推定量の「一致性」と「不偏性」についても解説を加えます。

■第４章　パーセプトロン：分類アルゴリズムの基礎

「分類アルゴリズム」の基礎となる「パーセプトロン」について解説します。数値計算を用いてパラメーターを修正してゆく「確率的勾配降下法」の手続きは、機械学習における数値計算手法の基礎となります。一般的な入門書ではあまり触れられていない点として、バイアス項の修正による収束速度の改善、および、アルゴリズムの幾何学的な解釈についても解説を行ないます。

■第５章　ロジスティック回帰とROC曲線：学習モデルの評価方法

最尤推定法を用いた分類アルゴリズムとして、「ロジスティック回帰」の解説を行ないます。ROC曲線を利用しながら、現実の問題に機械学習を適用する際の考え方、そして、複数の分類アルゴリズムを比較する方法を学びます。数学的な興味を持つ読者のために、数値計算でパラメーターを修正していく「IRLS法」の厳密な導出も行ないます。

■第６章　k平均法：教師なし学習モデルの基礎

教師なし学習によるクラスタリングの基礎として、「k平均法」のアルゴリズムを解説します。また、具体的な応用例として、画像ファイルの減色処理を実装します。文書データの自動分類など、単純ながらも応用範囲の広いアルゴリズムです。さらに参考として、怠惰学習モデルである「k近傍法」を紹介した上で、機械学習における「データのモデル化」の意義について考えます。

■第７章　EMアルゴリズム：最尤推定法による教師なし学習

教師なし学習によるクラスタリングのアルゴリズムとして、最尤推定法を利用した「EMアルゴリズム」を紹介します。比較的複雑なアルゴリズムとなるため、手書き文字の分類問題に対する、具体的な適用例を通して解説を進めます。現実世界でも活用される「画像分類」のアルゴリズムが、確率を用いた考え方でどのように実現されるのか、興味を引く実例です。

■第８章　ベイズ推定：データを元に「確信」を高める手法

モデルに含まれるパラメーターの値を確率的に推測する「ベイズ推定」の手法を解説します。理論的基礎となる「ベイズの定理」の解説からはじめて、第2章、第3章と同じ回帰分析の例題について、ベイズ推定を適用していきます。ベイズ推定には、計算で得られた結果の「確信度」がわかるという特徴があり、最尤推定法とは異なる、新たな知見を得られることがわかります。

目次 Contents