元线性回归:

$$y = ax + b \xrightarrow[\hat{a},\hat{b}\ \text{known}]{\text{Assume}} y_{pred} = \hat{a}x_i + \hat{b}$$

$$SSE = \sum (y_i - \hat{y}_{pred})^2$$

$$= \sum_i (y_i - \hat{a}x_i - \hat{b})^2$$

[Now, because (x_i, y_i) are known, this SSE is a function of $(\hat{a}, \hat{b})$, because they are unknown!!!]

Need To Minimise SSE. Note more proof is needed for SSE being convex.

$$\frac{\delta SSE}{\delta \hat{b}} = \frac{\delta \sum_i (y_i - \hat{a}x_i - \hat{b})^2}{\delta \hat{b}}$$

$$= 2\sum_i (y_i - \hat{a}x_i - \hat{b}) \cdot (-1) = 0$$

$$\sum_i y_i - \hat{a}\sum_i x_i - \sum_i \hat{b} = 0$$

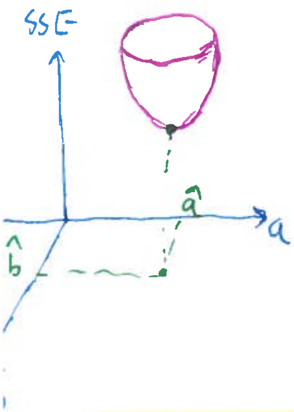
$$n\bar{Y} - \hat{a} \cdot n\bar{X} - n\hat{b} = 0$$

$$\hat{b} = \bar{Y} - \hat{a}\bar{X}$$

$$\frac{\delta SSE}{\delta \hat{a}} = \frac{\delta \sum_i (y_i - \hat{a}x_i - \hat{b})^2}{\delta \hat{a}}$$

$$= 2\sum_i (y_i - \hat{a}x_i - \hat{b}) \cdot (-x_i)$$

$$= (-2)\left[\sum x_i y_i - \hat{a}x_i^2 - \hat{b}x_i\right] = 0$$



$$\sum x_i y_i - \hat{a}\sum x_i^2 - \hat{b}\sum X_i = 0$$

$$\sum x_i y_i - \hat{a}\sum x_i^2 - \sum x_i (\bar{Y} - \hat{a}\bar{X}) = 0$$

$$\sum x_i y_i - \hat{a}\sum x_i^2 - n\bar{x}\bar{Y} + \hat{a}n\bar{X}^2 = 0$$

$$\hat{a} = \frac{\sum x_i y_i - n\bar{x}\bar{Y}}{\sum x_i^2 - n\bar{x}^2} = \frac{\sum (x_i - \bar{x})(Y_i - \bar{Y})}{\sum (x_i - \bar{x})(x_i - \bar{x})}$$

$$\hat{b} = \bar{Y} - \hat{a}\bar{X}$$

$$\sum x_i y_i - n\bar{x}\bar{Y} = \hat{a}\left[\sum x_i^2 - n\bar{x}^2\right]$$

$$\hat{a} = \frac{\sum x_i y_i - n\bar{x}\bar{Y}}{\sum x_i^2 - n\bar{x}^2}$$

元: $(x_1, y_1), (x_2, y_2) \cdots \quad (x_n, y_n)$

$$y = ax + b$$

$$\hat{y}_1 = \hat{a}x_1 + \hat{b}$$

$$\hat{y}_2 = \hat{a}x_2 + \hat{b}$$

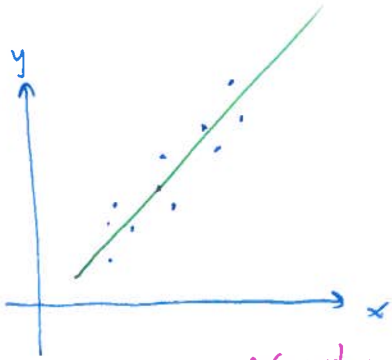
$$\vdots$$

$$\hat{y}_n = \hat{a}x_n + \hat{b}$$

X sample

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix}$$

n sample



元:

$$\hat{y}_1 = \hat{a}_1 x_{11} + \hat{a}_2 x_{12} + \cdots \hat{a}_p x_{1p} + b$$

$$\hat{y}_2 = \hat{a}_1 x_{21} + \hat{a}_2 x_{22} + \cdots \hat{a}_p x_{2p} + b$$

$$\vdots$$

$$\hat{y}_n = \hat{a}_1 x_{n1} + \hat{a}_2 x_{n2} + \cdots \hat{a}_p x_{np} + b$$

n samples

X samples

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & & \vdots \\ x_{31} & x_{32} & & \vdots \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

We have n samples, For each sample X_i,
X_i has p features $x_{i1} \cdots x_{ip}$

n samples we still have n ys, each y_i is just a number.

$$Y = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ x_{21} & \ddots & x_{2p} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix} \cdot \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \vdots \\ \hat{a}_p \end{bmatrix} + b$$

$n \times 1 \qquad n \times p \qquad p \times 1$

$$Y = X \cdot \Theta + b$$

$$Y = X \cdot \Theta + b$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_p \end{bmatrix} + \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}$$

$$= \begin{bmatrix} 1 \cdot b_1 + \sum_{j=1}^{p} x_{1j} \cdot \theta_j \\ 1 \cdot b_2 + \sum_{j=1}^{p} x_{2j} \cdot \theta_j \\ \vdots \\ 1 \cdot b + \sum_{j=1}^{p} x_{nj} \cdot \theta_j \end{bmatrix}$$

$$= \underbrace{\begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix}}_{n \times (1+p)} \underbrace{\begin{bmatrix} b \\ \theta_1 \\ \vdots \\ \theta_p \end{bmatrix}}_{(1+p) \cdot 1}$$

b could be rewrite as θ_0

$$= X \cdot \Theta$$

$$Y = \underset{n \times (1+p)}{X} \cdot \underset{(1+p) \times 1}{\Theta}$$

Assume we know Θ

$$Y_{pred} = X\hat{\Theta}$$

$$SSE = \sum_{i=1}^{n} (Y - Y_{pred})^2$$

---

ector:

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \qquad X^T = [x_1 \cdots x_n]$$

$$\underset{1\times n}{[x_1 \cdots x_n]} \underset{n\times 1}{\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}} = x_1^2 + x_2^2 + \cdots x_n^2$$

$$= \sum_{i=1}^{n} x_i^2$$

$$\Downarrow$$

$$\sum x_i^2 = X^T X$$

$$SSE = \sum (Y - Y_{pred})^2$$

$$= (Y - Y_{pred})^T (Y - Y_{pred})$$

$$= (Y - X\hat{\Theta})^T (Y - X\hat{\Theta})$$

$$= Y^T Y - \hat{\Theta}^T X^T Y - \underset{Y^T X\hat{\Theta}}{\cancel{X^T\hat{\Theta}}} + \hat{\Theta}^T X^T X \hat{\Theta}$$

$$\underset{n\cdot 1}{Y} = \underset{n\cdot(1+p)}{X} \cdot \underset{(1+p)\cdot 1}{\Theta} \qquad \underset{(1+p)\cdot n}{X^T} \qquad \underset{1\cdot(1+p)}{\Theta^T} \qquad \underset{1\cdot n}{Y^T}$$

$$SSE = Y^TY - \underline{\hat\Theta^T X^T Y} - \underline{Y^T X \hat\Theta} + \hat\Theta^T X^T X \hat\Theta$$

$\hat\Theta^T X^T Y$: $\{1\cdot(1+p)\}\{(1+p)\cdot n\}\{n\cdot 1\} = 1\cdot 1$

$Y^T X \hat\Theta$: $\{1\cdot n\}\{n\cdot(1+p)\}\{(1+p)\cdot 1\} = 1\cdot 1$

Both results in a number

Therefore $\hat\Theta^T X^T Y + Y^T X \hat\Theta = 2\hat\Theta^T X^T Y$

$$\frac{\delta SSE}{\delta \Theta} = \frac{\delta\left[Y^TY - 2\hat\Theta^T X^T Y + \hat\Theta^T X^T X \hat\Theta\right]}{\delta\Theta}$$

$$= -2X^TY + 2X^TX\hat\Theta = 0$$

$$\Downarrow$$

$$\hat\Theta = \frac{X^TY}{X^TX} = (X^TX)^{-1}X^TY$$

ck $1-D\Theta$ correct.

$$\hat\Theta = \frac{\begin{bmatrix} 1 & \cdots & 1 \\ x_{11} & & \\ \vdots & & \\ x_{1p} & \cdots & x_{np}\end{bmatrix}\cdot\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n\end{bmatrix}}{\begin{bmatrix} 1 & \cdots & 1 \\ x_{11} & & \\ \vdots & & \\ x_{1p} & \cdots & x_{np}\end{bmatrix}\cdot\begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & & x_{np}\end{bmatrix}}$$

when univariate($-$)
$p=1$

$$\hat\Theta = \frac{\begin{bmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n\end{bmatrix}\begin{bmatrix} y_1 \\ \vdots \\ y_n\end{bmatrix}}{\begin{bmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n\end{bmatrix}\begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n\end{bmatrix}} = \begin{bmatrix} n & \Sigma x_i \\ \Sigma x_i & \Sigma x_i^2\end{bmatrix}^{-1}\begin{bmatrix} \Sigma y_i \\ \Sigma x_i y_i\end{bmatrix}$$

$$= \frac{1}{n\Sigma x_i^2 - n^2\bar{x}^2}\begin{bmatrix} \Sigma x_i^2 & -\Sigma x_i \\ -\Sigma x_i & n\end{bmatrix}\begin{bmatrix} \Sigma y_i \\ \Sigma x_i y_i\end{bmatrix}$$

$$= \frac{1}{n\Sigma x_i^2 - n^2\bar{x}^2}\begin{bmatrix} \Sigma x_i^2 \cdot n\bar{y} - n\bar{x}\cdot\Sigma x_i y_i \\ -\Sigma x_i \Sigma y_i + n\Sigma x_i y_i\end{bmatrix} = \begin{bmatrix} \bar{y} - a\bar{x} \\ \dfrac{\Sigma x_i y_i - n\bar{x}\bar{y}}{\Sigma x_i^2 - n\bar{x}^2}\end{bmatrix} = \begin{bmatrix} \hat{b} \\ \hat{a}\end{bmatrix}$$

Inverse of a matrix

Normally, a $n \times n$ (square) matrix has inverse

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \qquad A^{-1} = \frac{1}{\det(A)} \begin{bmatrix} -a & c \\ b & -d \end{bmatrix}$$

$$AA^{-1} = AA^{-1} = I$$

If $A = \begin{bmatrix} 1 & 1 & 2 \\ 5 & 3 & 11 \\ 2 & 2 & 4 \end{bmatrix} \Longrightarrow \begin{bmatrix} -1 & 1 & 2 \\ 0 & 2 & -1 \\ 0 & 0 & 0 \end{bmatrix}$, rank is not full,

Then $\det(A) = 0$ A^{-1} does not exist

$$Y = X\Theta \qquad \hat{\Theta} = (X^T X)^{-1} X^T Y$$

$$X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix}$$

$\underset{n \times (p+1)}{X}$, $\underset{(p+1) \cdot n}{X^T}$ $\underset{(p+1 \times n)(n \times p+1)}{X^T X} = (p+1 \times p+1)$

Tip: For a matrix A having rank $= n$. $rank(A) = rank(A^T) = rank(A^T A) = rank(AA^T)$

So, if X has multicollinearity. Then $rank(X) < p+1$. $X^T X_{(p+1 \times p+1)}$ not full rank

if X does not have multicollinearity. Then ~~rank~~ rank $(X) = p+1$. $X^T X_{(p+1, p+1)}$ full rank

If $rank(X) < p+1$. Then $\underset{(p+1)(p+1)}{X^T X}$ rank not full, $\det(X^T X) = 0$

If $rank(X) = p+1$, Then $X^T X$ rank is full. $\det(X^T X) > 0$, $(X^T X)^{-1}$ exist.