最小二乘法

- 一、线性最小二乘法
- 二、非线性最小二乘法
- 1.改进的Gauss-Newton法
- 2.Levenberger-Marquart方法

最小二乘问题

例 已知某种化合物的生成 量 z 与 物质 A、B 的质量 x、y 及 温度 t 的 关系为

$$z=(a_1x^{1/2}+a_2y^{1/2})\ln(1+a_3t).$$

现通过测量得到 10 组数据 $(x_i,y_i,t_i,z_i)^T, i=1,\cdots,10$。试求 a_1,a_2,a_3 的值。

模型:

令 $$f(a_1,a_2,a_3)=\sum_{i=1}^{10}[z_i-(a_1x_i^{1/2}+a_2y_i^{1/2})\ln(1+a_3t_i)]^2$$

$$\min \quad f(a_1,a_2,a_3)$$

一、线性最小二乘法

1. (P) $\min S(x) = f^T(x) f(x)$

$$= \| Ax - b \|^2$$

这里 $f(x) = Ax - b,$

$A \in R^{m \times n}$, $x \in R^n$, $b \in R^m$

例.

$$\min \quad S(x) = (2x_1 + 3x_2 - 6)^2 + (x_1 + 4x_2 - 3)^2 + (3x_1 - 2x_2 + 2)^2$$

$$\therefore f(x) = \begin{pmatrix} 2x_1 + 3x_2 - 6 \\ x_1 + 4x_2 - 3 \\ 3x_1 - 2x_2 + 2 \end{pmatrix} = \begin{pmatrix} 2 & 3 \\ 1 & 4 \\ 3 & -2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} - \begin{pmatrix} 6 \\ 3 \\ -2 \end{pmatrix}$$

$$\therefore A = \begin{pmatrix} 2 & 3 \\ 1 & 4 \\ 3 & -2 \end{pmatrix}, x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, b = \begin{pmatrix} 6 \\ 3 \\ -2 \end{pmatrix}$$

2. 问题（P）的最优解 x^* 满足

$$A^T A x^* = A^T b$$

证明: "$\Rightarrow$" $S(x) = x^T A^T A x - 2b^T A x + b^T b$

$$\nabla S(x) = 2A^T A x - 2A^T b = 0$$

"$\Leftarrow$" $\forall\ x = x^* + \delta$

$$\|Ax - b\|^2 = \|A(x^* + \delta) - b\|^2$$

$$= [(Ax^* - b) + A\delta]^T [(Ax^* - b) + A\delta]$$

$$= \|Ax^* - b\|^2 + \|A\delta\|^2 + 2\delta^T A^T (Ax^* - b)$$

$$= \left\| Ax^* - b \right\|^2 + \left\| A\delta \right\|^2 + 2\delta^T (A^T Ax^* - A^T b)$$

$$= \left\| Ax^* - b \right\|^2 + \left\| A\delta \right\|^2$$

可见，当δ ＝0时取到最小值.

注：

当 $A^T A$ 可逆时，则线性最小二乘问题的最优解为

$$x^* = (A^T A)^{-1} A^T b$$

例 给定方程组 $\begin{cases} 2x_1+2x_2=3 \\ x_1-2x_2=1 \\ x_1+4x_2=3 \end{cases}$，试求此方程组的最小二 乘解。

解：令 $F(x)=(2x_1+2x_2-3)^2+(x_1-2x_2-1)^2+(x_1+4x_2-3)^2$，

则原问题化为 $\min\limits_{x\in R} F(x)$ 。

记 $A=\begin{bmatrix} 2 & 2 \\ 1 & -2 \\ 1 & 4 \end{bmatrix}, x=\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}, b=\begin{bmatrix} 3 \\ 1 \\ 3 \end{bmatrix}$,

则 $F(x)=(Ax-b)^T(Ax-b)$。

$$\therefore A^T A = \begin{bmatrix} 2 & 1 & 1 \\ 2 & -2 & 4 \end{bmatrix} \begin{bmatrix} 2 & 2 \\ 1 & -2 \\ 1 & 4 \end{bmatrix} = \begin{bmatrix} 6 & 6 \\ 6 & 24 \end{bmatrix},$$

$$A^T b = \begin{bmatrix} 2 & 1 & 1 \\ 2 & -2 & 4 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \\ 3 \end{bmatrix} = \begin{bmatrix} 10 \\ 16 \end{bmatrix}。$$

$$\therefore (A^T A)^{-1} = \begin{bmatrix} \frac{2}{9} & \frac{-1}{18} \\ \frac{-1}{18} & \frac{1}{18} \end{bmatrix}。$$

$$\therefore x^* = (A^T A)^{-1} A^T b = \begin{bmatrix} \frac{2}{9} & \frac{-1}{18} \\ \frac{-1}{18} & \frac{1}{18} \end{bmatrix} \begin{bmatrix} 10 \\ 16 \end{bmatrix} = \begin{bmatrix} \frac{4}{3} \\ \frac{1}{3} \end{bmatrix}。$$

二、非线性最小二乘法

1. 一般形式： $\min S(x) = f^T(x) f(x) = \|f(x)\|^2$

其中：$f(x) = (f_1(x), f_2(x), \ldots, f_m(x))^T$

$x = (x_1, x_2, \ldots, x_n)^T$;

思想：用线性函数来近 似非线性函数，再模仿 线性最小二乘法求解。

2. 分析求解

设已知第 k次的迭代点 $x^{(k)}$，求下一个迭代点 $x^{(k+1)}$。由泰勒公式可知

$$f_i(x) \approx f_i(x^{(k)}) + \nabla f_i(x^{(k)})^T (x - x^{(k)}), (i = 1,2,\ldots,m)。$$

所以 $f(x) \approx f(x^{(k)}) + A(x^{(k)})(x - x^{(k)})$,

其中 $$A(x^{(k)}) = \left. \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \vdots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \right|_{x=x^{(k)}} = \left(\frac{\partial f_i(x^{(k)})}{\partial x_j}\right)_{m\times n}。$$

例 设 $f(x)=\begin{pmatrix} x_1^2+2x_2^2 \\ \ln(1+x_1^2-x_2^2) \\ 2x_1^2+\sin\dfrac{x_2\pi}{2} \end{pmatrix}$, 已知 $x^{(1)}=(1,1)^T$ 。

则有: $x_1^2+2x_2^2 \approx 3+2\times(x_1-1)+4\times(x_2-1)$

$$\ln(1+x_1^2-x_2^2) \approx 0+2\times(x_1-1)+(-2)\times(x_2-1)$$

$$2x_1^2+\sin\frac{x_2\pi}{2} \approx 3+4\times(x_1-1)+0\times(x_2-1)$$

$$\therefore f(x) \approx \begin{pmatrix} 3 \\ 0 \\ 3 \end{pmatrix} + \begin{pmatrix} 2 & 4 \\ 2 & -2 \\ 4 & 0 \end{pmatrix} \begin{pmatrix} x_1-1 \\ x_2-1 \end{pmatrix}$$

记 $A_k = A(x^{(k)})$,则有

$$S(x) \approx \left\| f(x^{(k)}) + A_k(x - x^{(k)}) \right\|^2 = \left\| A_k d^k + f(x^{(k)}) \right\|^2$$

$$= [A_k d^{(k)} + f(x^{(k)})]^T [A_k d^{(k)} + f(x^{(k)})],$$

其中：$d^{(k)} = x - x^{(k)}$。

记 $\varphi(x) = [A_k d^{(k)} + f(x^{(k)})]^T [A_k d^{(k)} + f(x^{(k)})]$,则可用 $\min \varphi(x)$

问题的极小点近似原问 题的极小点。

对于 $\min \varphi(x)$问题，由线性最小二乘 法可得

$$A_k^T A_k d^{(k)} = -A_k^T f(x^{(k)}) \qquad (1)$$

当 $A_k{}^T A_k$ 可逆时则有

$$d^{(k)} = -(A_k{}^T A_k)^{-1} A_k^T f(x^{(k)}) \qquad (2)$$

令$x^{(k+1)}=x^{(k)}+d^{(k)}$，则有迭代公式

$$x^{(k+1)}=x^{(k)}+d^{(k)}=x^{(k)}-(A_k^TA_k)^{-1}A_k^Tf(x^{(k)}) \quad (3)$$

性质：若$f(x)$满足一定条件且$x^{(0)}$充分接近x^*，

则：(1)由迭代得到的$\{x^{(k)}\}$是收敛的；

(2)当$f(x^*)=0$,收敛阶至少是二阶的。

3.改进的Gauss-Newton法：

因为$\nabla S(x)=2\sum_{i=1}^{m}f_i(x)\nabla f_i(x)=2A^T(x)f(x)$,

记$H_k=2A_k^TA_k$，则H_k是$\varphi(x)$在点$x^{(k)}$的*Hesse*矩阵。

（1）式可改写为

$$H_kd^{(k)}=-\nabla S(x^{(k)}) \quad (4)$$

即 $$d^{(k)} = -H_k^{-1} \nabla S(x^{(k)}) \qquad (5)$$

所以 $$x^{(k+1)} = x^{(k)} - H_k^{-1} \nabla S(x^{(k)}) \qquad (6)$$

(6)式称为$Gauss-Newton$公式,

（5）式称为$Gauss-Newton$方向。

令 $g_k = \dfrac{\nabla S(x^{(k)})}{2} = A^T(x^{(k)}) f(x^{(k)}) = A_k^T f(x^{(k)})$。

若$(A_k^T A_k)^{-1}$存在,则由其正定性可知:

$\nabla S(x^{(k)})^T d^k = -2 g_k^T (A_k^T A_k)^{-1} g_k < 0 \ \Rightarrow d^k$为下降方向

否则，若 $A_k^T A_k$奇异，则解不出 $d^{(k)}$。此时令$d^{(k)} = -g_k$。

改进的 $Gauss-Newton$ 算法（共五步）

已知：$f(x)=(f_1(x),\cdots,f_m(x))^T, A(x)=\left(\dfrac{\partial f_i}{\partial x_j}\right)_{m\times n}$。

$Step\ 1$：给定初始点 $x^{(0)}$，精度 $\varepsilon>0$。令 $k:=0$。

$Step\ 2$：计算：$a_{ij}^{(k)}=\dfrac{\partial f_i(x^{(k)})}{\partial x_j}\Rightarrow A_k$；

$$g_j^{(k)}=\sum_{i=1}^{m} f_i(x^{(k)})\frac{\partial f_i(x^{(k)})}{\partial x_j}\Rightarrow g_k。$$

$Step\ 3$：令 $d^k=\begin{cases}-(A_k^T A_k)^{-1}A_k^T f(x^{(k)}) & \text{如果} rank(A)=n\\ -g_k=-A_k^T f(x^{(k)}) & \text{如果} rank(A)<n\end{cases}$

$Step4:$令$x^{(k+1)} = x^{(k)} + t_k d^{(k)}$,其中$t_k := \min\limits_{t} f(x^{(k)} + td^{(k)})$。

$Step5:$若$\| A_k{}^T f(x^{(k)}) \| < \varepsilon$, 则$x^* = x^{(k+1)}$,算法结束;否则,

$k := k+1$,转$Step2$。

三、Levenberger-Marquart方法

1. $Gauss-Newton$法的迭代中，当 $B=A^T(x)A(x)$为奇异或接近奇异时，迭代难以 继续!

2. $L-M$方法采用适当增加 $A^T(x)A(x)$的对角元的措施使迭代继续!

3. 当前的迭代点$x^{(k)}$满足$\nabla S(x^{(k)})\neq 0$, $L-M$方法求解如下方程组:

$$(A(x^{(k)})^T A(x^{(k)})+\lambda I)z := C(x^{(k)})z = -A(x^{(k)})^T f(x^{(k)}) \qquad (**)$$

然后令 $x^{(k+1)}=x^{(k)}+z$ 得到下一迭代点 $x^{(k+1)}$。

其中：λ是一个正数，它在迭代 中不断调整使 $C(x^{(k)})$为正定，且 $S(x^{(k+1)})<S(x^{(k)})$能够成立又不至于使 λ过大!

4. 设$\lambda > 0$，z是方程组（**）的解，那么有：

性质1: $\left\| f(x^{(k)}) + A(x^{(k)})y \right\|^2 \ge \left\| f(x^{(k)}) + A(x^{(k)})z \right\|^2$ （$\forall \|y\| = \|z\|$）

性质2：只要λ适当大，总成立：$S(x^{(k)} + z) < S(x^{(k)})$.

性质3：当λ充分大时，方向 z与方向 $-\nabla S(x^{(k)})$充分接近。

5. 若取$\lambda = 0$，得到的 z就是$Gauss - Newton$ 方向；随着 λ 的增大，上面的性质 3表明方向 z逐步向负梯度 方向 $-\nabla S(X)$偏移，当 λ充分大时便接近负梯度 方向。因而尽可能限制 λ的增加，否则将影响收 敛速度！

6.采用进退方法调整 λ：一次成功迭代后将 λ缩小，迭代遇到困难时将 λ放大。

7. λ调整算法：

初始：给定步长放大因 子$\alpha > 1$和步长缩小因子 $0 < \beta < 1$,
给定$f(x)$,初始点 x，λ的初值，控制终止常数 ε。

$Step1:$求解$(A(x)^T A(x) + \lambda I)z = -A^T(x) f(x)$得$z$ 。

$Step2:$若$S(x+z) \geq S(x)$,则令$\lambda := \alpha\lambda$并返回 $Step1$。

$Step3:$令$x := x+z, \lambda = \beta\lambda$;若$\left\| A(x)^T f(x) \right\| < \varepsilon$ 则迭代终止，
否则返回$Step1$。