

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

$$\frac{\partial J(w)}{\partial w} = 2(X^T X + \lambda I)w - 2X^T Y = 0$$

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

惩罚项

最小二乘估计 \Leftrightarrow 极大似然估计

LSE \Rightarrow MLE (noise 为 Gaussian 分布)

regularized LSE \Leftrightarrow MAP (最大后验)
(noise 和 prior 也是 GD)

(四) 线性分类

硬输出: {线性判别 (输出结果)}

软输出: {生成} $P(x|x) = \frac{P(x|y)P(y)}{P(x)}$

生成 $\Rightarrow \max P(x|y) \Rightarrow \max (P(x|y)P(y))$
后验 先验 相乘

1. 感知机 (Perceptron Learning Algorithm)

思想: 错误驱动

Vazyme

模型: $f(x) = \text{sign}(w^T x)$, $x \in \mathbb{R}^p, w \in \mathbb{R}^p$
Loss function:

$$L(w) = \sum_{i=1}^N I\{y_i w^T x_i < 0\}$$

D : {错误分类的样本}

$$L(w) = \sum_{x \in D} -y_i w^T x_i$$

$$\frac{\partial L(w)}{\partial w} = \sum_{x \in D} -y_i x_i$$

$$\text{SGD: } w^{(t+1)} = w^{(t)} - \lambda \frac{\partial L(w)}{\partial w} = w^{(t)} + \lambda y_i x_i \quad (\text{改变斜率})$$

(LDA) 线性判别分析 思想: 类内小, 类间大

投影方向 w , 为实践 w^T $\|w\|=1$

$$\bar{z}_i = w^T x_i$$

$$S_1 = \frac{1}{N} \sum_{i=1}^N z_i = \frac{1}{N} \sum_{i=1}^N w^T x_i$$

$$S_2 = \frac{1}{N} \sum_{i=1}^N (z_i - \bar{z})^T = \frac{1}{N} \sum_{i=1}^N (w^T x_i - \bar{z}) (w^T x_i - \bar{z})^T$$

$$C_1: \bar{z}_1 = \frac{1}{N} \sum_{i=1}^N w^T x_i \quad S_1 = \frac{1}{N} \sum_{i=1}^N (w^T x_i - \bar{z}_1) (w^T x_i - \bar{z}_1)^T$$

$$C_2: \bar{z}_2 = \frac{1}{N} \sum_{i=1}^N w^T x_i \quad S_2 = \frac{1}{N} \sum_{i=1}^N (w^T x_i - \bar{z}_2) (w^T x_i - \bar{z}_2)^T$$

类间: $(\bar{z}_1 - \bar{z}_2)^2 \uparrow$