最小二乘法

1.最小二乘法的原理：

目标函数 = $\sum$(观测值 - 理论值)^2

机器学习中的 loss function　　样本　　拟合函数的值.

例：线性回归 $(x^{(1)},y^{(1)})\cdots(x^{(m)},y^{(m)})$ 每个$x^{(i)}$只有一个特征x_1，那么样本的拟合函数：$h_\theta(x)=\theta_0+\theta_1x$

目标函数：$J(\theta_0,\theta_1)=\sum_{i=1}^{m}(y^{(i)}-h_\theta(x^{(i)}))^2=\sum_{i=1}^{m}(y^{(i)}-\theta_0-\theta_1x_1^{(i)})^2$ $\Rightarrow$ 求出使$J(\theta_0,\theta_1)$最小值的θ_0,θ_1，这样就得到了拟合函数.

2.最小二乘法的代数解法.

$J(\theta_0,\theta_1)$对θ_0 θ_1求导，令其等于0，二元一次方程 $\Rightarrow \theta_0,\theta_1$

多个样本特征，就是对$J(\theta_0,\theta_1,\ldots,\theta_n)$求多次导，令其等于0，得到$n+1$元一次方程，得$\theta_0,\ldots\theta_n$的解.

3.最小乘法的矩阵乘法.

多元线性回归：

假设函数：$h_\theta(x_1,x_2,\ldots,x_n)=\theta_0+\theta_1x_1+\cdots+\theta_n \Rightarrow h_\theta(X)=X\theta$　　$X-(m\times n)$　$\theta(n\times1)$　m—样本个数，n—样本特征数.

损失函数：$J(\theta)=\frac{1}{2}(X\theta-Y)^T(X\theta-Y)$　　Y样本输出向量　$Y—(m,1)$

$$\frac{\partial J(\theta)}{\partial\theta}=X^T(X\theta-Y)=0 \Rightarrow X^TX\theta=X^TY \Rightarrow \theta=(X^TX)^{-1}X^TY.$$

4.最小二乘法的局限性和使用场景

① 最小二乘法需要计算X^TX的逆矩阵，若逆矩阵不存在时，就不能用最小二乘法了，但仍可以用梯度下降。或去掉冗余特征，使X^TX的行列式不为0，然后用最小二乘法。

② 当特征n很大时X^TX逆矩阵求起来，很难而且很慢.

③ 拟合函数不是线性的，无法用最小二乘法？需要用技巧将其转化(x_1x_2, x_1^2) —— 可能θ不表示为$(X^TX^{-1})X^TY$