§ 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のか?

$\boxed{y = ax + b}$

aとbがわかれば
他のデータに対しても推定可能

最小二乗法

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

二乗の和を最小化する.

$L(a,b)$:　a, b について二次式

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

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

多変数関数の微分 → 偏微分

a, b 変数　$L(a,b)$ を最小にする.

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

偏微分の定義

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

b を固定して a について微分

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

a を固定して b について微分

(例)　$f(a,b) = a^2 + 3a\underline{b} - \underline{b^2} + 2\underline{b}$

$$\begin{cases} \dfrac{\partial f}{\partial a} = 2a + 3b \\[2ex] \dfrac{\partial f}{\partial b} = 3a - 2b + 2 \end{cases}$$

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

$(\vec{y} - (a\vec{x} + b\vec{u}))$ の i 番目の成分

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

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

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

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

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

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

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

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

2変数 連立線形方程式

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

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

注意） 元々 線形回帰の式は

$y_i = a x_i + b$ 　　　 $X = [\vec{x} \ \ \vec{u}]$ 　 $n \times 2$ 行列

$\vec{y} = X \begin{bmatrix} a \\ b \end{bmatrix}$ 　　　 未知変数 2コ

方程式の数 nコ

${}^tX$: Xの 転置行列 　 $2 \times n$ 行列

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

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

$\left\{ \begin{array}{l} \text{未知変数 2コ} \\ \text{方程式の数 2コ} \end{array} \right.$

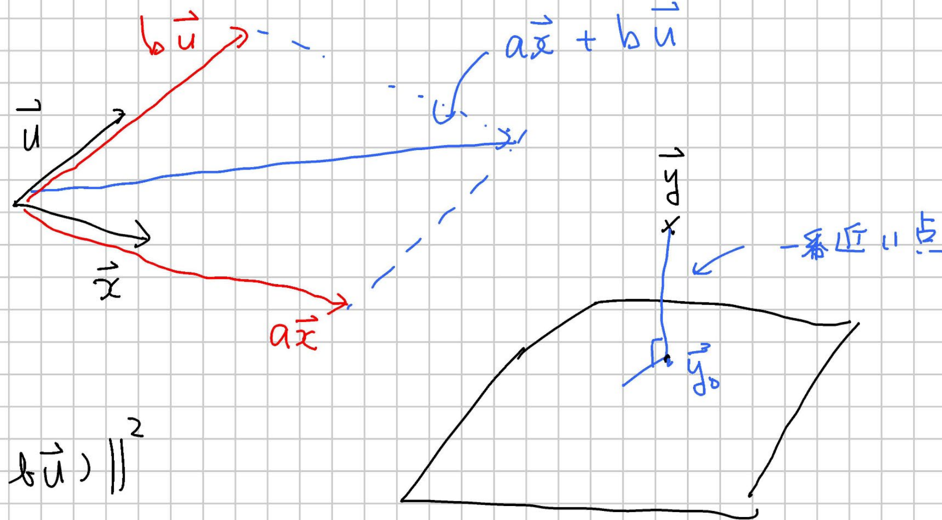
${}^{t}X \cdot X$: グラム行列

最小二乗法の幾何学的イメージ

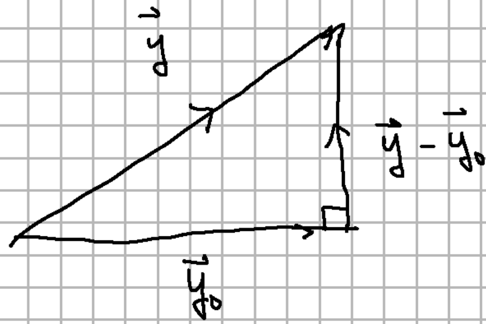
$\vec{y}$ と $a\vec{x} + b\vec{u}$ をできるだけ近づけるように a, b を選ぶ

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

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



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



三平方の定理

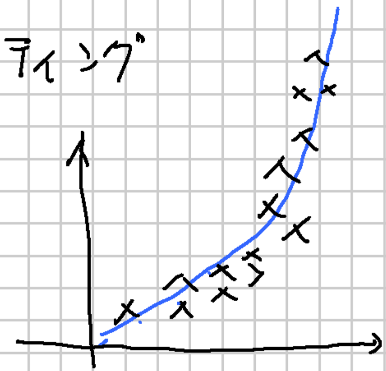
$$\vec{y} = (\vec{y}-\vec{y_0})^2 + \vec{y_0}^2$$

$\Updownarrow$ 同値

κ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

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

$$\vec{x}^{(0)} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \quad \vec{x}^{(1)} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \quad \vec{x}^{(2)} = \begin{bmatrix} (x_1)^2 \\ \vdots \\ (x_n)^2 \end{bmatrix}$$

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

誤差を最小化

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

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

$$\frac{\partial L}{\partial a_1} = -2\vec{x}^{(1)} \cdot (\quad \text{同じ} \quad) = 0$$

$$\frac{\partial L}{\partial a_2} = -2\vec{x}^{(2)} (\quad \text{同じ} \quad) = 0$$

$$X = \begin{bmatrix} \vec{x}^{(0)} & \vec{x}^{(1)} & \vec{x}^{(2)} \end{bmatrix} \quad : n \times 3 \text{ 行列}$$

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

${}^tX \cdot X$: 3×3 行列

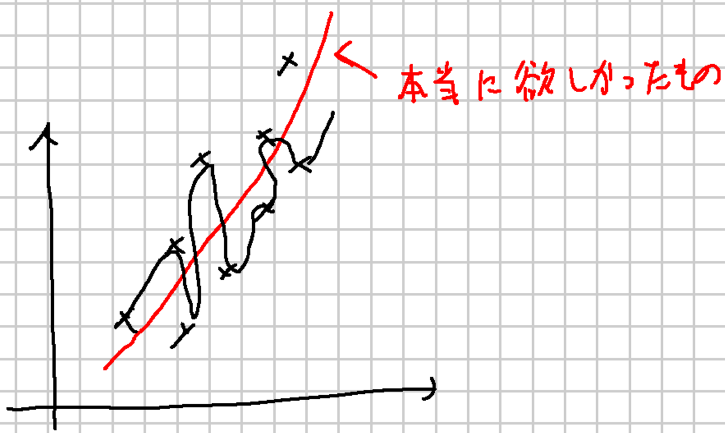
${}^tX \cdot X$: グラム行列

一般化　$y = a_0 + a_1 x + a_2 x^2 + \cdots + a_m x^m$

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

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

過学習



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

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

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

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

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