パターン認識と機械学習 まえがき

鈴木拓己

2024年11月20日

まえがき

パターン認識 (pattern recognition) は工学を起源とするが、機械学習 (machine learning) は計算機科学の分野から生じている。しかし、これらの研究活動内容は、同じ分野を2つの側面から見たものとみなせ、両分野ともこの10年間に大きく発展した。中でも、限られた専門家のものであったベイズ的手法は、主流の手法へと成長した。その一方で、グラフィカルモデルは、確立モデルを記述し、適用するための汎用的な枠組みとして進展してきた。また、ベイズ的手法は、変分ベイズ法やEP 法などの近似推論アルゴリズムが開発されたことで、より広範囲の現実の問題に適用されるようになった。同様に、カーネルを用いた新たなモデルも、アルゴリズムや応用分野に大きな影響を与えている。

この新しい教科書は、近年の研究成果を反映する一方、パターン認識と機械学習の包括的な入門書でもある。本書は、学部 3,4 年生や、博士課程の初年度の学生の他、研究者や開発者をも対象としている。本書を理解するのに、パターン認識や機械学習について、あらかじめ何も知らなくてよいが、多変数微積分や基礎的な線形代数についての知識は必要である。また、確率について精通していれば、本書の理解に役立つ。だが、本書には基礎的な確率論についての解説もあるので、知らなくてかまわない。

本書の扱う内容は広範囲にわたるため、すべての参考文献を網羅するのは不可能である。特に、いろいろな考えの正確な歴史的な意義について、説明しようとはしていない。その代わり、本書より詳細な内容を扱った。または、進展させた内容への入り口となるような、参考文献を挙げるようにした。そのため、原論文より、近年の教科書や解説論文を参考文献として主に挙げてある。

本書には、講義資料や本書で使われたすべての図表など、多数の追加資料が用意されている. これらについての最新情報を得るには次の Web ページを参照されたい.

http://research.microsoft.com/~cmbishop/PRML

演習問題

本書では、各章末の演習問題も重視している。問題には、本文で説明した概念を発展させたり、新たな手法を開発したり、手法を一般化するのに役立つようなものを注意深く選んだ。問題には難易度も示し、**(基本)**は数分で解けるような簡単なもの、**(難問)**は非常に複雑な演習を示している。

演習問題の解答を、どれくらい入手しやすくすべきかを決めるのは難しい.本書で独学する読者には解答はとても役立つだろう.だが、本書を教科書とする講師にとっては、演習問題を講義で利用できるように、解答は出版社から取り寄せられるようにしておく方が良いであろう.こうした相反する要求に応じるようとするため、本文の重要な点を拡充するのに役立つ演習問題や、重要な細部を補足するような問題についてのみ、本書の Web サイトから解答を PDF ファイルで入手できるようにした.こうした演習問題は www で示した.他

の演習問題の解答は、出版社に連絡すれば(詳細は Web ページを参照)、講師には入手できるようにする. だが、読者には、すべて独力でこれらの演習問題を解き、必要なときにのみ解答を見るようにすることを強く薦める.

本書は概念的・原理的な事柄を中心に執筆した.だが、できれば学生は適当なデータ集合を用いて、主なアルゴリズムのいくつかを実験してみるとよい.本書で示したほとんどのアルゴリズムを Matlab で実装したソフトウェアと、例題用データ集合は Web サイトから入手できるようにする.また、これらは、機械学習に現れる最適化問題を解く実用的アルゴリズムについての姉妹書 (Bishop and Nabney, 2008) にも収録する予定である.

謝辞

まず最初に、本書の図表や \LaTeX での組版の準備に多大な貢献をしてくれた Markus Svensén に心からの感謝を示したい. 彼の手助けは計りしれないものであった.

また,非常に刺激的な研究環境を提供し、本書を執筆できるよう取りはからってくれた Microsoft Research 社に謝意を表す(しかし、本書の立場や意見は私自身のもので、したがってそれらは必ずしも Microsoft 社やその関係団体のそれと同じではない).

Springer 社は、本書の執筆の最終段階を通して、すばらしい援助をしてくれた。担当編集者 John Kimmel には、彼の支援とプロ精神に対して、Joseph Piliero には、本書の表紙と体裁への手助けに対して、MaryAnn Brickner には、制作段階での多大な貢献に対して感謝したい。表紙のデザインは、Antonio Criminisi との議論に触発されたものである。

以前の教科書 Neural Networks for Pattern Recognition (Bishop, 1995) からの抜粋を許可してくれた Oxford University Press 社にも感謝したい. Mark 1 パーセプトロンと Frank Rosenblatt の画像は、Arvin Calspan Advanced Technology Center の許可を得て掲載した。図 13.1 のスペクトル図を描いてくれた Asela Gunawardana と、図 12.7 を描くためにカーネル PCA のコードを利用させてくれた Bernhard Schölkopf にも感謝したい。

また、本書の予稿を閲読し、助言や提言をしてくれた次の方々のお名前を挙げておきたい。Shivani Agarwal、Cedric Archambeau、Arik Azran、Andrew Blake、Hakan Cevikalp、Michael Fourman、Brendan Frey、Zoubin Ghahramani、Thore Graepel、Katherine Heller、Ralf Herbrich、Geoffrey Hinton、Adam Johansen、Matthew Johnson、Michael Jordan、Eva Kalyvianaki、Anitha Kannan、Julia Lasserre、David Liu、Tom Minka、Ian Nabney、Tonatiuh Pena、Yuan Qi、Sam Roweis、Balaji Sanjiya、Toby Sharp、Ana Costa e Silva、David Spiegelhalter、Jay Stokes、Tara Symeonides、Martin Szummer、Marshall Tappen、Ilkay Ulusoy、Chris Williams、John Winn、Andrew Zisserman.

最後に、本書の執筆に費やした数年間を通じて大きな支えとなってくれた妻 Jenna にお礼を述べたい.

2006 年 2 月ケンブリッジにて C.M. ビショップ

数式の表記

本書では、数学的な内容は、この分野を正しく理解するのに必要最小限に留めた。だが、この最小限は0ではなく、最近のパターン認識や機械学習を明確に理解するには、微積分、線形代数、確率論を、十分に把握しておくことは必須である。だが、本書では、数学的な厳密さより、背後の概念を説明することを重視している。本書を通して一貫した表記を用いるように努めた。そのため、該当する研究分野で使われる一般的な表記と

は、しばしば違ったものとなることもある.ベクトルは、x などの太字のローマン体小文字で記し、すべてのベクトルは列ベクトルと仮定する.上付きの \top は、行列やベクトルの転置を表す.よって、 x^\top は行ベクトル

になる. M などの太字のローマン体大文字は行列を表す. $(w_1, ..., w_M)$ は M 要素の行ベクトルであり,これに対応する列ベクトルは $\mathbf{w} = (w_1, ..., w_M)^{\top}$ と書く.

[a,b] の表記は、a から b への閉区間、すなわち、a と b の値を含む区間を記すのに用いる。一方、(a,b) は開区間、すなわち、a や b は含まない区間を記す。同様に、[a,b) は、a は含むが b は含まない区間を表す。しかし、ほとんどの場合、区間の端点を含むかどうかといった細部について、あまり考える必要はない。

 $M\times M$ の単位行列は I_M と記し、次元数が曖昧でなければ I と略記する.この行列の要素 I_{ij} は、i=j なら 1 で、 $i\neq j$ なら 0 である.

y(x) を関数として、汎関数を f[y] と記す.汎関数の概念については付録 D で述べる.g(x) = O(f(x)) は、|f(x)/g(x)| が、 $x \to \infty$ で有界であることを示す.例えば、 $g(x) = 3x^2 + 2$ なら、 $g(x) = O(x^2)$ である.

関数 f(x,y) の,確率変数 x についての期待値を, $BE_x[f(x,y)]$ と記す.どの変数で期待値をとるのかが曖昧でないときは,添え字を省略して, $\mathbb{E}[x]$ などと簡略化する.x の分布が,別の変数 z で条件付けされているなら,このときの条件付き期待値は $\mathbb{E}_x[f(x)|z]$ と書く.同様に,分散は $\mathrm{var}[f(x)]$ と記し,ベクトル変数に対する共分散は $\mathrm{cov}[x,y]$ と記す.また, $\mathrm{cov}[x,x]$ を短くした表記として $\mathrm{cov}[x]$ も用いる.期待値や共分散については,1.2.2 節で紹介する.

D 次元ベクトル $\mathbf{x}=(x_1,\ldots,x_D)^\top$ が, $\mathbf{x}_1,\ldots,\mathbf{x}_N$ のように N 個あるとき,これらの観測値を,第 n 行 が,行ベクトル \mathbf{x}_n^\top となるようなデータ行列 \mathbf{X} にまとめることがある.よって, \mathbf{X} の n,i 要素は,第 n 観 測値 \mathbf{x}_n の,第 i 要素に該当する.1 次元変数の場合では,このような行列は \mathbf{x} と記す.これは,第 n 要素 が \mathbf{x}_n であるような列ベクトルである.なお,(次元数が D の) \mathbf{x} と区別するため,(次元数が N の) \mathbf{x} には 違った書体を用いる.