Class 1 线性回归

2018年10月19日 15:05

- 模型假设:

$$h_\theta(x)=\sum_{i=1}^{m}\theta_i x_i=\theta^T x$$

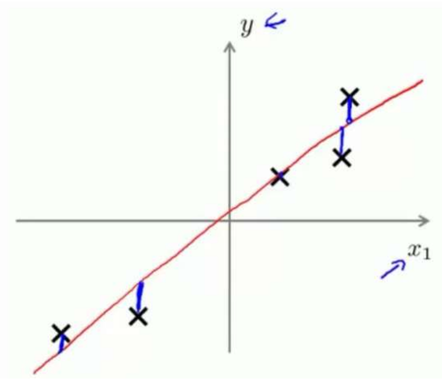
其中，θ 为参数，x为样本的特征，m为样本特征的维数。

- 损失函数：最小均方误差（LMS）

$$\begin{aligned}J(\theta)&=\frac{1}{2}\sum_{i=1}^{n}\left(h_\theta\left(x^{(i)}\right)-y^{(i)}\right)^2\\&=\frac{1}{2}\sum_{i=1}^{n}\left(\theta^T x^{(i)}-y^{(i)}\right)^2\end{aligned}$$

其中，n 为样本的个数。

- 优化目标:



$$\theta^* = \arg\theta\ \min J(\theta)$$

- 优化方法:
  - 标准方程法
    - **定义**

$$X=\begin{bmatrix}-\left(x^{(1)}\right)^{\mathrm{T}}-\\-\left(x^{(2)}\right)^{\mathrm{T}}-\\\vdots\\-\left(x^{(n)}\right)^{\mathrm{T}}-\end{bmatrix}\qquad\qquad y=\begin{bmatrix}y^{(1)}\\y^{(2)}\\\vdots\\y^{(n)}\end{bmatrix}$$

其中，X 表示所有样本的特征集合，y 表示每个样本对应的真实值。

- **然后有，**

$$X\theta-y=\begin{bmatrix}\left(x^{(1)}\right)^{\mathrm{T}}\theta\\\vdots\\\left(x^{(n)}\right)^{\mathrm{T}}\theta\end{bmatrix}-\begin{bmatrix}y^{(1)}\\\vdots\\y^{(n)}\end{bmatrix}=\begin{bmatrix}h_\theta(x^{(1)})-y^{(1)}\\\vdots\\h_\theta(x^{(n)})-y^{(n)}\end{bmatrix}$$

上式表示预测值与真实值之差的集合。

- **成本函数（最小二乘法）**

$$J(\theta)=\frac{1}{2}\sum_{i=1}^{n}\left(h_\theta\left(x^{(i)}\right)-y^{(i)}\right)^2=\frac{1}{2}(X\theta-y)^{\mathrm{T}}(X\theta-y)$$

- **通过矩阵微分计算梯度**

$$\begin{aligned}\nabla_\theta J(\theta)&=\nabla_\theta\frac{1}{2}(X\theta-y)^{\mathrm{T}}(X\theta-y)\\&=\frac{1}{2}\nabla_\theta\left(\theta^{\mathrm{T}}X^{\mathrm{T}}X\theta-\theta^{\mathrm{T}}X^{\mathrm{T}}y-y^{\mathrm{T}}X\theta+y^{\mathrm{T}}y\right)\\&=\frac{1}{2}\nabla_\theta\mathrm{tr}\left(\theta^{\mathrm{T}}X^{\mathrm{T}}X\theta-\theta^{\mathrm{T}}X^{\mathrm{T}}y-y^{\mathrm{T}}X\theta+y^{\mathrm{T}}y\right)\\&=\frac{1}{2}\nabla_\theta\left(\mathrm{tr}\theta^{\mathrm{T}}X^{\mathrm{T}}X\theta-2\mathrm{tr}y^{\mathrm{T}}X\theta\right)=X^{\mathrm{T}}X\theta-X^{\mathrm{T}}y\end{aligned}$$

- **通过令梯度等于零求得闭式解**

$$\theta^*=\boxed{(X^{\mathrm{T}}X)^{-1}}X^{\mathrm{T}}y$$

➢ 梯度下降法

- **简介：**
  - ❖ 梯度下降是求函数$f(\theta)$最小值的一阶迭代优化算法；
  - ❖ 梯度下降方向是函数值增加最快的方向。

- **梯度：**

$$\frac{\partial J_l(\theta)}{\partial \theta} = \frac{1}{2}\frac{\partial}{\partial \theta}\sum_{i=1}^{n}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)^2$$

$$= \frac{1}{2}\cdot 2\sum_{i=1}^{n}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)\cdot\frac{\partial}{\partial \theta}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)$$

$$= \sum_{i=1}^{n}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)\frac{\partial}{\partial \theta}\left(\theta^T x^{(i)}\right)$$

$$= \sum_{i=1}^{n}\boxed{\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)x^{(i)}}$$

“Error · Feature”

- **梯度下降优化：**

$$\theta := \theta - \alpha\frac{\partial}{\partial \theta}J_l(\theta)$$

$$= \theta - \alpha\sum_{i=1}^{n}\left(h_\theta\left(x^{(i)}\right) - y^{(i)}\right)x^{(i)}$$