# 最小二乘法及矩阵形式

July 26, 2014

## 1 问题

什么是最小二乘法？矩阵形式是什么样的？本文基于维基百科的Linear least squares (mathematics)词条写就。

## 2 解答

### 2.1 最小二乘法是求最优解的过程

最优解的标准是什么？——到所有样本的方差之和最小。

### 2.2 一个拟合直线的例子

#### 2.2.1 由样本列方程

有四个点：(1,6),(2,5),(3,7) 和 (4,10)。由四点来拟合直线。设：

$$y = \beta_1 + \beta_2 x \tag{1}$$

将四点代入，得到方程：

$$\beta_1 + 1 * \beta_2 = 6$$

$$\beta_1 + 2 * \beta_2 = 5$$

$$\beta_1 + 3 * \beta_2 = 7$$

$$\beta_1 + 4 * \beta_2 = 10$$

### 2.2.2 方差之和最小即是最优

右减左，平方，并求和：

$$
\begin{aligned}
S(\beta_1,\beta_2) &= [6-(\beta_1+1*\beta_2)]^2+[5-(\beta_1+2*\beta_2)]^2 \\
&= +[7-(\beta_1+3*\beta_2)]^2+[10-(\beta_1+4*\beta_2)]^2 \\
&= 4\beta_1^2+30\beta_2^2+20\beta_1\beta_2-56\beta_1-154\beta_2+210
\end{aligned}
$$

求最小值，还是要找老朋友，偏导：

$$\frac{\partial S}{\partial \beta_1}=0=8\beta_1+20\beta_2-56$$

$$\frac{\partial S}{\partial \beta_2}=0=20\beta+60\beta_2-154$$

解方程可得:

$$\beta_1=3.5$$

$$\beta_2=1.4$$

所以直线方程就是：

$$y=3.5+1.4x$$

## 2.3 矩阵形式

将上面的四个方程写成矩阵形式：

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{bmatrix}\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}=\begin{bmatrix} 6 \\ 5 \\ 7 \\ 10 \end{bmatrix}$$

简化为：

$$X\beta=y$$

已知的为：

- X，即样本数据
- $y$，样本数据

未知的是：

- $\beta$，系统方程的系数

### 2.3.1 更一般 X,y,$\beta$ 与求

如果所求的函数是直线，只有两个系数，$\beta_1$,$\beta_2$。如果是更复杂的函数，所需的参数就更多。写成更一般的形式，如下：

$$X = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{m1} & X_{m2} & \cdots & X_{mm} \end{bmatrix}, \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix}, y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

对于 $X\beta = y$，如何求 $\beta$？形式如下：

$$(X^T X)\hat{\beta} = X^T y \tag{2}$$

步骤：

- 左右都乘上一个 $X^T$，
- 再求 $(X^T X)^{-1}$，矩阵的逆，
- $\hat{\beta} = (X^T X)^{-1} * (X^T y)$。

编程时，求逆、转置与矩阵的乘法，最好都由库函数来完成。

### 2.3.2 证明

为什么最终求解是上述的形式？需要乘上转置，再求逆？

- 写出目标函数 $S(\beta) = \sum_{i=1}^{m} |y_i - \sum_{j=1}^{n} X_{ij}\beta_j| = ||y - X\beta||^2$
- 第 i 行为:$r_i = y_i - \sum_{j=1}^{n} X_{ij}\beta_j$（$r_i$ 的定义）
- 所以目标函数可以写为：$S = \sum_{i=1}^{m} r_i^2$

到这里，求最小值，又需要求助于偏导数。类比上面的非矩阵形式，最小值一定位于各偏导都为 0 的地方。

对 S 求偏导：

$$\frac{\partial S}{\partial \beta_j} = 2\sum_{i=1}^{m} r_i * \frac{\partial r_i}{\partial \beta_j}$$

再 $r_i$ 求偏导（参考上面 $r_i$ 的定义）：

$$\frac{\partial r_i}{\partial \beta_j} = -X_{ij}$$

代入代换：

$$\frac{\partial S}{\partial \beta_j} = 2\sum_{i=1}^{m} (y_i - \sum_{k=1}^{n} X_{ik}\beta_k) * (-X_{ij}) \qquad (j = 1, 2, \cdots, n)$$

乘入再移项，得到：

$$\sum_{i=1}^{m}\sum_{k=1}^{n} X_{ij}X_{ik}\hat{\beta}_k = \sum_{i=1}^{m} X_{ij}y_i \qquad (j = 1, 2, \cdots, n)$$

将上面的求和形式，写为矩阵相乘的形式，即得到：

$$(X^T X)\hat{\beta} = X^T y$$

以右侧为例：

$$\begin{aligned}
\sum_{i=1}^{m} X_{ij}y_i &= \begin{matrix} X_{1j} * y_1 + \\ X_{2j} * y_2 + \\ \vdots \\ X_{mj} * y_4 + \end{matrix} \\
&= \begin{bmatrix} X_{1j} & X_{2j} & \cdots & X_{mj} \end{bmatrix} * \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_4 \end{bmatrix} \\
&= X^T y
\end{aligned}$$

# 3 总结

- 最小二乘法中的“二乘”，就是“平方差求和”的意思。翻译的不好。“雅”，但不“信”。
- 最小二乘法的本质是“大家好才是真的好”，即，求一个目标函数，接近所有的样本。
- 一个应用例子，是求亚像素角点。