

線形単回帰・重回帰モデルの行列・ベクトルによる表現

樋口さぶろお <https://hig3.net>

龍谷大学 先端理工学部 数理・情報科学課程

多変量解析☆演習 L02(2026-10-01 Thu)

最終更新: Time-stamp: "2026-10-02 Fri 15:43 JST hig"

今日の目標

- 線形モデルとしての単回帰・重回帰を説明できる
- 正規方程式を行列とベクトルで書ける



ここまで来たよ

2. 線形単回帰・重回帰モデルの行列・ベクトルによる表現

- 正規モデル
- 線形単回帰モデル
- 線形重回帰モデル
- 線形回帰の行列とベクトルによる定式化

正規分布にしたがう重さのドーナツ

確率統計でよく見た状況

あるドーナツ製造機の作るドーナツの重さ Y は、母平均値 μ ，母分散 σ^2 の正規分布に従う．

式による表現

$$\begin{aligned} \text{確率変数 } Y &\sim N(\mu, \sigma^2), \\ \text{(確率密度関数)} \quad f(y \mid \mu, \sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}. \end{aligned}$$

確率的に y を生み出す数学的仕組みのことを (統計) モデルという．
標本 $y_1, \dots, y_n$ を抽出する n 回の独立な試行は、独立同一分布にしたがう $Y_1, \dots, Y_n$ と表現できる．
この状況で、先週は、 μ や σ^2 の尤度を求めた．

モデルの同値な表現

定義 (正規モデル)

$$Y = \mu + \epsilon,$$
$$\epsilon \sim N(0, \sigma^2).$$

Y, ϵ : 確率変数, μ, σ^2 : 母数.

大注意: 確率変数は通常大文字で表すが, ここでは誤差を表す記号として小文字の ϵ を用いることが多い (ϵ - δ とは無関係で, 「小さい誤差」の意味合いがある).

同値であるという説得

(0) 定数 μ で平行移動しただけ.

(1) $\epsilon = Y - \mu$ と定義すると, 確率密度関数 $f(y \mid \mu, \sigma^2)$ における $y - \mu$ を ϵ に置き換えることで, $N(0, \sigma^2)$ の密度が得られる.

(2) 期待値・分散を計算すると,

$$E[Y] = E[\mu + \epsilon] = \mu + E[\epsilon] = \mu + 0 = \mu,$$
$$\text{Var}(Y) = \text{Var}(\mu + \epsilon) = \text{Var}(\epsilon) = \sigma^2.$$

2 変数の最尤推定

$$\text{最大化する尤度} \quad L(\mu, \sigma^2) = \prod_{i=1}^n f(y_i | \mu, \sigma^2)$$

$$\frac{\partial}{\partial \mu} \log L(\hat{\mu}, \hat{\sigma}^2) = 0 \rightsquigarrow \frac{1}{\hat{\sigma}^2} \sum_{i=1}^n (y_i - \hat{\mu}) = 0 \rightsquigarrow (\text{先に解ける}) \hat{\mu} = \frac{1}{n} \sum_{i=1}^n y_i$$

$$\frac{\partial}{\partial (\sigma^2)} \log L(\hat{\mu}, \hat{\sigma}^2) = 0 \rightsquigarrow -\frac{n}{2\hat{\sigma}^2} + \frac{1}{2(\hat{\sigma}^2)^2} \sum_{i=1}^n (y_i - \hat{\mu})^2 = 0 \rightsquigarrow \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{\mu})^2.$$

$\frac{\partial}{\partial (\sigma^2)}$ は、2 階微分でなく、 $s = \sigma^2$ を塊と思った合成微分。

復習 (確率統計) $\hat{\theta}$ が母数 θ の**不偏推定量** $\Leftrightarrow E[\hat{\theta}] = \theta$.

例: 標本平均値 $\bar{X} = \frac{1}{n} \sum_i X_i$ は母平均値 $\mu = E[X]$ の**不偏推定量**

推定量 $\hat{\mu}, \hat{\sigma}$ は、 $Y_1, \dots, Y_n$ で書ける確率変数で、

- $E[\hat{\mu}] = \frac{1}{n} \sum_i E[Y_i] = \mu$. 不偏.
- $E[\hat{\sigma}^2] \neq \sigma^2$. 不偏でない. $\frac{n}{n-1} \hat{\sigma}^2$ が不偏.

ここまで来たよ

2. 線形単回帰・重回帰モデルの行列・ベクトルによる表現

- 正規モデル
- 線形単回帰モデル
- 線形重回帰モデル
- 線形回帰の行列とベクトルによる定式化

線形単回帰モデル

状況

あるドーナツ製造機の作るドーナツの重さ Y は、温度ダイヤル x に「比例して」増え、ダイヤルが決まっていれば、ある母平均値、母分散 σ^2 の正規分布に従う。

温度ダイヤル x を増やすと重さが増える、ことだけがわかっているときに、とりあえず「比例して」と仮定することも多い

この場合は、温度（説明変数）が重さ（目的変数）の原因になっている。数学-物理-情報との違いを感じて。

$Y \sim N(\beta_0 + \beta_1 x, \sigma^2)$ と書いても同じだが、次の書き方をよく見る。

定義 (線形単回帰モデル (simple linear regression model))

$$Y = \beta_0 + \beta_1 x + \epsilon,$$
$$\epsilon \sim N(0, \sigma^2)$$

線形: 右辺が x の 1 次, 単: $x = x_1$ のみ, 回帰 x で y を説明

- 確率変数
 - ▶ Y : 目的変数 (従属変数)
 - ▶ ϵ : 誤差, ノイズ
- 確率変数でない変数
 - ▶ x : 説明変数 (独立変数) 確率変数でない, 分析者が自由に設定できる.
- 母数 (パラメタ)
 - ▶ β_0, β_1 : 回帰係数
 - ▶ $\sigma > 0$: 誤差の大きさ

$X \sim \beta_0 + \beta_1 y + \epsilon$ とするのは, 今の状況とは異なる. x は人間が精密に (例えば等間隔にも) 設定できる. 誤差があるのは温度でなく重さ.

母数（回帰係数，ノイズの大きさ）の最尤推定

温度 x_i で製造したときの重さ Y_i ($i = 1, \dots, n$) を考える.

$\epsilon_i = y_i - \beta_1 x_i - \beta_0 \sim N(0, \sigma^2)$ は独立同一分布にしたがう.

3 変数の尤度関数

$$\begin{aligned} L(\beta_0, \beta_1, \sigma^2) &= f(y_1, y_2, \dots, y_n | \beta_0, \beta_1, \sigma^2) \\ &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y_i - \beta_0 - \beta_1 \cdot x_i)^2}{2\sigma^2}} \\ &= (2\pi\sigma^2)^{-n/2} e^{-\sum_{i=1}^n \frac{(y_i - \beta_0 - \beta_1 \cdot x_i)^2}{2\sigma^2}} \end{aligned}$$

を最大化したい.

3 変数の対数尤度関数 $\log L(\beta_0, \beta_1, \sigma^2)$ を最大化したい.

$$\log L(\beta_0, \beta_1, \sigma^2) = C - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 \cdot x_i)^2$$

$$0 = \frac{\partial}{\partial \beta_0} \log L = \frac{\partial}{\partial \beta_1} \log L = \frac{\partial}{\partial (\sigma^2)} \log L$$

の解 $(\hat{\beta}_0, \hat{\beta}_1, \hat{\sigma})$ を求めたい.

ここでは最初の2つの等式だけを見ると…**最小2乗法**やってるのと同じじゃん!

推定量 $\hat{y}_i = \hat{\beta}_0 + \beta_1 x_i$ と書いたとき,

残差平方和 $S_e = \sum_i (y_i - \hat{y}_i)^2 = \sum_i (y_i - \beta_0 - \beta_1 \cdot x_i)^2$ を最小化してる.

しよせん, β_0, β_1 の連立1次方程式

正規方程式

$$\begin{aligned} n\hat{\beta}_0 + \left(\sum_i x_i\right)\hat{\beta}_1 &= \sum_i y_i \\ \left(\sum_i x_i\right)\hat{\beta}_0 + \left(\sum_i x_i^2\right)\hat{\beta}_1 &= \sum_i x_i y_i \end{aligned}$$

加減法 $\rightsquigarrow$

$$\hat{\beta}_1 = \frac{s_{xy}}{s_{xx}},$$

$$\hat{\beta}_0 = \bar{y} - \frac{s_{xy}}{s_{xx}} \bar{x}$$

$$y = \frac{s_{xy}}{s_{xx}} x + \bar{y} - \frac{s_{xy}}{s_{xx}} \bar{x}$$

$$y - \bar{y} = \frac{s_{xy}}{s_{xx}} (x - \bar{x})$$

ここで,

$$\bar{x} = \frac{1}{n} \sum_i x_i \quad \text{標本平均値っぽい形}$$

$$\bar{y} = \frac{1}{n} \sum_i y_i \quad \text{標本平均値っぽい形}$$

$$s_{xy} = \frac{1}{n} \sum_i x_i y_i - \bar{x} \cdot \bar{y} = \frac{1}{n} \sum_i (x_i - \bar{x})(y_i - \bar{y}) \quad \text{共分散っぽい形}$$

$$s_{xx} = \frac{1}{n} \sum_i x_i^2 - \bar{x}^2 = \frac{1}{n} \sum_i (x_i - \bar{x})^2 \quad \text{分散っぽい形}$$

また, $\frac{1}{n-1}$ でなく $\frac{1}{n}$.

$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 \cdot x$. ($\hat{\beta}_0, \hat{\beta}_1$: 回帰係数) に代入.

回帰直線

$$\hat{y} - \bar{y} = \frac{s_{xy}}{s_{xx}}(x - \bar{x})$$

実は $\hat{\beta}_0, \hat{\beta}_1, \hat{y}$ は正規分布にしたがう. $\leadsto$ 信頼区間, 仮説検定
帰無仮説: $\hat{\beta}_1 = 0$ 「 y は x に無関係」

確率統計 II,III

3 番目の式を使って, σ^2 も推定できる. 来週.

ここまで来たよ

2. 線形単回帰・重回帰モデルの行列・ベクトルによる表現

- 正規モデル
- 線形単回帰モデル
- **線形重回帰モデル**
- 線形回帰の行列とベクトルによる定式化

線形重回帰モデル (multiple linear regression model)

状況

このドーナツ製造機の作るドーナツの重さ Y は、温度 x_1 と庫内湿度 x_2 に「比例して」増減し、決まったダイヤルならある正規分布にしたがう。

定義 (重回帰モデル)

$$Y = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \epsilon,$$
$$\epsilon \sim N(0, \sigma^2).$$

Y : 目的変数 (従属変数) 確率変数, ϵ : 確率変数

$x_1, \dots, x_p$: 説明変数 (独立変数) 確率変数でない

$\beta_0, \beta_1, \dots, \beta_p$: 偏回帰係数, 母数. $\sigma > 0$, 母数.

一般の p : 重回帰, multiple regression, $p = 1$: 単回帰 simple regression

重回帰は、ローテクだが、すぐ使えて、超強力な機械学習の予測手法

様々な n 組のダイヤル設定 $x_{i1}, \dots, x_{ip}$, $i = 1, \dots, n$ で製造したときの重さ Y_i ($i = 1, \dots, n$) を考える.

x_{ij} で, $i = 1, \dots, n$ は標本内データ番号, $j = 1, \dots, p$ は母数 (ダイヤル) 番号.

$$\begin{array}{rcccl} Y_1 & \Leftarrow & x_{11} & \dots & x_{1p} \\ & & & \vdots & \\ Y_n & \Leftarrow & x_{n1} & \dots & x_{np} \end{array}$$

$\epsilon_i = y_i - \beta_1 x_{i1} - \dots - \beta_p x_{ip} - \beta_0 \sim N(0, \sigma^2)$ は独立同一分布にしたがう.
最尤推定 $\leadsto$

残差平方和 $S_e = \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_{i1} + \dots \beta_p x_{ip}))^2$ を最小化.

線形重回帰のイメージ

n 個のデータ ($i = 1, \dots, n$) を取得.

説明変数の値 x_{ij} ($i = 1, \dots, n, j = 1, \dots, p$)

目的変数の値 y_i ($i = 1, \dots, n$)

しばらく $p = 2$ で. 注: p : 説明変数の個数 ($j = 1, \dots, p$)

回帰平面

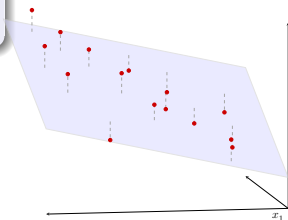
(x_1, x_2, y) 空間内の平面をデータ点のなるべく近く (上下方向に) にとりたい.

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 \cdot x_1 + \hat{\beta}_2 \cdot x_2$$

xyz 空間での平面の式と比べて

$$z = f(x, y) = c + ax + by$$

微積分 II



ここまで来たよ

2. 線形単回帰・重回帰モデルの行列・ベクトルによる表現

- 正規モデル
- 線形単回帰モデル
- 線形重回帰モデル
- 線形回帰の行列とベクトルによる定式化

ベクトルと行列の復習

定義 (行列 石村畑 p.21)

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1p} \\ a_{21} & a_{22} & \cdots & a_{2p} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{np} \end{pmatrix}$$

のように長方形状に実数を並べて書いたものを n 行 p 列の $((n, p)$ 型の) 行列 という.

$a_{ij} \in \mathbb{R}$ を A の (i, j) 成分という. i 行目, j 列目.

定義 (行, 列 石村畑 p.26)

- n 行 1 列の行列を縦ベクトル, 列ベクトルという 石村畑 p.220
- 行列 A 中の $\begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{nj} \end{pmatrix}$ を, 行列 A の第 j 列という
- A の第 j 列を n 行 1 列の行列すなわちベクトル (列ベクトル, 縦ベクトル) と見なした $\begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix}$ を, A の第 j 列ベクトルという
- 1 行 p 列の行列を横ベクトル, 行ベクトルという
- 行列 A 中の $a_{i1} a_{i2} \cdots a_{ip}$ を, 行列 A の第 i 行という
- A の第 i 行を 1 行 p 列の行列すなわちベクトル (行ベクトル, 横ベクトル) と見なした $(a_{i1} \ a_{i2} \ \cdots \ a_{ip})$ を, A の第 i 行ベクトルという

定義 (単位行列 石村畑 p.30)

n 次の単位行列 E or E_n とは,

n 次正方行列で (i, j) 成分が $e_{ij} = \delta_{ij} \stackrel{\text{def.}}{=} \begin{cases} 1 & (i = j) \\ 0 & (i \neq j) \end{cases}$ であるものをいう.

行列の積 石村畑 p.36

定義 (行列の積 石村畑 p.36)

- 行列 A ($\ell \times m$ 型, 成分 a_{ij} ($1 \leq i \leq \ell, 1 \leq j \leq m$))
- 行列 B ($m' \times n$ 型, 成分 b_{ij} ($1 \leq i \leq m', 1 \leq j \leq n$))

に対して, $m = m'$ のとき (だけ), 行列 A, B の積 AB が定義され,

- 行列 $C = AB$ ($\ell \times n$ 型, 成分 c_{ij} ($1 \leq \ell, 1 \leq j \leq n$))

$$c_{ij} = \sum_{k=1}^m a_{ik} b_{kj} = a_{i1} b_{1j} + a_{i2} b_{2j} + \cdots + a_{im} b_{mj}.$$

行列の積のルールの直観のおぼえ方

$$c_{ij} = \sum_{k=1}^m a_{ik} b_{kj} = a_{i1} b_{1j} + a_{i2} b_{2j} + \cdots + a_{im} b_{mj}.$$

$$\begin{pmatrix} \square & \square \\ \square & \square \\ \square & \square \\ \square & \square \end{pmatrix} = \begin{pmatrix} \square & \square & \square \\ \square & \square & \square \\ \square & \square & \square \\ \square & \square & \square \end{pmatrix} \begin{pmatrix} \square & \square \\ \square & \square \\ \square & \square \end{pmatrix}$$

$$\begin{pmatrix} 1 & 2 & 3 \\ -2 & 1 & 1 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ -1 & 2 \\ 3 & 0 \end{pmatrix}$$

横ベクトルと行列の積. $\begin{pmatrix} 1 & 2 & 3 \end{pmatrix} \begin{pmatrix} 10 & 20 \\ 30 & 40 \\ 50 & 60 \end{pmatrix} \rightsquigarrow$ 縦ベクトルと横ベクトルを区別する必要.

定義 (転置行列) 石村畑 §4.1(p.100)

m 行 n 列の行列 $A = \begin{pmatrix} a_{11} & \cdots & a_{1p} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{np} \end{pmatrix}$ に対して, n 行 m 列の行列 $\begin{pmatrix} a_{11} & \cdots & a_{n1} \\ \vdots & & \vdots \\ a_{1p} & \cdots & a_{np} \end{pmatrix}$ を A の転置行列 (transposed matrix) といい, $A^\top$ と書く.

A の成分を a_{ij} , $B = A^\top$ の成分を b_{ij} とすると, $a_{ij} = b_{ji}$.

例 (転置行列)

縦ベクトルの転置は横ベクトル, 横ベクトルの転置は縦ベクトル.

$$\begin{pmatrix} 3 \\ -1 \end{pmatrix}^\top = (3 \ -1), (3 \ -1)^\top = \begin{pmatrix} 3 \\ -1 \end{pmatrix}$$

$$C^\top = \begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix}^\top = D = \begin{pmatrix} a & d \\ b & e \\ c & f \end{pmatrix}, c_{12} = b = d_{21}.$$

$\mathbf{a}, \mathbf{b}$ が縦ベクトルのとき,

$$\mathbf{a} \cdot \mathbf{b} = \mathbf{a}^\top \mathbf{b},$$

縦ベクトルの内積 行列の積

行列とベクトルによる定式化 I

$$\begin{aligned} Y &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon \\ &= \alpha_0 + \beta_1 (x_1 - \bar{x}_{\bullet 1}) + \beta_2 (x_2 - \bar{x}_{\bullet 2}) + \epsilon \end{aligned}$$

y_i, x_{ij} データ番号 $i = 1, \dots, n$, 説明変数番号 $j = 1, \dots, p$.
 データの平均 $\bar{x}_{\bullet j} = \frac{1}{n} \sum_i x_{ij}$.

$$\begin{aligned} y_1 &= \alpha_0 + \beta_1 (x_{11} - \bar{x}_{\bullet 1}) + \beta_2 (x_{12} - \bar{x}_{\bullet 2}) + \epsilon_1 \\ &\vdots \\ y_n &= \alpha_0 + \beta_1 (x_{n1} - \bar{x}_{\bullet 1}) + \beta_2 (x_{n2} - \bar{x}_{\bullet 2}) + \epsilon_n \end{aligned}$$

行列とベクトルによる定式化 II

行列 X とベクトル $\mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\epsilon}$ で、また ϵ_i は独立より、次のようにとまとめられる。

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \epsilon_i \sim N(0, \sigma^2) \text{ (独立同一分布)}$$

$$\begin{bmatrix} \mathbf{y} \end{bmatrix} = \begin{bmatrix} X \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon} \end{bmatrix}$$

$$\mathbf{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, X = \begin{pmatrix} 1 & x_{11} - \bar{x}_{\bullet 1} & x_{12} - \bar{x}_{\bullet 2} \\ \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots \\ 1 & x_{n1} - \bar{x}_{\bullet 1} & x_{n2} - \bar{x}_{\bullet 2} \end{pmatrix}, \boldsymbol{\beta} = \begin{pmatrix} \alpha_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}, \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix}.$$

L02-Q1

Quiz(行列形の正規方程式の解)

授業の記号を使う． $\beta = \begin{pmatrix} \alpha_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}$ とする． 次のデータを得たとする．

x_1	x_2	y
4	0	2
8	4	4
4	8	6
0	8	8

- ① 行列 X を求めよう．
- ② 行列 $X\beta$ を書こう．
- ③ 行列 $X^T y$ を書こう．
- ④ 行列 $X^T X$ を求めよう．
- ⑤ 行列 $(X^T X)^{-1}$ を求めよう． Hint. ブロック対角行列の逆行列．

略解

$$\textcircled{1} \quad X = \begin{pmatrix} 1 & 4 & -4 & 0 & -5 \\ 1 & 8 & -4 & 4 & -5 \\ 1 & 4 & -4 & 8 & -5 \\ 1 & 0 & -4 & 8 & -5 \end{pmatrix}$$

$$\textcircled{2} \quad X\beta = \begin{pmatrix} 1 & 4 & -4 & 0 & -5 \\ 1 & 8 & -4 & 4 & -5 \\ 1 & 4 & -4 & 8 & -5 \\ 1 & 0 & -4 & 8 & -5 \end{pmatrix} \begin{pmatrix} \alpha_0 + 0 & -5\beta_2 \\ \alpha_0 + 4\beta_1 & -\beta_2 \\ \alpha_0 + 0 & 3\beta_2 \\ \alpha_0 + -4\beta_1 & 3\beta_2 \end{pmatrix}.$$

$$\textcircled{3} \quad X^\top \mathbf{y} = \begin{pmatrix} \sum_i y_i \\ 4y_2 - 4y_4 \\ -5y_1 - y_2 + 3y_3 + 3y_4 \end{pmatrix}.$$

$$\textcircled{4} \quad X^\top X = \begin{pmatrix} 4 & 0 & 0 \\ 0 & 32 & -16 \\ 0 & -16 & 44 \end{pmatrix}$$

$$\textcircled{5} \quad \text{ブロック対角形なので, } 4^{-1} \text{ と } \begin{pmatrix} 32 & -16 \\ -16 & 44 \end{pmatrix}^{-1} \text{ を考えて,}$$

$$(X^\top X)^{-1} = \begin{pmatrix} \frac{1}{4} & 0 & 0 \\ 0 & \frac{11}{288} & \frac{1}{72} \\ 0 & \frac{1}{72} & \frac{1}{36} \end{pmatrix}$$

偏微分の復習

縦ベクトル $\mathbf{b}, \mathbf{y} \in \mathbb{R}^n$, n 次正方行列 $M = (m_{ij})$. 転置 $\mathbf{b}^\top, M^\top$.

M : 対称行列 ($M^\top = M, m_{ij} = m_{ji}$) に対して, 実数値関数

$$f(\mathbf{b}) = \mathbf{b}^\top \mathbf{y} = \boxed{\mathbf{b}^\top} \boxed{\mathbf{y}} = \sum_{i=1}^n b_i y_i,$$

$$g(\mathbf{b}) = \mathbf{b}^\top M \mathbf{b} = \boxed{\mathbf{b}^\top} \boxed{M} \boxed{\mathbf{b}} = \sum_{i,j=1}^n b_i m_{ij} b_j.$$

$\frac{d}{db}(by) = (by)' = y, (mb^2)' = 2mb$ を多変数にした偏微分.

微積分 II

$$\frac{\partial f}{\partial b_k}(\mathbf{b}) = 0 + \cdots + 0 + (b_k)' y_k + 0 + \cdots + 0 = y_k$$

$$\frac{\partial g}{\partial b_k}(\mathbf{b}) = \sum_j (b_k)' m_{kj} b_j + \sum_i b_i m_{ik} (b_k)' = \sum_j m_{kj} b_j + \sum_i \mathbf{m}_{ki} b_i = 2(M\mathbf{b})_k$$

これを, $k = 1, \dots, n$ 集めて縦ベクトルを作る. このとき, 左辺を表す記号を導入. ベクトル解析

$$\nabla_{\mathbf{b}} f(\mathbf{b}) = \frac{\partial}{\partial \mathbf{b}} f(\mathbf{b}) = \mathbf{y}$$

$$\nabla_{\mathbf{b}} g(\mathbf{b}) = \frac{\partial}{\partial \mathbf{b}} g(\mathbf{b}) = M\mathbf{b}.$$

L02-Q2

Quiz(偏微分)

$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$, $\mathbf{y} = \begin{pmatrix} 3 \\ 8 \\ 9 \end{pmatrix}$, $M = \begin{pmatrix} 11 & 12 & 13 \\ 21 & 22 & 23 \\ 31 & 32 & 33 \end{pmatrix}$ とする.

$f(\mathbf{x}) = \mathbf{x}^\top \mathbf{y}$, $h(\mathbf{x}) = \mathbf{x}^\top \mathbf{x}$, $g(\mathbf{x}) = \mathbf{x}^\top M \mathbf{x}$ と定める.

ベクトル $\begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \frac{\partial f}{\partial x_3} \end{pmatrix}$, $\begin{pmatrix} \frac{\partial h}{\partial x_1} \\ \frac{\partial h}{\partial x_2} \\ \frac{\partial h}{\partial x_3} \end{pmatrix}$, $\begin{pmatrix} \frac{\partial g}{\partial x_1} \\ \frac{\partial g}{\partial x_2} \\ \frac{\partial g}{\partial x_3} \end{pmatrix}$ を求めよう.

略解

$$\textcircled{1} \quad \mathbf{y} = \begin{pmatrix} 3 \\ 8 \\ 9 \end{pmatrix}.$$

$$\textcircled{2} \quad 2\mathbf{x}.$$

$$\textcircled{3} \quad 2M\mathbf{x} = \begin{pmatrix} 11x_1 + 12x_2 + 13x_3 \\ 21x_1 + 22x_2 + 23x_3 \\ 31x_1 + 32x_2 + 33x_3 \end{pmatrix}.$$

これらにおいて，問題のベクトル（偏微分を縦ベクトルに集めたもの）を略記して，

$$\nabla_{\mathbf{x}} f(\mathbf{x}) = \frac{\partial f}{\partial \mathbf{x}}(\mathbf{x}) = \mathbf{y}, \nabla_{\mathbf{x}} h(\mathbf{x}) = \frac{\partial h}{\partial \mathbf{x}}(\mathbf{x}) = 2\mathbf{x}, \nabla_{\mathbf{x}} g(\mathbf{x}) = \frac{\partial g}{\partial \mathbf{x}}(\mathbf{x}) = 2M\mathbf{x}$$

のように書くことがある．

測定値 $\mathbf{y}$ と推定値 $\hat{\mathbf{y}}$ との差=残差 (縦) ベクトル $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - X\hat{\boldsymbol{\beta}}$.

$$\begin{aligned} \text{残差二乗和 } S_e(\alpha_0, \beta_1, \beta_2) &= \sum_i e_i e_i = \mathbf{e}^\top \mathbf{e} = (\mathbf{y} - X\hat{\boldsymbol{\beta}})^\top (\mathbf{y} - X\hat{\boldsymbol{\beta}}) \\ &= \mathbf{y}^\top \mathbf{y} - 2\hat{\boldsymbol{\beta}}^\top X^\top \mathbf{y} + \hat{\boldsymbol{\beta}}^\top X^\top X \hat{\boldsymbol{\beta}} \end{aligned}$$

各項はスカラー (1 次正方行列) なので, 転置しても同じことを使った.

$$\mathbf{y}^\top X \hat{\boldsymbol{\beta}} = (\mathbf{y}^\top X \hat{\boldsymbol{\beta}})^\top = \hat{\boldsymbol{\beta}}^\top X^\top \mathbf{y}.$$

最尤推定 $\leadsto S_e$ 最小化 $\leadsto k = 1, 2, 3$ 成分で微分

$$\frac{\partial S_e}{\partial \alpha_0} = 0 \leadsto -2(X^\top \mathbf{y} \text{ の第 1 行}) + 2(X^\top X \hat{\boldsymbol{\beta}} \text{ の第 1 行}) = 0$$

$$\frac{\partial S_e}{\partial \beta_1} = 0 \leadsto -2(X^\top \mathbf{y} \text{ の第 2 行}) + 2(X^\top X \hat{\boldsymbol{\beta}} \text{ の第 2 行}) = 0$$

$$\frac{\partial S_e}{\partial \beta_2} = 0 \leadsto -2(X^\top \mathbf{y} \text{ の第 3 行}) + 2(X^\top X \hat{\boldsymbol{\beta}} \text{ の第 3 行}) = 0$$

あわせて

$$\frac{\partial S_e}{\partial \boldsymbol{\beta}} = -2X^\top \mathbf{y} + 2X^\top X \hat{\boldsymbol{\beta}} = \mathbf{0}.$$

$$-2X^{\top} \mathbf{y} + 2X^{\top} X \hat{\boldsymbol{\beta}} = \mathbf{0}.$$

を $\boldsymbol{\beta}$ について解いて、 $X^{\top} X$ の逆行列が存在するなら、

命題 (線形重回帰モデルにおける回帰係数の最尤推定値)

$$\hat{\boldsymbol{\beta}} = (X^{\top} X)^{-1} X^{\top} \mathbf{y}.$$

$X^\top X$ とは？

$p = 2$ のとき

$$\frac{1}{n}X^\top X = \begin{pmatrix} 1 & 0 & 0 \\ 0 & s_{11} & s_{12} \\ 0 & s_{12} & s_{22} \end{pmatrix},$$

(1,2) 成分は, $\frac{1}{n} \sum_i 1 \cdot (x_{i1} - \bar{x}_{\bullet 1}) = \sum_i x_{i1} - n\bar{x}_{\bullet 1} = 0$. (1,3), (2,1), (3,1) も同様.

$$s_{jk} = \frac{1}{n} \sum_{i=1}^n (x_{ij} - \bar{x}_{\bullet j})(x_{ik} - \bar{x}_{\bullet k})$$

は, 不偏でない共分散 ($j = k$ なら標本分散) .

命題 (線形重回帰モデルにおける $X^\top X$)

$p+1$ 次正方行列で,

$$\frac{1}{n}X^\top X = \left(\begin{array}{c|c} 1 & 0^\top \\ \hline 0 & S \end{array} \right).$$

S は, $x_1, \dots, x_p$ の共分散行列.

$p = 1$ のとき

$$x_1 = x \text{ と書くと, } (X^\top X)^{-1} = \frac{1}{n} \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{s_{xx}} \end{pmatrix} .$$

$$\begin{aligned} \hat{\beta} &= (X^\top X)^{-1} X^\top \mathbf{y} = \frac{1}{n} \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{s_{xx}} \end{pmatrix} \begin{pmatrix} \sum y_i \\ \sum (x_i - \bar{x}_\bullet) y_i \end{pmatrix} \\ &= \begin{pmatrix} \frac{1}{s_{xx}} \frac{1}{n} \sum_i \frac{1}{n} y_i \\ \frac{1}{s_{xx}} \frac{1}{n} \sum_i (x_i - \bar{x}_\bullet) (y_i - \bar{y}_\bullet) \end{pmatrix} = \begin{pmatrix} \bar{y}_\bullet \\ \frac{s_{xy}}{s_{xx}} \end{pmatrix} . \end{aligned}$$

知っていた式を再現.