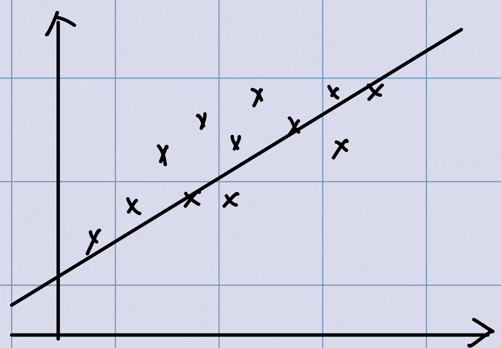
线性回归.

- 最小二乘法
- 概率角度：最小二乘法 $\Leftrightarrow$ noise为高斯的MLE
- 正则化 $\begin{cases} L_1 - \text{Lasso} \\ L_2 - \text{岭回归} \end{cases}$



$D: \{(x_1, y_1), (x_2, y_2), \ldots (x_n, y_n)\}$

$x_i \in \mathbb{R}^p$ p维列向量. $i = 1, 2 \cdots N$

$$X := (x_1 \ x_2 \ \ldots \ x_N)^T = \begin{pmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_N^T \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & & & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Np} \end{pmatrix}_{N \times p}$$

$$Y: \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix}_{N \times 1}$$

$$W: \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_p \end{pmatrix}$$

$$f(w) = w^T x + b \quad / \quad w^T x + w_0$$

- 最小二乘法，也叫参数估计.

$$L(w) = \sum_{i=1}^{N} \| w^T x_i - y_i \|^2 \qquad /* \text{差值的平方} */$$

$$= \sum_{i=1}^{N} (w^T x_i - y_i)^2$$

$$\xrightarrow{\text{矩阵}} \underbrace{\begin{pmatrix} w^T x_1 - y_1 & w^T x_2 - y_2 & \dots & w^T x_N - y_N \end{pmatrix}}_{(w^T x_1 \;\; w^T x_2 \; \dots \; w^T x_N) - (y_1 \;\; y_2 \dots y_N)} \underbrace{\begin{pmatrix} w^T x_1 - y_1 \\ w^T x_2 - y_2 \\ \vdots \\ w^T x_N - y_N \end{pmatrix}}_{XW - Y}$$

$$w^T (x_1 \;\; x_2 \dots x_N) - (y_1 \;\; y_2 \dots y_N)$$

$$w^T X^T - Y^T$$

$$\Rightarrow L(w) = (w^T X^T - Y^T)(WX - Y)$$

$$= w^T X^T X w - w^T X^T Y - Y^T w X + Y^T Y \qquad (\text{每一项皆一维实数})$$

$$= w^T X^T X w - 2 w^T X^T Y + Y^T Y$$

• $\hat{w}$ (估计的 w) $= \underset{w}{\arg\min}\, L(w)$

$$\Rightarrow \frac{\partial L}{\partial w} = 2wX^TX - 2X^TY + 0$$

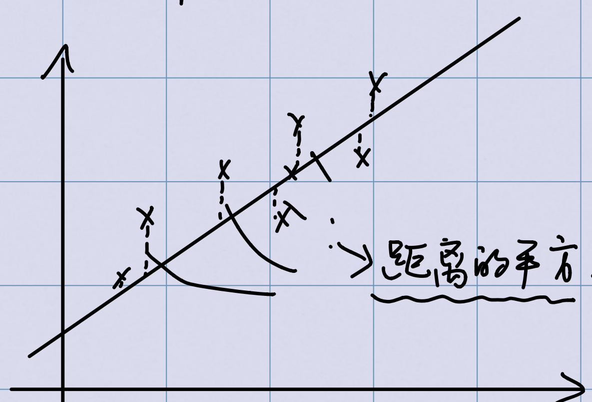
$$= 0$$

$$\Rightarrow wX^TX = X^TY$$

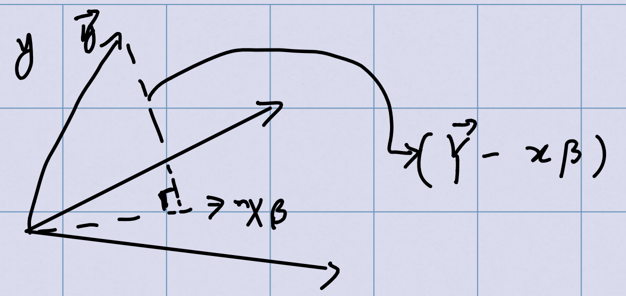
$$\Rightarrow \boxed{w = \underbrace{(X^TX)^{-1}X^T}_{\text{伪逆}}Y}$$

矩阵解释 ■

• 几何解释:



$$f(w) = w^Tx = x^T\beta \qquad \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & & & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{Np} \end{pmatrix}_{N\times p} \rightarrow p\text{维的子空间}$$



$$\Rightarrow X^T(Y - X\beta) \Rightarrow x^T Y \cdot \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = 0 \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$\Rightarrow X^T Y = X^T X \beta$$

$$\Rightarrow \beta = (X^T X)^{-1} X^T Y$$

- <u>概率视角的线性回归.</u>

$$D = \{(x_1, y_1) \cdots (x_N, y_N)\}$$

$x_i \in \mathbb{R}^p$. $y_i \in \mathbb{R}^1$. $i = 1, 2, \ldots N$.

$$X = (x_1 \; x_2 \; x_3 \ldots x_N)^T = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ \vdots & \vdots & & \vdots \\ x_{N1} & & & x_{Np} \end{pmatrix} \quad (N \times p)$$

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{pmatrix} \qquad w = \begin{pmatrix} w_1 \\ \vdots \\ w_p \end{pmatrix}$$

x：样本，y：值.

$$\begin{cases} L(w) = \sum_{i=1}^{N} \| w^T x_i - y_i \|^2 \\ \hat{w} = \arg\min_w L(w) \end{cases}$$

结果： $\hat{w} = (X^T X)^{-1} X^T Y$.

- 噪声： $\varepsilon \sim N(0, \sigma^2)$

$$y = f(w) + \varepsilon$$
$$= w^T x + b + \varepsilon$$
$$= w^T x + \varepsilon \qquad \bullet\ (\varepsilon \sim N(0, \sigma^2))$$

- $y|x;w \sim N(w^T x, \sigma^2)$

- MLE:

$$\mathcal{L}(w) = \log P(Y|X_i, w)$$

log-likelihood

$$= \log \prod_{i=1}^{N} P(y_i|x_i, w)$$

$$= \sum_{i=1}^{N} \log P(y_i|x_i, w)$$

$$\frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y - w^T x)^2}{2\sigma^2}\right)$$

(随机变量 - 均值)^2

$$= \sum_{i=1}^{N} \log \frac{1}{\sqrt{2\pi}\sigma} + \left(-\frac{(y_i - (w^T x_i))^2}{2\sigma^2}\right)$$

$$= \sum_{i=1}^{N} \log (\sqrt{2\pi}\sigma)^{-1} + \left(\left(-\frac{1}{2\sigma^2}\right)(y_i - w^T x_i)^2\right)$$

$$\underset{w}{\operatorname{argmax}}\, \mathcal{L}(w) : \quad \frac{\partial \mathcal{L}}{\partial w}$$

$$= \frac{\partial}{\partial w}\left(-\frac{1}{2\sigma^2}(y_i - w^T x_i)^2\right)$$

$$\Rightarrow \underset{w}{\operatorname{argmin}}\ \frac{1}{2\sigma^2}(y_i - w^T x_i)^2 .$$

$$\boxed{\Rightarrow \underset{w}{\operatorname{argmin}}\ (y_i - w^T x_i)^2}$$

$\Rightarrow$ 最小二乘分布.

且隐含了噪声 $\sim N(0, \sigma^2)$

$\mathcal{L}SE$(最小二乘估计) $\Leftrightarrow$ MLE(最大似然) 且 ε (noise) $\sim N(0, \sigma^2)$