

线性回归正规方程推导

nghuyong

December 25, 2018

1 问题定义

设 $X \in \mathbb{R}^{m \times n}$ 为设计矩阵, 其中 m 为样本数, n 为每个样本的特征数, $Y \in \mathbb{R}^m$. 我们的线性回归模型为:

$$X\theta + b = Y \quad (1)$$

模型的参数有两个: $\theta \in \mathbb{R}^n$ 和 $b \in \mathbb{R}$

为了形式上的简洁, 给每个样本最后加入一个特征, 值为 1, 同时将 b 添加到 θ 的末尾, 这样线性回归模型可以改写成:

$$X\theta = Y \quad (2)$$

这里的 $X \in \mathbb{R}^{m \times n+1}$, $\theta \in \mathbb{R}^{n+1}$

对于线性回归问题, 目标函数是均方误差最小, 即:

$$\theta = \arg \min_{\theta} J(\theta) = \frac{1}{2m} \|X\theta - Y\|^2 \quad (3)$$

2 正规方程求解模型参数

$$\begin{aligned} \nabla_{\theta} J(\theta) &= \nabla_{\theta} \frac{1}{2m} \|X\theta - Y\|^2 \\ &= \frac{1}{2m} \nabla_{\theta} (X\theta - Y)^T (X\theta - Y) \\ &= \frac{1}{2m} \nabla_{\theta} (\theta^T X^T X\theta - \theta^T X^T Y - Y^T X\theta + Y^T Y) \\ &= \frac{1}{2m} \nabla_{\theta} (\theta^T X^T X\theta - \theta^T X^T Y - Y^T X\theta) \end{aligned} \quad (4)$$

因为 $\theta^T X^T Y$ 是一个标量, 所以 $\theta^T X^T Y = (\theta^T X^T Y)^T = Y^T X \theta$, 所以

$$\begin{aligned}\nabla_{\theta} J(\theta) &= \nabla_{\theta} \frac{1}{2m} (\theta^T X^T X \theta - \theta^T X^T Y - Y^T X \theta) \\ &= \frac{1}{2m} \nabla_{\theta} (\theta^T X^T X \theta - 2Y^T X \theta)\end{aligned}\quad (5)$$

因为 $(\theta^T X^T X \theta - 2Y^T X \theta)$ 是一个标量, 对于一个标量 A , $A = \text{tr}(A)$, 并且迹运算, 满足 $\text{tr}(A + B) = \text{tr}(A) + \text{tr}(B)$ 所以

$$\begin{aligned}\nabla_{\theta} J(\theta) &= \frac{1}{2m} \nabla_{\theta} (\theta^T X^T X \theta - 2Y^T X \theta) \\ &= \frac{1}{2m} \nabla_{\theta} \text{tr}((\theta^T X^T X \theta - 2Y^T X \theta)) \\ &= \frac{1}{2m} \nabla_{\theta} (\text{tr}(\theta^T X^T X \theta) - \text{tr}(2Y^T X \theta))\end{aligned}\quad (6)$$

对于矩阵的迹运算求导有一个结论: $\nabla_A \text{tr}(ABA^T C) = B^T A^T C^T + BA^T C$, 所以, 令 $A = \theta^T$, $B = (X^T X)$, $C = I$, 有:

$$\begin{aligned}\nabla_{\theta} \text{tr}(\theta^T X^T X \theta) &= \nabla_{\theta} \text{tr}(\theta^T (X^T X) \theta I) \\ &= (X^T X)^T \theta I^T + (X^T X) \theta I \\ &= 2X^T X \theta\end{aligned}\quad (7)$$

对于矩阵的迹运算求导还有一个结论: $\nabla_{\theta} \text{tr}(W \theta) = W^T$, 所以:

$$\nabla_{\theta} \text{tr}(Y^T X \theta) = (Y^T X)^T = X^T Y \quad (8)$$

将上面两个结论, 代入(6)式:

$$\begin{aligned}\nabla_{\theta} J(\theta) &= \frac{1}{2m} \nabla_{\theta} (\text{tr}(\theta \theta^T X^T X) - 2Y^T X \theta) \\ &= \frac{1}{2m} (2X^T X \theta - 2X^T Y) \\ &= \frac{1}{m} (X^T X \theta - X^T Y)\end{aligned}\quad (9)$$

最后令上式为 0:

$$\begin{aligned}X^T X \theta - X^T Y &= 0 \\ X^T X \theta &= X^T Y \\ \theta &= (X^T X)^{-1} X^T Y\end{aligned}\quad (10)$$