

# 機械学習のための数学

博士(理学) 飯塚 秀明 [著]

コロナ社

# ま え が き

人工知能 (AI, Artificial Intelligence) は、交通事故の削減や交通の効率化のための自動運転技術、医用画像の解析による病変や疾患の検出、株価予測や不正検知に挙げられるような交通、医療、金融を含むさまざまな分野で活躍している。会話型人工知能や動画生成人工知能の急速な技術革新により、人工知能が我々の生活においてより身近な存在になっている。

人工知能開発のための手法の一つに、機械学習 (machine learning) がある。機械学習は、与えられたデータからルールやパターンを学習し、それらを用いて未知のデータに対する予測や判断を行うことができる。例えば、自動運転技術に関する行動予測においては、過去の交通データを学習することで、他の車両や歩行者の行動を予測するのに機械学習が利用される。

機械学習 [7章] を理解するためには“数学”が必要である。数学の土台は論理 [1章] と集合 [2章] である。機械学習を議論するための土俵がユークリッド空間 [3章] である。機械学習で扱うユークリッド空間の次元数はとてつもなく大きい。例えば、カラー画像のデータセットからルールを学習するための機械学習に現れる空間の次元数は一千万を超える。このことから、一般の次元数を有するユークリッド空間 [3.2節] の性質を理解することが、機械学習を理解するための第一歩といえる。一方で、1次元ユークリッド空間 [3.1節]、つまり、実数直線を軽んじてはいけない。なぜならば、機械学習などに関する数理的手法の性能は数値化して評価されることが多いからである。

4次元ユークリッド空間のベクトルは2行2列の行列 ( $2 \times 2$  行列) として表現することができる。この事実は一般のベクトルと行列に対しても成り立つので、ユークリッド空間 [3章] と行列に関する理論を展開する線形代数 [4章] には密接な関係がある [4.1節]。最適化理論や機械学習を理解するうえで重要

な行列といえば対称行列 [4.1.2 項] であろう。機械学習に現れる関数の 2 回微分係数 [例 4.3(3), 例 4.6(3)] や確率変数の共分散 [例 4.6(4)] は対称行列である。行列特有の距離 (ノルム) の測り方 [4.2 節] は最適化理論や機械学習のための手法の解析で役立つ。機械学習のための手法の多くは, 対象の関数 [7.1 節] の微分情報を利用し, その解析のために積分の概念を利用する [5 章]。ユークリッド空間 [3.2 節] 上で定義される関数の微分係数 (勾配) [5.1 節] は, 勾配降下法 [7.2 節] や確率的勾配降下法 [7.3 節] といった機械学習のための最適化手法を定義するうえで必要な概念である。さらに, 微分可能な凸関数とその 2 回微分係数 (対称行列) の関係 [5.3.2 項] は微分積分と線形代数の架け橋となる結果を示す。確率・統計 [6 章] は, データの確率分布の理解とその推定に必要な概念であり, 確率変数の独立性, および, その期待値と分散の性質 [6.3.2 項] を駆使することで, 機械学習のための最適化手法の解析が可能となる [7.3 節]。機械学習のための最適化手法で利用する特有のパラメータ (ステップサイズとバッチサイズ) の設定による多峰性を有する関数の大域的最小化 [7.4 節] は機械学習に限らず, 最適化理論の観点から見ても興味深い。

本書の読者対象は理工系の大学 2 年生程度の知識をもつ方を想定している。本書が機械学習関連の論文を読み解こうとする方の一助を担うことができれば, 著者にとって大きな喜びである。なお, 章末問題の解答は, 明治大学理工学部情報科学科数理最適化研究室が管理する GitHub Organization の Mathematical Optimization Lab. (<https://github.com/iiduka-researches>) からダウンロードすることができる。

本書の原稿に関して, 適切な指摘と校正に協力いただいた明治大学数理最適化研究室の佐藤尚樹氏と大和田佳生氏に深く感謝する。また, 本書の出版に関して, 大変お世話になったコロナ社の皆様に深く御礼申し上げる。

最後に, 理科系の道を志す我が息子たちにこの本を贈る。

2024 年 6 月 生田にて

飯塚秀明

# 目 次

## 1. 論 理

1.1 命題と論理記号 .....	1
1.1.1 論 理 和 .....	1
1.1.2 論 理 積 .....	2
1.1.3 否 定 .....	3
1.1.4 条 件 命 題 .....	4
1.1.5 双 条 件 命 題 .....	5
1.2 同 値 と 推 論 .....	6
1.2.1 真 理 関 数 .....	6
1.2.2 命 題 の 同 値 .....	7
1.2.3 命 題 の 推 論 .....	8
1.3 命題関数による命題表現 .....	9
1.3.1 全称命題と存在命題 .....	10
1.3.2 全称命題と存在命題の否定 .....	11
1.4 証 明 .....	13
1.4.1 対 偶 法 .....	13
1.4.2 背 理 法 .....	13
1.4.3 反 例 .....	14
1.4.4 数学的帰納法 .....	14
章 末 問 題 .....	14

## 2. 集 合

2.1 集合と演算	15
2.1.1 和 集 合	16
2.1.2 積 集 合	17
2.1.3 補 集 合	18
2.1.4 差 集 合	19
2.2 写像と集合族	20
2.2.1 写像と直積	20
2.2.2 写像の合成	24
2.2.3 集 合 族	24
章 末 問 題	25

## 3. ユークリッド空間

3.1 1次元ユークリッド空間	26
3.1.1 実数全体の集合と上限・下限	26
3.1.2 実数列の収束性	30
3.1.3 実数列の上極限と下極限	37
3.2 有限次元ユークリッド空間	43
3.2.1 ベクトル空間	43
3.2.2 内積とノルム	44
3.2.3 点列の収束性	51
3.2.4 閉集合と開集合	55
3.2.5 連 続 写 像	61
章 末 問 題	71

## 4. 線形代数

4.1 行列全体からなるベクトル空間	73
4.1.1 正方行列	78
4.1.2 対称行列	85
4.1.3 半正定値行列と正定値行列	90
4.2 行列全体からなるユークリッド空間	96
4.2.1 内積とノルム	96
4.2.2 行列点列の収束性	103
章末問題	104

## 5. 微分積分

5.1 微分可能関数	105
5.1.1 偏微分係数・微分可能性・導関数	105
5.1.2 連続的微分可能性・平滑性	114
5.1.3 平均値の定理	119
5.1.4 2次偏微分係数・2回微分可能性・ヘッセ行列	120
5.1.5 2回連続的微分可能性	126
5.2 積分可能関数	128
5.2.1 微分積分学の基本定理	132
5.2.2 Taylor の定理	137
5.3 凸関数	140
5.3.1 微分不可能凸関数と劣勾配	144
5.3.2 微分可能凸関数とヘッセ行列：微分積分と線形代数の架け橋	155
章末問題	161

## 6. 確率・統計

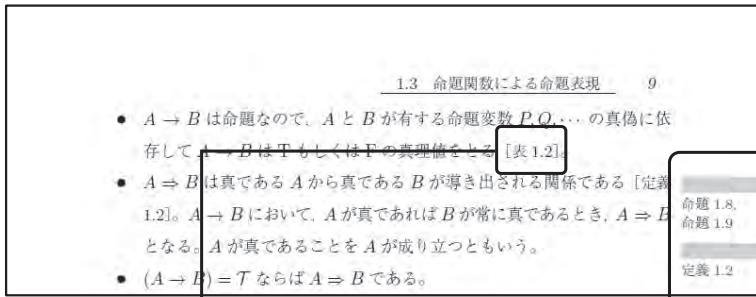
6.1 確率	162
6.1.1 条件付き確率と独立性	164
6.1.2 同時確率と周辺確率	166
6.1.3 事前確率と事後確率	166
6.2 確率分布	167
6.2.1 確率変数と確率分布	167
6.2.2 多次元確率変数と多次元確率分布	170
6.3 期待値・分散・共分散	175
6.3.1 条件付き期待値・条件付き分散	182
6.3.2 確率変数の独立性	186
6.4 統計	190
6.4.1 母数と統計量	190
6.4.2 極限定理	192
6.4.3 推定	194
章末問題	198

## 7. 機械学習と最適化

7.1 確率・統計に基づく経験損失最小化	199
7.2 勾配降下法	205
7.2.1 探索方向	207
7.2.2 ステップサイズ	208
7.3 確率的勾配降下法	210
7.3.1 確率的勾配	210

7.3.2	確率的勾配降下法の構成	212
7.3.3	確率的勾配降下法の利点	214
7.4	経験損失の大域的最小化のための手法	220
7.4.1	減少ステップサイズ	221
7.4.2	増加バッチサイズ	223
引用・参考文献		224
索引		226

## 本書の使い方



見開きページ上にある「式」「命題」「定義」「定理」「例」「注意」を表しています。

この文章で参照しているものを表しています。  
適宜参照しながら読み進めてください。



$$\forall \mathbf{x} \in D \forall \mathbf{y} \in D \forall \mathbf{g}_x \in \partial f(\mathbf{x}) \forall \mathbf{g}_y \in \partial f(\mathbf{y})$$

$$\langle \mathbf{g}_x - \mathbf{g}_y, \mathbf{x} - \mathbf{y} \rangle \geq c \|\mathbf{x} - \mathbf{y}\|^2$$

式 (5.47)

### 5.3.2 微分可能凸関数とヘッセ行列：微分積分と線形代数の架け橋

機械学習に現れる関数は一般には高次元ユークリッド空間の部分集合を定義域にもつ非凸関数 [7.1 節] であるため、これまでに詳解した凸関数の性質や解析を実問題に現れる最適化問題へ適用ができなれないと感じるかもしれない。しかしながら、この命題は真ではないことを示す [7.4.2 項, 注意 7.5]。

命題 5.18,  
命題 5.19

定理 5.8

初めに、2 回連続的微分可能凸関数  $f$  の必要十分条件がヘッセ行列  $\mathbf{H}_f(\mathbf{x})$  ( $= \nabla^2 f(\mathbf{x})$ ) の性質により表現できることを示す。定理 5.8 の証明は Taylor の定理 [定理 5.6] を利用する。

**定理 5.8** (2 回連続的微分可能凸関数の必要十分条件)  $c > 0$  とする。ユークリッド内積 [式 (3.11)] から誘導されるユークリッドノルム [式 (3.15)] を有する  $d$  次元ユークリッド空間  $\mathbb{R}^d$  の開凸集合  $D$  で定義される  $C^2$  級関数  $f: D \rightarrow \mathbb{R}$  に対して、(1)~(3) が成り立つ。

- (1) [凸関数の必要十分条件]  $f$  が凸  $\Leftrightarrow \forall \mathbf{x} \in D (\mathbf{H}_f(\mathbf{x}) \in \mathbb{S}_+^d)$
- (2) [狭義凸関数の十分条件]  $\forall \mathbf{x} \in D (\mathbf{H}_f(\mathbf{x}) \in \mathbb{S}_{++}^d) \Rightarrow f$  が狭義凸
- (3) [強凸関数の必要十分条件]  $f$  が  $c$ -強凸  $\Leftrightarrow \forall \mathbf{x} \in D (\mathbf{H}_f(\mathbf{x}) - c\mathbf{I} \in \mathbb{S}_+^d) \Leftrightarrow \forall \mathbf{x} \in D (\lambda_{\min}(\mathbf{H}_f(\mathbf{x})) = c)$

**証明** ヘッセ行列  $\mathbf{H}_f(\mathbf{x})$  は  $\mathbf{H}_f(\mathbf{x}) \stackrel{[定理 5.3(2)]}{=} \nabla^2 f(\mathbf{x}) \stackrel{[定理 5.4]}{\in} \mathbb{S}^d$  を満たす。

(1)  $\Rightarrow$   $\mathbf{x} \in D \stackrel{[命題 3.12]}{\Rightarrow} \exists \delta > 0 (N_2(\mathbf{x}; \delta) \subset D)$  から、 $\mathbf{y} (\neq \mathbf{0}) \in \mathbb{R}^d \Rightarrow \forall \alpha \in (0, \delta/\|\mathbf{y}\|_2) (\mathbf{x} + \alpha\mathbf{y} \in N_2(\mathbf{x}; \delta) \subset D)$  を満たす。Taylor の定理 [定理 5.6] から、 $\tau \in (0, 1)$  が存在して

$$f(\mathbf{x} + \alpha\mathbf{y}) = f(\mathbf{x}) + \alpha \langle \nabla f(\mathbf{x}), \mathbf{y} \rangle_2 + \frac{\alpha^2}{2} \langle \mathbf{y}, \nabla^2 f(\mathbf{z}_\tau(\alpha)) \mathbf{y} \rangle_2 \quad (5.47)$$

を満たす。ただし、 $\mathbf{z}_\tau(\alpha) = \mathbf{x} + \tau\alpha\mathbf{y}$  は  $\mathbf{z}_\tau(\alpha) \rightarrow \mathbf{x} (\alpha \downarrow 0)$  を満たす。

$$\begin{aligned}
 f \text{ が } D \text{ で凸} &\stackrel{[\text{定理 } 5.7(3)]}{\Rightarrow} f(\mathbf{x} + \alpha\mathbf{y}) \geq f(\mathbf{x}) + \alpha\langle \nabla f(\mathbf{x}), \mathbf{y} \rangle_2 \\
 &\stackrel{[\text{命題 } 5.16(1)]}{\Rightarrow} \langle \mathbf{y}, \nabla^2 f(\mathbf{z}_\tau(\alpha))\mathbf{y} \rangle_2 \geq 0 \\
 &\stackrel{[\text{式 } (5.47)]}{\Rightarrow} \langle \mathbf{y}, \nabla^2 f(\mathbf{x})\mathbf{y} \rangle_2 \geq 0 \Leftrightarrow \nabla^2 f(\mathbf{x}) \in \mathbb{S}_+^d \\
 &\stackrel{[\text{例 } 3.5(1)]}{\Rightarrow} \stackrel{[\text{定義 } 5.7]}{\Rightarrow} \stackrel{[\text{定理 } 4.2(1)]}{\Leftrightarrow}
 \end{aligned}$$

$\Leftrightarrow \mathbf{x}, \mathbf{y} \in D$  とする。Taylor の定理 [定理 5.6] から、 $\tau \in (0, 1)$  が存在して、 $\mathbf{z}_\tau = \mathbf{x} + \tau(\mathbf{y} - \mathbf{x}) = \tau\mathbf{y} + (1 - \tau)\mathbf{x} \in D$  は以下を満たす。

$$\begin{aligned}
 f(\mathbf{y}) &\stackrel{[\text{定理 } 5.6]}{=} f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle_2 + \frac{1}{2} \langle \mathbf{y} - \mathbf{x}, \nabla^2 f(\mathbf{z}_\tau)(\mathbf{y} - \mathbf{x}) \rangle_2 \\
 &\geq f(\mathbf{x}) + \langle \nabla f(\mathbf{x}), \mathbf{y} - \mathbf{x} \rangle_2 \stackrel{[\text{定理 } 5.7(3)]}{\Rightarrow} \stackrel{[\text{命題 } 5.16(1)]}{\Rightarrow} f \text{ が } D \text{ で凸} \\
 &\stackrel{[\text{定理 } 4.2(1)]}{\geq}
 \end{aligned}$$

(2), (3) 定理 5.8(1) の証明方法と同様の議論 (Taylor の定理を利用) と命題 5.18(1), 命題 5.19(1) を用いることで、定理 5.8(2), (3) を示すことができる (詳細については、文献13) の演習問題 2.16, 2.17 解答例を参照されたい)。□

定理 5.8 は、2 回連続的微分可能関数の凸性とヘッセ行列の正定値性 [4.1.3 項] が同等の力を有することを示しており、まさに、微分積分と線形代数の架け橋を担っている。

**注意 5.11** : [1 変数凸関数の必要十分条件] 関数  $f: \mathbb{R} \rightarrow \mathbb{R}$  の (1 次) 導関数  $f'$  と 2 次導関数  $f''$  の情報を用いて

$$\begin{aligned}
 f \text{ が凸} &\Leftrightarrow f' \text{ が単調増加 (つまり, } x_1 \leq x_2 \Rightarrow f'(x_1) \leq f'(x_2)) \\
 &\Leftrightarrow f'' \text{ が非負値 (つまり, } f''(x) \geq 0)
 \end{aligned}$$

となる。 $f'$  の単調増加性は、定理 5.7(3) と命題 5.16(2) から得られる

$$\langle \nabla f(\mathbf{x}) - \nabla f(\mathbf{y}), \mathbf{x} - \mathbf{y} \rangle \geq 0$$

の一例 ( $(f'(x_2) - f'(x_1))(x_2 - x_1) \geq 0$ ) であり、 $f''$  の非負値性 (半正定値性) は定理 5.8(1) の一例となる。例えば、1 変数 2 次凸関数  $f(x) = ax^2 + bx + c$  (ただし、 $a \geq 0$ ) において、 $f'(x) = 2ax + b$  は  $x$  に関して単調増加であり、 $f''(x) = 2a$  の固有値  $\lambda$  と固有ベクトル  $\mathbf{y} (\neq 0)$  は以下を満たす [式 (4.10)]。

$$f''(x)\mathbf{y} = \lambda\mathbf{y} \Leftrightarrow 2a\mathbf{y} = \lambda\mathbf{y} \Rightarrow \lambda = 2a \geq 0 \Leftrightarrow f''(x) \in \mathbb{S}_+^1$$

**注意 5.12** : [狭義凸関数のヘッセ行列の非負固有値] 定理 5.8(2) の逆命題が成り立たないことを反例 [1.4.3 項] を用いて示す。つまり,  $\exists \mathbf{x} (\nabla^2 f(\mathbf{x}) \notin \mathbb{S}_{++}^d)$  を満たす狭義凸関数  $f$  が存在することを示す。 $f(x) = x^4$  とすると,  $f'(x) = 4x^3$ ,  $f''(x) = 12x^2$  となる。 $x_1 \neq x_2$  となる  $x_1, x_2 \in \mathbb{R}$  に対して,  $(f'(x_2) - f'(x_1))(x_2 - x_1) = 4(x_2 - x_1)^2(x_2^2 + x_1x_2 + x_1^2) > 0$  から,  $f(x) = x^4$  は狭義凸である [命題 5.18(2)]。一方で,  $\exists x (=0) (f''(x) = 12x^2 = 0)$  から,  $f''(0) \notin \mathbb{S}_{++}^1$  となる。

定理 5.8 を用いて, 凸関数の例を挙げる。

**例 5.6** (凸関数の例 (その 2))

(1) 比例公平関数 [例 5.4(1)] :  $\forall i \in \{1, 2, \dots, d\} (w_i > 0)$  とする。

$\mathbf{x} = (x_1, x_2, \dots, x_d)^\top \in \mathbb{R}_{++}^d$  に対して, 比例公平関数は  $g(\mathbf{x}) = \sum_{i=1}^d w_i \log x_i$  で定義される。このとき,  $f = -g: \mathbb{R}_{++}^d \rightarrow \mathbb{R}$  は狭義凸関数である。

例 5.6  
注意 5.11,  
注意 5.12

(2) 2 次関数 [例 5.1(1), 例 5.2(1)] :  $A \in \mathbb{S}^d$  とする。例 5.1(1) で定義される 2 次関数  $f(\mathbf{x}) = (1/2)\langle \mathbf{x}, A\mathbf{x} \rangle_2 + \langle \mathbf{b}, \mathbf{x} \rangle_2 + c$  の 2 次微分係数

$$\nabla^2 f(\mathbf{x}) \underset{[\text{例 5.2(1)}]}{=} \frac{1}{2}(A + A^\top) \underset{[A \in \mathbb{S}^d]}{=} A$$

により,  $A \in \mathbb{S}_+^d$  のとき,  $f$  は凸である (1 変数 2 次凸関数については, 注意 5.11 を参照)。さらに,  $A \in \mathbb{S}_{++}^d$  のとき,  $f$  は  $\lambda_{\min}(A)$ -強凸, かつ,  $\lambda_{\max}(A)$ -平滑となる。

(3) 連続写像の合成 [例 5.1(2), 例 5.2(2)] : 例 5.1(2) で定義される関数  $f(\mathbf{x}) = (1/2)\|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2$  の 2 次微分係数

$$\nabla^2 f(\mathbf{x}) \underset{[\text{例 5.2(2)}]}{=} A^\top A \underset{[\text{例 4.3(2)}]}{\in} \mathbb{S}^n$$

により,  $A^\top A \in \mathbb{S}_+^n$  のとき,  $f$  は凸であり, また,  $A^\top A \in \mathbb{S}_{++}^n$  のとき,  $f$  は  $\lambda_{\min}(A^\top A)$ -強凸, かつ,  $\lambda_{\max}(A^\top A)$ -平滑となる。

**証明**

(1)  $f$  は  $\mathbb{R}_{++}^d$  で  $C^2$  級なので,  $\forall \mathbf{x} \in \mathbb{R}_{++}^d (\nabla^2 f(\mathbf{x}) \underset{[\text{定理 5.3(2)}]}{=} H_f(\mathbf{x}) =$

$\text{diag}(w_i/x_i^2) \stackrel{[\text{例 } 4.6(2)]}{\Rightarrow} \forall \mathbf{x} \in \mathbb{R}_{++}^d \quad (\nabla^2 f(\mathbf{x}) \in \mathbb{S}_{++}^d) \stackrel{[\text{定理 } 5.8(2)]}{\Rightarrow} f \text{ は狭義凸}$

(2)  $\nabla^2 f(\mathbf{x}) = A \in \mathbb{S}_+^d$  のとき、定理 5.8(1) から、 $f$  は凸である。 $\nabla^2 f(\mathbf{x}) = A \in \mathbb{S}_{++}^d$  のとき、 $\lambda_{\min}(A) > 0$  に注意して

$$\forall \mathbf{x} \quad (\lambda_{\min}(A) \|\mathbf{x}\|_2^2 \leq \langle \mathbf{x}, A\mathbf{x} \rangle_2 \Leftrightarrow \langle \mathbf{x}, (A - \lambda_{\min}(A)I)\mathbf{x} \rangle_2 \geq 0)$$

[式 (4.17)]

$$\stackrel{[\text{定理 } 4.2(1)]}{\Leftrightarrow} A - \lambda_{\min}(A)I \in \mathbb{S}_+^d \stackrel{[\text{定理 } 5.8(3)]}{\Leftrightarrow} f \text{ は } \lambda_{\min}(A)\text{-強凸}$$

となる。例 5.1(1) から、 $f$  の  $\lambda_{\max}(A)$ -平滑性が示される。

(3) 例 5.6(2) と同様の議論により証明ができる。 □

最小化問題の最小解 [定義 3.14] について定義をする。

**定義 5.12** (大域的最小解と局所的最小解) ノルム空間  $\mathbb{R}^d$  で定義される関数  $f: \mathbb{R}^d \rightarrow \mathbb{R}$  に関する最小化問題を考察する<sup>†</sup>。このとき、 $f$  の値を最も小さくするベクトル  $\mathbf{x}^*$ 、つまり

$$\forall \mathbf{x} \in \mathbb{R}^d \quad (f(\mathbf{x}^*) \leq f(\mathbf{x})) \Leftrightarrow \mathbf{x}^* \in \underbrace{\left\{ \mathbf{x}^* \in \mathbb{R}^d : f(\mathbf{x}^*) = \min_{\mathbf{x} \in \mathbb{R}^d} f(\mathbf{x}) \right\}}_{\text{argmin } f(\mathbf{x})}$$

を満たす  $\mathbf{x}^*$  を最小化問題の**大域的最小解** (global minimum solution) という。ベクトル  $\mathbf{x}_*$  が最小化問題の**局所的最小解** (local minimum solution) であるとは、 $\mathbf{x}_*$  の局所的な範囲 [例 3.4(1)] に限った場合、その範囲で  $f$  の値を最も小さくするベクトル  $\mathbf{x}_*$ 、つまり、以下を満たすときをいう。

$$\exists \delta > 0 \quad \forall \mathbf{x} \in \mathbb{R}^d \quad (\mathbf{x} \in N(\mathbf{x}_*; \delta) \Rightarrow f(\mathbf{x}_*) \leq f(\mathbf{x}))$$

定義 5.12 から、最小化問題の大域的最小解はその問題の局所的最小解となる。2 回連続的微分可能関数に関する最小化問題の局所的最小解となるための十分

<sup>†</sup> ある関数  $g: \mathbb{R}^d \rightarrow \mathbb{R}$  の最大化を考察する。このとき、 $f = -g$  で定義される関数  $f$  を最小化することと  $g$  を最大化することは同値である。よって、最適化においては最小化について考察すれば事が足りる。

条件を, Taylor の定理 [定理 5.6] を利用することで定理 5.9 のように与えることができる。

式 (5.48)

**定理 5.9** (局所的最小解のための十分条件) ユークリッド内積 [式 (3.11)] から誘導されるユークリッドノルム [式 (3.15)] を有する  $d$  次元ユークリッド空間  $\mathbb{R}^d$  で定義される  $C^2$  級関数  $f$  の最小化問題を考察する。 $\mathbf{x}_* \in \mathbb{R}^d$  とする。このとき

定義 5.12

定理 5.9

$$\nabla f(\mathbf{x}_*) = \mathbf{D}_f(\mathbf{x}_*) = \mathbf{0} \wedge \nabla^2 f(\mathbf{x}_*) = \mathbf{H}_f(\mathbf{x}_*) \in \mathbb{S}_{++}^d$$

ならば,  $\mathbf{x}_*$  は  $f$  に関する最小化問題の局所的最小解である。

**証明**  $\mathbf{d} \neq \mathbf{0}$  を満たす  $\mathbf{d} \in \mathbb{R}^d$  を任意にとると,  $\nabla^2 f(\mathbf{x}_*) = \mathbf{H}_f(\mathbf{x}_*) \in \mathbb{S}_{++}^d \Rightarrow$  [定理 4.4(1)]  $\varepsilon = \langle \mathbf{d}, \nabla^2 f(\mathbf{x}_*) \mathbf{d} \rangle_2 > 0$  となる。 $\nabla^2 f$  の連続性 [定義 5.7] から,  $F(\alpha) = \nabla^2 f(\mathbf{x}_* + \alpha \mathbf{d})$  で定義される  $F: \mathbb{R}_+ \rightarrow \mathbb{R}^{d \times d}$  は連続である。内積の連続性 [例 3.5(1)] から

$$\begin{aligned} \exists \delta > 0 \forall \alpha \geq 0 \left( \alpha \in [0, \delta) \right. \\ \Rightarrow \left| \langle \mathbf{d}, \nabla^2 f(\mathbf{x}_* + \alpha \mathbf{d}) \mathbf{d} \rangle_2 - \langle \mathbf{d}, \nabla^2 f(\mathbf{x}_*) \mathbf{d} \rangle_2 \right| < \frac{\varepsilon}{2} \\ \Rightarrow \underbrace{\langle \mathbf{d}, \nabla^2 f(\mathbf{x}_* + \alpha \mathbf{d}) \mathbf{d} \rangle_2}_{\nabla^2 f(\mathbf{x}_* + \alpha \mathbf{d}) \in \mathbb{S}_{++}^d} > \frac{\varepsilon}{2} \Big) \end{aligned} \tag{5.48}$$

となる。Taylor の定理 [定理 5.6] から,  $\forall \alpha \in [0, \delta) \exists \tau \in (0, 1)$ :

$$\begin{aligned} f(\mathbf{x}_* + \alpha \mathbf{d}) &= f(\mathbf{x}_*) + \underbrace{\langle \nabla f(\mathbf{x}_*), \mathbf{d} \rangle_2}_{=0} \alpha + \frac{1}{2} \underbrace{\langle \mathbf{d}, \nabla^2 f(\mathbf{x}_* + \tau \alpha \mathbf{d}) \mathbf{d} \rangle_2}_{> 0 \cdot (5.48)} \alpha^2 \\ &\geq f(\mathbf{x}_*) \end{aligned}$$

となる。 $\|\mathbf{d}\|_2 \leq 1$  を満たす  $\mathbf{d} (\neq \mathbf{0})$  に対しては,  $\mathbf{x}_* + \alpha \mathbf{d} \in N_2(\mathbf{x}_*; \delta)$  となる。よって,  $\alpha$  と  $\mathbf{d}$  の任意性により,  $\forall \mathbf{y} \in N_2(\mathbf{x}_*; \delta) (f(\mathbf{x}_*) \leq f(\mathbf{y}))$  を得る。□

定理 5.9 から,  $\mathbf{x}_*$  が局所的最適解になるためには, 勾配 ((1 次) 微分係数)  $\nabla f(\mathbf{x}_*)$  が零ベクトルになることを要求しているが, これは, 1 変数微分可能

関数  $f$  でいえば、 $x_*$  における  $f$  の接線の傾き  $f'(x_*)$  が零となることなので直観的にも理解ができる。 $\nabla f(\mathbf{x}) = \mathbf{0}$  を満たす  $\mathbf{x}$  は局所的最小解、もしくは、局所的な最大解のいずれかであろう（図 5.3 の  $\mathbf{x}_*$  と  $\mathbf{z}$  を参照）。ここで、最小化問題の局所的最適解のためのもう一つの十分条件「2 次微分係数（ヘッセ行列） $\nabla^2 f(\mathbf{x}_*)$  の正定値性」が重要な役割を演じる。具体的には

$$\begin{aligned} \nabla^2 f(\mathbf{x}_*) &= \mathbf{H}_f(\mathbf{x}_*) \in \mathbb{S}_{++}^d \\ \Rightarrow \exists \delta > 0 \forall \mathbf{y} \in N_2(\mathbf{x}_*; \delta) (\nabla^2 f(\mathbf{y}) &= \mathbf{H}_f(\mathbf{y}) \in \mathbb{S}_{++}^d) \\ \text{[式 (5.48)]} & \\ \Rightarrow f \text{ は } \mathbf{x}_* \text{ の } \delta\text{-近傍 } N_2(\mathbf{x}_*; \delta) \text{ で狭義凸} & \\ \text{[定理 5.8(2)]} & \\ \Rightarrow \mathbf{x}_* \text{ は最小化問題の局所的な最小解} & \\ \text{[}\nabla f(\mathbf{x}_*) = \mathbf{0}\text{]} & \end{aligned}$$

である。この結果により、局所的な最小解の近傍においては関数の凸性が保証される（図 5.3 の  $\mathbf{x}_*$  と  $\mathbf{x}^*$  の近傍を参照）。

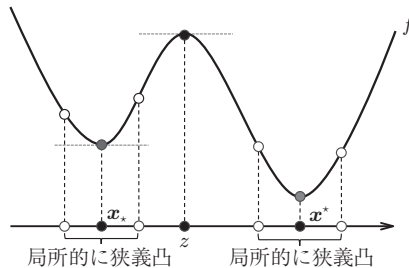


図 5.3 2 回連続的の微分可能関数の局所的凸性

一般の非凸関数に関する最小化問題を考察するとき、対象とする非凸関数の最小解の近傍では凸関数になることから、最小解の周りでの解析をするうえで凸関数の性質や解析を適用することができる。例えば、ある最適化手法（例えば、勾配降下法 [7.2 節]）により得られた近似解  $\mathbf{x}_A$ （勾配降下法を利用するとき、十分大きい反復回数  $K$  に対して、 $\mathbf{x}_A = \mathbf{x}_{K+1} = \mathbf{x}_K - \alpha \nabla f(\mathbf{x}_K)$ （ただし、 $\mathbf{x}_0 \in \mathbb{R}^d$  とし、 $\alpha > 0$  とする））が、十分小さい正数  $\varepsilon > 0$  に対して

$$\|D_f(\mathbf{x}_A)\|_2 < \varepsilon \wedge H_f(\mathbf{x}_A) \in \mathbb{S}_{++}^d \tag{5.49}$$

式 (5.49)

を満たすならば、定理 5.9 により、 $\mathbf{x}_A$  は正しく局所的最小解  $\mathbf{x}_*$  を近似していることがわかる。式 (5.49) の検証は、例えば、 $\varepsilon = 10^{-3}$  として

$$\|D_f(\mathbf{x}_A)\|_2^2 = \sum_{i=1}^d \left( \frac{\partial f(\mathbf{x}_A)}{\partial x_i} \right)^2 < (10^{-3})^2 \wedge \lambda_{\min}(H_f(\mathbf{x}_A)) > 0$$

であることを計算機を用いて調べればよい。

## 章 末 問 題

- 【5.1】** 命題 5.2 (微分作用素  $\nabla$  の線形性) を証明せよ。
- 【5.2】** 命題 5.3 (方向微分係数と導関数の関係) 内の式 (5.7) を式 (5.6) を用いて証明せよ。
- 【5.3】** 命題 5.6 (微分作用素  $\nabla^2$  の線形性) を証明せよ。
- 【5.4】** 命題 5.10 (凸関数の性質) を証明せよ。
- 【5.5】** 命題 5.14 (劣微分作用素  $\partial$  の加法性と正数同次性) の証明内にある  $\forall \alpha > 0 \forall \mathbf{x} \in D_1 (\partial(\alpha f_1))(\mathbf{x}) = \alpha \partial f_1(\mathbf{x})$  と  $\forall \mathbf{x} \in D_1 \cap D_2 (\partial(f_1 + f_2))(\mathbf{x}) \supset \partial f_1(\mathbf{x}) + \partial f_2(\mathbf{x})$  が成り立つことを示せ。

# 索引

## 【あ】

値	20
アフィン関数	140
アフィン写像	64
アフィン変換	200

## 【い】

一様乱数	170
一致推定	194
一致推定量	195

## 【う】

上に有界	27, 30
裏	7
裏命題	7

## 【え】

エポック	214
エルミート行列	85

## 【お】

凹関数	140
-----	-----

## 【か】

開 球	55
開集合	55
概収束	193
階 数	93
過学習	204
下極限	38
学習率	205
確 率	162
確率空間	162
確率収束	193

確率的勾配	211
確率的勾配降下法	213
確率分布	168
確率変数	167
確率密度関数	169
確率 1 で収束	193
下 限	28
過剰適合	204
型	73
下半三角行列	78
加法定理	163
関 数	20

## 【き】

機械学習	199
期待値	175
基 底	83
逆	7
逆行行列	79
逆 像	20
逆命題	7
吸収律	19
狭義単調性	154
狭義凸関数	153
強単調性	154
強凸関数	153
共分散	176
行ベクトル	74
行 列	73
行列式	81
極 限	31, 51, 103
局所的小解	158
局所の Lipschitz 連続	142

## 【く】

空 間	16
空事象	163
空集合	16
グリッドサーチ	209
訓練データ	202

## 【け】

経験損失	202
経験損失最小化	202
下 界	27
仮言三段論法	9
結 論	8
元	15
減衰ステップサイズ	209
限定記号	10

## 【こ】

項	30
降下方向	206
降下補題	117
恒偽式	6
交差エントロピー誤差	201
合 成	24
勾 配	23, 108
勾配降下法	206
固有値	85
固有値分解	89
固有ベクトル	85
根元事象	163
コンパクト集合	60

## 【さ】

最急降下法	206
-------	-----



最小元	29	上半三角行列	78	全順序集合	27
最小値	68	乗法定理	165	全称記号	10
最小要素	29	証 明	13	全称命題	10
再生性	191	真部分集合	16	全体集合	16
最大元	29	真理関数	6	前 提	8
最大値	67	真理式	6	全分散の法則	185
最大値ノルム	49	真理値	1		
最大要素	29	真理値表	1	<b>【そ】</b>	
最適化手法	205			像	20
最尤推定	194	<b>【す】</b>		相 似	80
最尤推定量	196	推 定	194	双条件命題	5
差集合	19	推定量	194	添 字	30
作用素ノルム	99	推 論	8	属する	15
三段論法	9	数学的帰納法	14	ソフトマックス関数	201
三段論法肯定式	9	スカラー倍	43, 73	存在記号	10
サンプリング	190	ステップサイズ	205	存在命題	10
		スペクトルノルム	100	損失関数	202
<b>【し】</b>		スペクトル半径	100		
シグモイド関数	201			<b>【た】</b>	
試 行	163	<b>【せ】</b>		大域的最小解	158
事後確率	167	正規化線形関数	201	対角化	86, 89
事 象	162	正規直交基底	83	対角行列	78
事前確率	167	正規直交系	82	対角成分	78
従 う	168	正規母集団	190	対角和	80
下に有界	27, 30	正則化付き最小化問題	204	対 偶	7
実数列	30	正則行列	79	対偶法	13
写 像	20	正定値行列	90	対偶命題	7
集 合	15	成 分	44	対称行列	85
集合演算	16	正方向列	78	大数の法則	192
集合族	24	積	74	多次元確率変数	171
集合値写像	146	積集合	17, 25	多次元正規分布	174
集積点	52	積 分	130	多次元離散一様分布	172
収 束	31, 51, 103	積分可能	130	多次元離散型確率変数	171
十分条件	8	積率母関数	191	多次元連続一様分布	173
周辺確率	166	全確率の法則	165	多次元連続型確率変数	173
順伝播	200	全期待値	189	単位行列	79
上 界	27	全期待値の法則	184	単位ベクトル	83
上極限	38	線形結合	83	段階最適化	221
上 限	27	線形作用素	98	探索方向	205
条件付き確率	164	線形写像	22	単調減少	33
条件付き確率分布	182	線形独立	83	(実数列の) 単調性	33
条件付き期待値	182	選 言	1	(劣微分の) 単調性	152
条件付き分散	183	全事象	163	単調増加	33
条件命題	4	全順序	27		

<b>【ち】</b>		<b>【に】</b>		フルバッチサイズ	211
値域	20	二乗誤差	117	分散	175
中心極限定理	193	ニューラルネットワーク	199	分配律	2, 19
中線定理	46			分布収束	194
直積	20	<b>【の】</b>		<b>【へ】</b>	
直交行列	81	ノイズ	215	平滑	114
<b>【て】</b>		ノルム	46	閉球	56
定義域	20	ノルム空間	46	平均	175
定数ステップサイズ	209			平均値の定理	119
テストデータ	203	<b>【は】</b>		平行四辺形の法則	46
点	46	背理法	13	閉集合	55
点推定	194	発散	31, 32	ベイズの定理	165
転置	21	バッチ勾配降下法	214	べき集合	24
転置行列	75	バッチサイズ	211	ベクトル	44
点列	51	汎化性能	203	ベクトル空間	44
<b>【と】</b>		半正定値行列	90	ヘッセ行列	121
等価なノルム	49	反復法	205	変数	20
導関数	108	反例	14	偏導関数	105
統計量	190	<b>【ひ】</b>		偏微分可能	105
同時確率	166	非線形変換	200	偏微分係数	105
同時確率密度関数	173	必要十分条件	9	<b>【ほ】</b>	
同値	7	必要条件	8	包含関係	16
同値関係	80	否定	3	方向微分係数	110
トートロジー	6	微分可能	107	補集合	18
特異値	93	微分係数	108	母集団	190
特異値分解	93	微分積分学の基本定理	132	母集団分布	190
(確率変数の) 独立	186	標本	162, 190	母数	190
(事象の) 独立	164	標本共分散	191	母分散	190
独立同一分布	187	標本空間	162	母平均	190
凸関数	65, 140	標本調査	190	<b>【み】</b>	
凸結合	70	標本分散	191	ミニバッチ確率的勾配降下法	
凸集合	70	標本平均	191		213
<b>【な】</b>		比例公平関数	141	ミニバッチ勾配	211
内積	44	<b>【ふ】</b>		ミニバッチサイズ	211
内積から誘導されるノルム	47	符号関数	149	<b>【む】</b>	
		負の対数尤度関数	196	無作為抽出	169
内積空間	44	部分集合	16	無作為標本	190
内点	142	部分列	34, 52	矛盾律	7
		不偏推定	194	無相関	186
		不偏推定量	194		

<b>【め】</b>	
命題	1
命題関数	9
命題変数	6
<b>【ゆ】</b>	
有界	27, 30, 51, 60, 103
ユークリッド空間	47
ユークリッド内積	45
ユークリッドノルム	48
尤度	167
誘導ノルム	99
尤度関数	196
<b>【よ】</b>	
要素	15

<b>【り】</b>	
離散一様分布	168
離散型確率分布	168
離散型確率変数	168
離散型周辺確率分布	172
離散型同時確率分布	171
<b>【れ】</b>	
零行列	74
劣勾配	146
劣乗法性	102
劣微分	145
劣微分可能	146
列ベクトル	74
連言	2
連鎖律	113
連続	61

連続一様分布	170
連続型確率分布	169
連続型確率変数	169
連続型周辺確率分布	173
連続型同時確率分布	173
連続関数	62
連続的微分可能	114

<b>【ろ】</b>	
論理記号	1
論理式	2
論理積	2
論理和	1

<b>【わ】</b>	
和	43, 73
和集合	16, 25

<b>【B, C】</b>	
Bolzano–Weierstrass の定理	68
Cauchy–Schwarz の不等式	46
Chebyshev の不等式	192
Cholesky 分解	94
$C^1$ 関数	114
$C^1$ 級	114
$C^2$ 関数	127
$C^2$ 級	127
<b>【D】</b>	
Darboux の定理 (集合に関する)	129
de Morgan の法則 (命題に関する)	18
de Morgan の法則	3
<b>【F】</b>	
Fréchet 微分可能	109
Frobenius ノルム	97

<b>【G】</b>	
Gâteaux 微分可能性	108
Gram–Schmidt の直交化法	83
<b>【H】</b>	
Hölder の不等式	72
$H$ -内積	45
$H$ -ノルム	48
<b>【K, L】</b>	
Kronecker のデルタ	82
Lipschitz 定数	114
Lipschitz 連続	114, 127
<b>【M】</b>	
Mahalanobis 距離	48
Minkowski の不等式	72
<b>【P, S】</b>	
$p$ -ノルム	48
Schur 分解	87, 89

<b>【T】</b>	
Taylor 展開	137
Taylor の定理	137

<b>【数字】</b>	
1 次元ユークリッド空間	26
1-ノルム	48
2 回微分可能	121
2 回偏微分可能	120
2 回連続的微分可能	126
2 次形式	89
2 次導関数	121
2 次偏導関数	121
2 次偏微分係数	121
2 次微分係数	121

<b>【ギリシャ文字】</b>	
$\varepsilon$ -近傍	29

—— 著者略歴 ——

- 2005年 東京工業大学大学院情報理工学研究科博士後期課程修了（数理・計算科学専攻）  
博士（理学）  
東京工業大学大学院補佐員
- 2007年 日本学術振興会特別研究員（PD）
- 2008年 九州工業大学ネットワークデザイン研究センター准教授
- 2013年 明治大学理工学部准教授
- 2019年 明治大学理工学部教授
- 現在に至る

## 機械学習のための数学

Mathematics for Machine Learning

© Hideaki Iiduka 2024

2024年8月26日 初版第1刷発行



検印省略

著者 <sup>い</sup>飯 <sup>づか</sup>塚 <sup>ひで</sup>秀 <sup>あき</sup>明  
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 978-4-339-06132-1 C3041 Printed in Japan

(田中)



<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。