平成24年度　東京大学理学部情報科学科

月曜2限10:30～12:00

理7-214

知識処理論

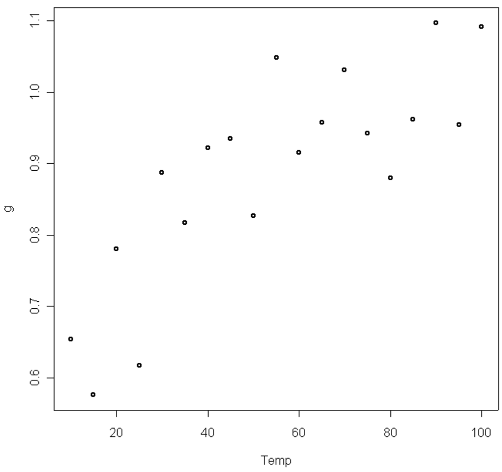
2012年4月16日(月)

井元清哉
東京大学医科学研究所
ヒトゲノム解析センター
DNA情報解析分野
http://bonsai.hgc.jp/~imoto
imoto@ims.u-tokyo.ac.jp

- 回帰モデル
- 正規分布
- 最小二乗法
- 尤度
- 最尤法
- 正規方程式
- ベクトルでの微分
- 数値最適化：ニュートン・ラフソン法

- 正則化法
- 正則化パラメータ
- ベイズアプローチ
- スパース学習

回帰分析



回帰モデル (加法ノイズ)

データ: $(x_1, y_1), \ldots, (x_n, y_n)$; n 個の観測値

モデル:

$$y_i = m(x_i) + e_i \qquad i = 1, \ldots, n$$

[観測値]=[シグナル]+[ノイズ] に分解

線形回帰モデル

データ: $(x_1, y_1), \ldots, (x_n, y_n)$; n 個の観測値

シグナル:

$$m(x_i) = E[Y_i \mid x_i] = a_0 + a_1 x_i$$

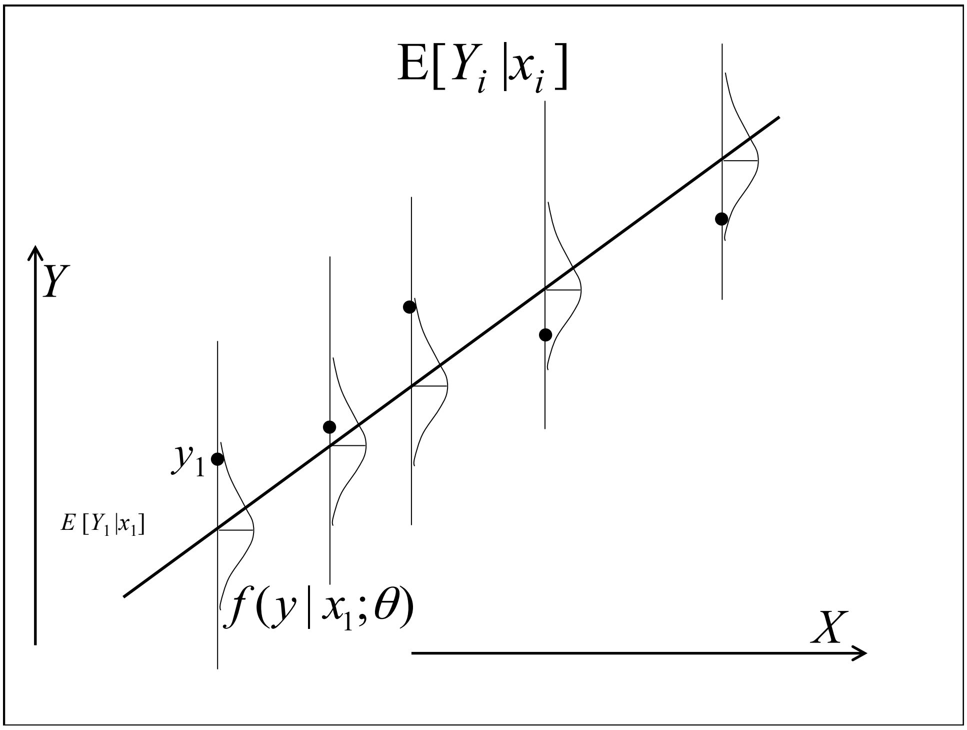
線形回帰モデル:

$$y_i = \underbrace{a_0 + a_1 x_i}_{x\text{ の線形式}} + e_i \qquad E[e_i] = 0,\ V[e_i] = \sigma^2$$

a_0, a_1 : パラメータ

正規線形回帰モデル:

$$e_i \sim N(0, \sigma^2): \quad \text{正規分布 平均0, 分散}\sigma^2$$



回帰モデル (一般化)

データ: $(x_1, y_1),\ldots,(x_n, y_n)$; n 個の観測値

モデル:

確率的成分: $f(y\,|\,x;\boldsymbol{\theta})$

系統的成分: $E[Y_i\,|\,x_i] = m(x_i)$

例

データ: $(x_1, y_1), \ldots, (x_n, y_n)$; n 個の観測値

モデル:

確率的成分:

$$f(y_i \mid x_i; \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y_i - \mu_i)^2}{2\sigma^2}\right\}$$

系統的成分:

$$\mu_i = E[Y_i \mid x_i] = \beta_0 + \beta_1 x_i$$

正規線形回帰モデル:

$$f(y_i \mid x_i; \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}\right\}$$

モデルの当てはめ:最尤法

データ: $(x_1, y_1), \ldots, (x_n, y_n)$; n 個の観測値

正規線形回帰モデル:

$$f(y_i \mid x_i; \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}\right\}$$

パラメータ: β_0, β_1, σ^2

尤度:

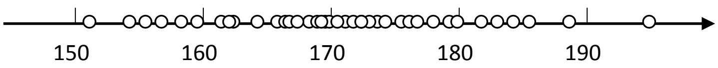
$$L(\beta_0, \beta_1, \sigma^2) = \prod_{i=1}^{n} f(y_i \mid x_i; \beta_0, \beta_1, \sigma^2)$$

$y_1, \ldots, y_n$ は互いに独立

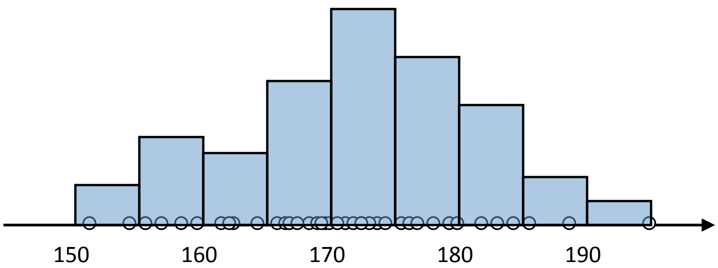
最尤法

データ: $x_1, \ldots, x_n$

（例えば，東京大学の男子学生の身長）



ヒストグラム



確率分布モデル

データ $x_1,\ldots,x_n$ は次の確率密度関数を持つ分布から得られたと仮定する

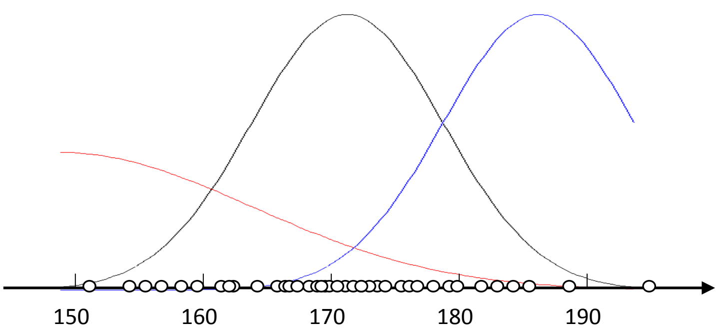
$$f(x\,|\,\theta),$$

ただし, θ はパラメータ.

例えば, 正規分布

$$f(x\,|\,\mu,\sigma^2)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

正規分布の当てはめ



どれがより良くデータに当てはまっているか？

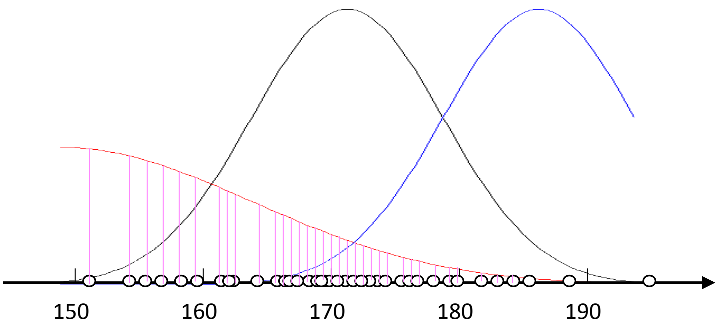
尤度

データ: $x_1,\ldots,x_n$

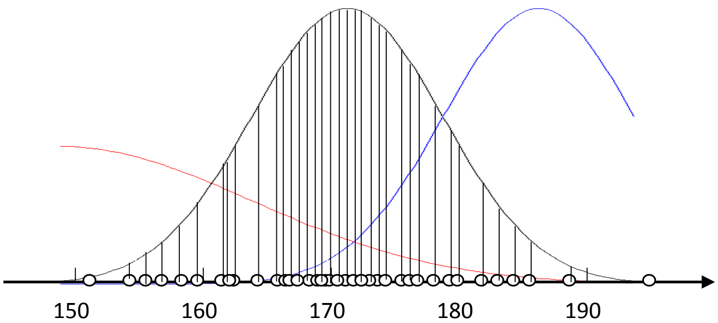
モデル: $f(x\,|\,\theta)$

尤度関数: $L(\theta)=\prod_{i=1}^{n} f(x_i\,|\,\theta)$

赤のモデルは？



黒のモデルは？



もっともデータに良く当てはまるパラメータ値は？

最尤推定量 $\hat{\theta}$

$$\hat{\theta} = \arg\max_{\theta} L(\theta)$$

$$= \arg\max_{\theta} \prod_{i=1}^{n} f(x_i \mid \theta)$$

通常は対数尤度関数の最大化

$$l(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log f(x_i \mid \theta)$$

正規モデルの最尤推定量

モデル: $$f(x_i \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(x_i - \mu)^2}{2\sigma^2} \right\}$$

対数尤度関数:

$$l(\mu, \sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

最尤推定量:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2$$

正規線形モデルの最尤推定

データ: $(x_1, y_1), \ldots, (x_n, y_n)$; n 個の観測値

正規線形回帰モデル:

$$f(y_i \mid x_i; \theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2} \right\}$$

パラメータ: β_0, β_1, σ^2

尤度: $$L(\beta_0, \beta_1, \sigma^2) = \prod_{i=1}^{n} f(y_i \mid x_i; \beta_0, \beta_1, \sigma^2)$$

正規線形モデルの最尤推定は最小2乗法と等価である

$$l(\beta_0,\beta_1,\sigma^2) = \log L(\beta_0,\beta_1,\sigma^2)$$
$$= \sum_{i=1}^{n} \log f(y_i \mid x_i;\beta_0,\beta_1,\sigma^2)$$
$$= \sum_{i=1}^{n} \log \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y_i-\beta_0-\beta_1 x_i)^2}{2\sigma^2}\right\}$$
$$= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\underline{\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 x_i)^2}$$
$$= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\underline{S(\beta_0,\beta_1)}$$

β_0,β_1 に対する $l(\beta_0,\beta_1,\sigma^2)$の最大化⇔ $S(\beta_0,\beta_1)$ の最小化

最小2乗法

データ: $(x_1, y_1),\ldots,(x_n, y_n)$; n 個の観測値

線形回帰モデル:

$y_i = \beta_0 + \beta_1 x_i + e_i \qquad E[e_i]=0,\ Var[e_i]=\sigma^2$

x の線形式

$\beta_0,\ \beta_1$; パラメータ

パラメータ推定:

$$S(\beta_0,\ \beta_1) = \sum_{i=1}^{n}\{y_i - (\beta_0 + \beta_1 x_i)\}^2$$

を最小化

正規線形モデルに対する

最尤推定量(=最小2乗推定量)

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

$$= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\boxed{(\boldsymbol{y} - \boldsymbol{X\beta})^T(\boldsymbol{y} - \boldsymbol{X\beta})}$$

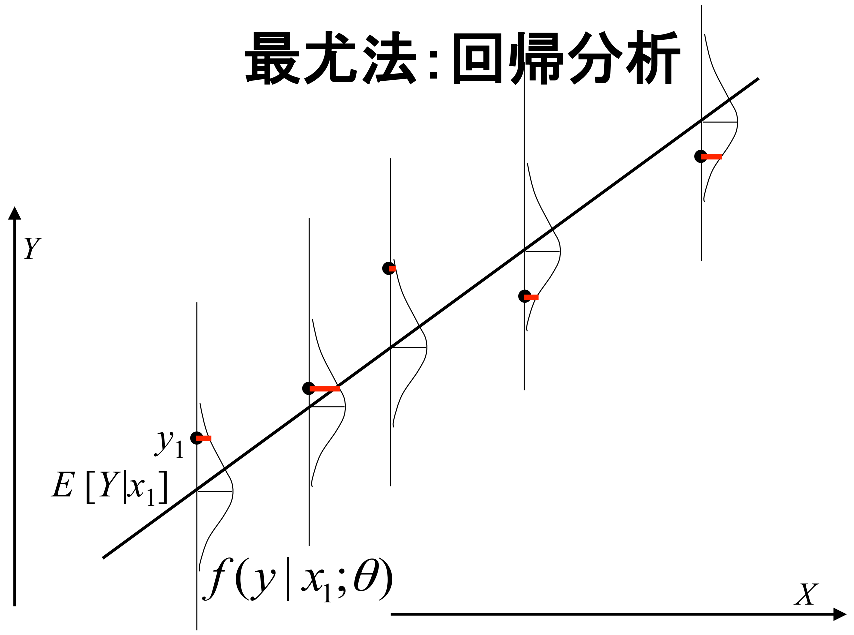
ただし

$$\boldsymbol{y} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad \boldsymbol{X} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

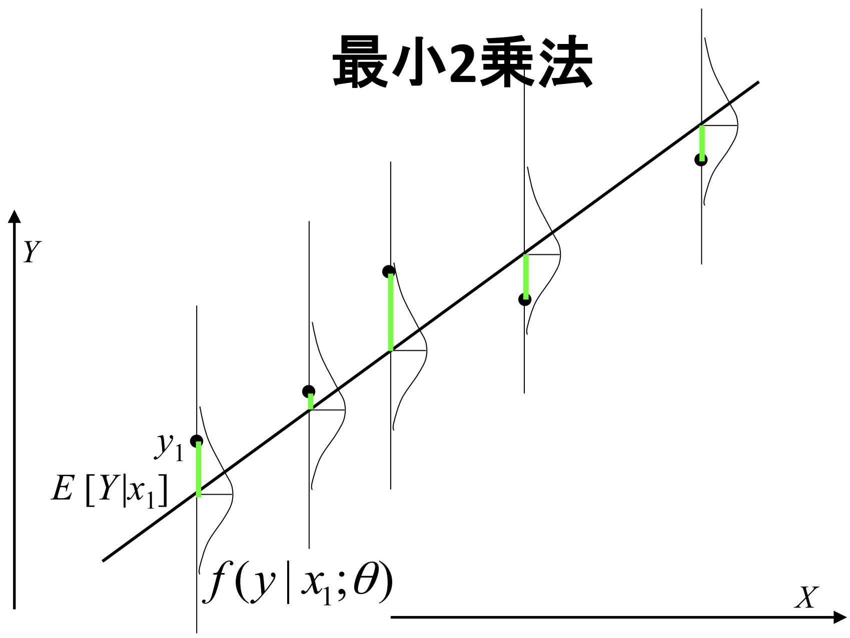
$$\boxed{(\boldsymbol{y} - \boldsymbol{X\beta})^T(\boldsymbol{y} - \boldsymbol{X\beta})} = \{\boldsymbol{\beta} - (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}\}^T(\boldsymbol{X}^T\boldsymbol{X})\{\boldsymbol{\beta} - (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}\} + \boldsymbol{y}^T\{\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\}\boldsymbol{y}$$

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$$

最尤法:回帰分析



最小2乗法



多変量データ

データ: $(x_{11}, x_{12}, \ldots, x_{1p} y_1),$

$(x_{21}, x_{22}, \ldots, x_{2p}, y_2),$

$\cdots\cdots\cdots,$

$(x_{n1}, x_{n2}, \ldots, x_{np}, y_n)$

線形重回帰モデル:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{22} + \cdots + \beta_p x_{ip} + \varepsilon_i$$

正規線形重回帰モデルの最尤推定

線形重回帰モデルの尤度も単回帰と同様

$$l(\boldsymbol{\beta},\sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$$
$$= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$$

とかける

ベクトルでの微分

定義　$\boldsymbol{\theta} = (\theta_1,\ldots,\theta_p)^T$ としたとき

$$\frac{\partial f(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \left(\frac{\partial f(\boldsymbol{\theta})}{\partial \theta_1},\quad \cdots,\quad \frac{\partial f(\boldsymbol{\theta})}{\partial \theta_p} \right)^T$$

公式

$\boldsymbol{a} = (a_1,\ldots,a_p)^T,\quad \boldsymbol{x} = (x_1,\ldots,x_p)^T,\quad \boldsymbol{B} = (b_{ij})_{p\times p}$ のとき

$$(I)\ \frac{\partial \boldsymbol{a}^T\boldsymbol{x}}{\partial \boldsymbol{x}} = \boldsymbol{a}$$

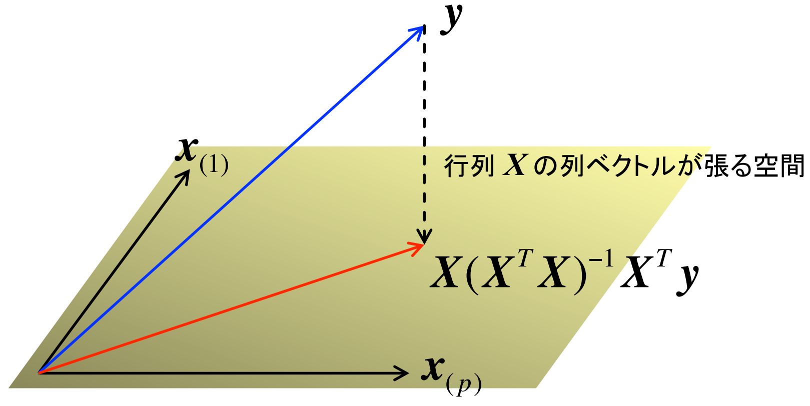
$$(II)\ \frac{\partial \boldsymbol{x}^T\boldsymbol{x}}{\partial \boldsymbol{x}} = 2\boldsymbol{x},\ \frac{\partial \boldsymbol{x}^T\boldsymbol{B}\boldsymbol{x}}{\partial \boldsymbol{x}} = (\boldsymbol{B}+\boldsymbol{B}^T)\boldsymbol{x},\ \frac{\partial \boldsymbol{x}^T\boldsymbol{B}\boldsymbol{y}}{\partial \boldsymbol{x}} = \boldsymbol{B}\boldsymbol{y}$$

$$l(\boldsymbol{\beta},\sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$$
$$= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^T(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})$$

$$\left.\frac{\partial l(\boldsymbol{\beta},\sigma^2)}{\partial \boldsymbol{\beta}}\right|_{\beta=\hat{\beta}} = -\frac{1}{2\sigma^2}(2\boldsymbol{X}^T\boldsymbol{X}\hat{\boldsymbol{\beta}} - 2\boldsymbol{X}^T\boldsymbol{y}) = 0$$
$$\Rightarrow \boldsymbol{X}^T\boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}^T\boldsymbol{y}$$
$$\Rightarrow \hat{\boldsymbol{\beta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$$

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$$

幾何的考察



$$\boldsymbol{X} = (\boldsymbol{x}_{(0)},\ldots,\boldsymbol{x}_{(p)})$$

$\boldsymbol{X}$ の列ベクトルは一次独立

説明変数間の相関、データ数

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{22} + \cdots + \beta_p x_{ip} + \varepsilon_i$$

$$\boldsymbol{X} = (\boldsymbol{x}_{(0)}, \ldots, \boldsymbol{x}_{(p)})$$ ：$n \times (p+1)$ 行列

$$\boldsymbol{X}^T \boldsymbol{X}$$ ：$(p+1) \times (p+1)$ 行列

$n < p+1$ であったり、独立な説明変数でない場合

$$rank(\boldsymbol{X}^T \boldsymbol{X}) < p+1$$

$\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$ の $(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ が存在しない