Linear Regression

一. 最小二乘法(OLS)：最小化平方误差

一元: Square error $= \sum_{(x,y)\in S}(mx+\theta-y)^2$,

对参数 m、θ 求导:

正规方程 $\begin{cases} \dfrac{\partial L}{\partial m} = \sum_{(x,y)\in S} 2(mx+\theta-y)x = 0 \\ \dfrac{\partial L}{\partial \theta} = \sum_{(x,y)\in S} 2(mx+\theta-y) = 0 \end{cases}$

正规方程的解即可最小化平方误差!

从线性代数的角度看:

训练集 $(\vec{x_1}^d, y_1), (\vec{x_2}^d, y_2) \cdots (\vec{x_n}^d, y_n)$

$$A = n\left\{\begin{bmatrix} \leftarrow & \vec{x_1}^T & \rightarrow \\ \leftarrow & \vec{x_2} & \rightarrow \\ & \vdots & \\ \leftarrow & \vec{x_n} & \rightarrow \end{bmatrix}\right., \quad \underbrace{}_{d} \quad \vec{w}^d \qquad b = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

最小化: $\|A\vec{w}-\vec{b}\|_2^2$

$$\Rightarrow L = (A\vec{w}-\vec{b})^T(A\vec{w}-\vec{b}) = \vec{w}^TA^TAw - 2\vec{w}^TA^T\vec{b} + \vec{b}^T\vec{b}.$$

$$\therefore \frac{\partial L}{\partial w} = 2A^TAw - 2A^T\vec{b} = 0.$$

$$\Rightarrow (A^TA)\vec{w} = A^T\vec{b}$$

$$\therefore \vec{w} = (A^TA)^{-1}A^T\vec{b}$$ ☆ 解析解, 在A的行空间内.

∴ 但是, 若 A^TA 不可逆(矩阵A列不满秩), 该怎么办呢?

$\Rightarrow$ $\vec{w}$ 有无穷多解.

二. Normal Linear Regression Model

假设: $Y \mid \vec{X} = \vec{x} \sim N(\vec{w}\cdot\vec{x}, \sigma^2)$

求极大似然估计.

$$f(y_i \mid \vec{x_i} = \vec{x_i}) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{1}{2\sigma^2}(y_i - \vec{w_i}\vec{x_i})^2\right)$$

$$\therefore \ln f(y_i \mid \vec{x_i} = \vec{x_i}) = -\ln(\sqrt{2\pi}\,\sigma) - \frac{1}{2\sigma^2}(y_i - \vec{w_i}\vec{x_i})^2$$

$$\text{maximize } \ln f(y_i \mid \vec{x_i} = \vec{x_i}) \Leftrightarrow \min (y_i - \vec{w_i}\vec{x_i})^2.$$

∴对 Normal Linear Regression model 来说，

极大似然估计 (MLE) = 最小平方误差 (MSE) = 最小二乘 (OLS)

三. Regularization

当 A 列不满秩，有无穷解，如何选择解呢？

⇒ 引入 inductive bias 来 break the tie

目标函数: $L = \|A\vec{w} - \vec{b}\|_2^2 + \lambda \|\vec{w}\|_2^2$

理解: 相当于做了数据增强:

$$\left\{\begin{array}{ll} \vec{x_1} = (\sqrt{\lambda}, 0, 0, \cdots 0) & y_1 = 0 \\ \vec{x_2} = (0, \sqrt{\lambda}, 0, \cdots 0) & \Rightarrow \quad y_2 = 0 \\ \vdots & \quad\quad \vdots \\ \vec{x_d} = (0, 0, \cdots \sqrt{\lambda}) & y_d = 0 \end{array}\right.$$

增强数据的 Loss: $\lambda w_1^2 + \lambda w_2^2 + \cdots \lambda w_d^2 = \lambda \|\vec{w}\|_2^2.$

解析解: $\underbrace{(A^TA + \lambda I)}_{\text{一定可逆}} \vec{w} = A^T\vec{b}$

一定可逆 ⇒ 必有唯一解