第十八讲　全面最小二乘法

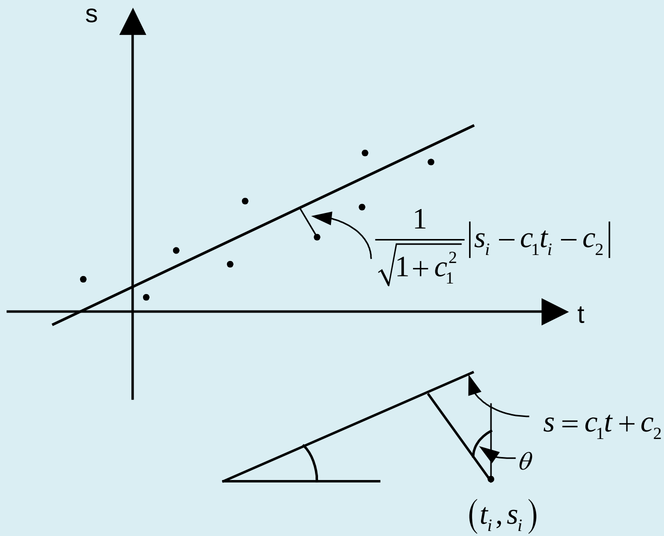
一、 法向回归

一组测量数据$\left(t_i, s_i\right)$，欲拟和直线

$$s = c_1 t + c_2$$

最小二乘法采取目标函数：$E\left(c_1, c_2\right) = \sum_{i=1}^{n} \left| s_i - c_1 t_i - c_2 \right|^2 = \min$

它隐含了在测量中，t_i是精确测量的，只有s_i才测得不准确，而在实际测量中，t_i，s_i都是无法准确测量的，因此，采用法向回归更有可能。



点(t_i, s_i)到直线$s = c_1 t + c_2$的距离为

$$\frac{1}{\sqrt{1+c_1^2}}|s_i - c_1 t_i - c_2|$$

故法向回归的目标函数为

$$E(c_1,c_2)=\left(\frac{1}{\sqrt{1+c_1^2}}\right)^2\sum_{i=1}^{n}\left|s_i-c_1t_i-c_2\right|^2=\min$$

$$\frac{\partial E}{\partial c_2}=\frac{1}{1+c_1^2}\sum_{i=1}^{n}(-2)(s_i-c_1t_i-c_2)=0\rightarrow c_2=\frac{1}{n}\sum_{i=1}^{n}s_i-c_1t_i$$

$$\frac{\partial E}{\partial c_1}=-\frac{2c_1}{\left(1+c_1^2\right)^2}\sum_{i=1}^{n}(s_i-c_1t_i-c_2)^2+\frac{2}{1+c_1^2}\sum_{i=1}^{n}(-t_i)(s_i-c_1t_i-c_2)$$

$$=\frac{2}{1+c_1^2}\sum_{i=1}^{n}(c_1c_2-c_1s_i-t_i)(s_i-c_1t_i-c_2)$$

$$= \frac{2}{1+c_1^2}\left\{c_1 c\sum_{i=1}^{n}\left(s_i - c_1 t_i - c_2\right) - c_1\sum_{i=1}^{n} s_i\left(s_i - c_1 t_i - c_2\right) - \sum_{i=1}^{n} t_i\left(s_i - c_1 t_i - c_2\right)\right\}$$

$$= \frac{-2}{1+c_1^2}\left\{c_1\sum_{i=1}^{n} s_i\left(s_i - c_1 t_i - c_2\right) + \sum_{i=1}^{n} t_i\left(s_i - c_1 t_i - c_2\right)\right\}$$

$$= 0$$

将c_2代入之，可得

$$\begin{cases} c_1 = \dfrac{\left(l_{ss} - l_{tt}\right) + \sqrt{\left(l_{ss} - l_{tt}\right)^2 + 4l_{st}^2}}{2l_{st}} \\ c_2 = \bar{s} - c_1 \bar{t} \end{cases}$$

其中

$$
\begin{cases}
\bar{s} = \dfrac{1}{n}\sum\limits_{i=1}^{n} s_i \\
\bar{t} = \dfrac{1}{n}\sum\limits_{i=1}^{n} t_i \\
l_{ss} = \sum\limits_{i=1}^{n}\left(s_i - \bar{s}\right)^2 = \sum\limits_{i=1}^{n} s_i^2 - \dfrac{1}{n}\left(\sum\limits_{i=1}^{n} s_i\right)^2, \\
l_{st} = \sum\limits_{i=1}^{n}\left(s_i - \bar{s}\right)\left(t_i - \bar{t}\right) = \sum\limits_{i=1}^{n} s_i t_i - \dfrac{1}{n}\left(\sum\limits_{i=1}^{n} s_i\right)\left(\sum\limits_{i=1}^{n} t_i\right) \\
l_{tt} = \sum\limits_{i=1}^{n}\left(t_i - \bar{t}\right)^2 = \sum\limits_{i=1}^{n} t_i^2 - \dfrac{1}{n}\left(\sum\limits_{i=1}^{n} t_i\right)^2
\end{cases}
$$

另一种推导方法：

$$E\left(c_1,c_2\right)=\frac{1}{1+c_1^2}\sum_{i=1}^{n}\left(s_i-c_1t_i-c_2\right)^2$$

$$\frac{\partial E}{\partial c_2}=0\rightarrow c_2=\frac{1}{n}\sum_{i=1}^{n}\left(s_i-c_1t_i\right)=\bar{s}-c_1\bar{t}\Rightarrow E\left(c_1,c_2\right)=\frac{\sum_{i=1}^{n}\left[\left(s_i-\bar{s}\right)-c_1\left(t_i-\bar{t}\right)\right]^2}{1+c_1^2}$$

$$E\left(c_1,c_2\right)=\frac{l_{ss}-2c_1l_{st}+l_{tt}c_1^2}{1+c_1^2}$$

$$\frac{\partial E}{\partial c_1}=0\rightarrow c_1=\frac{\left(l_{ss}-l_{tt}\right)\pm\sqrt{\left(l_{ss}-l_{tt}\right)^2+4l_{st}^2}}{2l_{st}}$$

“±”中，“－”对应的 E 的最大值

作为比较，最小二乘法$\sum_{i=1}^{n}\left|s_i - c_1 t_i - c_2\right|^2 = \min$给出

$$\begin{cases} c_1 = {l_{st}}\big/{l_{tt}} \\ c_2 = \bar{s} - c_1 \bar{t} \end{cases}$$

例1.　7 点测量

$$(t_i, s_i) = (0,3.1),(0.5,3.9),(1,5.2),(1.5,6.0),(2,6.9),(2.5,8.0),(3.0,9.1)$$

拟合直线$c_1 t + c_2 = s$

解：计算结果$\bar{t} = 1.5, \bar{s} \approx 6.02857, l_{tt} = 7, l_{ss} = 27.8743, l_{st} = 13.95$

最小二乘法给出 $c_1 = 1.99286, c_2 = 3.03929$

全面最小二乘法（法向回归）给出 $c_1 = 1.99709, c_2 = 3.03293$

测量数据误差小，分布合理时，两种方法效果非常接近。

二 、全面最小二乘法（Totally Least Square Method）

当方程 $Ax = b$ 成为矛盾方程时，采用最小二乘法求解的观点实际上认为 b 存在误差，而 A 不存在误差，故应有 ε，使得

$$Ax = b + \varepsilon$$

ε 应尽量小以使得不至于严重得破坏方程 $\rightarrow \|\varepsilon\|_2 = \min$

全面最小二乘法采取如下观点解决矛盾方程的问题，不仅 b 存在误差，A 也存在误差，故，存在 E 和 ε，使

$$(A+E)x=b+\varepsilon$$

E、ε 也应该尽量小，以使得不至于严重偏离原方程

$$\rightarrow \left\|[E\,|\,\varepsilon]\right\|_F = \min$$

$$(A+E)x=b+\varepsilon \Leftrightarrow ([A|b]+[E\,|\,\varepsilon])\begin{bmatrix} x \\ -1 \end{bmatrix}=0$$

记 $C=[A|b], \Delta=[E\,|\,\varepsilon], v=\begin{bmatrix} x \\ -1 \end{bmatrix}$，则全面最小二乘解即求如下方程

$$(C+\Delta)v=0$$

的非零解 v，且 v 的最后分量不能为零，而其中 Δ 应满足 $\|\Delta\|_F=\min$

引理：设 $X\in C_r^{m\times n}$，且存在奇异值分解，

$$X = U\begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & 0 \end{bmatrix}_{m\times n} V^H$$，其中$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$。又设

$$Y = U\begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_s & \\ & & & 0 \end{bmatrix}_{m\times n} V^n \qquad (s<r)$$则

$$\|X-Y\|_F = \min_{\substack{z\in C^{m\times n}\\ rankz=s}} \|X-Z\|_F$$

首先来考虑 F-范数。设$P_{m\times n} = UQV^H$,U、V 分别为 m 阶、n 阶酉

矩阵。Q 为$m\times n$阶矩阵（上式不一定是奇异值分解）。则

$$\|P\|_F^2=\sum_{i,j}|p_{ij}|^2=\sum_{i=1}^{m}\left(\sum_{j=1}^{n}p_{ij}\overline{p_{ij}}\right)=\sum_{i=1}^{m}\left(PP^H\right)_{ii}=tr\left(PP^H\right)=tr\left(P^HP\right)$$

$$=tr\left(UQV^HVQ^HU^H\right)=tr\left(UQQ^HU^H\right)=tr\left(Q^HU^HUQ\right)$$

$$=tr(Q^HQ)=tr\left(QQ^H\right)=\|Q\|_F^2$$

（按照教材上的说法，正交相抵或酉相抵的矩阵与 F 范数相同）

$\|X-Y\|_F^2=\sum_{i=s+1}^{r}\sigma_i^2$，又令$Z=UTV^H\leftarrow T=U^HZV$，则

$$\|X-Z\|_F^2=\sum_{i=1}^{r}|t_{ii}-\sigma_i|^2+\sum_{\substack{i=1\\ i\neq j}}^{r}\sum_{j=1}^{n}|t_{ij}|^2+\sum_{i=r+1}^{m}\sum_{j=1}^{n}|t_{ij}|^2$$

对任意 Z 矩阵而言，各t_{ij}之间完全独立，则$\|X-Z\|_F$是可能等于零的。但是$rank(Z)=s<r$。故$\|X-Z\|_F$不可能为零。详细论证可知$t_{ij}=0(i\neq j),t_{ii}=0(i>s),t_{ii}=\sigma_i(i=1,2,\cdots,s)$时，$\|X-Z\|_F$最小

下面仅考虑在实际应用中非常常见的一种情况：$A\in C_n^{m\times n}$，$[A|b]\in C_{n+1}^{m\times n}$，即 A 是列满秩的，$[A|b]$也是列满秩的。这样，系数矩阵与增广矩阵的秩不相等，方程$Ax=b$不相容。

定理 1： 设$A\in C_n^{m\times n}$,$\left[A|b\right]\in C_{n+1}^{m\times(n+1)}$具有如下的奇异值分解

$$C=U\begin{bmatrix}\sigma_1 & & \\ & \ddots & \\ & & \sigma_{n+1} \\ & 0 & \end{bmatrix}V^H,(\sigma_1\geq\sigma_2\geq\cdots\geq\sigma_{n+1})$$

则使方程$(C+\Delta)v=0$具有非零解，且 F 范数最小的Δ存在，并且

$$\|\Delta\|_F=\sigma_{n+1}$$

证明：方程$(C+\Delta)v=0$要有非零解，必须$rank(C+\Delta)<n+1$，故由引理知

$$\min \|\Delta\|_F = \min_{rank(C+\Delta)<n+1} \left\| C - (C+\Delta) \right\|_F$$

$$= \min_{rank(C+\Delta)=n} \left\| C - (C+\Delta) \right\| = \sigma_{n+1}$$

显然满足

$$\Delta = U \begin{bmatrix} 0 & & & \\ & \ddots & & \\ & & 0 & \\ & & & \sigma_{n+1} \\ & & O & \end{bmatrix} V^H$$

定理 2： 设σ_{n+1}为 C 的 n-k+1 重奇异值，且$v_{k+1}, v_{k+2}, \cdots v_{n+1}$相应的为$C^H C$的属于$(n-k+1)$重特征值$\sigma_{n+1}^2$的正交归一特征向量，则使方程$(C+\Delta)v=0$具有非零的解且 F 范数最小的$\Delta$为

$$\Delta = -C v_s v_s^H \big/ v_s^H v_s$$

而方程的解则为$v=v_s$，其中$v_s \in S_c \stackrel{\Delta}{=} span\{v_{k+1}, v_{k+2}, \cdots v_{n+1}\}$

证：（1） 显然$CC^H v_s = \sigma_{n+1}^2 v_s$

$$\|\Delta\|_F^2 = \|C v_s v_s^H\|_F^2 \Big/ \left(v_s^H v_s\right)^2 = tr\left(v_s v_s^H C^H C v_s v_s^H\right) \Big/ \left(v_s^H v_s\right)^2$$

$$= \frac{\sigma_{n+1}^2}{\left(v_s^H v_s\right)^2} tr\left(v_s v_s^H v_s v_s^H\right) = \frac{\sigma_{n+1}^2}{v_s^H v_s} tr\left(v_s v_s^H\right) = \frac{\sigma_{n+1}^2}{v_s^H v_s} tr\left(v_s^H v_s\right)$$

$$= \sigma_{n+1}^2$$

（2） $(C+\Delta)v_s = Cv_s - \dfrac{Cv_s v_s^H v_s}{v_s^H v_s} = 0$

（3） $\forall v \in C^{n+1}$,有

$$v = \left(\sum_{i=1}^{n-k+1} v_{k+i} v_{k+i}^H\right) v + \left(I_{n+1} - \sum_{i=1}^{n-k+1} v_{k+i} v_{k+i}^H\right) v \overset{\Delta}{=} v_s + v_T$$

虽然，$\left(C - C\dfrac{vv^H}{v^H v}\right) v = 0$,但 $\left\| -C\dfrac{vv^H}{v^H v} \right\|_F > \sigma_{n+1}$

$$\because \; C^H C v = C^H C\left(v_s + v_T\right) = \sigma_{n+1}^2 v_s + C^H C v_T > \sigma_{n+1}^2 v_s + \sigma_{n+1}^2 v_T$$

$$= \sigma_{n+1}^2\left(v_s + v_T\right) = \sigma_{n+1}^2 v$$

$\therefore$

$$\left\| C\frac{vv^H}{v^H v} \right\|_F^2 = \frac{1}{\left(v^H v\right)^2} tr\left(vv^H C^H C vv^H\right) > \frac{\sigma_{n+1}^2}{\left(v^H v\right)^2} tr\left(vv^H vv^H\right) = \sigma_{n+1}^2$$

定理 3 ：在定理 2 的条件下，全面最小二乘解存在的充要条件为：向量 $e_{n+1} = \begin{bmatrix} \underbrace{0 \quad \cdots \quad 0}_{n个} & 1 \end{bmatrix}^T$ 不正交于 S_c 。此时，

$\forall v \in \left\{ q = \begin{bmatrix} y \\ \alpha \end{bmatrix} \middle| q \in S_c, \alpha \neq 0 \right\}$，则最小二乘解为

$$x = \frac{1}{\alpha} y$$

说明：（1）最小二乘解一定存在，但全面最小二乘解不一定

（2）存在全面最小二乘解时，若σ_{n+1}为 C 的单重奇异值，全面最小二乘解唯一，否则，解不唯一

例 2. 采用全面最小二乘法重新研究（上例）法向回归的问题

$$Ax = b \qquad A = \begin{bmatrix} t_1 & 1 \\ t_2 & 1 \\ \vdots & \\ t_n & 1 \end{bmatrix} \qquad x = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \qquad b = \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{bmatrix}$$

$$C = \begin{bmatrix} t_1 & 1 & s_1 \\ t_2 & 1 & s_2 \\ \vdots & \vdots & \vdots \\ t_n & 1 & s_n \end{bmatrix}$$

$$C^T C = \begin{bmatrix} t_1 & t_2 & \cdots & t_n \\ 1 & 1 & \cdots & 1 \\ s_1 & s_2 & \cdots & s_n \end{bmatrix} \begin{bmatrix} t_1 & 1 & s_1 \\ t_2 & 1 & s_2 \\ \vdots & \vdots & \vdots \\ t_n & 1 & s_n \end{bmatrix} = \begin{bmatrix} \sum t_i^2 & \sum t_i & \sum s_i t_i \\ \sum t_i & n & \sum s_i \\ \sum s_i t_i & \sum s_i & \sum s_i^2 \end{bmatrix}$$

$$C^T C = \begin{bmatrix} 22.75 & 10.5 & 77.25 \\ 10.5 & 7 & 42.2 \\ 77.25 & 42.2 & 282.28 \end{bmatrix} ,$$

$$\lambda = 309.7754, 2.249389, 0.0051987257$$

$$v_3 = \begin{bmatrix} 1.990944 & 3.044416 & -1 \end{bmatrix}^T$$

$$\begin{array}{ccc} \downarrow & \downarrow & \\ c_1 & c_2 & \end{array} \quad (\text{对应}\,\lambda_3)$$

与法向回归结果并不相同，但亦十分接近。值得注意的是$\bar{s} \neq c_1 \bar{t} + c_2$（全面最小二乘解）