

■ 監修にあたって

目下、人工知能ブームの真っ只中です！ 情報・ウェブ系の企業は当然として、情報技術と直接関係のないさまざまな企業も人工知能に興味をもち、ネットニュースや新聞にも毎日のように人工知能の文字が飛び交っています。また、学术界においても、情報系だけでなく、理学、工学、医学、農学、薬学などのさまざまな理系分野で人工知能に関する研究が行われており、さらには、法学、倫理学、経済学などの文系分野でも、人工知能に関する議論が活発化しています。このような人工知能分野の盛り上がりを受け、私自身もこの分野の研究者の1人として、非常にエキサイティングな毎日を過ごしています。

しかし、世間で見かける人工知能に関する情報の多くは、人工知能によって何ができるようになるか、そして、人工知能は私達の日常生活にどんな影響を与えるか、といった非技術系のものです。一方、本書を手にした好奇心の豊かな皆さんは、そのような人工知能がどのように作られているのか、という人工知能の技術的な側面に興味をおもちだと思います。本書は、そのような皆さんの知的好奇心を満たし、実用的な技術の獲得を支援します。

現在の最先端の人工知能は、コンピュータにヒトのような学習能力をもたせる機械学習と呼ばれる技術によって支えられています。機械学習の研究はこの20年余りで飛躍的に発展し、カーネル法、ベイズ推論、深層学習など、実用的な技術が開発されました。近年、これらの技術の詳細を記したさまざまな和書が出版され、日本語で機械学習の専門知識を得る環境が整いつつあります。しかし、これらの専門書は数学的に高度な内容を含むことが多く、この分野への大きな参入障壁となっています。

本書はベイズ推論に特化した入門書であり、機械学習の基礎から先端的なベイズ推論アルゴリズムの詳細まで、わかりやすく解説しています。大学の教養レベルの数学の知識さえあれば、高度な機械学習手法の数学的な導出を1つ1つ追っていくことができ、その原理を理解することができます。自分が学生の頃にこのような教科書があったら良かったのに、と感じさせる一冊です。

2017年6月

理化学研究所 革新知能統合研究センター センター長
東京大学 大学院新領域創成科学研究科 教授
杉山 将

■ まえがき

近年、計算機と通信技術の飛躍的な進歩により、いまだかつてないほどの多次元・多量のデータが取り扱える環境が整ってきました。それに伴って、自動車や工場などに取り付けられている各種のセンサーから取得されるデータを解析して機器の異常を検知したり、ウェブ上に蓄積されているテキストデータの内容を要約することによってマーケティング戦略に生かすなど、実践的なデータ解析に関する要望が非常に高まっています。一方で、こういったデータは実態としては単なる整然とした数値の羅列であるため、何かしらの「知的な」処理を施さなければ有益な解析を行うことができません。このような背景から、機械学習と呼ばれる技術はデータに対して「意味」を抽出し、まだ観測されていない現象に対して「予測」を行うことを目的とする方法論として発展してきました。

しかし、このような大量データを使った解析に対する要望の高まりと技術の目覚ましい進歩があるにもかかわらず、機械学習を駆使してうまく現実の課題解決に取り組むことのできる技術者はそれほど多くはないのが現状です。この背景には、機械学習における基礎研究が急速な勢いで発展してきているために、応用分野における技術習得が間に合っていないという現実があるのはもちろんですが、その一方で機械学習という技術領域自体が多様な「ツール群」あるいは「アルゴリズム群」として認識されているために、各構成技術をバラバラに勉強し理解しなければならないということにも原因があると思われます。新しいアルゴリズム、新しい方法論が国際学会や産業界で毎年のように開発されているため、技術者は多大な時間をかけて1つ1つのアルゴリズムを勉強することに終始してしまい、データを使った問題解決に関する本質的な原理・原則を理解するのは非常に困難な状況になっています。

本書は、このような問題意識を起点として執筆されたデータ解析手法の解説書です。データ解析の分野は歴史的な経緯からいくつかのコミュニティが存在しますが、現在では対象とする課題設定から大まかに統計学と機械学習に分けられ、さらにデータから推定・予測を行う方法論として頻度主義とベイズ主義に分けられます。本書では、これらのコミュニティの中でも近年特に注目を集めているベイズ主義機械学習（ベイズ学習）に基づいた実践的な

データ解析アルゴリズムの構築法に関して解説をします。また、既存の機械学習の参考書の多くが1つ1つのツールあるいはアルゴリズムの動作原理と使い方に焦点を当てている一方で、本書は技術者が自らアルゴリズムを「デザイン」することに重点をおいた非常にユニークな内容になっています。作曲家は頭に思い描いた美しいメロディーを譜面に落とし込むことができます。また、優れたプログラマーは自分が実現したい機能をロジカルにソースコードに書き上げていくことができます。それとまったく同じようなことが機械学習アルゴリズムの開発においても可能であり、データを使って実現したい目標やデータの特性をある程度把握できれば、あとは確率モデリングと推論アルゴリズムを使った一貫したアプローチにより解決法を導くことができるようになります。さらに、解いている課題に合わせて既存のアルゴリズムに「ちょっとした工夫」を加えて予測精度や計算速度を向上させたり、同じモデルで複数のアプリケーションを同時に実現するなどの柔軟な拡張も可能になります。

本書では次のような方を標準的な読者対象としています。

- これから機械学習やベイズ学習を研究で使いたい学生
- データに関わる業務や基礎科学に携わっており、最新の機械学習技術を使いこなしたい技術者および研究者
- すでにいくつかの機械学習技術を使っているが、もっと問題に合わせて自由にアルゴリズムを構築・改良したい技術者および研究者

本書では、線形代数、微分積分、統計学やプログラミングなどといった、主に理数系の大学1、2年生で習うような基本的な数学とコンピュータの知識が必要になってきます。ただし、必ずしもこれらを完全に理解した状態で読み始める必要はなく、内容の理解に詰まった際にこれらに関する教科書を適宜参考にする程度でよいでしょう。本書の数式の量は比較的多いような印象を受けられるかもしれませんが、これはアルゴリズムの導出にあたってしつこいくらい1つ1つ丁寧に計算手順を書き下しているためです。また、本書ではベイズ学習の基本原則に則り、一貫して「モデルの構築→推論の導出」という手順でアルゴリズムを作っていきます。計算が得意な方から見ると繰り返しばかりが目立ってしまうかもしれませんが、これはこのような丁寧な式展開が行われていることと、アルゴリズム導出の基本ステップが一貫していることが理由でもあります。

本書では、最終的に「目的や状況に合わせて自由にアルゴリズムを作れる

ようになること」を目標とし、余計な脱線はなるべく省き、基礎となる考え方を1つ1つ順を追って説明していきます。したがって、本書はリファレンス的に利用することも可能ではありますが、基本的には最初の章から順番に読み進めていくことを想定して構成されています。図0.1に本書で学ぶ内容や手順の概要を示しています。1章では、機械学習や確率推論の概要、アルゴリズムを構築するうえで基本となる確率計算、グラフィカルモデル、意思決定などの重要な概念を解説します。2章および3章では、学習アルゴリズムをデザインするために必須のパーツとなる各種の確率分布（ガウス分布など）を紹介し、さらにそれらを使って回帰（連続値の予測）などのもっともシンプルな学習アルゴリズムを作っていくことにします。4章では、2章で紹介した種々の確率分布を使って、混合モデルと呼ばれる少しだけ複雑な確率モデルを構築してみます。また、混合モデルから計算効率の高い推論アルゴリズムを導くために、ギブスサンプリングや変分推論（変分ベイズ）と呼ばれる実用性の高い近似アルゴリズムを導入します。続く5章はここまでで紹介した方法をふんだんに使ったショーケースのような章で、4章で紹介したモデルの構築手段や近似アルゴリズムの導出方法をそのまま幅広い問題に適用していきます。ここでは、画像圧縮や時系列モデル、自然言語処理、推薦技術、ニューラルネットワークなど、近年急速に応用が進められている各種アルゴリズムをすべてベイズ推論の枠組みで導出してみます。これら全体の

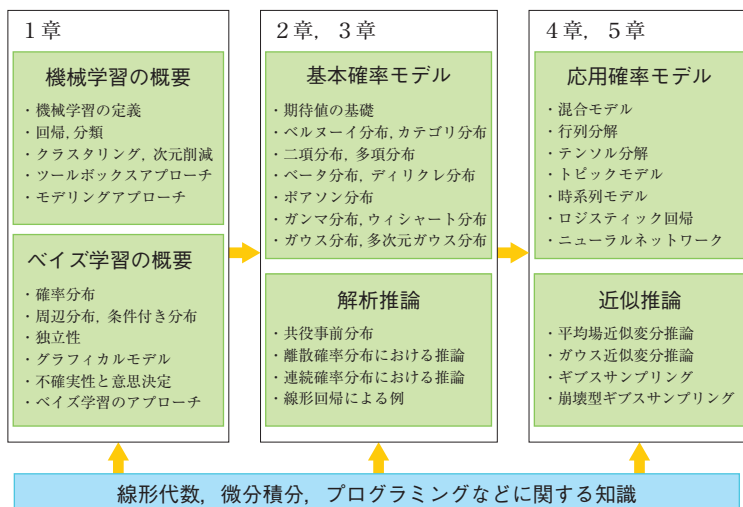


図0.1 本書でカバーする内容.

章を通して土台となる数学（線形代数，微積分）やプログラミングなどの実装手段に関しては，補足が必要であると思われる部分のみ付録や脚注，コーヒブレイクなどで適宜説明していきます。

一方，本書では説明を簡略化したい都合から，誤解を招かない範囲で数学的に厳密な議論は避けていることに注意してください。また，強化学習やベイジアンノンパラメトリクスの理論，深層学習に関わる最新のテクニックは応用上非常に重要ではありますが，本書では詳細には踏み込まず，関連のある個所で適宜触れる程度に留めています。この理由としては，本書が内容の網羅性よりも，アプローチの一貫性を保持することと，最小限の知識から最大限の応用範囲をカバーすることを目指しているためです。

本書では次のような記号表記を使っています。ただし，文脈によっては例外や新しい表記を設ける場合もあるので，その都度，章や節のはじめの文字定義をご確認ください。

- 1 次元の値は通常の字体 x を用います。ベクトルや行列など，複数の値を内部にもっていることを強調したい場合は \mathbf{x} や \mathbf{X} などの太字を用います。また，ベクトルは縦ベクトルとして扱い，要素を例えば $\mathbf{x} = (x_1, x_2, x_3)^\top$ といったように表します。ゼロベクトルも太字を使って $\mathbf{0}$ と表記します。
- 波括弧は集合を表します。例えば， N 個のベクトル $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ の集合を $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ と表します。
- \mathbb{R} は実数の集合で， \mathbb{R}^+ は非負の実数の集合を表します。
- 开区間は括弧を使って表すことにします。例えば，実数 a が $0 < a < 1$ を満たすことを $a \in (0, 1)$ と表します。
- 有限個の離散シンボルは波括弧を使って表すことにします。例えば， a が 0 または 1 の 2 値しかとらない場合は $a \in \{0, 1\}$ と表します。
- サイズが D の実ベクトルを $\mathbf{x} \in \mathbb{R}^D$ と表します。また，サイズが $M \times N$ の実行列を $\mathbf{A} \in \mathbb{R}^{M \times N}$ と表します。
- 行列 \mathbf{A} の転置は \mathbf{A}^\top と表現します。
- D 次元の単位行列を \mathbf{I}_D と表記します。
- e は自然対数の底であり，指数関数は $\exp(x) = e^x$ のように表記します。
- 各種確率分布は，ガウス分布 \mathcal{N} ，ベルヌーイ分布 Bern ，二項分布 Bin ，カテゴリ分布 Cat ，多項分布 Mult ，ポアソン分布 Poi ，ベータ分布 Beta ，ディリクレ分布 Dir ，ガンマ分布 Gam ，ウィシャート分布 \mathcal{W} などと表

記します.

- 分布 $p(x)$ からサンプルを得ることを $x \sim p(x)$ と表記します.
- 記号 \approx は近似を意味します. 例えば, $\mathbf{X} \approx \mathbf{Y}$ は, 「 \mathbf{X} を \mathbf{Y} で近似する」という意味になります.
- 式中の const. は計算に不必要な項をすべてまとめたものです. 例えば $x \propto y$ の対数を $\ln x = \ln y + \text{const.}$ と表記したりします.
- 式中の s.t. は subject to の略で, 「 \sim の条件で」といった意味になります.

本書で紹介する一部のアルゴリズムのソースコードや内容の正誤表に関しては, 下記のページで公開する予定です.

<https://github.com/sammy-suyama/>

本書を執筆するにあたって, 東京大学大学院教授・理化学研究所革新知能統合研究センター長の杉山 将先生には, 全体の構成から詳細な数式・用語の確認に至るまで丁寧にご指導いただきました. ケンブリッジ大学教授のズービン・ガラマーニ先生には, 本書の執筆開始時から盛り込むべき内容を一緒に議論していただき, 理論的な側面から数多くのアドバイスをいただきました. また, 市川 清人氏, 伊藤 真人氏, 小山 裕一郎氏, 近藤 玄大氏, 只野 太郎氏, 鶴野 瞬氏, フェイファン・チェン氏, 寶理 翔太郎氏には, 原稿の改善に多大なるご協力をいただきました. 最後に, 講談社サイエンティフィクの横山 真吾氏には, 執筆作業全般にわたって大変お世話になりました. 皆様の熱心なご支援がなければ本書の刊行は成しえなかったと思います. 心より感謝申し上げます.

2017 年 7 月

須山 敦志

目次

■ 監修にあたって	iii
■ まえがき	iv
第 1 章 機械学習とベイズ学習	1
1.1 機械学習とは	1
1.2 機械学習の代表的なタスク	2
1.2.1 回帰	2
1.2.2 分類	4
1.2.3 クラスタリング	5
1.2.4 次元削減	6
1.2.5 その他の代表的なタスク	8
1.3 機械学習の 2 つのアプローチ	8
1.3.1 ツールボックスとしての機械学習	9
1.3.2 モデリングとしての機械学習	10
1.4 確率の基本計算	12
1.4.1 確率分布	12
1.4.2 確率分布の推論	14
1.4.3 赤玉白玉問題	15
1.4.4 観測データが複数個ある場合	18
1.4.5 逐次推論	21
1.4.6 パラメータが未知である場合	22
1.5 グラフィカルモデル	23
1.5.1 有向グラフ	23
1.5.2 ノードの条件付け	25
1.5.3 マルコフブランケット	28
1.6 ベイズ学習のアプローチ	29
1.6.1 モデルの構築と推論	29
1.6.2 各タスクにおけるベイズ推論	30
1.6.3 複雑な事後分布に対する近似	34
1.6.4 不確実性に基づく意思決定	36
1.6.5 ベイズ学習の利点と欠点	38
第 2 章 基本的な確率分布	44
2.1 期待値	44
2.1.1 期待値の定義	44
2.1.2 基本的な期待値	45
2.1.3 エントロピー	46
2.1.4 KL ダイバージェンス	47
2.1.5 サンプリングによる期待値の近似計算	47
2.2 離散確率分布	48
2.2.1 ヘルヌーイ分布	48
2.2.2 二項分布	51
2.2.3 カテゴリ分布	52

	2.2.4 多項分布	53
	2.2.5 ポアソン分布	55
2.3	連続確率分布	56
	2.3.1 ベータ分布	57
	2.3.2 ディリクレ分布	58
	2.3.3 ガンマ分布	60
	2.3.4 1次元ガウス分布	62
	2.3.5 多次元ガウス分布	64
	2.3.6 ウィンシャート分布	68

第 3 章	ベイズ推論による学習と予測	71
3.1	学習と予測	71
	3.1.1 パラメータの事後分布	71
	3.1.2 予測分布	72
	3.1.3 共役事前分布	74
	3.1.4 共役でない事前分布の利用	75
3.2	離散確率分布の学習と予測	76
	3.2.1 ベルヌーイ分布の学習と予測	76
	3.2.2 カテゴリ分布の学習と予測	82
	3.2.3 ポアソン分布の学習と予測	85
3.3	1次元ガウス分布の学習と予測	87
	3.3.1 平均が未知の場合	87
	3.3.2 精度が未知の場合	91
	3.3.3 平均・精度が未知の場合	94
3.4	多次元ガウス分布の学習と予測	97
	3.4.1 平均が未知の場合	98
	3.4.2 精度が未知の場合	100
	3.4.3 平均・精度が未知の場合	102
3.5	線形回帰の例	104
	3.5.1 モデルの構築	104
	3.5.2 事後分布と予測分布の計算	107
	3.5.3 モデルの比較	108

第 4 章	混合モデルと近似推論	115
4.1	混合モデルと事後分布の推論	115
	4.1.1 混合モデルを使う理由	116
	4.1.2 混合モデルのデータ生成過程	117
	4.1.3 混合モデルの事後分布	120
4.2	確率分布の近似手法	121
	4.2.1 ギブスサンプリング	121
	4.2.2 変分推論	124
4.3	ポアソン混合モデルにおける推論	128
	4.3.1 ポアソン混合モデル	128
	4.3.2 ギブスサンプリング	130
	4.3.3 変分推論	133
	4.3.4 崩壊型ギブスサンプリング	137
	4.3.5 簡易実験	142
4.4	ガウス混合モデルにおける推論	144
	4.4.1 ガウス混合モデル	145

4.4.2	ギブスサンプリング	145
4.4.3	変分推論	149
4.4.4	崩壊型ギブスサンプリング	153
4.4.5	簡易実験	156
第 5 章 応用モデルの構築と推論 161		
5.1	線形次元削減	161
5.1.1	モデル	162
5.1.2	変分推論	163
5.1.3	データの非可逆圧縮	166
5.1.4	欠損値の補間	168
5.2	非負値行列因子分解	170
5.2.1	モデル	171
5.2.2	変分推論	174
5.3	隠れマルコフモデル	177
5.3.1	モデル	178
5.3.2	完全分解変分推論	182
5.3.3	構造化変分推論	186
5.4	トピックモデル	191
5.4.1	モデル	191
5.4.2	変分推論	195
5.4.3	崩壊型ギブスサンプリング	198
5.4.4	LDA の応用と拡張	201
5.5	テンソル分解	202
5.5.1	協調フィルタリング	202
5.5.2	モデル	205
5.5.3	変分推論	206
5.5.4	欠損値の補間	213
5.6	ロジスティック回帰	215
5.6.1	モデル	215
5.6.2	変分推論	216
5.6.3	離散値の予測	219
5.7	ニューラルネットワーク	221
5.7.1	モデル	221
5.7.2	変分推論	223
5.7.3	連続値の予測	225

付録 A	計算に関する補足	227
A.1	基本的な行列計算	227
A.1.1	転置	227
A.1.2	逆行列	227
A.1.3	トレース	228
A.1.4	行列式	229
A.1.5	正定値行列	229
A.2	特殊な関数	229
A.2.1	ガンマ関数とディガンマ関数	229
A.2.2	シグモイド関数とソフトマックス関数	230
A.3	勾配法	231
A.3.1	関数の勾配	231

	A.3.2 最急降下法	232
	A.3.3 座標降下法	232
A.4	周辺尤度の下限	233
	A.4.1 周辺尤度と ELBO	233
	A.4.2 ポアソン混合分布の例	234
■	参考文献	237
■	索引	240