

# データ科学のための基礎数理

— 情報数理・確率統計・パターン認識 —

博士(工学) 後藤 正幸 著

コロナ社

## まえがき

近年、人々の生活や組織の諸活動では、あらゆる場面で情報技術が活用されるようになりました。人々はインターネットの SNS（ソーシャルネットワーキングサービス）を通じて人々とオンラインで繋がるだけでなく、さまざまなアプリケーションを使うようになっています。電車に乗る際にも IC カードやスマートフォンを使って駅に入場しますし、小売店でモノを購入する際にもポイントカードやクレジットカードを利用しています。インターネットで宅配サービスを頼むこともできますし、スマートフォンでタクシーを呼ぶこともできます。企業の従業員が自らの仕事を遂行する際には、コンピュータを用いてシステムにログインして作業をしたり、電子メールのほかにチャットアプリを用いて従業員間でコミュニケーションを取るようなことも普通になっています。店舗やイベント会場にはビデオカメラが設置され、店内の顧客の行動が映像で記録されていますし、モノの生産工場もさまざまな機械が情報技術を使って制御されています。これらの情報技術の活用は、あらゆる場面で大規模で多様なデータが記録されることを意味しています。これらの大量のデータをうまく活用し、さまざまな課題解決に結び付けることが期待されており、データサイエンスの重要性が高まっているのです。ビジネス分野におけるデータ活用やデータ分析はビジネスアナリティクスとも呼ばれ、おもにビジネス現場で活躍するデータサイエンティストにとっては必須の技術になりつつあります。近年では、従来の伝統的な統計分析手法だけでなく、パターン認識と機械学習の理論がビジネスアナリティクス分野に適用されつつあり、さまざまな場面で活用されるようになりました。これらの技術はデータ分析のツールとして一般的になりつつあり、データサイエンティストが学んでおくべき内容といえます。

このようなパターン認識と機械学習とその応用を学ぶ書籍としてすでに『入門 パターン認識と機械学習』を発行しましたが、これらの手法の本質を深く理解するためには情報理論、確率論、統計学といった情報数理について基礎を積み上げる必要があります。また、データサイエンティストには、統計学の知識だけでなく、さまざまなデータを分析可能な形に加工したり、形式を変換したりといったデータエンジニアリングの知識も必要です。このデータエンジニアリングの素養はプログラミング力と見なされがちですが、それだけではなく、きちんと情報を扱うための数理を理解しておくことが、本質を理解する上では大変重要です。

「情報数理」というと、さまざまな定義やイメージがあって、一つの共通した枠組みが確立しているわけではありません。情報数理は文字通り解釈すれば“情報を扱うための数理”であるけれども、このような意味でいうと、その範囲はとてつもなく広いものになります。それは、コンピュータで使われる 2 進演算や論理演算も含まれるでしょうし、情報量の概念や情報圧縮の

数理も、確率論や統計学も、アルゴリズムとデータ構造も、システム論やシミュレーション技術も、パターン認識と機械学習も、すべて情報数理に含まれるでしょう。CD や DVD の誤り訂正符号などで重要な有限体（ガロア体）などの代数学も重要な一部です。しかしながら、このような膨大な知識を学ぶことは困難ですので、本書では最終的に、データサイエンティストがパターン認識や機械学習の本質を学ぶために必要となる基礎的な情報数理に絞り、原理についてのイメージの理解、かつ将来に向けた学習意欲を持たせることを到達点として、特に知っておいたほうがよいと考えられる基礎的な事項に絞って解説をしています。

昨今、対話型のチャット AI の飛躍的な性能向上が注目され、これらの生成 AI をいかに活用すべきか、あるいは、その脅威にいかに対応すべきかといった議論が盛んになされていますが、一人のユーザとして AI を活用しているだけでは見えてこない景色もあります。また、これからの AI を含むデータ科学技術の進歩は目覚ましく、10 年後はまったく違った状況になっていることでしょう。それらの技術の根底にある基礎的な理論をしっかり押さえておくことは、非常に速い技術進歩に流されず、本質を理解する上できっと将来に渡って役に立ってくれるはず

です。

本書の内容は、データサイエンスやデータエンジニアリングを少し学んでから、基礎を固めるために勉強してもよいでしょう。しかし、これらの情報数理の基礎は、じつはさまざまな発展的な内容に通じています。まだ、自分自身の専門性の方向が決まっていなくても、本書の内容をしっかり固めておくことは大きな意味があります。本書の内容が、読者の皆様にとって、早い段階で情報数理の基礎を固めることに寄与し、その上に積み上げていく知識や技能が安定していくことに貢献できれば幸いです。

最後となりますが、本書をまとめる機会を与えていただいた株式会社コロナ社の皆様に感謝いたします。また、日頃よりご指導をいただいている早稲田大学理工学術院名誉教授の平澤茂一先生、同大学教授の松嶋敏泰先生、教学や研究面でも大変お世話になっている早稲田大学創造理工学部経営システム工学科の教職員すべての皆様に深く感謝の意を表します。本書の内容は、早稲田大学に着任してからの研究活動や授業での活動を通じて、大学の低学年に学ぶことが望ましい基礎的な情報数理について少しずつまとめてきたものです。その意味で、本書は経営システム工学科と後藤研究室で研究活動に携わってくれている大学院生や学部生の皆さんの協力なくして完成はありませんでした。日頃から著者の研究教育活動を温かく見守ってくれている家族も含め、著者を支えていただいているすべての皆様のご協力で深く感謝いたします。

2023 年 8 月

後藤 正幸

# 目 次

## 1. 多次元関数の基礎

1.1 ユークリッド空間と関数	1
1.2 1次関数と2次関数の勾配ベクトル	3
1.3 2次関数の性質	4
1.3.1 固有値と固有ベクトル	4
1.3.2 2次関数の標準形	5
1.3.3 2次関数の形状	8
1.4 関数の極値	10
1.5 制約付き最適化問題：ラグランジュの未定乗数法	12
1.5.1 制約付き最適化問題の例	12
1.5.2 問題の一般定式化	13
1.5.3 ラグランジュの未定乗数法	13
1.5.4 ラグランジュの未定乗数法の解釈	15
章末問題	16

## 2. 関数の最適化手法

2.1 2次関数の最適化	18
2.2 勾配法と最急降下法	21
2.3 確率的勾配降下法	23
2.4 ニュートン法	25
2.5 共役勾配法	26
2.6 さまざまな改良アルゴリズム	31
章末問題	32

## 3. 確率論の基礎

3.1 事象と確率	33
3.1.1 標本空間と事象	33

3.1.2 $\sigma$ -集合族	35
3.1.3 確率空間	36
3.2 条件付き確率	37
3.3 ベイズの定理	39
章末問題	41

#### 4. 確率変数と確率分布

4.1 確率変数	43
4.1.1 1次元確率変数	43
4.1.2 多次元確率変数	44
4.2 離散型確率変数と確率分布	45
4.2.1 離散型確率変数	45
4.2.2 期待値	46
4.2.3 多次元離散型確率変数	46
4.2.4 代表的な1次元離散型確率分布	47
4.3 連続型確率変数と確率分布	48
4.3.1 連続型確率変数	48
4.3.2 期待値	49
4.3.3 多次元連続型確率変数	49
4.3.4 代表的な1次元連続型確率分布	50
4.3.5 多次元正規分布	52
4.4 期待値と分散の性質	57
章末問題	58

#### 5. 統計的推定と統計的予測

5.1 確率モデル	60
5.1.1 表記法の定義	60
5.1.2 経験分布	62
5.1.3 パラメトリックモデル族	62
5.2 統計的推定と予測	64
5.3 一般的なパラメータの推定法	65
5.4 簡単な予測問題：確率分布が既知のときの予測	69
5.5 ベイズ推定とベイズ予測	75

5.5.1	ベイズの公式	75
5.5.2	パラメータの推定問題	75
5.5.3	統計的予測問題	78
5.6	ベルヌーイ試行のベイズ推定と予測	82
	章末問題	88

## 6. 統計的モデル選択

6.1	具体例	90
6.1.1	統計的仮説検定の問題	90
6.1.2	統計的モデル選択問題としての捉え方	92
6.2	階層モデル族	93
6.3	統計的モデル選択問題	96
6.4	モデル選択規準	98
6.4.1	赤池情報量規準 (AIC)	99
6.4.2	Schwarz の BIC	101
6.4.3	Rissanen の MDL 規準	102
6.5	モデル選択規準に関するさまざまな議論	103
6.6	クロスバリデーション	104
	章末問題	106

## 7. 統計的決定理論とベイズ決定

7.1	統計的決定の問題	108
7.1.1	具体例	108
7.1.2	一般論	113
7.2	決定理論の幾何的考察	117
7.2.1	リスクセット	118
7.2.2	ベイズ決定の幾何的意味	119
7.2.3	ミニマックス決定の幾何的意味	120
7.3	決定問題の性質	121
7.3.1	ベイズ決定の許容性	121
7.3.2	非ランダム決定の最適性	122
7.3.3	ミニマックス定理	123
	章末問題	124

## 8. 情報表現と2値演算

8.1 2進数と演算	125
8.1.1 デジタル情報と2進数表示	125
8.1.2 2進数と10進数の関係	126
8.1.3 2進数表現の桁数	128
8.2 2進整数の演算	128
8.2.1 整数の加算・減算と補数	128
8.2.2 整数の乗算・除算とシフト演算	131
8.3 2進小数の表現	133
8.4 文字の表現	134
8.5 16進数	134
章末問題	135

## 9. 情報エントロピー論

9.1 情報とはなにか	137
9.2 情報量はどのように測れるか? —自己情報量の概念	138
9.2.1 Hartleyの自己情報量	139
9.2.2 Shannonの自己情報量	140
9.3 エントロピー	141
9.3.1 エントロピーの定義	141
9.3.2 定常無記憶情報源とエントロピー	143
9.4 各種情報量	144
9.4.1 二つの情報源	144
9.4.2 結合エントロピー	145
9.4.3 条件付きエントロピー	145
9.4.4 相互情報量	148
9.5 エントロピーの性質	149
章末問題	151

## 10. 情報源符号化の理論

10.1 情報源モデル	153
-------------	-----

10.1.1	情報源アルファベット	153
10.1.2	情報源モデル	154
10.1.3	定常無記憶情報源	154
10.1.4	マルコフ情報源	155
10.2	情報源符号化と符号長	157
10.2.1	2元符号における情報源符号化の例	158
10.2.2	記号列に対する情報源符号化の例	159
10.3	符号化の持つべき性質	160
10.3.1	一意復号可能な符号	160
10.3.2	瞬時符号	161
10.3.3	瞬時符号の調べ方	162
10.3.4	クラフトの不等式	163
10.4	ハフマン符号	164
10.5	符号化の限界 (符号長の下限・上限と情報源符号化定理)	166
10.5.1	符号長の下限	166
10.5.2	符号長の上限	167
10.5.3	拡大情報源	168
10.5.4	情報源符号化定理	170
10.6	漸近等分割性	170
	章末問題	172

## 11. 通信路符号化の方法

11.1	通信のモデル	175
11.2	通信路のモデル	176
11.2.1	通信路における誤り	176
11.2.2	2元対称通信路 (BSC)	176
11.2.3	2元消失通信路 (BEC)	177
11.3	通信路容量	177
11.3.1	通信路容量の定義	177
11.3.2	2元対称通信路の通信路容量	178
11.3.3	2元対称通信路の通信路容量の計算例	179
11.3.4	通信路符号化定理	179
11.4	誤り訂正符号	181
11.4.1	誤り混入のモデル	181



11.4.2	最も簡単な例：繰り返し符号	182
11.4.3	組織符号とパリティ検査符号	182
11.4.4	ハミング符号	184
11.4.5	誤り検出能力と誤り訂正能力	186
11.4.6	線形符号	187
章末問題		189

## 12. パターン認識と統計的学習の概要

12.1	パターン認識とは	192
12.1.1	パターン認識問題の例	192
12.1.2	パターン認識問題の基本モデル	193
12.1.3	特徴空間と特徴ベクトル	194
12.2	ベイズ決定による最適判別	194
12.2.1	一般論	195
12.2.2	生成モデルと識別モデル	196
12.2.3	正規分布を仮定した生成モデルと識別関数との関係	198
12.3	線形識別関数によるパターン判別	200
12.3.1	線形判別モデルとパーセプトロン	200
12.3.2	パーセプトロンの学習	201
12.4	特徴パターンとの照合によるパターン判別	202
12.4.1	テンプレートマッチング	202
12.4.2	k-最近傍識別法	204
12.5	さまざまな機械学習アルゴリズム	205
章末問題		206
引用・参考文献		208
索引		209

(章末問題の解答は、コロナ社 <https://www.coronasha.co.jp/np/isbn/9784339029376/>内の関連資料に掲載しています。)

# 1 | 多次元関数の基礎

統計モデルや機械学習では、与えられたデータに対してモデルが持つパラメータを調整することでモデルを構築することが多い。その具体的な方法としては、データに対する当てはまりの程度を表した損失関数を定義し、これを最小化、もしくは極小化するパラメータを求めることで、統計モデルを構築する。一般に、パラメータは多次元であることがほとんどであるため、これは多次元関数の最適化問題に帰着する。本章では、このような多次元関数の最適化に必要な数学的な基礎知識について解説する。

## 1.1 ユークリッド空間と関数

一般に、 $d$ 次元のベクトル  $\mathbf{x} = (x_1, x_2, \dots, x_d)^T$  を考え、その各要素である  $x_1, x_2, \dots, x_d$  がそれぞれ実数の値をとり得るものとする。ベクトルの右肩にある  $T$  は転置を表しており、この転置記号を用いずに  $\mathbf{x}$  を記述すると

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{pmatrix} \tag{1.1}$$

と縦ベクトルで定義されていることを意味する。縦ベクトルによる定義は分野の慣習に従っているが、この式の例でわかるように縦ベクトルをそのまま記述すると紙面を要するため、転置記号を用いてベクトルを定義している。

いま、実数全体の集合を  $\mathbb{R}$  と書くものとし、 $d$ 個の実数を要素に持つベクトルの全体の集合、すなわち

$$\mathbb{R}^d = \{(x_1, x_2, \dots, x_d)^T \mid x_1, x_2, \dots, x_d \in \mathbb{R}\} \tag{1.2}$$

を考える。 $\mathbb{R}^d$  に属する任意の  $\mathbf{x} = (x_1, x_2, \dots, x_d)^T \in \mathbb{R}^d$  と  $\mathbf{y} = (y_1, y_2, \dots, y_d)^T \in \mathbb{R}^d$  に対し、これらの距離として

$$d(\mathbf{x}, \mathbf{y}) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_d - y_d)^2} \tag{1.3}$$

が定義されているとき、この  $\mathbb{R}^d$  を  **$d$ 次元ユークリッド空間**という<sup>†</sup>。また、式(1.3)で定義される距離を**ユークリッド距離**という。

ある微分可能な関数  $f(\mathbf{x})$  は、 $\mathbf{x} \in \mathbb{R}^d$  を一つの値に写像する。いま、 $c$  をある定数として、 $f(\mathbf{x}) = c$  で描かれる図形を考えてみる。 $f(\mathbf{x})$  が微分可能であることから、この図形は  $d$  次元空間上の曲面 (2次元の場合は曲線) を表す。ここでは

$$\frac{\partial f(\mathbf{x})}{\partial x_1} = \frac{\partial f(\mathbf{x})}{\partial x_2} = \dots = \frac{\partial f(\mathbf{x})}{\partial x_d} = 0 \tag{1.4}$$

となる点が  $f(\mathbf{x}) = c$  上には存在しないものとする。このとき、 $f(\mathbf{x}) = c$  で表される曲面 (等値面という。2次元の場合は等高線とも呼ばれる) で区分される片側の領域では  $f(\mathbf{x}) > c$ 、反対側の領域では  $f(\mathbf{x}) < c$  である (図 1.1)。特に、閉曲面であれば、その内側と外側で  $f(\mathbf{x}) - c$  の符号が異なることになる。

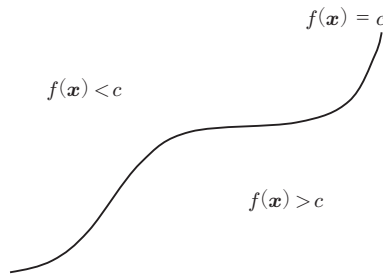


図 1.1  $f(\mathbf{x}) = c$  の曲面で分けられる領域

ある点  $\mathbf{x} \in \mathbb{R}^d$  における**勾配ベクトル**  $\nabla f(\mathbf{x})$  は

$$\nabla f(\mathbf{x}) = \left( \frac{\partial f(\mathbf{x})}{\partial x_1}, \frac{\partial f(\mathbf{x})}{\partial x_2}, \dots, \frac{\partial f(\mathbf{x})}{\partial x_d} \right)^T \tag{1.5}$$

で定義される。 $f(\mathbf{x}) = c$  を満たす点  $\bar{\mathbf{x}}$  における勾配ベクトル  $\nabla f(\bar{\mathbf{x}})$  は、曲面  $f(\mathbf{x}) = c$  に対する**法線ベクトル**となっている。すなわち、ある点  $\bar{\mathbf{x}}$  における勾配ベクトル  $\nabla f(\bar{\mathbf{x}})$  は、その点における接線と直交し (図 1.2)、その点において関数の値が最も増大する方向となっている。逆に、勾配ベクトルにマイナスを付けた  $-\nabla f(\bar{\mathbf{x}})$  は、その点において関数の値が最も減少する方向となっている。そのため、 $-\nabla f(\bar{\mathbf{x}})$  は**最急降下方向**と呼ばれており、関数の極小解を探索的に求めるアルゴリズムにおいて、探索の手がかりとして使われることが多い。

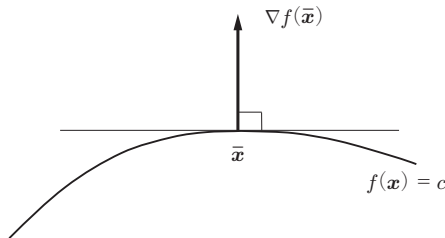


図 1.2  $f(\mathbf{x}) = c$  の勾配ベクトルと接線

<sup>†</sup> ユークリッド空間は、加算やスカラー情報の演算が自然に定義されている**線形空間**の一種であるが、単に線形空間というだけでなく、任意の二つの要素 (ベクトル) の**内積**を計算したり、それらの角度や距離を測ることができる。

## 1.2 1次関数と2次関数の勾配ベクトル

変数の1次の項のみからなる式を1次形式、または1次関数という。 $d$ 次元のベクトル  $\mathbf{x} = (x_1, x_2, \dots, x_d)^T$  の1次式は

$$\begin{aligned} f(\mathbf{x}) &= a_1x_1 + a_2x_2 + \dots + a_dx_d \\ &= \sum_{i=1}^d a_ix_i \end{aligned} \quad (1.6)$$

で与えられる。いま、ベクトル  $\mathbf{a}$  を  $\mathbf{a} = (a_1, a_2, \dots, a_d)^T$  と定義すれば

$$f(\mathbf{x}) = \mathbf{a}^T \mathbf{x} \quad (1.7)$$

と書くことができる。

式(1.6)を  $x_1, x_2, \dots, x_d$  で微分すると、 $\partial f(\mathbf{x})/\partial x_1 = a_1, \partial f(\mathbf{x})/\partial x_2 = a_2, \dots, \partial f(\mathbf{x})/\partial x_d = a_d$  であることから、式(1.7)の勾配ベクトルは

$$\nabla f(\mathbf{x}) = \mathbf{a} \quad (1.8)$$

で与えられる。

一方、つぎのように変数の2次の項のみからなる式を2次形式、または2次関数という。 $d$ 次元ベクトル  $\mathbf{x} = (x_1, x_2, \dots, x_d)^T$  の2次式は

$$\begin{aligned} f(\mathbf{x}) &= a_{11}x_1^2 + a_{12}x_1x_2 + a_{13}x_1x_3 + \dots + a_{21}x_2x_1 + a_{22}x_2^2 + \dots + a_{dd}x_d^2 \\ &= \sum_{i=1}^d \sum_{j=1}^d a_{ij}x_ix_j \end{aligned} \quad (1.9)$$

で与えられる。ただし、 $a_{ij} = a_{ji}$  としても一般性を失わないため、以後はすべての  $i, j = 1, 2, \dots, d$  に対して  $a_{ij} = a_{ji}$  とする<sup>†</sup>。いま、行列  $\mathbf{A}$  を

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1d} \\ a_{21} & a_{22} & \dots & a_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ a_{d1} & a_{d2} & \dots & a_{dd} \end{pmatrix} \quad (1.10)$$

と定義すると

$$f(\mathbf{x}) = \mathbf{x}^T \mathbf{A} \mathbf{x} \quad (1.11)$$

と記述することができる。 $a_{ij} = a_{ji}$  であることから、 $\mathbf{A}$  は対称行列であることに注意する。

<sup>†</sup>  $a_{ij} \neq a_{ji}$  であった場合、新たに  $\bar{a}_{ij} = \bar{a}_{ji} = (a_{ij} + a_{ji})/2$  と計算した  $\bar{a}_{ij}, \bar{a}_{ji}$  を  $a_{ij}$  と  $a_{ji}$  に置き直してもまったく等価な関数が表現できる。

式 (1.9) を  $x_i$  で微分すると,  $a_{ij} = a_{ji}$  であることから

$$\frac{\partial f(\mathbf{x})}{\partial x_i} = 2a_{i1}x_1 + 2a_{i2}x_2 + 2a_{i3}x_3 + \cdots + 2a_{id}x_d = 2 \sum_{j=1}^d a_{ij}x_j$$

であることがわかる。これをまとめてベクトル表現した勾配ベクトルは,  $\mathbf{A}$  を用いて

$$\nabla f(\mathbf{x}) = 2\mathbf{A}\mathbf{x} \quad (1.12)$$

と与えられることがわかる。

### 1.3 2次関数の性質

$d$  次元 2 次関数は, 一般の関数の性質を調べる上でも大変重要である。ここでは, 式 (1.11) で与えられる 2 次関数の性質について述べる。

#### 1.3.1 固有値と固有ベクトル

式 (1.12) から, 式 (1.11) で与えられる 2 次関数は原点  $\mathbf{0} = (0, 0, \dots, 0)^T$  において

$$\nabla f(\mathbf{x}) = \mathbf{0} \quad (1.13)$$

となる。この点が最小値なのか, 最大値なのか, あるいは最小値でも最大値でもないのかという疑問は, 2 次関数の最適化問題を解くという観点から重要である。原点  $\mathbf{0}$  以外の点で  $f(\mathbf{x}) > 0$  となるのであれば, 式 (1.11) で与えられる 2 次関数は原点  $\mathbf{0}$  で最小値をとることになるが, 逆に  $f(\mathbf{x}) < 0$  となるのであれば原点  $\mathbf{0}$  は最大値である。じつは, 多次元 2 次関数の場合, 原点  $\mathbf{0}$  以外の点で  $f(\mathbf{x})$  が正と負の両方の点をとるような関数も作ることができるため, どのような場合に最小値や最大値を持つのかを明らかにしよう。

1 次元の変数  $x \in \mathbb{R}$  の場合

$$f(x) = ax^2 \quad (1.14)$$

という 2 次関数を考えると  $a$  の符号を見ればすべてが解決する。すなわち,  $a > 0$  のとき, 関数  $f(x) = ax^2$  は  $x = 0$  で最小値  $0$  をとる。 $a < 0$  のとき, 関数  $f(x) = ax^2$  は  $x = 0$  で最大値  $0$  をとる。 $a = 0$  のときは, 恒等的に  $f(x) = 0$  となる関数である。式 (1.11) で与えられる 2 次関数についても, 行列  $\mathbf{A}$  によって関数形が規定されることは容易に想像がつくが,  $\mathbf{A}$  は  $d \times d$  行列であるため, 1 次元 2 次関数の場合のように話が単純ではない。そこで, この 2 次関数の性質を捉える上で大変重要になるのが固有値, 固有ベクトルである。

## 固有値と固有ベクトル

$d \times d$  行列  $\mathbf{A}$  ( $d$  次正方行列) に対し,  $\lambda$  をスカラ,  $\mathbf{u}$  を  $d$  次元ベクトルとして

$$\mathbf{A}\mathbf{u} = \lambda\mathbf{u} \quad (1.15)$$

が成り立つとき,  $\lambda$  を  $\mathbf{A}$  の固有値,  $\mathbf{u}$  を  $\mathbf{A}$  の固有ベクトルという。

式 (1.15) より

$$(\mathbf{A} - \lambda\mathbf{I})\mathbf{u} = \mathbf{0}$$

が成り立つ。ただし,  $\mathbf{I}$  は  $d \times d$  単位行列である。これが,  $\mathbf{u} \neq \mathbf{0}$  である解を持つためには,  $\mathbf{A} - \lambda\mathbf{I}$  の行列式が 0 でなければならない。すなわち,  $|\cdot|$  が行列式を表すものとして

$$|\mathbf{A} - \lambda\mathbf{I}| = 0 \quad (1.16)$$

を満たす必要がある。この式 (1.16) を  $\mathbf{A}$  の固有方程式, または特性方程式という。したがって, 実際に固有値と固有ベクトルを求める際には, 式 (1.16) の固有方程式を解けばよい。 $\mathbf{A}$  がその要素がすべて実数である  $d \times d$  対称行列 ( $d \times d$  実対称行列) のとき,  $\mathbf{A}$  は  $d$  個の実数の固有値  $\lambda_1, \lambda_2, \dots, \lambda_d$  を持つことが知られている<sup>†</sup>。また,  $d$  個の固有値  $\lambda_1, \lambda_2, \dots, \lambda_d$  に対応する固有ベクトル  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_d$  は要素がすべて実数で, 互いに直交する。固有ベクトルは方向のみが定められ, 定数倍しても式を満たすが, 大きさ 1 の単位ベクトルになるように選ぶことが一般的である。このように,  $d$  個のベクトル  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_d$  がすべて単位ベクトルで, かつ互いに直交しているとき, これらのベクトルを正規直交系という。

## 実対称行列の固有ベクトルと正規直交系

$d \times d$  実対称行列  $\mathbf{A}$  は重解も含めて  $d$  個の実数の固有値  $\lambda_1, \lambda_2, \dots, \lambda_d$  を持つ。また, 対応する固有ベクトル  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_d$  の大きさを 1 としたとき, これらは正規直交系となる。

## 1.3.2 2次関数の標準形

式 (1.11) の 2 次関数の性質を理解しやすくするために, 行列  $\mathbf{A}$  を変形することを考え, そのために固有ベクトルを活用しよう。いま,  $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_d$  を行列  $\mathbf{A}$  の正規直交系とし, これらを列として並べた  $d \times d$  行列

$$\mathbf{U} = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_d \end{pmatrix} \quad (1.17)$$

を定義する。このように正規直交系を並べた正方行列は直交行列と呼ばれる。 $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_d$  は正規直交系であることから

<sup>†</sup> 対称行列とは限らない一般の  $d$  次正方行列  $\mathbf{A}$  の場合,  $d$  次方程式の解は虚数解や重解も含めて  $d$  個あることから, 固有値は虚数解や重解も含めて  $d$  個存在することになる。

# 索引

<b>【あ】</b>	
赤池情報量規準	99
アナログ情報	125
誤り	176
誤り確率	114, 176
誤り検出	182
誤り検出可能性	187
誤り検出能力	186
誤り訂正	182
誤り訂正可能性	187
誤り訂正能力	186
誤り訂正符号	181
誤り率	70
アルファベット	45, 143
<b>【い】</b>	
イェンセンの不等式	123
一意復号可能な符号	160, 161, 163
一意復号不可能な符号	161
一元配置	91
一様分布	44, 50
一致性	104
一般化加法混合モデル	205
一般化加法モデル	205
一般化線形混合効果モデル	205
一般化線形モデル	205
入れ子型モデル族	93, 95
インジケータ関数	62
<b>【う】</b>	
ウェイト	200
<b>【え】</b>	
エルゴード情報源	154
エントロピー	80, 141, 143
——の性質	149
<b>【お】</b>	
オーバーフィッティング	198
重み	183
重み係数	200
音声認識	192
オンライン学習	24

<b>【か】</b>	
回帰木	205
概収束	51
階層型ニューラルネットワーク	31
階層型のニューラルネットワーク	97
モデル	97
階層ベイズモデル	205
階層モデル族	93, 94
概念	191
カイ 2 乗分布	50
ガウス・ニュートン法	26
ガウスの消去法	25
過学習	97, 206
学習	23, 191
学習アルゴリズム	206
学習機械	206
学習係数	22
学習データ	19
学習率	22
拡大情報源	168
確率	33, 36
——の加法定理	37
——の乗法公式	38
——の 3 公理	36
確率空間	36
確率収束	51
確率測度	36
確率的勾配降下法	24
確率分布	43, 45, 48
確率変数	43
確率密度関数	37, 48, 49
確率モデル	60
確率モデル族	60
下限	116
可算無限アルファベット	153
可算無限集合	33
可算有限アルファベット	153
可算有限集合	33
仮説検定	91
画像生成 AI	206
可測関数	46
可測空間	35
過適合	198

カテゴリ	191
カーネル手法	206
カーネル多変量解析モデル	205
慣性項	31
慣性項付き確率的勾配降下法	31
観測系列	60
ガンマ関数	52, 83

## 【き】

木	162, 165
記憶に基づく学習	204
機械学習	21, 23, 205
機械学習アルゴリズム	205
機械学習モデル	31, 198
棄却	91
擬距離	80, 151
記述長最小規準	103
記述長最小原理	102
期待値	46, 49, 57, 73
基本事象	33
帰無仮説	91, 114
教師付き学習	194
教師なし学習	194
共分散	50, 55, 56
共分散行列	46, 50
共役勾配法	26, 28
共役事前分布	82, 84
行列式	5, 19
極小値	10, 18
極大値	10, 18
極値	10-12
許容性	114
許容的	115, 121
距離	1, 80, 198
寄与率	104

## 【く】

偶重み符号	183
空事象	34
偶数パリティ検査符号	183
組合せ最適化問題	21
クラス	191
クラスタ分析	194
クラスタリング	194

クラフトの不等式 163  
 繰り返し符号 182  
 グリッドサーチ 106  
 クロスバリデーション 104

## 【け】

経験分布 62, 100  
 結合エントロピー 145  
 決定 109, 113  
 決定関数 109, 113  
 決定木 205  
 決定空間 110, 113  
 ゲームの理論 117  
 検査記号 183

## 【こ】

交差検証法 104  
 構造化データ 206  
 行動 109  
 行動空間 113  
 勾配降下法 21  
 勾配ブースティング 205  
 勾配ベクトル 2, 11  
 勾配法 21  
 固有値 5, 19, 54, 55  
 固有ベクトル 5, 54, 55  
 固有方程式 5  
 根元事象 33  
 混合正規分布 197  
 混合モデル 82  
 コンパクト符号 165

## 【さ】

最悪の事前分布 117, 120, 124  
 最急降下法 21  
 最急降下方向 2, 21  
 最近傍識別法 204  
 最小距離 186, 187  
 最小値 4, 10, 19  
 最小 2 乗法 19  
 最大値 4, 10, 19  
 最適化問題 18  
 最適な予測 71, 73, 74  
 最適パラメータ 63, 94  
 最頻値 71, 72, 74, 77, 79  
 最尤推定量 63, 65-67, 76, 96  
 —の漸近正規性 100  
 サポートベクトルマシン 206  
 算術シフト 131  
 算術平均 67

## 【し】

識別 193  
 識別関数 197  
 識別関数法 197  
 識別境界面 201  
 識別モデル 196, 197  
 自己回帰移動平均モデル 97  
 自己回帰モデル 97  
 事後確率 41  
 事後確率最大推定量 96  
 事後確率最大のベイズ識別 195, 196  
 事後確率密度関数 75, 82  
 自己情報量 138  
 事後分布 75, 84, 85  
 事象 33, 34  
 指数分布 52  
 事前確率 41  
 事前確率密度関数 75  
 事前分布 75, 82, 84, 85, 112  
 実数 35  
 シフト演算 131  
 重回帰分析 20, 97  
 重回帰モデル 96  
 集団学習モデル 205  
 自由度 50, 51  
 自由度調整済み寄与率 104  
 瞬時符号 161-163  
 準ニュートン法 26  
 上限 116  
 条件付きエントロピー 145  
 条件付き確率 37  
 条件付きベイズリスク 77, 79  
 冗長度 150  
 情報 137  
 情報圧縮 161  
 情報源 60, 143, 153  
 情報源アルファベット 45, 60, 153  
 情報源記号 143  
 情報源符号化 153, 157, 175  
 情報源符号化定理 170  
 情報源符号器 175  
 情報源モデル 154  
 情報縮約 161  
 情報伝送速度 180  
 情報伝送量 177  
 情報量 138  
 情報理論 60  
 初期値 23  
 初期値依存性 23  
 初期分布 155

人工知能 191, 205  
 深層学習モデル 24, 206  
 深層ニューラルネットワーク 31, 206  
 シンドローム 185, 188  
 シンボル 143

## 【す】

スカラ 45  
 ステップ幅 22  
 スペクトル分解 6

## 【せ】

正規化 193  
 正規直交系 5  
 正規分布 50, 58, 63, 67, 198  
 —の確率密度関数 55  
 生成行列 188  
 生成モデル 196  
 正定値対称行列 7-9, 18, 19  
 正の相関 54  
 制約付き最適化問題 12  
 積事象 34  
 絶対誤差損失 70, 74, 77, 79  
 説明変数 19, 97  
 遷移確率 155  
 遷移確率行列 155  
 全角文字 134  
 漸近正規性 100  
 漸近等分割性 170, 172  
 線形回帰モデル 63, 96  
 線形空間 2  
 線形識別関数 200  
 線形符号 187  
 線形分離可能 202  
 線形変換 54  
 センサ 193

## 【そ】

相関係数 57  
 相互情報量 148  
 組織符号 182  
 損失関数 1, 69, 114, 195  
 損失最小推定量 96

## 【た】

対角化 6  
 大規模言語モデル 206  
 対数尤度関数 64, 66  
 ダイバージェンス 150  
 代表的系列 171  
 代表ベクトル 203



対立仮説 91, 114  
 竹内の情報量規準 101  
 多項分布 47, 63  
 多次元確率変数 44  
 多次元正規分布 52, 56  
 多次元離散型確率変数 46  
 多次元連続型確率変数 49  
 多重共線性 20  
 単回帰分析 19  
 単純マルコフ情報源 155

**【ち】**

知識 137  
 知識情報処理 191  
 中央値 77, 79  
 中間ノード 162  
 中心極限定理 100  
 直線探索 22, 27  
 直交行列 5

**【つ】**

通信のモデル 175  
 通信路 153  
 通信路符号化 153, 175  
 通信路符号化定理 179, 180  
 通信路符号器 175  
 通信路容量 177

**【て】**

デジタル情報 125  
 デジット 135  
 定常確率分布 155, 156  
 定常情報源 154  
 定常無記憶情報源 143, 144, 154  
 低密度パリティ検査符号 183  
 テイラー展開 11  
 停留点 10  
 手書き文字認識 192  
 データ 137  
 データ空間 60  
 データ系列 60  
 テンプレート 203  
 テンプレートマッチング 198, 202, 203

**【と】**

統計的学習 191, 193  
 統計的決定 108  
 統計的推定 64  
 統計的推定問題 65, 76  
 統計的モデル選択 90  
 統計的モデル選択問題 93, 96

統計的予測 64  
 統計的予測問題 78  
 等高線 2  
 同時確率分布 53  
 等値面 2  
 特性方程式 5  
 特徴空間 194  
 特徴抽出 193  
 特徴ベクトル 194  
 特徴量 193  
 独立 38  
 凸集合 118  
 凸部分集合 118

**【な】**

内積 2  
 ナイブベイズ法 197

**【に】**

二項分布 47  
 二重調整済み寄与率 104  
 ニュートン法 25  
 ニュートン・ラフソン法 25  
 ニューラルネットワーク 202, 206

**【の】**

ノード 162  
 ノンパラメトリックな認識 196

**【は】**

バイト 127  
 排反 34  
 背理法 92  
 パーセプトロン 201  
 パターン認識 191-193  
 パターン分類問題 194  
 バックプロパゲーション学習アルゴリズム 31, 206  
 バッチ学習 24  
 葉ノード 162  
 ハフマン木 165  
 ハフマン符号 164, 165  
 ハミング距離 186  
 ハミング符号 184  
 パラメータ 1, 62, 113  
 —の推定問題 75  
 パラメトリックな認識 196  
 パラメトリックモデル族 62, 93  
 パリティ検査行列 185, 188  
 パリティ検査符号 182, 183  
 半角文字 127, 134  
 半順序関係 115

半正定値対称行列 8  
 半負定値対称行列 8

**【ひ】**

非可算無限集合 33  
 非許容的 115  
 非構造化データ 206  
 非瞬時符号 162  
 左シフト 132  
 ビット 127  
 ビットパターン 134  
 非負定値行列 8  
 微分エントロピー 81  
 標準系列 171  
 標準正規分布 50, 53  
 標本空間 33  
 標本分散 68  
 非ランダム決定 122, 123  
 非ランダムリスクセット 118  
 品質特性 90

**【ふ】**

符号 159  
 符号化の限界 166  
 符号木 162  
 符号語 159  
 符号長の下限 166  
 符号長の上限 167  
 負定値対称行列 8, 18, 19  
 負の相関 54  
 不偏推定量 66, 68  
 不偏分散 68  
 フルモデル 92, 97  
 分散 46, 49, 55, 57  
 分散共分散行列 46, 50, 55  
 分散分析 91  
 文書分類 197  
 分布予測問題 78, 80  
 分類 193

**【へ】**

平均 46, 49, 55, 61, 73, 75, 77, 79  
 平均情報量 141  
 平均損失 71, 113  
 平均符号長 158, 159  
 平均予測誤差 114  
 ベイズ解 117  
 ベイズ決定 116, 117, 194  
 ベイズ決定関数 112  
 ベイズ最適化 106

ベイズ最適解 77, 79, 81, 112, 117, 121	ミニマックス解 111, 116, 120, 124	予測問題 69
ベイズ識別 195	ミニマックス規準 116	<b>【ら】</b>
ベイズ情報量規準 102	ミニマックス決定 116	ラグランジュの未定乗数法 12, 13
ベイズ推定 69, 75, 82, 85	ミニマックス戦略 111	ランク 20
ベイズ統計 75	ミニマックス定理 123	ランダム決定 113, 115, 123
ベイズの公式 75	<b>【む】</b>	ランダムフォレスト 205
ベイズの定理 39, 40	無限系列 60	ランレングス・ハフマン符号 166
ベイズ予測 75, 82, 87	無理数 35	<b>【り】</b>
ベイズリスク 76, 116	<b>【め】</b>	離散型確率分布 45
ヘシアン 12	メディアン 73-75, 77, 79	離散型確率変数 45, 46
ベータ分布 83	<b>【も】</b>	リスク関数 71, 110, 114
ヘッセ行列 12, 25	目的関数 21	リスクセット 118
ベルヌーイ試行 47, 82	目的変数 19, 97	リーマン積分 37
<b>【ほ】</b>	文字コード 134	<b>【る】</b>
ポアソン分布 47	文字認識 192	類似度 198
法線ベクトル 2	モデル 93	(累積) 分布関数 48
母集団 60	モデル選択規準 93, 98	ルベグ積分 37, 61, 71, 76, 113
補数 129	モード 71, 72, 74, 77, 79	<b>【れ】</b>
母相関係数 57	モメンタム法 31	連続型確率分布 48
ホールドアウト法 104	<b>【ゆ】</b>	連続型確率変数 48, 49
ボレル集合族 35	優越する 114	連立 1 次方程式 25
<b>【ま】</b>	有効推定量 66, 68	<b>【ろ】</b>
前処理 193	尤度関数 65, 76	ロジスティック回帰モデル 26
マハラノビスの距離 199	有理数 35	論理シフト 131
マルコフ情報源 155	ユークリッド距離 2	<b>【わ】</b>
——のエントロピー 156	ユークリッド空間 2, 35	ワイブル分布 52
マルチコ 20	ユニバーサル符号 166	和事象 34
<b>【み】</b>	<b>【よ】</b>	
右シフト 132	余事象 34	
ミニバッチ 24	予測分布 79, 81-83, 86	
ミニバッチ学習 24		

<b>【A】</b>	<b>【B】</b>	<b>【F】</b>
AdaGrad 31	BART 206	F 値 104
Adam 32	BIC 98, 101, 102	F 分布 91
AI 191, 205	binary erasure channel 177	Fisher 情報行列 100
AIC 98, 99, 104	binary symmetric channel 176	Fletcher-Reeves の式 31
analog information 125	BP 学習 31	<b>【G】</b>
AR モデル 97	<b>【D】</b>	GPT 206
ARMA モデル 97	d 次元確率変数 44	<b>【H】</b>
asymptotic equipartition property 172	d 次元正規分布 56	Hartley の自己情報量 139
Attention 機構 206	d 次元ユークリッド空間 2	
	differential entropy 81	
	digital information 125	

**【I】**

independent and identically distributed sequence	82
independent and identically distributed source	144, 154
i.i.d. 系列	82
i.i.d. 情報源	144, 154

**【K】**

KL 情報量	80, 151
Kullback-Leibler 情報量	64, 80, 151
$k$ -最近傍識別法	198, 204
$k$ -分割交差検証法	105
$k$ -NN 法	198, 204
$k$ -fold cross-validation method	105

**【L】**

$l$ 次の有限マルコフモデル	95
$l$ 次マルコフ情報源	155
LDPC code	183
leave-one-out 交差検証法	105
LOOCV	105
LU 分解法	25

**【M】**

Mahalanobis' distance	199
MDL 規準	98, 102, 103
mixture model	82
modulo two addition	181
mod 2	181

**【P】**

Polak-Ribière-Polyak の式	31
-------------------------	----

**【Q】**

$q$ 元符号	159
---------	-----

**【S】**

Shannon の自己情報量	140, 141
Shannon の補助定理	80, 150
Sorenson-Wolfe の式	31
supervised learning	194

**【T】**

$t$ 分布	51
TIC	101
Transformer	206
tree	162

**【U】**

unsupervised learning	194
-----------------------	-----

**【数字】**

0-1 損失	69, 77, 79, 114
1 次関数	3
1 次形式	3
1 点予測問題	78
1 の補数	130
10 進数	126, 135
16 進数	134, 135
2 元消失通信路	177
2 元対称通信路	176
2 元符号	159
2 次関数	3, 4, 18, 27
——の標準形	7
2 次近似	11, 25
2 次形式	3
2 次元正規分布	52
2 乗誤差損失	70, 73, 77, 79, 114
2 進小数	133
2 進数	126, 135
2 進数表現	128
2 進数変換アルゴリズム	127
2 の補数	129
2 を法とする加法	181
$\chi^2$ 分布	50
$\sigma$ -集合族	35

— 著者略歴 —

1992年 武蔵工業大学（現東京都市大学）工学部経営工学科卒業  
1994年 武蔵工業大学大学院工学研究科修士課程修了（経営工学専攻）  
1997年 早稲田大学助手  
2000年 博士（工学）（早稲田大学）  
2000年 東京大学助手  
2002年 武蔵工業大学助教授  
2008年 早稲田大学准教授  
2011年 早稲田大学教授  
現在に至る

データ科学のための基礎数理

— 情報数理・確率統計・パターン認識 —

Fundamental Mathematics for Data Science

— Information Mathematics, Probability Statistics, Pattern Recognition —

© Masayuki Goto 2023

2023年10月6日 初版第1刷発行

★

検印省略

著者 後藤正幸  
発行者 株式会社 コロナ社  
代表者 牛来真也  
印刷所 三美印刷株式会社  
製本所 有限会社 愛千製本所

112-0011 東京都文京区千石 4-46-10

発行所 株式会社 コロナ社  
CORONA PUBLISHING CO., LTD.

Tokyo Japan

振替 00140-8-14844・電話(03)3941-3131(代)

ホームページ <https://www.coronasha.co.jp>

ISBN 978-4-339-02937-6 C3055 Printed in Japan

(新宅)



JCOPY <出版者著作権管理機構 委託出版物>

本書の無断複製は著作権法上での例外を除き禁じられています。複製される場合は、そのつど事前に、出版者著作権管理機構（電話 03-5244-5088、FAX 03-5244-5089、e-mail: info@jcopy.or.jp）の許諾を得てください。

本書のコピー、スキャン、デジタル化等の無断複製・転載は著作権法上での例外を除き禁じられています。購入者以外の第三者による本書の電子データ化及び電子書籍化は、いかなる場合も認めていません。落丁・乱丁はお取替えいたします。