概率视角

$D = \{(x_1, y_1), (x_2, y_2) \cdots (x_N, y_N)\}$　　样本独立同分布

$x_i \in R^p \quad y_i \in R \quad i = 1, 2 \cdots N$

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

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

最小二乘估计

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

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

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

假设数据噪声：　$\varepsilon \sim N(0, \sigma^2)$

$y = f(w) + \varepsilon$

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

$y = w^T x + \varepsilon$

$\Rightarrow \quad y|x; w \sim N(w^T x, \sigma^2) \quad \longrightarrow \quad P(y_i|x_i; w) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y_i - w^T x)^2}{2\sigma^2}\right)$

MLE：

定义 $\mathcal{L}(w) = \log P(Y|X; w) = \log \prod_{i=1}^{N} P(y_i|x_i; w) = \sum_{i=1}^{N} \log P(y_i|x_i; w)$

log似然

log-likehood

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

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

$$\hat{w} = \arg\max_w \mathcal{L}(w)$$

$$= \arg\max_w -\frac{1}{2\sigma^2}(y_i - w^T x_i)^2$$

$$= \boxed{\arg\min_w (y_i - w^T x_i)^2}$$

$\Rightarrow$ LSE $\Longleftrightarrow$ MLE

最小二乘估计 (noise 服从高斯分布)