

クレジット:

Mathematics and Informatics Center 文科系のための線形代数・解析Ⅱ  
2020 藤堂 眞治・松尾 泰・藤原 毅夫

ライセンス:

利用者は、本講義資料を、教育的な目的に限ってページ単位で利用することができます。特に記載のない限り、本講義資料はページ単位でクリエイティブ・コモンズ 表示-非営利-改変禁止 ライセンスの下に提供されています。

<http://creativecommons.org/licenses/by-nc-nd/4.0/>

本講義資料内には、東京大学が第三者より許諾を得て利用している画像等や、各種ライセンスによって提供されている画像等が含まれています。個々の画像等を本講義資料から切り離して利用することはできません。個々の画像等の利用については、それぞれの権利者の定めるところに従ってください。



# § 線形回帰と最小二乗法

データ

$$\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

$$\vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

例  $x_i$ : i番目の人の講義の出席率

$y_i$ : 期末試験の成績

講義の出席率から 期末試験の成績を推定できるか?



$a$  と  $b$  がわかれば

他のデータに対しても推定可能

最小 = 乗法

誤差  $L(a, b) = \sum_{i=1}^n (y_i - (ax_i + b))^2$

= 乗の和を最小化する。

$L(a, b)$ :  $a, b$  について = 2次式

一変数の場合 変数  $x$   $y = ax^2 + bx + c$

( $a, b, c$  は定数)

$y$  が最小 / 最大となる  $x$  の値は何か

$$\frac{dy}{dx} = 2ax + b = 0 \quad \Rightarrow \quad x = -\frac{b}{2a}$$

多変数関数の微分  $\rightarrow$  偏微分

a, b 変数

$L(a, b)$  を最小に  $\vec{a}$ .

$$\frac{df}{dx} = \lim_{\epsilon \rightarrow 0} \frac{f(x+\epsilon) - f(x)}{\epsilon}$$

偏微分の定義

$$\frac{\partial L}{\partial a} = \lim_{\epsilon \rightarrow 0} \frac{L(a+\epsilon, b) - L(a, b)}{\epsilon}$$

$b$  を固定して  $a$  について  
微分

$$\frac{\partial L}{\partial b} = \lim_{\epsilon \rightarrow 0} \frac{L(a, b+\epsilon) - L(a, b)}{\epsilon}$$

$a$  を固定して  $b$  について  
微分

(例)  $f(a, b) = a^2 + 3ab - b^2 + 2b$

$$\left\{ \begin{array}{l} \frac{\partial f}{\partial a} = 2a + 3b \\ \frac{\partial f}{\partial b} = 3a - 2b + 2 \end{array} \right.$$

$$L(a, b) = \sum_{i=1}^n (y_i - (ax_i + b))^2$$

$(\vec{y} - (a\vec{x} + b\vec{u}))$   $n$  個目の成分

$$\vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad \vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}, \quad \vec{u} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

$$= \|\vec{y} - (a\vec{x} + b\vec{u})\|^2$$

$$\sum_i (x_i)^2 \Rightarrow \frac{\partial}{\partial a} \sum_i x_i^2 = \sum_i 2 \frac{\partial x_i}{\partial a} x_i$$

$$\frac{\partial L}{\partial a} = -2 \sum_{i=1}^n x_i (y_i - (ax_i + bu_i))$$

$$= -2 \vec{x} \cdot (\vec{y} - (a\vec{x} + b\vec{u})) = 0$$

$$\left\{ \begin{array}{l} x_i = y_i - (ax_i + bu_i) \\ \frac{\partial x_i}{\partial a} = -x_i \quad \frac{\partial x_i}{\partial b} = -u_i \end{array} \right.$$

$$\frac{\partial L}{\partial b} = -2 \vec{u} \cdot (\vec{y} - (a\vec{x} + b\vec{u})) = 0$$

$$\left\{ \begin{array}{l} \vec{x} \cdot \vec{y} - (a\vec{x} \cdot \vec{x} + b\vec{x} \cdot \vec{u}) = 0 \\ \vec{u} \cdot \vec{y} - (a\vec{u} \cdot \vec{x} + b\vec{u} \cdot \vec{u}) = 0 \end{array} \right.$$

2変数 連立線形方程式

$$\begin{bmatrix} \vec{x} \cdot \vec{x} & \vec{x} \cdot \vec{u} \\ \vec{u} \cdot \vec{x} & \vec{u} \cdot \vec{u} \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \vec{x} \cdot \vec{y} \\ \vec{u} \cdot \vec{y} \end{bmatrix} \quad *$$

$$\text{解は} \quad \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \vec{x} \cdot \vec{x} & \vec{x} \cdot \vec{u} \\ \vec{u} \cdot \vec{x} & \vec{u} \cdot \vec{u} \end{bmatrix}^{-1} \begin{bmatrix} \vec{x} \cdot \vec{y} \\ \vec{u} \cdot \vec{y} \end{bmatrix}$$

注意)  $n$  元  $n$  線形回帰の式は

$$y_i = ax_i + b$$

$$X = \begin{bmatrix} \vec{x} & \vec{u} \end{bmatrix} \quad n \times 2 \text{ 行列}$$

$$\vec{y} = X \begin{bmatrix} a \\ b \end{bmatrix}$$

未知変数 2

方程式の数  $n$

${}^t X$  :  $X$  の転置行列  $2 \times n$  行列

$$\begin{bmatrix} \vec{x} \cdot \vec{x} & \vec{x} \cdot \vec{u} \\ \vec{u} \cdot \vec{x} & \vec{u} \cdot \vec{u} \end{bmatrix} = {}^t X \cdot X, \quad \begin{bmatrix} \vec{x} \cdot \vec{y} \\ \vec{u} \cdot \vec{y} \end{bmatrix} = {}^t X \cdot \vec{y}$$

$${}^t X \cdot X \begin{bmatrix} a \\ b \end{bmatrix} = {}^t X \cdot \vec{y}$$

} 未知変数 2コ  
} 方程式の数 2コ

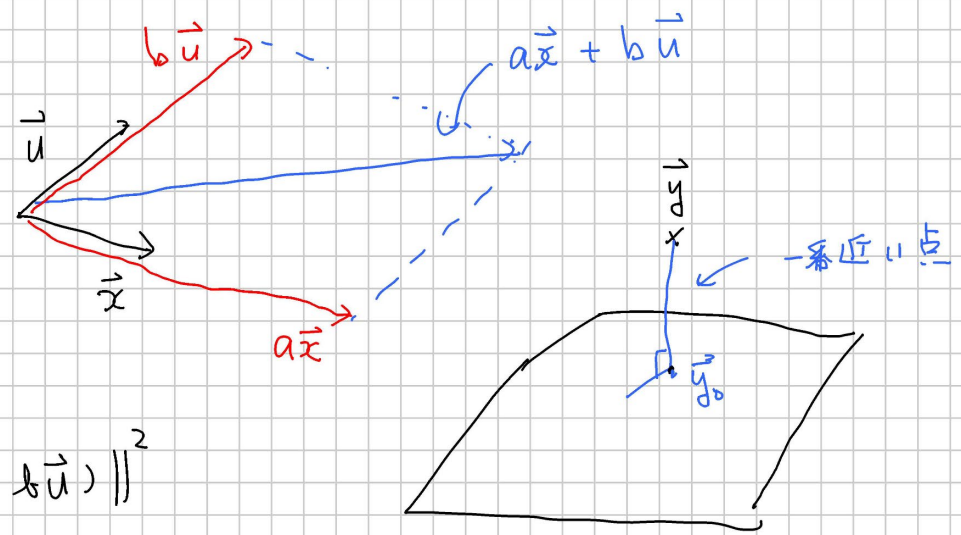
${}^t X \cdot X$  : グラム行列

最小二乗法の幾何学的イメージ

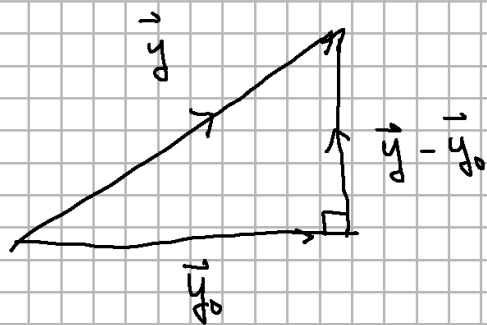
$\vec{y}$  と  $a\vec{x} + b\vec{u}$  をできるだけ近づけるために  $a, b$  を選ぶ

$\vec{x}, \vec{u} \in \mathbb{R}^n$       $n$ 次元ベクトル

$a\vec{x} + b\vec{u}$



$$L(a, b) = \| \vec{y} - (a\vec{x} + b\vec{u}) \|^2$$



≡ 平方の定理

$$\vec{y} = (\vec{y} - \vec{y}_0)^2 + \vec{y}_0^2$$

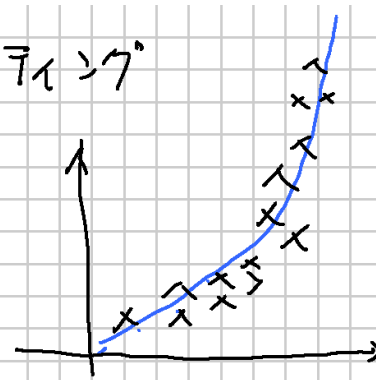
⇔ 同値

\* の方程式



▷ より複雑な関数を用いたフィッティング

= 二次関数を用いたフィッティング



$$y = \underline{a_0} + \underline{a_1}x + \underline{a_2}x^2$$

$a_0, a_1, a_2$  をうまく選んで 誤差の二乗を最小化

$$L(a_0, a_1, a_2) = \sum_{i=1}^n (y_i - (a_0 + a_1 x_i + a_2 x_i^2))^2$$

$$\vec{x}^{(0)} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

$$\vec{x}^{(1)} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

$$\vec{x}^{(2)} = \begin{bmatrix} (x_1)^2 \\ \vdots \\ (x_n)^2 \end{bmatrix}$$

$$L = \left\| \vec{y} - (a_0 \vec{x}^{(0)} + a_1 \vec{x}^{(1)} + a_2 \vec{x}^{(2)}) \right\|^2$$

誤差を最小化

$$\frac{\partial L}{\partial a_0} = \frac{\partial L}{\partial a_1} = \frac{\partial L}{\partial a_2} = 0$$

$$\frac{\partial L}{\partial a_0} = -2\vec{x}^{(0)} \cdot (\vec{y} - (a_0\vec{x}^{(0)} + a_1\vec{x}^{(1)} + a_2\vec{x}^{(2)})) = 0$$

$$\frac{\partial L}{\partial a_1} = -2\vec{x}^{(1)} \cdot (\vec{y} - (a_0\vec{x}^{(0)} + a_1\vec{x}^{(1)} + a_2\vec{x}^{(2)})) = 0$$

$$\frac{\partial L}{\partial a_2} = -2\vec{x}^{(2)} \cdot (\vec{y} - (a_0\vec{x}^{(0)} + a_1\vec{x}^{(1)} + a_2\vec{x}^{(2)})) = 0$$

$$X = [\vec{x}^{(0)} \quad \vec{x}^{(1)} \quad \vec{x}^{(2)}] : n \times 3 \text{ 行列}$$

$$\boxed{{}^t X \cdot X \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = {}^t X \cdot \vec{y}} \Rightarrow \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = ({}^t X \cdot X)^{-1} {}^t X \cdot \vec{y}$$

$${}^t X \cdot X : 3 \times 3 \text{ 行列}$$

$${}^t X \cdot \vec{y} : 1 \times n \text{ 行列}$$

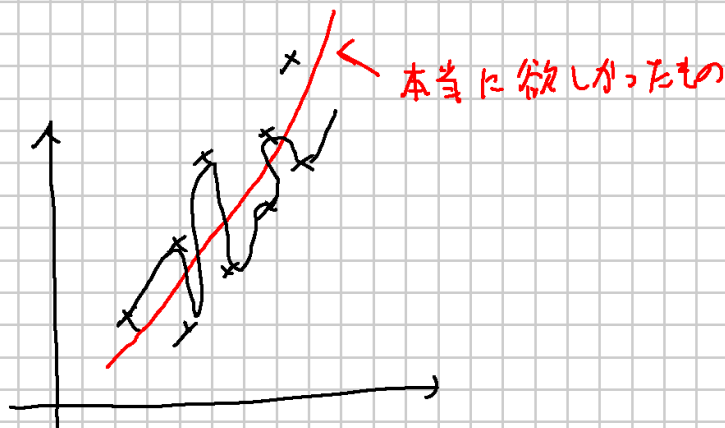
- 一般化

$$y = a_0 + a_1 x + a_2 x^2 + \dots + a_m x^m$$

$$X = \begin{bmatrix} 1 & x_1 & x_1^2 & \dots & (x_1)^m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \dots & (x_n)^m \end{bmatrix}$$

$${}^t X X \begin{bmatrix} a_0 \\ \vdots \\ a_m \end{bmatrix} = {}^t X \cdot \vec{y}$$

図学習



正則化最小二乗法

正則化項を新たに加えて過学習を防ぐ

$$L = \sum_{i=1}^n (y_i - \sum_{r=0}^m a_r (x_i)^r)^2 + \lambda \sum_{r=0}^m a_r^2$$

$$\frac{\partial L}{\partial a_0} = \frac{\partial L}{\partial a_1} = \dots = \frac{\partial L}{\partial a_m} = 0$$

$$({}^t X \cdot X + \lambda E) \begin{bmatrix} a_0 \\ \vdots \\ a_m \end{bmatrix} = {}^t X \cdot \underline{y}$$

$$\begin{bmatrix} a_0 \\ \vdots \\ a_m \end{bmatrix} = \underline{\underline{({}^t X \cdot X + \lambda E)^{-1}}} \cdot {}^t X \cdot \underline{y}$$