线性方程组最小二乘法的推导

Yijun WU

2022 年 5 月 25 日

推导过程

对于有 n 个未知数的 m 行线性方程组:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \cdots\cdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{cases}$$

设该方程组的最优解为:

$$\begin{cases} x_1 = c_1 \\ x_2 = c_2 \\ \cdots\cdots \\ x_m = c_m \end{cases}$$

则可以定义该线性方程组的方差:

$$\begin{aligned} \sigma^2 &= (a_{11}c_1 + a_{12}c_2 + \cdots + a_{1n}c_n - b_1)^2 \\ &\quad + (a_{21}c_1 + a_{22}c_2 + \cdots + a_{2n}c_n - b_2)^2 \\ &\quad + \cdots + (a_{m1}c_1 + a_{m2}c_2 + \cdots + a_{mn}c_n - b_m)^2 \end{aligned}$$

用矩阵的语言表示: 记 $A \in M_{m\times n}$ 为 m 行 n 列的矩阵, $\vec{x}$ 为 n 行的列向量, $\vec{b}$ 为 m 行的列向量。则原线性方程组可以表示为:

$$A\vec{x} = \vec{b}$$

对于任意一个解

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{pmatrix}$$

方差可以表示为:

$$\begin{aligned} \sigma^2 &= (A\vec{x} - \vec{b})^T (A\vec{x} - \vec{b}) \\ &= (\vec{x}^T A^T - \vec{b}^T)(A\vec{x} - \vec{b}) \\ &= \vec{x}^T A^T A\vec{x} - \vec{x}^T A^T \vec{b} - \vec{b}^T A\vec{x} + \vec{b}\vec{b}^T \end{aligned}$$

又由于 $\vec{x}^T A^T \vec{b}$ 是 1×1 矩阵，因此 $\vec{x}^T A^T \vec{b} = (\vec{x}^T A^T \vec{b})^T = \vec{b}^T A\vec{x}$，于是

$$\sigma^2 = \vec{x}^T (A^T A)\vec{x} - (2\vec{b}^T A)\vec{x} + \|\vec{b}\|^2$$

不妨令 $S := A^T A \in M_{n\times n}$ 为 n 阶方阵，$\vec{t} := 2\vec{b}^T A$ 为 n 列的行向量，$c := \|\vec{b}\|^2$ 为常数。则 $f(x_1, x_2, \cdots, x_n) := \sigma^2 = \vec{x}^T S\vec{t} - \vec{t}\vec{x} + c$，当 f 取得最值时，有 $f'_{x_1} = f'_{x_2} = \cdots = f'_{x_n} = 0$。

不妨记 $\vec{e_i}$ 为单位阵 I_n 的第 i 列这个列向量，记 $\vec{S}^{(i)}$ 为 S 的第 i 列的这个列向量，$\vec{S}_{(i)}$ 为 S 的第 i 行的这个行向量，记 t_i 为 $\vec{t}$ 的第 i 个元素。则

$$\begin{aligned}
\forall i \in \{1,2,3,\cdots,n\}, f'_{x_i} &= \frac{\partial \vec{x}^T}{\partial x_i}(S\vec{x}) + (\vec{x}^T S)\frac{\partial \vec{x}}{\partial x_i} - (\vec{t})\frac{\partial \vec{x}}{\partial x_i} \\
&= \vec{e_i}^T (S\vec{x}) + (\vec{x}^T S)\vec{e_i} - \vec{t}\vec{e_i} \\
&= (\vec{e_i}^T S)\vec{x} + \vec{x}^T (S\vec{e_i}) - t_i \\
&= \vec{S}_{(i)}\vec{x} + \vec{x}^T \vec{S}^{(i)} - t_i \\
&= \vec{S}_{(i)}\vec{x} + (\vec{S}^{(i)})^T \vec{x} - t_i \\
&= 0
\end{aligned}$$

因此有：$(\vec{S}_{(i)} + (\vec{S}^{(i)})^T)\vec{x} = t_i, \forall i \in \{1,2,3,\cdots,n\}$ 。也即：

$$\begin{pmatrix} \vec{S}_{(1)} \\ \vec{S}_{(2)} \\ \cdots \\ \vec{S}_{(n)} \end{pmatrix}\vec{x} + \begin{pmatrix} (\vec{S}^{(1)})^T \\ (\vec{S}^{(2)})^T \\ \cdots \\ (\vec{S}^{(n)})^T \end{pmatrix}\vec{x} = \begin{pmatrix} t_1 \\ t_2 \\ \cdots \\ t_n \end{pmatrix}$$

由 $\vec{S}_{(i)}$ 和 $\vec{S}^{(i)}$ 和 t_i 的定义知：

$$\begin{aligned}
& & S\vec{x} + S^T\vec{x} =& \vec{t}^T \\
\Rightarrow & & (S + S^T)\vec{x} =& \vec{t}^T \\
\Rightarrow & & (AA^T + (AA^T)^T)\vec{x} =& (2\vec{b}^T A)^T \\
\Rightarrow & & 2AA^T\vec{x} =& 2A^T\vec{b} \\
\Rightarrow & & \vec{x} =& (AA^T)^{-1}A^T\vec{b}
\end{aligned}$$

当 $\vec{x} = (AA^T)^{-1}A^T\vec{b}$ 时，线性方程组取最优解。

其它说明 有一点需要注意的事情：
当 $m <= n$ 时，线性方程的数量不大于未知数的数量，因此存在解使得等

号同时成立，也就是 $\sigma^2 \equiv 0$。此时该结论没有意义。|
当 $m > n$，且 AA^T 可逆，也就是 $rank(AA^T) = n$ 时，该结论适用最优解。
当原线性方程组有线性相关的方程时，我也不知道适不适用 (笑)