

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、バイズ的な考え方は、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、本文中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」になるわけではなく、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年の山岡大輝さんをお願いした。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⁾を利用していただいた。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}^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 $\mathbf{x}^T \mathbf{y}$ とかける。もちろん、 $\mathbf{x}^T \mathbf{y} = \mathbf{y}^T \mathbf{x}$ 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、 $\|\mathbf{x}\| \equiv \sqrt{\mathbf{x}^T \mathbf{x}}$ 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}^T は、 \mathbf{M} の転置行列を表わす。
- (a, b) は开区間を、 $[a, b]$ 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 $\mathbf{x}_1, \dots, \mathbf{x}_N$ に対し、 \mathbf{X} は、集合 $\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ を表わす。ただし、 \mathbf{X} は、第 i 行が \mathbf{x}_i^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 \mathbf{R} 、 D 次元実ベクトル全体は \mathbf{R}^D で表わす。ただし、データの集合は 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 $\phi(\mathbf{x})$ ）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 I 部 入門的基礎	1
第 1 章 機械学習入門	2
1.1 はじめに	2
1.2 データをもとにした予測モデル	2
1.2.1 線形関係の決定問題	2
1.2.2 最小 2 乗法：誤差が最も小さくなるように	3
1.2.3 モデル選択	5
1.2.4 過学習	8
1.2.5 正則化	10
1.2.6 確認用集合	12
1.2.7 モデル選択ふたたび	13
1.3 不確実性への対応：確率の導入	15
1.3.1 確率による予測モデルの表現	20
1.3.2 独立同分布とデータ	21
1.3.3 確率の解釈：頻度主義とベイズ主義	23
1.4 パラメータの推定	24
1.4.1 最尤推定：頻度主義の立場から	24
1.4.2 パラメータの事後分布：ベイズ推論	26
1.5 損失と誤差関数	34
1.6 決定理論	36
1.6.1 回帰への適用	37

viii 目次

1.6.2	分類への適用	38
1.6.3	一般の損失に対する決定理論	41
1.7	評価指標	42
	演習問題	47
第 II 部 パラメトリックモデル		49
第 2 章 確率密度関数の推定：パラメトリック		50
2.1	はじめに	50
2.2	最尤推定による密度推定	55
2.2.1	1次元ガウス分布の最尤推定	55
2.2.2	多次元ガウス分布の最尤推定	56
2.3	ベイズ推定による密度推定	62
2.3.1	事後分布計算の困難性	63
2.3.2	共役事前分布	63
2.3.3	1次元ガウス分布の平均に対するベイズ推論	64
2.3.4	1次元ガウス分布の精度に対するベイズ推論	66
2.3.5	1次元ガウス分布の平均と精度に対するベイズ推論	69
2.3.6	多次元ガウス分布の平均に対するベイズ推論	71
2.3.7	多次元ガウス分布の精度に対するベイズ推論	72
2.3.8	多次元ガウス分布の平均と精度に対するベイズ推論	74
	演習問題	75
第 3 章 線形回帰		79
3.1	はじめに	79
3.2	単純線形回帰の拡張	79
3.2.1	入力の多次元化	79
3.2.2	基底関数の導入	80
3.2.3	線形回帰モデル	81
3.3	最尤推定によるモデルパラメータの推定	82
3.3.1	線形回帰モデルの最尤推定	82

3.3.2	ムーア・ベンローズの一般逆行列	85
3.3.3	バイアスパラメータの役割	86
3.3.4	最尤推定解をもちいた予測	86
3.4	ベイズ統計による線形回帰	87
3.4.1	ベイズ線形回帰	87
3.4.2	ベイズ予測分布	93
3.5	ベイズモデル比較	96
3.6	エビデンス近似	100
3.6.1	ベイズ線形回帰予測分布のエビデンス近似	101
3.6.2	エビデンス最大解の性質	104
	演習問題	109
第 4 章	一般化線形モデルによる分類	111
4.1	はじめに：分類問題	111
4.1.1	目標変数値とラベルの表現	112
4.1.2	分類問題へのアプローチ	112
4.2	特徴空間	113
4.3	一般化線形モデル	115
4.4	識別関数による 2 クラス分類	117
4.4.1	線形識別関数による分類	117
4.4.2	決定面の特徴づけ	118
4.4.3	ダミー入力 of 導入	120
4.4.4	パーセプトロン	120
4.5	確率的識別モデルによる 2 クラス分類	122
4.5.1	ロジスティックシグモイド関数	122
4.5.2	ロジスティック回帰	123
4.5.3	ロジスティック回帰の学習	124
4.5.4	最適化計算と凸関数	126
4.5.5	ロジスティック回帰の最尤推定の問題点	132
4.6	確率的生成モデルによる 2 クラス分類	134

x 目 次

4.6.1	クラスの事後確率：一般化線形モデルで表現できる場合	134
4.6.2	決定境界	136
4.6.3	クラスの事前確率の影響	137
4.6.4	確率的生成モデルの学習	138
4.6.5	確率的生成モデルから導かれたロジスティック回帰の 一般性	140
4.6.6	識別モデル vs. 生成モデル：ロジスティック回帰	141
4.7	確率的識別モデルによる多クラス分類	142
4.7.1	ソフトマックス関数	142
4.7.2	多クラスロジスティック回帰モデル	143
4.7.3	多クラスロジスティック回帰モデルの学習	144
4.8	確率的生成モデルによる多クラス分類	145
4.8.1	クラスの事後確率	145
4.8.2	決定境界	145
4.8.3	共分散行列が異なる場合	146
	演習問題	149
第 5 章	ニューラルネットワーク：非線形パラメトリックモデル	152
5.1	はじめに	152
5.2	3層パーセプトロン：ニューラルネットワークの基礎	153
5.2.1	前向き計算：関数としてのニューラルネットワーク	154
5.2.2	ニューラルネットワークの学習	158
5.3	たたみこみニューラルネットワーク	171
5.3.1	たたみこみ層	174
5.3.2	プーリング層	180
5.3.3	CNN の学習	183
5.4	学習の促進	186
	演習問題	189

第 II 部の付録	191
A ガンマ関数	191
B 行列, アドバンスト要点	192
B.1 行列の重要な性質	192
B.2 行列の微分	193
B.3 計画行列をもちいた表現	200

第 2 巻 ノンパラメトリックモデル / 潜在モデル

第 III 部 ノンパラメトリックモデル	201
第 6 章 訓練データ保持型の学習	202
6.1 はじめに	202
6.2 確率密度関数の推定: ノンパラメトリック	202
6.2.1 ヒストグラム密度推定法	203
6.2.2 カーネル密度推定法	204
6.3 Nadaraya-Watson モデル	206
6.4 k 近傍法	210
6.4.1 ノンパラメトリック密度推定の利害得失	214
演習問題	215
第 7 章 カーネル法	216
7.1 はじめに	216
7.2 カーネル関数	216
7.2.1 定義	217
7.2.2 代表的なカーネル	218
7.2.3 新たなカーネルの構築	220
7.3 ガウス過程	221
7.3.1 集合上の確率分布	221

xii 目次

7.3.2	ガウス過程への架け橋	223
7.3.3	ガウス過程	225
7.3.4	ガウス過程による回帰	231
7.4	サポートベクトルマシン	237
7.4.1	マージンの最大化による分類	242
7.4.2	線形分離可能でないデータに対する SVM	249
7.4.3	逐次最小問題最適化法	255
7.4.4	SVM の理論的側面	259
7.4.5	多クラス SVM と SVM 回帰	263
	演習問題	264
第 8 章	アンサンブル学習	267
8.1	はじめに	267
8.2	バギング	268
8.2.1	ブートストラップ法	268
8.2.2	バギング	268
8.3	ランダムフォレスト	269
8.3.1	決定木	269
8.3.2	ランダムフォレスト	277
8.4	アンサンブル学習の期待損失	280
	演習問題	282
第 IV 部	潜在モデル	285
第 9 章	次元圧縮	286
9.1	はじめに	286
9.2	主成分分析	287
9.2.1	PCA の定式化	290
9.2.2	アルゴリズム	292
9.2.3	分散最大化としての PCA	297
9.2.4	高次元データのあつかい	297

9.2.5 次元の決定	299
9.3 t-SNE : t分布確率的近傍埋めこみ	300
9.3.1 t-SNE	304
演習問題	310
第 10 章 混合ガウス分布と EM アルゴリズム	313
10.1 はじめに	313
10.2 混合ガウス分布	314
10.3 潜在変数	317
10.3.1 潜在変数による混合要素の指定	317
10.3.2 潜在変数による混合ガウス分布の表現	318
10.3.3 負担率 : 潜在変数の事後確率	320
10.4 EM アルゴリズム	321
10.4.1 混合ガウス分布の最尤推定の困難性	321
10.4.2 完全データの尤度	322
10.4.3 EM アルゴリズム : 考え方	323
10.4.4 EM アルゴリズム	324
10.5 混合ガウス分布のパラメータ推定	324
10.5.1 E ステップ	324
10.5.2 M ステップ	327
10.5.3 まとめ	329
10.5.4 混合分布のパラメータ推定における特異性の問題	330
10.6 EM アルゴリズムの適用性と収束性	331
10.6.1 EM アルゴリズムの適用性	331
10.6.2 EM アルゴリズムの収束性	333
演習問題	336
第 11 章 深層生成モデル	338
11.1 はじめに	338
11.2 自己符号化器	338
11.3 変分自己符号化器	342

xiv 目次

11.3.1	変分自己符号化器の構成	342
11.3.2	変分自己符号化器の学習	343
11.3.3	VAE の特徴	348
	演習問題	350

第3巻 数学事項：機械学習のいしずえ／演習問題解答

第V部 数学事項：機械学習のいしずえ 351

第12章 確率・統計ダイジェスト 352

12.1	確率	352
12.1.1	日常生活で現われる確率の例	352
12.1.2	標本空間	352
12.1.3	事象	353
12.1.4	確率の公理	354
12.1.5	確率の解釈	356
12.2	確率変数	356
12.2.1	複数の確率変数	362
12.2.2	確率変数の期待値と分散	376
12.2.3	確率変数の関数	378
12.2.4	変数変換	382
12.3	ガウス分布（正規分布）	385
12.3.1	1次元ガウス分布	385
12.3.2	2次元ガウス分布	386
12.3.3	多次元ガウス分布	388
12.3.4	確率変数の分布表記	388
12.4	サンプル	388
12.4.1	サンプル（標本）	388
12.4.2	橋渡し：数値と確率変数の間で	389

12.4.3	確率変数としてのサンプル	390
12.4.4	統計量	391
12.4.5	推定量	391
12.4.6	大数の法則：統計量に関する重要定理 I	392
12.4.7	中心極限定理：統計量に関する重要定理 II	393
付 記		394
演習問題		397
第 13 章	ガウス分布の性質	400
13.1	確率密度関数であること：1次元	400
13.2	1次元ガウス分布の期待値と分散	401
13.3	確率密度関数であること：多次元	402
13.4	多次元ガウス分布の期待値と分散	405
13.5	分割多次元ガウス分布	408
13.5.1	分割多次元ガウス分布：条件つき分布	408
13.5.2	分割多次元ガウス分布：周辺分布	414
13.5.3	分割多次元ガウス分布：ベイズの定理	418
演習問題		423
第 14 章	行列, アドバンスト	426
14.1	行列, アラカルト	426
14.1.1	ベクトル・スカラー・行列	426
14.1.2	逆行列	427
14.1.3	転置行列とその重要な性質	427
14.1.4	行列のランク (階数)	427
14.1.5	行列式	428
14.1.6	余因子行列	429
14.1.7	分割行列・ブロック行列	430
14.1.8	トレース	431
14.2	行列の微分	433
14.2.1	なぜ行列の微分が必要なのか	433

xvi 目 次

14.2.2	行列の微分 3 種	434
14.2.3	対称行列の微分	446
14.3	計画行列	446
14.3.1	計画行列 (design matrix)	446
14.3.2	計画行列による和の表現	447
14.4	実対称行列の固有値問題のまとめ	449
演習問題解答		454