1.引言

言归正传，在此先列举一下最小二乘家族成员。最小二乘法**直线拟合**，最小二乘法**多项式（曲线）拟合**，机器学习中**线性回归**的最小二乘法，**系统辨识**中的最小二乘辨识法，参数估计中的最小二乘法，等等。由此可见，我们每次碰到的都是最小二乘法这个多面体的其中一个面。**如果只从单个面研究，就看不到它的整体，也就不能理解它的内涵。通过最小化误差的平方和，使得拟合对象无限接近目标对象**，这就是最小二乘的核心思想。

2.最小二乘法解析

1. 怎么列出误差方程?
2. 怎么最小化误差方程？
3. 怎么验证结果的准确性？

最小二乘法的核心，其实就是针对上述提出的三个问题的解法。

除此之外，**还有一个重要的内容：**

最小二乘应该说是一种思想，而只有结合了具体对象，才变成最小二乘法。这也就导致了多种多样的最小二乘公式、推导、证明等等。但是，其核心是最小二乘的思想，只是展示形式不同。那么，这个不同在哪里呢?

用一句话来说：基底不同。那么什么又是基底呢?

$$\boxed{y = \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n}$$

上式中的x1,x2,.....,xn就是基底。对于曲线拟合而言（见后面应用部分），x1=1，x2=x，x3=x^2.....。对于其他应用而言，x的含义又有所不同。看到这里，可能大家有一个疑问：都说**最小二乘求解的是线性问题，为什么曲线拟合方程中出现幂次方这个非线性的环节了呢?**

这就又涉及到了一个重要概念，那就是：**所谓线性方程，是有针对的对象的，也就是有所指的。**就像一个坐标点（x,y），其含义是针对x坐标轴和y坐标轴而言的。那么这里，对于曲线拟合过程。**虽然方程针对x是非线性的，但是针对各个参数θ而言，这个方程就是线性的了。**那么，如果所求是x，那么就无法用最小二乘法。**换句话说，如果针对所求参数，方程是线性的，那么就可以使用最小二乘法求解。**

2.1 误差方程

先说误差方程，就是用目标函数减去拟合函数，再取其平方即可

2.2 最小化误差函数

最小二乘的误差函数形式多样，但其解决方法与求一元二次方程极值的方法相同。无非就是将原来对变量求导，变成了对向量求偏导（就是在向量某一个维度上的导数）。如果不是对向量求导，而是对函数求导，就是复变函数的变分法。求导之后，就是另求导的式子为0，解出极值点。再判断该极值点是极大值点还是极小值点，这样就得到了使误差函数最小化的向量值。

2.3 矩阵求导解最小二乘

$$x^{(i)}=\begin{pmatrix} x_0^{(i)} \\ x_1^{(i)} \\ \vdots \\ x_n^{(i)} \end{pmatrix} \quad \theta=\begin{pmatrix} \theta_0 \\ \theta_1 \\ \vdots \\ \theta_n \end{pmatrix} \quad X=\begin{pmatrix} (x^{(1)})^T \\ (x^{(2)})^T \\ \vdots \\ (x^{(m)})^T \end{pmatrix} \quad Y=\begin{pmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(m)} \end{pmatrix}$$

$$X\theta-Y=\begin{pmatrix} (x^{(1)})^T\theta-y^{(1)} \\ (x^{(2)})^T\theta-y^{(2)} \\ \vdots \\ (x^{(m)})^T\theta-y^{(m)} \end{pmatrix}=\begin{pmatrix} h_\theta(x^{(1)})-y^{(1)} \\ h_\theta(x^{(2)})-y^{(2)} \\ \vdots \\ h_\theta(x^{(m)})-y^{(m)} \end{pmatrix}$$

接下来会涉及到矩阵求导，因为

$$J(\theta)=\frac{1}{2}\sum_{i=1}^{m}\left(h_\theta(x^{(i)})-y^{(i)}\right)^2=\frac{1}{2}tr\left[(X\theta-Y)^T(X\theta-Y)\right]$$

那么进一步利用矩阵求导并利用上述定理，得到

$$\begin{aligned}\frac{\partial J(\theta)}{\partial\theta}&=\frac{1}{2}\cdot\frac{\partial tr\left(\theta^TX^TX\theta-\theta^TX^TY-Y^TX\theta+Y^TY\right)}{\partial\theta}\\&=\frac{1}{2}\cdot\left[\frac{\partial tr(\theta I\theta^TX^TX)}{\partial\theta}-\frac{\partial tr(\theta^TX^TY)}{\partial\theta}-\frac{\partial tr(\theta Y^TX)}{\partial\theta}\right]\\&=\frac{1}{2}\cdot\left[X^TX\theta I+(X^TX)^T\theta I^T-X^TY-(Y^TX)^T\right]\\&=X^TX\theta-X^TY\end{aligned}$$

我们知道在极值点处梯度值为零，即

$$\frac{\partial J(\theta)}{\partial \theta} = X^T X\theta - X^T Y = 0$$

上述得到的方程组叫做**正规方程组**，那么最终得到

$$\theta = (X^T X)^{-1} X^T Y$$

这样最小二乘问题只需解一个线性方程组即可，不再需要像梯度下降那样迭代了。

https://blog.csdn.net/ACdreamers/article/details/44662633

3.拟合直线

平面直线的通用方程可以表示为

A+Bx-y=0

其中，A是直线的截距，B是直线的斜率。对于测量的二维坐标(x,y)，x是精确分布的，而y是观测值。基于最小二乘的理论，我们要得到观测值的误差的平方和的最小值。我们设定目标函数为：

$$F(A,B) = \sum_{i=1}^{N} (A + Bx_i - y_i)^2 \qquad (2)$$

分别对A和B求偏导数，并命其为零，得到以下方程组：

$$\begin{cases} A \times N + B \times \sum_{i=1}^{N} x_i = \sum_{i=1}^{N} y_i \\ A \times \sum_{i=1}^{N} x_i + B \times \sum_{i=1}^{N} x_i^2 = \sum_{i=1}^{N} x_i y_i \end{cases} \qquad (3)$$

解此方程组，则可以得到A和B。

$\hat{k} = \frac{\sum_{i=1}^{m} xy - \frac{1}{m}\sum_{i=1}^{m} x * \sum_{i=1}^{m} y}{\sum_{i=1}^{m} x^2 - \frac{1}{m}(\sum_{i=1}^{m} x)^2}$，$b = \hat{y} - \hat{k}x$。