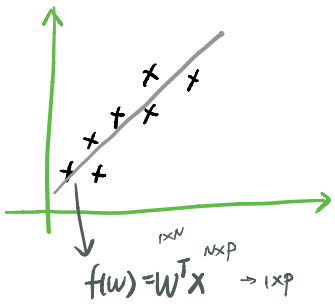
线性回归

总览

<1> 最小二乘法 { 矩阵表达 / 几何意义

<2> 概率角度：最小二乘法 ⟺ 噪声 (noise) 为 Gaussian (MLE) ？

<3> 正则化 { L1：Lasso / L2：Ridge regression 岭回归



$D = \{(x_1, y_1), (x_2, y_2) \cdots (x_N, y_N)\}$

$x_i \in R^p$（p维，列向量）　$y_i \in R$　$i = 1, 2 \cdots N$

- $X = (x_1, x_2, \cdots x_N)^T \Rightarrow N \times p$

$$\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & & & \\ x_{N1} & \cdots & & x_{Np} \end{pmatrix}_{N \times p}$$

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

a. 矩阵表达：

最小二乘估计：$L(w) = \sum_{i=1}^{N} \| w^T x_i - y_i \|^2$

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

$$= (w^T x_1 - y_1 \quad w^T x_2 - y_2 \cdots w^T x_N - y_N) \begin{pmatrix} w^T x_1 - y_1 \\ w^T x_2 - y_2 \\ \vdots \\ w^T x_N - y_N \end{pmatrix}$$

$$= \underline{w^T (x_1\ x_2 \cdots x_N) - (y_1\ y_2 \cdots y_N)} \begin{pmatrix} w^T x_1 - y_1 \\ w^T x_2 - y_2 \\ \vdots \\ w^T x_N - y_N \end{pmatrix}$$

$w^T X^T - Y^T$ $\quad (w^T X^T - Y^T)^T = Xw - Y$

$$\Rightarrow L(w) = (w^T X^T - Y^T)(Xw - Y)$$
$$= w^T X^T X w - w^T X^T Y - Y^T X w + Y^T Y$$
$$= w^T X^T X w - 2 w^T X^T Y + Y^T Y$$

?

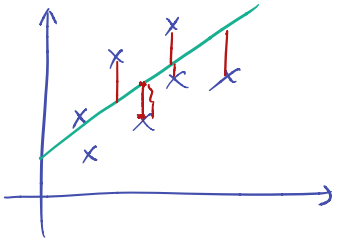
$$\hat{w} = \arg\min L(w)$$

$$\hookrightarrow \frac{\partial L(w)}{\partial w} = 2X^T X w - 2X^T Y = 0 \Rightarrow X^T X w = X^T Y$$
$$\Rightarrow w = (X^T X)^{-1} X^T Y$$

X^+ 伪逆

b. n维表达

①



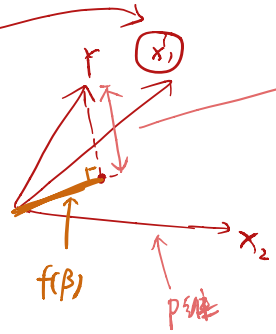
②

$$f(w) = w^T x = x^T \beta$$

$1 \times N \quad N \times p \quad p \times N \to p \times 1$

$$\begin{pmatrix} x_{11} & x_{21} & \cdots & x_{N1} \\ x_{12} & x_{22} & \cdots & x_{N2} \\ \vdots & & & \\ x_{1p} & x_{np} & \cdots & x_{Np} \end{pmatrix} \Rightarrow p \times 1$$

// Todo



$$x^T (Y - X\beta) = 0 \quad \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}_{p \times 1}$$

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

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

eg:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} (1 \ 2 \ 3) = \begin{pmatrix} x_1 \\ 2x_2 \\ 3x_3 \end{pmatrix}$$