最小二乘法

Dezeming Family

2021 年 5 月 12 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。

本书的售价是 3 元（电子版），我们不直接收取任何费用，如果本书对大家学习有帮助，可以往我们的支付宝账户（17853140351）进行支持，您的赞助将是我们 Dezeming Family 继续创作各种计算机视觉、图形学、机器学习、以及数学原理小册子的动力！

目录

一　要解决的问题

有些时候我们需要利用特征来估计目标量 l，比如根据天气和温度估计水库的水蒸发量；比如根据一个人的年龄和购物水平估计他的工资等；这里的目标量都是我们未知的，但我们已经有一些观测值了。

假设一个待估计目标有 k 个特征，第 i 个特征的值为 a_i，第 i 个特征的权重设为 x_i，我们就可以估计一个线性模型：

$$\sum_{i=1}^{k} a_i x_i = l \tag{一.1}$$

比如我们估计食品浪费因子 a_1 与服务员人数 a_2 和餐馆年盈利 l（单位万元）之间的关系，我们就可以估计一个线性模型：

$$l = 20 - 2 * a_1 + 0.3 * a_2,\ 3 < a_2 < 5 \tag{一.2}$$

服务员需要小于五个人，当食物浪费因子为 2，服务员人数为 4 时，餐馆盈利为 17.2 万元。

当然，有时候我们有 k 个实际特征，我们还会加一个偏移特征，称为截距。只不过我们令所有样本的该特征值为固定值，一般为 1。后面为了方便起见，我们不再单独叙述偏移特征。

我们现在已经有很多样本了，也有实际观测值，我们如何估计线性模型的每个特征的权重呢？为了方便叙述，我们定义一些符号。

我们设所有样本的第 i 个特征的值构成列向量 $\overrightarrow{a}_i$，样本 i 的观测值为 b_i 构成列向量 $\overrightarrow{b}$，每个特征的权重定义为 x，一共有 k 个特征构成特征向量 $\overrightarrow{x}$。我们可以得到矩阵：

$$A = \begin{bmatrix} \overrightarrow{a}_1 & \overrightarrow{a}_2 & \cdots & \overrightarrow{a}_k \end{bmatrix} \tag{一.3}$$

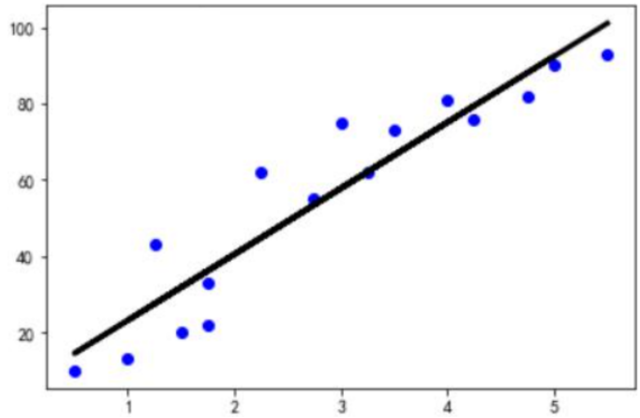
$$\begin{bmatrix} \overrightarrow{a}_1 & \overrightarrow{a}_2 & \cdots & \overrightarrow{a}_k \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_k \end{bmatrix} = \overrightarrow{b} \tag{一.4}$$

注意这里的 $\overrightarrow{a}_i$ 都是列向量，即相当于：

$$x_1\overrightarrow{a}_1 + x_2\overrightarrow{a}_2 + \cdots + x_k\overrightarrow{a}_k = \overrightarrow{b} \tag{一.5}$$

注意矩阵 A 的第 i 行向量代表的是第 i 个样本的所有特征的值构成的向量，第 i 列代表的是所有样本的第 i 个特征的值构成的向量。

通常情况下，这个 x 是解不出来的，大家可以考虑二维平面上无法用一条线来完全拟合样本：



因此我们需要找一个办法来得到最优的线性近似，最小二乘法就是直接得到最优线性近似的方法。(机器学习中的线性回归等也是可以逐步优化的，只是需要样本的不断训练逼近最优拟合，但可以用多次拟合线，而最小二乘法只用于线性拟合。)

二　问题分析

什么是最优拟合呢？

我们定义误差为 e，然后我们或许找不到一个 $\overrightarrow{x}$ 满足 $A\overrightarrow{x} = \overrightarrow{b}$，但是我们可以假定一个 $\overrightarrow{x}^*$，然后定义向量：$\overrightarrow{e} = \overrightarrow{b} - A\overrightarrow{x}^*$，其中 $\overrightarrow{e}$ 向量长度为样本数，误差 e 的值就是 $\overrightarrow{e}$ 中每个值求平方再相加。

所谓最优拟合，就是找一个 $\overrightarrow{x}^*$，使得 e 的值最小。

所以我们现在目标就是最小化 e 了。

三　误差的矩阵形式和求解

我们可以把误差 e 写为：

$$e = \overrightarrow{e}^T\overrightarrow{e} = \left(\overrightarrow{b} - A\overrightarrow{x}^*\right)^T\left(\overrightarrow{b} - A\overrightarrow{x}^*\right) \tag{三.1}$$

$$= \left(A\overrightarrow{x}^* - \overrightarrow{b}\right)^T\left(A\overrightarrow{x}^* - \overrightarrow{b}\right) \tag{三.2}$$

$$= \left(\overrightarrow{x}^{*T}A^TA\overrightarrow{x}^* - \overrightarrow{b}^TA\overrightarrow{x}^* - (A\overrightarrow{x}^*)^T\overrightarrow{b} + \overrightarrow{b}^T\overrightarrow{b}\right) \tag{三.3}$$

上式还可以继续化简，我们先展开里面的一些项：

$$(A\overrightarrow{x}^*)^T\overrightarrow{b} = \overrightarrow{x}^{*T}A^T\overrightarrow{b} \tag{三.4}$$

同时我们注意到上式的结果其实是一个值，因此可以得到下式：

$$\vec{b}^T A \vec{x}^* = \left(\vec{b}^T A \vec{x}^*\right)^T = \left(\vec{x}^{*T} A^T\right) \vec{b} \tag{三.5}$$

因此 e 就可以简化为：

$$e = \left(\vec{x}^{*T} A^T A \vec{x}^* - 2\vec{x}^{*T} A^T \vec{b} + \vec{b}^T \vec{b}\right) \tag{三.6}$$

我们需要 e 进行求导，让导数为 0 时则说明达到了极值点：

$$\frac{\partial}{\partial \vec{x}^*} e = 2A^T\left(A\vec{x}^* - \vec{b}\right) = 0 \tag{三.7}$$

求导的含义我们可以理解为，当 $\vec{x}^*$ 中有任意元素发生变化时，e 也会发生变化。关于矩阵求导的内容可以参考 DezemingFamily 的《函数对向量求导-通俗易懂的描述》一书。上式结果可以化简得到：

$$\vec{x}^* = (A^T A)^{-1} A^T \vec{b} \tag{三.8}$$

四 几何含义

我们回到第一章的描述：

$$\begin{bmatrix} \vec{a}_1 & \vec{a}_2 & \cdots & \vec{a}_k \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdots \\ x_k \end{bmatrix} = \vec{b} \tag{四.1}$$

$$x_1\vec{a}_1 + x_2\vec{a}_2 + \cdots + x_k\vec{a}_k = \vec{b} \tag{四.2}$$

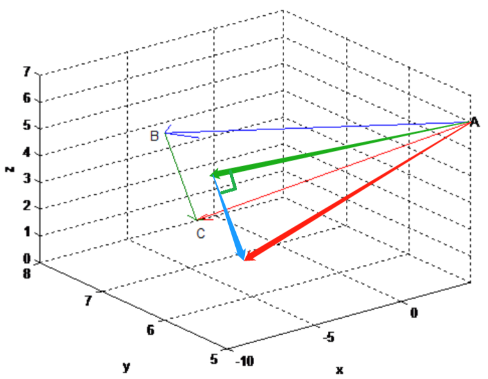
其实向量 $\vec{b}$ 就是 $\vec{a}_i$ 的线性组合。

我们假设有两个特征，表示为 $y = a_{i,1}x_1 + x_2$，i 表示第 i 个样本。所有样本构成矩阵：

$$\begin{bmatrix} a_{1,1} & 1 \\ a_{2,1} & 1 \\ a_{3,1} & 1 \\ a_{4,1} & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} \tag{四.3}$$

我们可以隐约感受到，样本越多，特征量越少时，越难线性组合成样本 $\vec{b}$，具体原理大家可以参考 DezemingFamily 的《矩阵与方程组的解》。

我们假设只有三个样本，即列向量维度为 3，就可以画一个三维示意图：



我们可以理解为想用向量 $\overrightarrow{AB}$ 和 $\overrightarrow{AC}$ 拟合出红色的向量 $\vec{b}$，但我们知道，$\overrightarrow{AB}$ 和 $\overrightarrow{AC}$ 只能拟合出在 ABC 三点构成的平面上的向量，因此无法拟合出 $\vec{b}$。为了让拟合结果更接近 $\vec{b}$，我们希望拟合出 $\vec{b}$ 在 ABC 平面的投影，即绿色向量。我们设蓝色向量为 $\overrightarrow{Blue}$，绿色向量为 $\overrightarrow{Green}$。

根据前面的描述，$\overrightarrow{AB}$ 和 $\overrightarrow{AC}$ 分别可以代表 $\vec{a}_1$ 和 $\vec{a}_2$。由投影我们可以得到向量的垂直关系：

$$\vec{a}_1^T \overrightarrow{Blue} = 0 \tag{四.4}$$

$$\vec{a}_2^T \overrightarrow{Blue} = 0 \tag{四.5}$$

即：

$$A^T\overrightarrow{Blue} = 0 \tag{四.6}$$

$$A^T(\overrightarrow{b} - \overrightarrow{Green}) = 0 \tag{四.7}$$

$$\overrightarrow{Green} = A\overrightarrow{x}^* \tag{四.8}$$

$$\tag{四.9}$$

分析一下上式就可以得到：

$$A^T(\overrightarrow{b} - A\overrightarrow{x}^*) = 0 \tag{四.10}$$

$$\overrightarrow{x}^* = (A^TA)^{-1}A^T\overrightarrow{b} \tag{四.11}$$

参考文献

[1] 本书由于比较简单，没有参考任何文献资料。