线性回归的正则化.

- Loss function

$$\mathcal{L}(w) = \sum_{i=1}^{N} \|w^T x_i - y_i\|^2, \quad \text{where } i = 1, 2, \cdots N.$$

(最小二乘估计的结果)：$\hat{w} = \underbrace{(X^T X)^{-1}}_{①} X^T Y$.

1. 加数据
2. 特征提取
3. 正则化.(对w的约束)

①：有时是不可逆的.

此处，N个样本，$x_i \in \mathbb{R}^p$.

若 N 不是 >> p . ⇒ (X^TX) 不可逆、→ 过拟合.

若 N >> p ⇒ (X^TX) 可逆.

- 所以，引入正则化.(1. 不可逆问题 2. 过拟合问题)

---

1. 正则化的框架

〈对 w 的约束〉

$$\mathcal{L}(w) + \lambda \underbrace{P(w)}_{\text{惩罚项}} \Rightarrow \arg\min_{w} \left[\mathcal{L}(w) + \lambda P(w) \right]$$

1. $\mathcal{L}_1$ 正则化 / $\mathcal{L}_2$ 岭回归（权值衰减）

$P(w)=\|w\|_1$ 　　　　$P(w)=\|w\|_2=w^Tw$

（二范数）

- $J_2(w)=\mathcal{L}(w)+\lambda P(w)$

$$=\sum_{i=1}^{N}\|w^Tx_i-y_i\|^2+\lambda w^Tw$$

$$=(w^TX^T-Y^T)(Xw-Y)+\lambda w^Tw$$

$$=w^TX^TXw-w^TX^TY-Y^TXw+Y^TY+\lambda w^Tw$$

实数、

$$=w^TX^TXw-2w^TX^TY+Y^TY+\lambda w^Tw$$

$$=w^T(X^TX+\lambda I)w-2w^TX^TY+Y^TY$$

单位矩阵、

$$\Rightarrow \underset{w}{\arg\min}\, J_2(w)$$

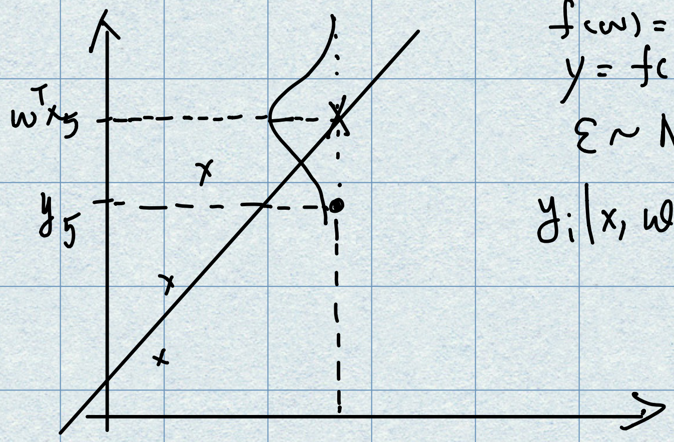
$$\Rightarrow \frac{\partial J_2(w)}{\partial (w)}=2(X^TX+\lambda I)w-2X^TY+0=0$$

$$\Rightarrow (X^TX+\lambda I)w=X^TY$$

$$\Rightarrow w=(X^TX+\lambda I)^{-1}X^TY$$

岭回归下的解析解

$\blacksquare$



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

$y = f(w) + \varepsilon$

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

$y_i | x, w \sim N(w^T x, \sigma^2)$

贝叶斯角度认正则化

- 当y看作证据.
- 给w一个先验. $w \sim N(0, \sigma^2)$

则后验 → $p(w|y) = \dfrac{p(y|w)\,p(w)}{p(y)}$

与w无关系,省略掉.

- MAP(最大后验)

$$\arg\max_{w} p(w|y)$$

$$= \arg\max_{w} p(y|w)\,p(w)$$

已知: $p(y|w) = \dfrac{1}{\sqrt{2\pi}\sigma} \exp\left(-\dfrac{(y-w^Tx)^2}{2\sigma^2}\right)$

$$p(w) = \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left(-\frac{(w-0)^2}{2\sigma_0^2}\right)$$

$\sim N(0, \sigma_0^2)$

则有：$p(y|w) \cdot p(w)$

$$= \frac{1}{\sqrt{2\pi}\sigma} \cdot \frac{1}{\sqrt{2\pi}\sigma_0} \exp\left\{-\frac{(y-w^Tx)^2}{2\sigma^2} - \frac{\|w\|^2}{2\sigma_0^2}\right\}$$

$$\hat{w} = \log \max_{w} p(y|w)p(w)$$

$$\Rightarrow \arg\max_{w} \log\frac{1}{\sqrt{2\pi}\sigma} + \log\frac{1}{\sqrt{2\pi}\sigma} + \log\exp\{\cdots\}$$

$\Rightarrow$ 保留与 w 相关

$$\Rightarrow \arg\max_{w} -\left(\frac{(y-w^Tx)^2}{2\sigma^2} + \frac{\|w\|^2}{2\sigma_0^2}\right)$$

$$\Rightarrow \arg\min_{w} \frac{(y-w^Tx)^2}{2\sigma^2} + \frac{\|w\|^2}{2\sigma_0^2} \qquad \underbrace{\times 2\sigma^2}$$

$$\Rightarrow (y-w^Tx)^2 + \frac{\sigma^2}{\sigma_0^2}\|w\|^2$$

而我们的定义：

$$\hat{w}_{MAP} = \arg\min_{w} \sum_{i=1}^{N}(y_i - w^Tx_i)^2 + \frac{\sigma^2}{\sigma_0^2}\|w\|_2^2$$

$$J_2(w) = \sum_{i=1}^{N}(w^Tx_i - y_i)^2 + \lambda w^Tw$$

$\downarrow$

L_2

☆ 正则化的LSE(最小二乘) $\Longleftrightarrow$ MAP(最大后验) $\begin{cases} \text{先验} \sim \text{高斯分布} \\ \text{其中 noise} \sim N(0,\sigma^2) \end{cases}$