はじめに

本書は，大学学部の２年生から３年生むけの機械学習の教科書である．現在も発展し，深化しつづけている機械学習の全体を網羅することは，（少なくとも筆者には）不可能であり，本書はそれを意図しない．かといって，深層学習を中心にすえ，画像や言語・音声などの処理への対処法を記述したものでもない．また，機械学習に関するプログラミングの技法を紹介するものでもない．

むしろ，本書は古典機械学習ともよぶべき題材にまとをしぼり，考え方をできるだけ詳細に記述した．手法や技術は，その多くが時間とともに陳腐化するのに対し，考え方を学ぶことは，新たな課題に挑戦するときに役にたつと考えるからである．大量のデータが存在する対象，あるいはその近傍の対象に対しては，深層学習はきわめて高性能を発揮する．しかし，少数のデータしか得ることができない対象も多く，本書で紹介する古典的な機械学習の手法は今後も随所で活躍するであろう．とりわけ，ベイズ的な考え方は，予測の損失最小を保証するという意味で重要である．

多くの大学理工系の学部で初年次に学ぶ，多変数の微積分と固有値問題の基本を含む線形代数は既知とした．確率と統計の基本も既習であることがのぞましいが必須ではない．確率と統計や，対称行列に関する固有値問題などの数学的事項の要点は，必要となる箇所の直前でまとめてある．また，行列の微分の公式も第 II 部の後ろに付録としてまとめた．

本書は当初，第 I 部〜第 V 部までを１冊として出版することを意図していたが，読者の便宜を考えて，第 I 部〜第 V 部を分冊化させ，３巻構成とした（電子版では，当初の意図のとおり，分冊化させずに１冊として出版する）．これにより，自身のレベルに応じた必要な箇所が手に取りやすくなることと思う．分冊化にあたっては，各巻の位置づけを明確にしておくために，以下の

ようにそれぞれを掲載している．

- はじめに・記法：全巻共通
- 目次：全巻共通
- ページ番号：全巻通し
- 索引：全巻共通（ただし，該当巻のページ番号は下線を入れて示す）

第 1 巻は，入門的基礎とパラメトリックモデルを含み，第 2 巻は，ノンパラメトリックモデルと潜在モデルで構成され，第 3 巻は，機械学習に必要な数学的基礎事項と演習の解答例からなる．この構成からわかるとおり，機械学習の考え方や理論・モデルに早めに取りくめるよう，本来ならば導入部におくべき確率と統計の入門的事項，および，行列の微分を含むアドバンストな章は，第 3 巻にまわしてある．適宜参照していただいてもいいし，第 1 章の読後に目をとおしていただいてもよい．各章には，少ないながら演習問題を配置し，第 3 巻に，それらに対する詳細な解答例をあげた．演習問題には，本文では省いた重要事項や数式の導出も含まれており，それらについては，本文中の該当するところに演習問題の番号をしるした．

本書の執筆では，多くのすぐれた書物を参照させていただいた．とりわけ，『パターン認識と機械学習［上・下]』（C. M. ビショップ，丸善，2007）の影響は随所にみられると思う．数学的記法も同書に準拠した．本書は，第 I 部 基礎，第 II 部 パラメトリックモデル，第 III 部 ノンパラメトリックモデル，第 IV 部 潜在モデルの 4 部からなっている．この構成は，*Probabilistic Machine Learning: An Introduction* (K. P. Murphy, MIT Press, 2022) に影響をうけている．Murphy の本では，深層学習を 1 つの部としているが，本書では，深層学習の部はもうけず，ニューラルネットワークの基礎的事項をパラメトリックモデルの部へ，また，深層生成モデル（の 1 つである VAE）を潜在モデルの部へおいた．ベイズ推論の重要性に鑑み，潜在モデルを第 4 部としたことは本書の特徴の 1 つである．

数学的事項の復習箇所を講義にいれないのであれば，1 週 90 分 1 コマが 15 週の講義で，若干詰めこみすぎになるが，すべての章を終えることができると考える．比較的高度な話題である「エビデンス近似」や「ガウス過程」，データマイニングなどの授業で講義する可能性のある「主成分分析」など，いくつ

かの章や節を省略すれば余裕をもたせることができると思う．また，1 週 90 分 2 コマが 15 週，あるいは 1 週 90 分 1 コマが 30 週の講義であれば，数学的事項を含め，すべてを丁寧にカバーできるであろう．ただし，章によって長さがかなり異なるので，残念ながら，1 コマの講義は，章ごとの「読み切り」になるわけでなく，章の途中で次回につづくこともある．逆に，1 コマに複数の章がはいる状況も起こると思われる．

TeX による清書や，図表の作成では，関西学院大学工学部課程秘書の堀口恵子さんにお世話になった．また，VAE のプログラムと画像の生成は，関西学院大学理工系研究科修士 1 年の山岡大輝さんにお願いした．共立出版の山内千尋さんには，出版の計画時から世にでるまですべての段階で相談にのっていただいた．いくつかの図の作成には，scikit-learn[1)]の API や例プログラムを，ガウス過程回帰の節の図は，GPy[2)]を利用させていただいた．あわせてお礼申しあげる．

2022 年 7 月

岡留　剛

1) Pedregosa, F., *et al.* (2011). Scikit-learn: Machine Learning in Python, *JMLR*, **12**, pp.2825–2830.

2) http://sheffieldml.github.io/GPy/

記法　Notation

- $\equiv$ は左辺が右辺で定義されることを表わす．たとえば，$n! \equiv n \cdot (n-1)!$, $n > 1$, は，$n!$ が，$n > 1$ なる n に対し，$n \cdot (n-1)!$ で定義されることを表わす．
- イタリック体の小文字（たとえば x）はスカラーを表わす．
- 立体で太字の小文字（たとえば $\mathbf{x}$）は列ベクトルを表わす．ベクトル（や行列）の右肩につけた T は転置を表わし，たとえば，$\mathbf{x}^{\mathrm{T}}$ は行ベクトルとなる．
- この表記のもとで，2 つのベクトル $\mathbf{x}$ と $\mathbf{y}$ の通常の内積は $\mathbf{x}^{\mathrm{T}}\mathbf{y}$ とかける．もちろん，$\mathbf{x}^{\mathrm{T}}\mathbf{y} = \mathbf{y}^{\mathrm{T}}\mathbf{x}$ が成りたつ．
- ベクトル $\mathbf{x}$ に対し，$\|\mathbf{x}\|$ は，そのノルム（大きさ）を表わし，$\|\mathbf{x}\| \equiv \sqrt{\mathbf{x}^{\mathrm{T}}\mathbf{x}}$ で定義される．
- 立体で太字の大文字（たとえば $\mathbf{M}$）は行列を表わす．とくに，$\mathbf{I}$ は単位行列を表わす．また，$\mathbf{M}^{\mathrm{T}}$ は，$\mathbf{M}$ の転置行列を表わす．
- (a, b) は開区間を，$[a, b]$ は閉区間を表わす．ただし，x 座標が a で，y 座標は b の 2 次元平面上の点の座標表示など，実数 a，b の組も (a, b) で表わす．
- 行ベクトルの成分表示は，$(a_1 \ \cdots \ a_D)$ のように，カンマのない表現とする．
- N 個の D 次元ベクトルの観測値 $\mathbf{x}_1, \ldots, \mathbf{x}_N$ に対し，$\mathbf{X}$ は，集合 $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ を表わす．ただし，$\mathbf{X}$ は，第 i 行が $\mathbf{x}_i^{\mathrm{T}}$ である行列を表わすこともある．また，N 個のスカラー観測値をならべた 1 次元ベクトルは，$\mathbf{x}$（N 次元ベクトル；フォント注意）とかく．
- 観測値の集合以外の一般的な集合は，イタリック体の大文字（たとえば S）で表わす．とくに，実数全体の集合は $\boldsymbol{R}$，D 次元実ベクトル全体は $\boldsymbol{R}^D$ で表わす．ただし，データの集合は $\mathcal{D}$ で表わす．
- スカラー値をとる関数はイタリック体（たとえば $f(\mathbf{x})$）で表わす．また，ベクトル値をとる関数は太字（たとえば $\boldsymbol{\phi}(\mathbf{x})$）で表わす．

目　次

第 1 巻　入門的基礎／パラメトリックモデル

第 3 巻　数学事項：機械学習のいしずえ／演習問題解答