# 递归最小二乘法

Dezeming Family

2021 年 5 月 15 日

DezemingFamily 系列书和小册子因为是电子书，所以可以很方便地进行修改和重新发布。如果您获得了 DezemingFamily 的系列书，可以从我们的网站 [https://dezeming.top/] 找到最新版。对书的内容建议和出现的错误欢迎在网站留言。

20210721：完成第一版。虽然 5 月份就想写了，但因为科研任务较重，所以一直推迟到了 7 月中旬。

## 目录



## 一 介绍

关于最小二乘法已经在 DezemingFamily 的《最小二乘法》中进行了讲解。为了保证读者阅读方便，我们还是使用《最小二乘法》中定义的符号。

但是很多时候最小二乘法并不能很好应用，因为存在以下情况：

- 当估计值与真实值相差很大时，应该降低对估计结果的影响，而估计值与真实值相差很小时，应该提高对估计结果的影响。
- 当有新样本被添加进来以后，最好不要重新计算整个逆再重新求权重参数，而是可以进行迭代地计算和更新权重参数。

解决第一个问题的方法被称为加权最小二乘法，解决第二个问题的方法被称为递归最小二乘法。

## 二 加权最小二乘法简介

我们在《最小二乘法》中已知：

$$e = \overrightarrow{e}^T\overrightarrow{e} = \left(\overrightarrow{b} - A\overrightarrow{x}^*\right)^T\left(\overrightarrow{b} - A\overrightarrow{x}^*\right) \tag{二.1}$$

$$\overrightarrow{x}^* = (A^TA)^{-1}A^T\overrightarrow{b} \tag{二.2}$$

其中，$e$ 表示误差，$A$ 表示样本参数矩阵，$\overrightarrow{x}^*$ 表示待估计样本权重。

加权最小二乘估计的表示为 $e$ 加一个误差权重矩阵：

$$e=\overrightarrow{e}^T W\overrightarrow{e}=\left(\overrightarrow{b}-A\overrightarrow{x}^*\right)^T W\left(\overrightarrow{b}-A\overrightarrow{x}^*\right) \tag{二.3}$$

采用矩阵求导可得（见 DezemingFamily 的《矩阵导数》）：

$$\overrightarrow{x}^*=(A^TWA)^{-1}A^TW\overrightarrow{b} \tag{二.4}$$

分析一下误差权重矩阵，我们完全可以让误差权重矩阵 $W$ 为一个对角矩阵，对角线上的元素 $a_{i,i}$ 就相当于为第 $i$ 个样本加权。但其实在实际运算中我们可以求比较理想的 $W$ 误差权重矩阵。

下面讲解一种计算误差权重的方法：

- 先使用普通的最小二乘法，计算得到样本权重 $\overrightarrow{x}^*$。
- 使用重新计算估计样本值与实际样本值之间的误差。
- 用第 $i$ 个样本的误差的绝对值取倒数，作为误差权重 $W$ 的对角线第 $i$ 个元素值。

至于什么线性无偏最小方差估计（马尔科夫估计）这里就先不提了，我会放在《加权最小二乘法-马尔科夫估计》，是具有一定理论依据的误差权重生成方法。

注意加权最小二乘法还有一些应用，例如我们企图让近期的数据权重更大，远期的数据权重更小，这个时候也可以使用加权最小二乘法，对角线上的元素可以从上往下依次递减。

## 三　递归最小二乘法的基本思想

不论是递归最小二乘还是加权递归最小二乘，都需要获得全部测量值才能计算，非常不方便。有些时候，我们会获得成千上万的数据，对这些数据的矩阵运算量非常巨大。

递归最小二乘法则是可以将数据一个一个送入，然后不断根据新数据来更新模型；或者可以一批一批送入然后更新模型。这个思想可以参考求平均值的过程，如果我们已经有了前 $k$ 个数据的平均值 $\overrightarrow{a}_k$，又来了一个新的数据 $a_{k+1}$，应该怎么得到这 $k+1$ 个数据的平均值呢？其实很简单：

$$\overrightarrow{a}_{k+1}=\frac{\overrightarrow{a}_k\times k+a_{k+1}}{k+1} \tag{三.1}$$

$$=\frac{\overrightarrow{a}_k\times(k+1)+a_{k+1}-\overrightarrow{a}_k}{k+1} \tag{三.2}$$

$$=\overrightarrow{a}_k+\frac{a_{k+1}-\overrightarrow{a}_k}{k+1} \tag{三.3}$$

将新添加的数据根据公式三.3计算生成新的平均数即可。

## 四　递归最小二乘法的详细推导

有人觉得递归最小二乘法的推导比较复杂所以选择了跳过，其实它的推导过程挺简单的，只是我们需要记好这些矩阵和向量及其转置的形式。

假如我们进行了 $k$ 次估计，得到：

$$\overrightarrow{b}'_k=A_k\overrightarrow{x}^*_k \tag{四.1}$$

$$e_k=\left(\overrightarrow{b}_k-\overrightarrow{b}'_k\right)^T\left(\overrightarrow{b}_k-\overrightarrow{b}'_k\right) \tag{四.2}$$

其中，$\overrightarrow{b}_k$ 表示样本实际值构成的向量，$\overrightarrow{b}'_k$ 表示样本估计值构成的向量；$\overrightarrow{x}^*_k$ 表示我们 $k$ 次估计得到的权重向量。

我们的目标是，希望第 $k$ 次得到的权重估计为**上一次估计的权重**与**当前测量误差**的线性组合：

$$\overrightarrow{x}^*_k=\overrightarrow{x}^*_{k-1}+K\left(\overrightarrow{b}_k-A_k\overrightarrow{x}^*_{k-1}\right) \tag{四.3}$$

现在开始正式推导。根据最小二乘法，我们知道 $\vec{x}^*_k$ 的计算公式是：

$$\vec{x}^*_k = (A_k^T A_k)^{-1} A_k^T \vec{b}_k \tag{四.4}$$

这里的 $\vec{b}_k$ 表示 $k$ 个观测值构成的向量。

我们将上式拆成两部分：$(A_k^T A_k)^{-1}$ 和 $A_k^T \vec{b}_k$。

**递推一**

设 $P_k^{-1} = A_k^T A_k$，我们再重新定义一下数据格式，设样本一共 t 个特征，第 i 个样本表示为：

$$b_i = \begin{bmatrix} a_{i,1} & a_{i,2} & \dots & a_{i,t} \end{bmatrix} \begin{bmatrix} x_1^* \\ x_2^* \\ \dots \\ x_t^* \end{bmatrix} = \vec{a}_i^T \vec{x}^*_k \tag{四.5}$$

设样本特征构成的矩阵 $A_k$ 为：

$$A_k = \begin{bmatrix} \vec{a}_1^T \\ \vec{a}_2^T \\ \dots \\ \vec{a}_k^T \end{bmatrix} \tag{四.6}$$

注意这里 $\vec{a}_i$ 表示的是第 i 个样本的特征构成的向量，而《最小二乘法》中 $\vec{a}_i$ 定义的是所有样本的第 i 个特征构成的向量，虽然表示不同，但构成的矩阵 $A$ 是一样的。

因此：

$$P_k^{-1} = A_k^T A_k \tag{四.7}$$

$$= \begin{bmatrix} \vec{a}_1 & \vec{a}_2 & \dots & \vec{a}_k \end{bmatrix} \begin{bmatrix} \vec{a}_1^T \\ \vec{a}_2^T \\ \dots \\ \vec{a}_k^T \end{bmatrix} \tag{四.8}$$

$$= \sum_{i=1}^{k} \vec{a}_i \vec{a}_i^T = \sum_{i=1}^{k-1} \vec{a}_i \vec{a}_i^T + \vec{a}_k \vec{a}_k^T \tag{四.9}$$

$$= P_{k-1}^{-1} + \vec{a}_k \vec{a}_k^T \tag{四.10}$$

**递推二**

我们再看一下 $A_k^T \vec{b}_k$：

$$A_k^T \vec{b}_k = \begin{bmatrix} \vec{a}_1 & \vec{a}_2 & \dots & \vec{a}_k \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \dots \\ b_k \end{bmatrix} \tag{四.11}$$

$$= \sum_{i=1}^{k} \vec{a}_i b_i = \sum_{i=1}^{k-1} \vec{a}_i b_i + \vec{a}_k b_k \tag{四.12}$$

$$= A_{k-1}^T \vec{b}_{k-1} + \vec{a}_k b_k \tag{四.13}$$

有了上述两个递推式，我们就可以继续向下推导了。

我们把当前的已知公式列出：

$$\vec{x}^*_k = (A_k^T A_k)^{-1} A_k^T \vec{b}_k \tag{四.14}$$

$$A_k^T A_k = A_{k-1}^T A_{k-1} + \vec{a}_k \vec{a}_k^T \tag{四.15}$$

$$A_k^T \vec{b}_k = A_{k-1}^T \vec{b}_{k-1} + \vec{a}_k b_k \tag{四.16}$$

因此：

$$\vec{x}_k^* = (A_k^T A_k)^{-1} A_k^T \vec{b}_k = P_k A_k^T \vec{b}_k = P_k (A_{k-1}^T \vec{b}_{k-1} + \vec{a}_k b_k) \tag{四.17}$$

又因为：

$$\vec{x}_{k-1}^* = P_{k-1} A_{k-1}^T \vec{b}_{k-1} \tag{四.18}$$

$$P_{k-1}^{-1} \vec{x}_{k-1}^* = A_{k-1}^T \vec{b}_{k-1} \tag{四.19}$$

因此可以得到：

$$\vec{x}_k^* = P_k (A_{k-1}^T \vec{b}_{k-1} + \vec{a}_k y_k) = P_k (P_{k-1}^{-1} \vec{x}_{k-1}^* + \vec{a}_k y_k) \tag{四.20}$$

$$= P_k \Big( (P_k^{-1} - \vec{a}_k \vec{a}_k^T) \vec{x}_{k-1}^* + \vec{a}_k b_k \Big) \tag{四.21}$$

$$= \vec{x}_{k-1}^* + P_k \vec{a}_k (b_k - \vec{a}_k^T \vec{x}_{k-1}^*) \tag{四.22}$$

因此我们就得到了根据 $\vec{x}_{k-1}^*$ 和新的样本来更新得到 $\vec{x}_k^*$ 的形式。

# 五 递归最小二乘法的使用步骤

我们定义一些符号来描述这个过程：

$$\varepsilon_k = b_k - \vec{a}_k^T \vec{x}_{k-1}^* \tag{五.1}$$

$$K_k = P_k \vec{a}_k \tag{五.2}$$

$$P_k = (P_{k-1}^{-1} + \vec{a}_k \vec{a}_k^T)^{-1} \tag{五.3}$$

$$\vec{x}_k^* = \vec{x}_{k-1}^* + K_k \varepsilon_k \tag{五.4}$$

尽管 $P_{k-1}^{-1} + \vec{a}_k \vec{a}_k^T$ 只是 $k \times k$ 的矩阵，但是求逆也不是那么方便，而我们有新的方法，首先大家需要知道一个定理，当矩阵 $A$、$C$ 和 $BCD$ 都是可逆矩阵时，则：

$$[A + BCD]^{-1} = A^{-1} - A^{-1} B \big[C^{-1} + D A^{-1} B\big]^{-1} D A^{-1} \tag{五.5}$$

我们令 $A = P_{k-1}^{-1}$，$C = E$，$B = \vec{a}_k$，$D = \vec{a}_k^T$，代入上式之后，推导出得到 $P_k^{-1}$ 的公式：

$$P_k = P_{k-1} - \frac{P_{k-1} \vec{a}_k \vec{a}_k^T P_{k-1}}{1 + \vec{a}_k^T P_{k-1} \vec{a}_k} \tag{五.6}$$

这样我们就不需要求逆了！

因此，最终递归最小二乘法的使用步骤是：

$$\varepsilon_k = b_k - \vec{a}_k^T \vec{x}_{k-1}^* \tag{五.7}$$

$$K_k = P_k \vec{a}_k \tag{五.8}$$

$$P_k = P_{k-1} - \frac{P_{k-1} \vec{a}_k \vec{a}_k^T P_{k-1}}{1 + \vec{a}_k^T P_{k-1} \vec{a}_k} \tag{五.9}$$

$$\vec{x}_k^* = \vec{x}_{k-1}^* + K_k \varepsilon_k \tag{五.10}$$

# 参考文献


[1] https://blog.csdn.net/weixin_38898944/article/details/

[2] https://zhuanlan.zhihu.com/p/59532437

[3] https://zhuanlan.zhihu.com/p/53883828

[4] https://blog.csdn.net/weixin_45072297/article/details/

[5] http://www.personal.reading.ac.uk/ sis01xh/