# 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，大学の低学年に学んでおく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に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



### 2.　関数の最適化手法



### 3.　確 率 論 の 基 礎

## 4.　確率変数と確率分布



## 5.　統計的推定と統計的予測

## 6.　統計的モデル選択



## 7.　統計的決定理論とベイズ決定

## 8.　情報表現と2値演算



## 9.　情報エントロピー論



## 10.　情報源符号化の理論

## 11.　通信路符号化の方法

## 12.　パターン認識と統計的学習の概要



（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となる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 $\boldsymbol{x} = (x_1, x_2, \cdots, x_d)^T$ を考え，その各要素である $x_1, x_2, \cdots, x_d$ 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 $\boldsymbol{x}$ を記述すると

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

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

$$\mathbb{R}^d = \left\{ (x_1, x_2, \cdots, x_d)^T | x_1, x_2, \cdots, x_d \in \mathbb{R} \right\} \tag{1.2}$$

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

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

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[†]。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 $f(\boldsymbol{x})$ は，$\boldsymbol{x} \in \mathbb{R}^d$ を一つの値に写像する。いま，$c$ をある定数として，$f(\boldsymbol{x}) = c$ で描かれる図形を考えてみる。$f(\boldsymbol{x})$ 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

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

となる点が $f(\boldsymbol{x}) = c$ 上には存在しないものとする。このとき，$f(\boldsymbol{x}) = c$ 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 $f(\boldsymbol{x}) > c$，反対側の領域では $f(\boldsymbol{x}) < c$ である（**図 1.1**）。特に，閉曲面であれば，その内側と外側で $f(\boldsymbol{x}) - c$ の符号が異なることになる。

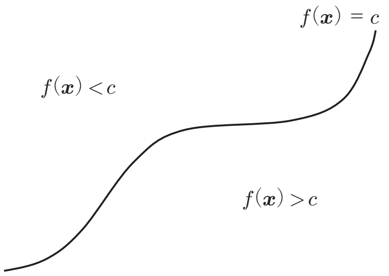


**図 1.1**　$f(\boldsymbol{x}) = c$ の曲面で分けられる領域

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

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

で定義される。$f(\bar{\boldsymbol{x}}) = c$ を満たす点 $\bar{\boldsymbol{x}}$ における勾配ベクトル $\nabla f(\bar{\boldsymbol{x}})$ は，曲面 $f(\boldsymbol{x}) = c$ に対する**法線ベクトル**となっている。すなわち，ある点 $\bar{\boldsymbol{x}}$ における勾配ベクトル $\nabla f(\bar{\boldsymbol{x}})$ 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 $-\nabla f(\bar{\boldsymbol{x}})$ 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，$-\nabla f(\bar{\boldsymbol{x}})$ 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。

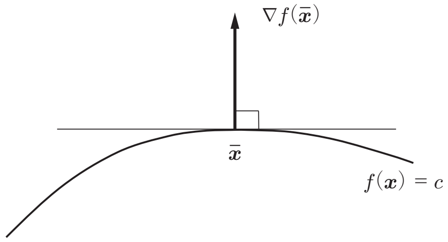


**図 1.2**　$f(\boldsymbol{x}) = c$ の勾配ベクトルと接線

[†] ユークリッド空間は，加算やスカラ倍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 $\boldsymbol{x} = (x_1, x_2, \cdots, x_d)^T$ の 1 次式は

$$f(\boldsymbol{x}) = a_1x_1 + a_2x_2 + \cdots + a_dx_d = \sum_{i=1}^{d} a_ix_i \tag{1.6}$$

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

$$f(\boldsymbol{x}) = \boldsymbol{a}^T\boldsymbol{x} \tag{1.7}$$

と書くことができる。

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

$$\nabla f(\boldsymbol{x}) = \boldsymbol{a} \tag{1.8}$$

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 $\boldsymbol{x} = (x_1, x_2, \cdots, x_d)^T$ の 2 次式は

$$f(\boldsymbol{x}) = a_{11}x_1^2 + a_{12}x_1x_2 + a_{13}x_1x_3 + \cdots + a_{21}x_2x_1 + a_{22}x_2^2 + \cdots + a_{dd}x_d^2 = \sum_{i=1}^{d}\sum_{j=1}^{d} a_{ij}x_ix_j \tag{1.9}$$

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 $i, j = 1, 2, \cdots, d$ に対して $a_{ij} = a_{ji}$ とする[†]。いま，行列 $\boldsymbol{A}$ を

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1d} \\ a_{21} & a_{22} & \cdots & a_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ a_{d1} & a_{d2} & \cdots & a_{dd} \end{pmatrix} \tag{1.10}$$

と定義すると

$$f(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} \tag{1.11}$$

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

---

† $a_{ij} \neq a_{ji}$ 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

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

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，$\boldsymbol{A}$ を用いて

$$\nabla f(\boldsymbol{x}) = 2\boldsymbol{A}\boldsymbol{x} \tag{1.12}$$

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 $\mathbf{0} = (0, 0, \cdots, 0)^T$ において

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

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 $f(\boldsymbol{x}) > 0$ 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 $f(\boldsymbol{x}) < 0$ 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 $f(\boldsymbol{x})$ 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

$$f(x) = ax^2 \tag{1.14}$$

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

$$\boldsymbol{A}\boldsymbol{u} = \lambda \boldsymbol{u} \tag{1.15}$$

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

式 (1.15) より

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

が成り立つ。ただし，$\boldsymbol{I}$ は $d \times d$ 単位行列である。これが，$\boldsymbol{u} \neq \boldsymbol{0}$ である解を持つためには，$\boldsymbol{A} - \lambda \boldsymbol{I}$ 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すると

$$|\boldsymbol{A} - \lambda \boldsymbol{I}| = 0 \tag{1.16}$$

を満たす必要がある。この式 (1.16) を $\boldsymbol{A}$ 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。$\boldsymbol{A}$ がその要素がすべて実数である $d \times d$ 対称行列（$\boldsymbol{d \times d}$ **実対称行列**）のとき，$\boldsymbol{A}$ は $d$ 個の実数の固有値 $\lambda_1, \lambda_2, \cdots, \lambda_d$ を持つことが知られている†。また，$d$ 個の固有値 $\lambda_1, \lambda_2, \cdots, \lambda_d$ に対応する固有ベクトル $\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_d$ 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 $\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_d$ 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 $\boldsymbol{A}$ は重解も含めて $d$ 個の実数の固有値 $\lambda_1, \lambda_2, \cdots, \lambda_d$ を持つ。また，対応する固有ベクトル $\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_d$ 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 $\boldsymbol{A}$ 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，$\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_d$ を行列 $\boldsymbol{A}$ の正規直交系とし，これらを列として並べた $d \times d$ 行列

$$\boldsymbol{U} = \begin{pmatrix} \boldsymbol{u}_1 & \boldsymbol{u}_2 & \cdots & \boldsymbol{u}_d \end{pmatrix} \tag{1.17}$$

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。$\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_d$ は正規直交系であることから

† 対称行列とは限らない一般の $d$ 次正方行列 $\boldsymbol{A}$ 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を含めて $d$ 個存在することになる。

# 索　　引

―― 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。