

第 III 部 線型モデル

2 誤差の二乗和を最小化する方法

ポイント

- 微分
- 導関数

- 微分の公式の導出
- 偏微分

2.1 はじめに

- 山を登り続けると一番上にたどり着きます。
- 谷を下り続けると一番下にたどり着きます。
- 上りか下りを決めているのは傾きです。

2.2 最小二乗法

n 個ある観測点に直線

$$\hat{y}_i = \alpha + \beta x_i \quad (1)$$

をあてはめることを考える。ここで y 軸上における各点 (y_i) と直線上の点 ($\hat{y}_i$) の差として誤差 (e_i) を定義する。

$$e_i = y_i - \hat{y}_i \quad (2)$$

$$= y_i - (\alpha + \beta x_i) \quad (3)$$

$$= y_i - \alpha - \beta x_i \quad (4)$$

そして誤差の二乗和

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \alpha - \beta x_i)^2 \quad (5)$$

$$= n\alpha^2 + 2\alpha\beta \sum_{i=1}^n x_i + \beta^2 \sum_{i=1}^n x_i^2 - 2\alpha \sum_{i=1}^n y_i - 2\beta \sum_{i=1}^n x_i y_i + \sum_{i=1}^n y_i^2 \quad (6)$$

を最小化するようにパラメータ (α, β) を推定する方法を『最小二乗法』という。

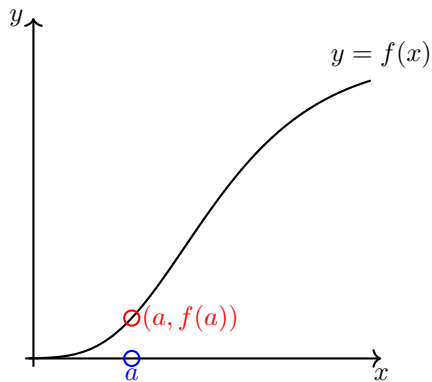
2.3 微分

誤差の二乗和を最小化するために微分を使う。ここで具体的に誤差の二乗和を微分する前に微分の定義を確認する。そして関連する知識は導関数そして偏微分である。

関数 $f(x)$ の $x = a$ の点

連続な関数 $f(x)$ が与えられたとき、定義域内における点 $x = a$ を考える。 $x = a$ に対応する関数 $f(x)$ 上の点の座標は $(a, f(a))$ である。

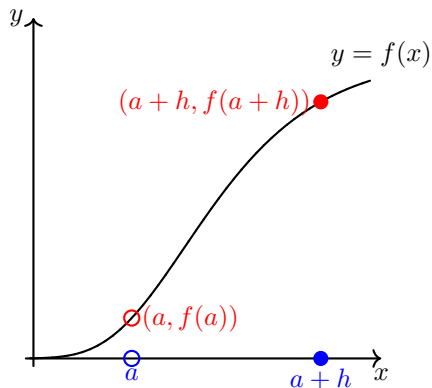
図1 $f(x)$ と x 軸上の点 a



a から h 離れた点

定義域内において、 x 軸上で a から h 離れた点をとる。ここで $h \neq 0$ とする。 $x = a + h$ に対応する $f(x)$ 上の点は $(a + h, f(a + h))$ である。

図2 $x = a + h$ に対応する $f(x)$ の点



x の変化量と y の変化量

ここで x 軸上の変化量は

$$h$$

であり、 y 軸上の変化量は

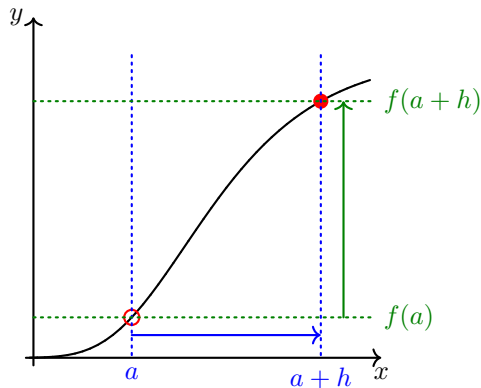
$$f(a+h) - f(a)$$

である。

(7)

(8)

図3 x の変化量と y の変化量



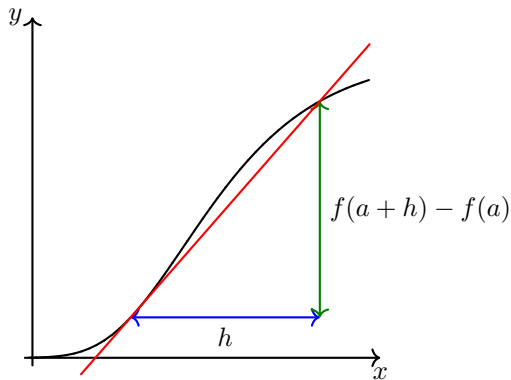
x と y の変化の割合

ここで x 軸上の変化量と y 軸上の変化量の比をとり変化の割合

$$\frac{f(a+h) - f(a)}{h} \quad (9)$$

をもとめる。この値は、点 $(a, f(a))$ と点 $(a+h, f(a+h))$ を通る直線の傾きである。

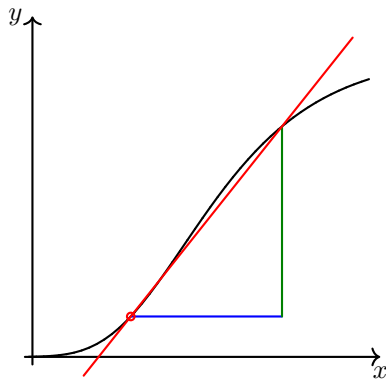
図4 x と y の変化の割合



$h \rightarrow 0$ の時の傾き

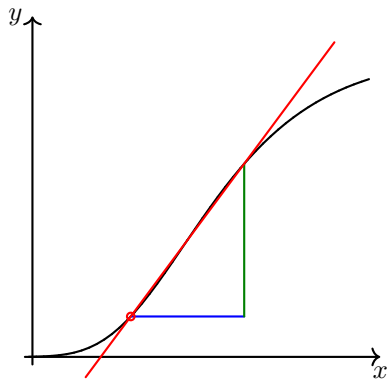
ここで h を h が 0 になることなく、限りなく 0 に近づくことを考える。

h が変化することにより変化の割合も変化する。

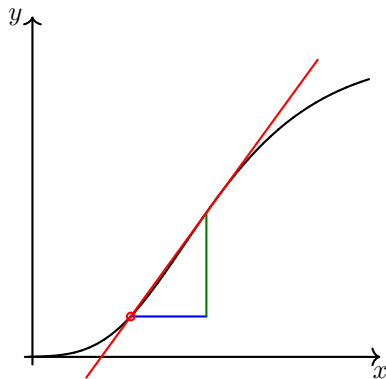
図5 h が少し 0 に近づいた時の傾き

h を少し 0 に近づけると、変化の割合は h の変化により影響を受ける。

図6 h がもう少し 0 に近づいた時

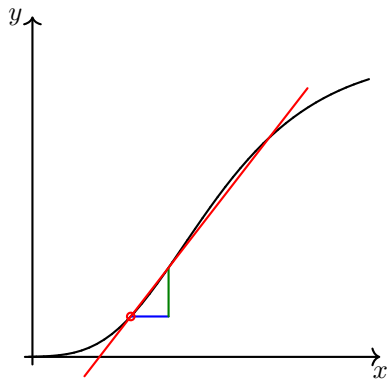


h を更に 0 に近づけると、 $(a, f(a))$ と $(a + h, f(a + h))$ は関数上において近づいていき、それに伴い、変化の割合も変化してゆく。

図7 h が更に 0 に近づいた時

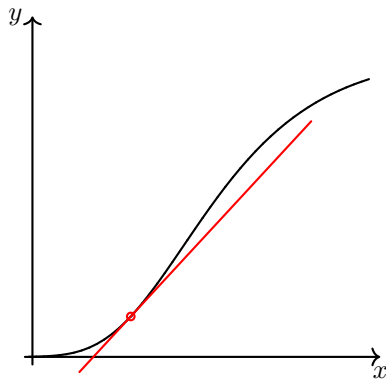
h をかなり 0 に近づけると、関数上の 2 点は相当に近づいてゆく。

図 8 h がかなり 0 に近づいた時



h を限りなく 0 に近づけると、2 点は限りなく近づき、(9) は $(a, f(a))$ における変化の割合を示す。

図 9 h が限りなく 0 に近づいた時



2.3.1 微分係数と導関数

関数 $f(x)$ が与えられたとき、定義域内において、点 a と、 a から h 離れた点 $a+h$ を考える。ここで $h \neq 0$ とする。 x の変化量 h と x が h 変化したことによってもたらされる関数 $f(x)$ の変化量 $f(a+h) - f(a)$ の比を用いて変化の割合をもとめ、 h が 0 になることなく h を限りなく 0 に近づける。

$$\lim_{h \rightarrow 0} \frac{f(a+h) - f(a)}{h} \quad (10)$$

この時に (10) が値を取るならば、関数 $f(x)$ は点 a で微分可能といい、その値を $x = a$ の微分係数という。

微分係数は、 $\left. \frac{dy}{dx} \right|_{x=a}$ という記号を使ってあらわす。

定義

微分係数は a が変れば変化する。従って (10) 式は a の関数である。そこで (10) の a を x で置き換えた式を導関数という。

$$\lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} \quad (11)$$

$f(x)$ から導関数を求めることを『微分する』という。また $f(x)$ を『原始関数』と呼ぶ。

通常の微分の定義において与えられた関数に名前を付けることは行われませんが、後に積分を紹介することを想定し、『原始関数』という語を採用する。

2.3.2 導関数の表記法

導関数は様々な記号であらわされることが多い。これらの表現は、その時に説明しようとする内容によって使い分けられている。 $y = f(x)$ とすると

$$\begin{array}{ll} (f(x))' & \frac{dy}{dx} \\ \frac{d}{dx} f(x) & f'(x) \end{array}$$

はすべて導関数をあらわす記号である。

2.3.3 $f(x) = x^2$ の導関数 $x = a$ のとき

$$f(a) = a^2 \quad (12)$$

 $x = a + h$ のとき

$$f(a + h) = (a + h)^2 \quad (13)$$

なので

$$\left. \frac{df}{dx} \right|_{x=a} = \lim_{h \rightarrow 0} \frac{(a + h)^2 - a^2}{h} \quad (14)$$

$$= \lim_{h \rightarrow 0} \frac{a^2 + 2ah + h^2 - a^2}{h} \quad (15)$$

$$= \lim_{h \rightarrow 0} \frac{2ah + h^2}{h} \quad (16)$$

$$= \lim_{h \rightarrow 0} (2a + h) \quad (17)$$

$$= 2a \quad (18)$$

である。 $x = a$ で値をとるので、 $f(x) = x^2$ は点 $x = a$ で微分可能である。

したがって (18) の a を x で置き換えた式が $f(x)$ の導関数なので

$$(x^2)' = 2x \quad (19)$$

である。かなり冗長に記述したが通常は別解のように、より簡便にあらわすのが一般的である。

別解

$$(x^2)' = \lim_{h \rightarrow 0} \frac{(x+h)^2 - x^2}{h} \quad (20)$$

$$= \lim_{h \rightarrow 0} \frac{\cancel{x^2} + 2xh + h^2 - \cancel{x^2}}{h} \quad (21)$$

$$= \lim_{h \rightarrow 0} \frac{2x\cancel{h} + h^{\cancel{2}}}{\cancel{h}} \quad (22)$$

$$= \lim_{h \rightarrow 0} (2x + h) \quad (23)$$

$$= 2x \quad (24)$$

2.3.4 $f(x) = x$ の導関数

$$(x)' = \lim_{h \rightarrow 0} \frac{(x+h) - x}{h} \quad (25)$$

$$= \lim_{h \rightarrow 0} \frac{x + h - x}{h} \quad (26)$$

$$= \lim_{h \rightarrow 0} \frac{h}{h} \quad (27)$$

$$= 1 \quad (28)$$

2.3.5 $f(x) = c$ の導関数

$$(c)' = \lim_{h \rightarrow 0} \frac{c - c}{h} \quad (29)$$

$$= \lim_{h \rightarrow 0} \frac{0}{h} \quad (30)$$

$$= 0 \quad (31)$$

2.3.6 $f(x) + g(x)$ の導関数

$f(x)$, $g(x)$ はともに微分可能とする。

$$(f(x) + g(x))' = \lim_{h \rightarrow 0} \frac{(f(x+h) + g(x+h)) - (f(x) + g(x))}{h} \quad (32)$$

$$= \lim_{h \rightarrow 0} \frac{\textcolor{blue}{f}(x+h) + \textcolor{green}{g}(x+h) - \textcolor{blue}{f}(x) - \textcolor{green}{g}(x)}{h} \quad (33)$$

$$= \lim_{h \rightarrow 0} \left(\frac{\textcolor{blue}{f}(x+h) - \textcolor{blue}{f}(x)}{h} + \frac{\textcolor{green}{g}(x+h) - \textcolor{green}{g}(x)}{h} \right) \quad (34)$$

$$= f'(x) + g'(x) \quad (35)$$

2.3.7 $\alpha f(x)$ の導関数

$f(x)$ は微分可能、 α は定数とする。

$$(\alpha f(x))' = \lim_{h \rightarrow 0} \frac{\alpha f(x+h) - \alpha f(x)}{h} \quad (36)$$

$$= \lim_{h \rightarrow 0} \alpha \frac{f(x+h) - f(x)}{h} \quad (37)$$

$$= \alpha f'(x) \quad (38)$$

2.3.8 $f(x) g(x)$ の導関数

$f(x)$, $g(x)$ は共に微分可能とする。

$$(f(x) g(x))' = \lim_{h \rightarrow 0} \frac{f(x+h) g(x+h) - f(x) g(x)}{h} \quad (39)$$

ここで極限内分子に $-f(x) g(x+h) + f(x) g(x+h) = 0$ を加えると、

$$= \lim_{h \rightarrow 0} \frac{f(x+h) g(x+h) - f(x) g(x+h) + f(x) g(x+h) - f(x) g(x)}{h} \quad (40)$$

$$= \lim_{h \rightarrow 0} \left(\frac{f(x+h) - f(x)}{h} g(x+h) + f(x) \frac{g(x+h) - g(x)}{h} \right) \quad (41)$$

$$= f'(x) g(x) + f(x) g'(x) \quad (42)$$

2.3.9 微分の公式

$f(x)$, $g(x)$ は共に微分可能とする。導関数の定義より以下の公式を導き出すことができる。

$$(f(x) + g(x))' = f'(x) + g'(x) \quad (43)$$

$$(f(x) g(x))' = f'(x) g(x) + f(x) g'(x) \quad (44)$$

$$(\alpha f(x))' = \alpha f'(x) \quad ; \alpha \text{は定数} \quad (45)$$

$$(x^2)' = 2x \quad (46)$$

$$(x)' = 1 \quad (47)$$

$$(c)' = 0 \quad ; c \text{は定数} \quad (48)$$

問題 III-2-1

以下の関数を微分しなさい。

1. $f(x) = 2x^2 + 3x + 4$

2. $f(x) = -3x^2 + 2x - 1$

3. $f(x) = \frac{x^2 - 2x - 3}{2}$

解例 III-2-1-1

$$f'(x) = (2x^2 + 3x + 4)'$$

(43) により和の導関数を導関数の和としてあらわし

$$= (2x^2)' + (3x)' + (4)'$$

(45) により各項の係数を掃きだす

$$= 2(x^2)' + 3(x)' + (4)'$$

(46) (47) (48) より

$$= 2(2x) + 3(1) + (0)$$

$$= 4x + 3$$

解例 III-2-1-2

$$\begin{aligned}f'(x) &= (-3x^2 + 2x - 1)' \\&= (-3x^2)' + (2x)' + (-1)' \\&= -3(x^2)' + 2(x)' + (-1)' \\&= -3(2x) + 2(1) + (0) \\&= -6x + 2\end{aligned}$$

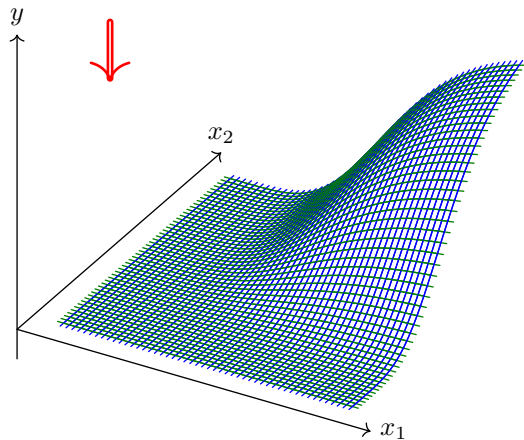
解例 III-2-1-3

$$\begin{aligned}f'(x) &= \left(\frac{x^2 - 2x - 3}{2} \right)' \\&= \frac{(x^2 - 2x - 3)'}{2} \\&= \frac{2x - 2}{2} \\&= x - 1\end{aligned}$$

2.4 偏微分

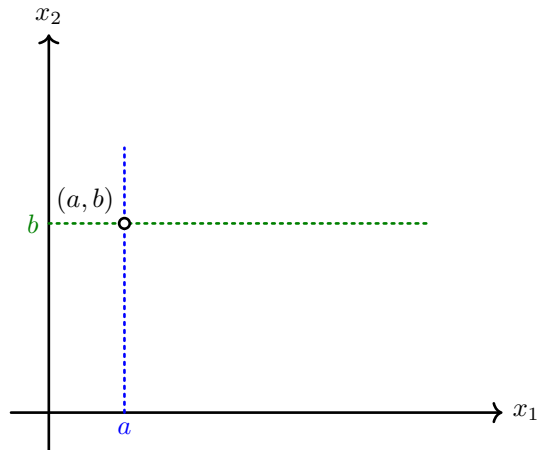
$f(x_1, x_2)$ のように独立変数が 2 つある関数を考える。この関数 $f(x_1, x_2)$ の定義域は平面であらわすことができる。そして $f(x_1, x_2)$ の値域は定義域平面上の点から延ばされた垂直線上の高さとしてあらわすことができる。

図 10 $y = f(x_1, x_2)$ の描く曲面



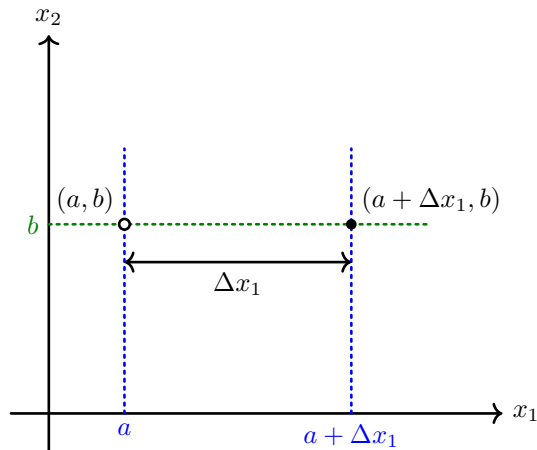
定義域上に点 (a, b) をとる。

図 11 (a, b) の点



点 (a, b) から x_1 軸上で Δx_1 離れた点 $(a + \Delta x_1, b)$ をとる。

図 12 (a, b) から x_1 軸方向に Δx_1 離れた点

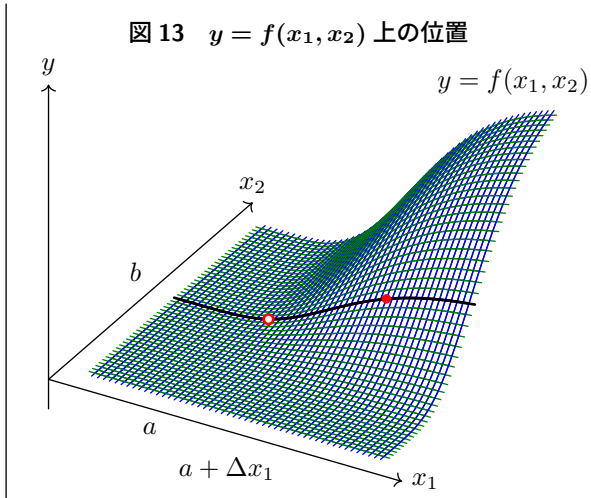


二つの点に対応する関数上の点は

$$y = f(a, b)$$

$$y = f(a + \Delta x_1, b)$$

である。



$x_2 = b$ に対応する関数上の点のみを表示し、 $x_1 \times y$ 平面からみる。

ここで表示されているのは

$$y = f(x_1, b)$$

である。

図 14 定義域上の点と関数上の点の対応

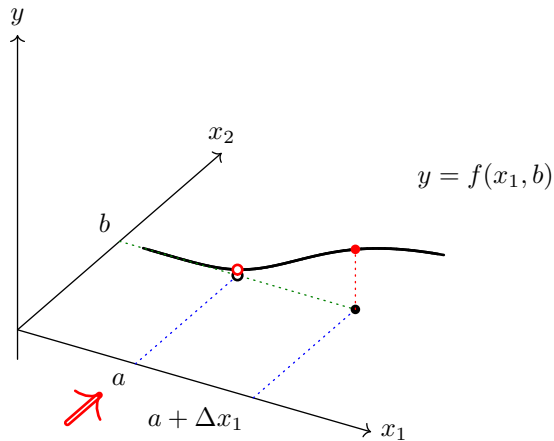


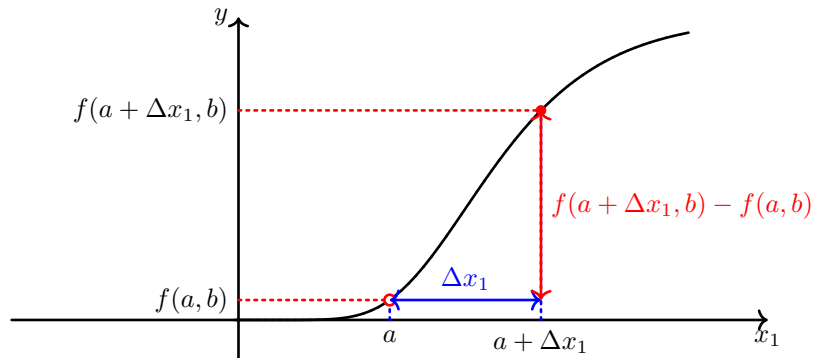
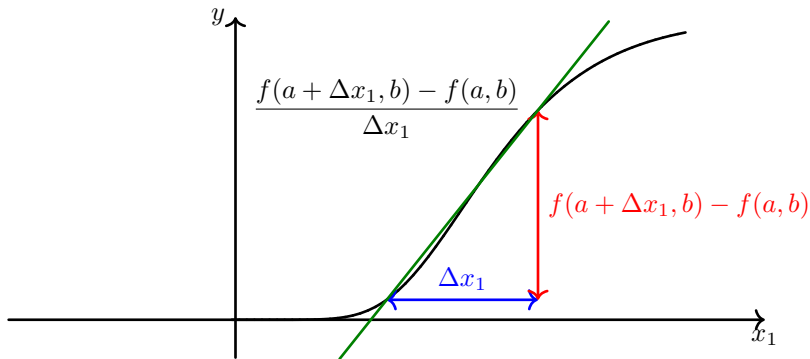
図 15 $x_1 \times y$ 上の点

図 16 x_1 と y の変化の割合

定義

2 変数関数 $f(x_1, x_2)$ において (a, b) から $(a + \Delta x_1, b)$ へ変化したときを考え、 Δx_1 と $f(x_1, x_2)$ の変化の割合を求める。ここで $\Delta x_1 \neq 0$ とする。このとき Δx_1 が 0 になることなく、限りなく 0 に近づけた時に値が存在するならば、その値を『点 (a, b) における x_1 の偏微分係数』と呼び $\frac{\partial f(a, b)}{\partial x_1}$ とあらわす。

$$\frac{\partial f(a, b)}{\partial x_1} = \lim_{\Delta x_1 \rightarrow 0} \frac{f(a + \Delta x_1, b) - f(a, b)}{\Delta x_1} \quad (49)$$

∂ は「ラウンド」「ラウンド d 」「ラウンド・マーク」等と呼ばれる。

定義

偏微分係数 $\frac{\partial f(a,b)}{\partial x_1}$ は (a,b) の関数なので、 (a,b) を (x_1, x_2) に置き換えた式を偏導関数と呼び $\frac{\partial f}{\partial x_1}$ であらわす。

$$\frac{\partial f}{\partial x_1} = \lim_{\Delta x_1 \rightarrow 0} \frac{f(x_1 + \Delta x_1, x_2) - f(x_1, x_2)}{\Delta x_1} \quad (50)$$

関数 $f(x_1, x_2)$ から $\frac{\partial f}{\partial x_1}$ を求めることを『 x_1 で偏微分する』という。具体的に偏導関数を求めるときは $(f(x_1, x_2))_{x_1}$ とあらわす。 x_2 に関しても同様である。

2.4.1 偏微分と「1 変数の微分」の関係

定義から明らかなように、 x_1 で偏微分するときは x_2 は変化していない。したがって、 x_1 で偏微分するときに x_2 は一定であるとして計算し、 x_2 で偏微分するときに x_1 は一定であるとして計算すればよい。つまり、1 変数の微分の公式はそのまま使える。

2.4.2 3 変数以上の多変数関数への拡張

m 変数関数 $y = f(x_1, x_2, \dots, x_m)$ を考える。独立変数を識別する添え字を j とする。 $y = f(x_1, x_2, \dots, x_m)$ の x_j に関する偏導関数を $\frac{\partial y}{\partial x_j}$ であらわし、以下のように定義する。

$$\frac{\partial y}{\partial x_j} = \lim_{\Delta x_j \rightarrow 0} \frac{f(x_1, x_2, \dots, x_j + \Delta x_j, \dots, x_m) - f(x_1, x_2, \dots, x_j, \dots, x_m)}{\Delta x_j} \quad (51)$$

偏導関数は独立変数の数 (m) だけある。

問題 III-2-2

以下の関数を x_1, x_2 で偏微分しなさい。

$$f(x_1, x_2) = 2x_1^2 + 3x_1x_2 + 4x_2^2 - 5x_1 - 6x_2 + 7$$

解例 III-2-2

$$\begin{aligned}\frac{\partial f}{\partial x_1} &= (2x_1^2 + 3x_1x_2 + 4x_2^2 - 5x_1 - 6x_2 + 7)_{x_1} \\&= (2x_1^2)_{x_1} + (3x_1x_2)_{x_1} + (4x_2^2)_{x_1} + (-5x_1)_{x_1} + (-6x_2)_{x_1} + (7)_{x_1} \\&= 2(x_1^2)_{x_1} + 3x_2(x_1)_{x_1} + (4x_2^2)_{x_1} - 5(x_1)_{x_1} - (6x_2)_{x_1} + (7)_{x_1} \\&= 2(2x_1) + 3x_2(1) + (0) - 5(1) - (0) + (0) \\&= 4x_1 + 3x_2 - 5 \\ \frac{\partial f}{\partial x_2} &= 3x_1 + 8x_2 - 6\end{aligned}$$

2.5 2変数二次関数の形状

1変数二次関数は下に凸の放物線を描くか上に凸の放物線を描く。1変数二次関数を(52)とすると凸の上下を決めているのは α_2 の符号である。

$$y = \alpha_0 + \alpha_1 x + \alpha_2 x^2 \quad (52)$$

2変数二次関数は最大値を持つ場合と最小値を持つ場合とどちらも持たない場合がある。 y を従属変数、 x_1, x_2 を独立変数、 β_j をパラメータとして2変数二次関数を以下のように定式化する。

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_1 x_2 + \beta_5 x_2^2 \quad (53)$$

ここで、 $x_2 = c$ として、(53) を表現すると

$$y = \beta_0 + \beta_1 x_1 + \beta_2 c + \beta_3 x_1^2 + \beta_4 x_1 c + \beta_5 c^2 \quad (54)$$

$$= (\beta_0 + \beta_2 c + \beta_5 c^2) + (\beta_1 + \beta_4 c) x_1 + \beta_3 x_1^2 \quad (55)$$

(55) は x_1 を変数とする 1 変数二次関数である。(55) は $\beta_3 > 0$ のとき下に凸であり、 $\beta_3 < 0$ のとき上に凸である。

同様に $x_1 = c$ とすると、(53) は

$$y = (\beta_0 + \beta_1 c + \beta_3 c^2) + (\beta_2 + \beta_4 c) x_2 + \beta_5 x_2^2 \quad (56)$$

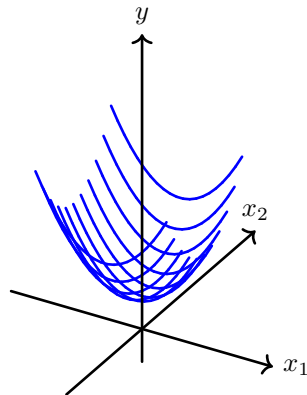
となる 1 変数二次関数である。この場合は β_5 の符号が (56) の凸の上下を決める。

β_3 は x_1^2 の係数であり、 β_5 は x_2^2 の係数である。

$\beta_3 > 0, \beta_5 > 0$ として、 x_2 ごとに (53) を描く。

x_2 ごとに描かれる曲線は下に凸の放物線である。

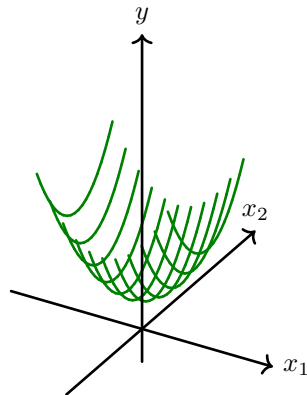
図 17 $\beta_3 > 0, \beta_5 > 0$ の放物線



同様に x_1 ごとに (53) を描く。 x_1 ごとに描かれる曲線は下に凸の放物線であり、これらの放物線は x_2 軸上の放物線に沿って描かれる。

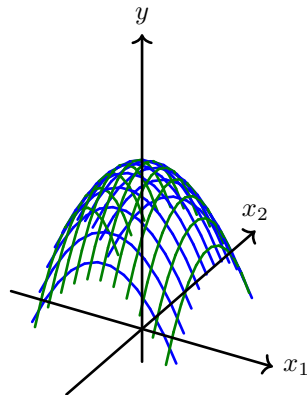
したがって、 $\beta_3 > 0, \beta_5 > 0$ の場合、(53) は下に凸であり、最小値を持つ。

図 18 $\beta_3 > 0, \beta_5 > 0$ の放物線



$\beta_3 < 0, \beta_5 < 0$ の場合は、上に凸の放物線が、上に凸の放物線に沿って描かれるので最大値を持つ。

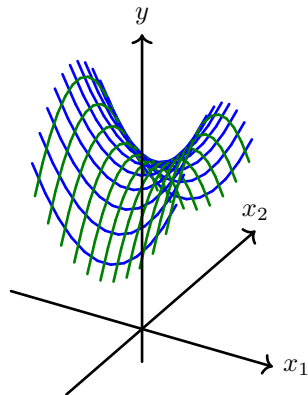
図 19 $\beta_3 < 0, \beta_5 < 0$ の放物線



$\beta_3 < 0, \beta_5 > 0$ の場合は、上に凸の放物線が、下に凸の放物線に沿って描かれるため最大値も最小値も持たない。

$\beta_3 > 0, \beta_5 < 0$ の場合も同様である。

図 20 $\beta_3 < 0, \beta_5 > 0$ の放物線



2.6 関数の最小化と極値

関数における局所的な最大値または最小値を極値という。極値における関数の傾きは 0 である。

誤差の二乗和は

$$\sum_{i=1}^n e_i^2 = n\alpha^2 + 2\alpha\beta \sum_{i=1}^n x_i + \beta^2 \sum_{i=1}^n x_i^2 - 2\alpha \sum_{i=1}^n y_i - 2\beta \sum_{i=1}^n x_i y_i + \sum_{i=1}^n y_i^2 \quad (6)$$

である。(6)において、 α^2 の係数と β^2 の係数はともに正である。したがって、(6)は最小値を持つ2変量二次関数である。最小値を持つ2変量二次関数の極値は最小値である。

(6)はパラメータ α , β の関数である。したがって誤差の二乗和を最小化する α , β を求めるには(6)の偏導関数を0と置いた連立方程式を解けばよい。

問題 III-2-3

次の関数を α, β で偏微分しなさい

$$f(\alpha, \beta) = n\alpha^2 + 2\alpha\beta \sum_{i=1}^n x_i + \beta^2 \sum_{i=1}^n x_i^2 - 2\alpha \sum_{i=1}^n y_i - 2\beta \sum_{i=1}^n x_i y_i + \sum_{i=1}^n y_i^2$$

解例 III-2-3

$$\frac{\partial f}{\partial \alpha} = 2n\alpha + 2\beta \sum_{i=1}^n x_i - 2 \sum_{i=1}^n y_i$$

$$\frac{\partial f}{\partial \beta} = 2\alpha \sum_{i=1}^n x_i + 2\beta \sum_{i=1}^n x_i^2 - 2 \sum_{i=1}^n x_i y_i$$

2.6.1 誤差の二乗和の偏導関数

$$f(\alpha, \beta) = n\alpha^2 + 2\alpha\beta \sum_{i=1}^n x_i + \beta^2 \sum_{i=1}^n x_i^2 - 2\alpha \sum_{i=1}^n y_i - 2\beta \sum_{i=1}^n x_i y_i + \sum_{i=1}^n y_i^2 \quad (57)$$

を α, β で偏微分すると

$$\frac{\partial f}{\partial \alpha} = 2n\alpha + 2\beta \sum_{i=1}^n x_i - 2 \sum_{i=1}^n y_i \quad (58)$$

$$\frac{\partial f}{\partial \beta} = 2\alpha \sum_{i=1}^n x_i + 2\beta \sum_{i=1}^n x_i^2 - 2 \sum_{i=1}^n x_i y_i \quad (59)$$

を得る。

2.6.2 偏導関数を 0 と置いた連立方程式

偏導関数はそれぞれのパラメータにおける傾きを返す関数である。ここで求めているのはそれぞれのパラメータ軸において傾きを持たない点である。誤差の二乗和が最小値を持つので、その点が誤差の二乗和を最小化する点である。この値は偏導関数をそれぞれ 0 とおいた連立方程式を解く事によって得られる。

$$\left\{ \begin{array}{l} 2n\alpha + 2\beta \sum_{i=1}^n x_i - 2 \sum_{i=1}^n y_i = 0 \\ 2\alpha \sum_{i=1}^n x_i + 2\beta \sum_{i=1}^n x_i^2 - 2 \sum_{i=1}^n x_i y_i = 0 \end{array} \right. \quad (60)$$

2.7 まとめ

- 微分はある点における傾きを返す関数を求める演算である。
- 元の関数 $f(x)$ を原始関数と呼び、微分された関数 $f'(x)$ を導関数と呼ぶ。
- $$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h}$$
- 定義に基づいて計算した結果は公式として利用可能である。
- 複数の変数がある場合は偏微分という。
- 偏微分も 1 変数の微分も計算方法は同じ。
- ∂ は「ラウンド」とよむ。
- 誤差の二乗和を偏微分して 0 と置いた連立方程式の解が、最小二乗推定値である。