第五章
最小二乘法与曲线拟合

§5.0 问题的提出

§5.1 用最小二乘法求解矛盾方程组

§5.2 多项式拟合

§5.0 问题的提出

如果实际问题要求解在[a,b]区间的每一点都“很好地” 逼近$f(x)$的话，运用插值函数有时就要失败。另外，插值所需的数据往往来源于观察测量，本身有一定的误差。要求插值曲线通过这些本身有误差的点，势必使插值结果更加不准确。

如果由试验提供的数据量比较大，又必然使得插值多项式的次数过高而效果不理想。

从给定的一组试验数据出发，寻求函数的一个近似表达式$y=\varphi(x)$，要求近似表达式能够反映数据的基本趋势而又不一定过全部的点(x_i,y_i)，这就是**曲线拟合问题**，函数的近似表达式$y=\varphi(x)$称为**拟合曲线**。本章介绍用**最小二乘法**求拟合曲线。

§5.1 用最小二乘法求解矛盾方程组

一、矛盾方程组的定义

设线性方程组

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \vdots \\ a_{N1}x_1 + a_{N2}x_2 + \cdots + a_{Nn}x_n = b_N \end{cases}$$

或写为

$$\sum_{j=1}^{n} a_{ij}x_j = b_i \qquad (j = 1,2,\cdots,N)$$

其矩阵形式为

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

当方程组的系数矩阵合增广矩阵的秩不相等时，方程组无解，此时方程组称为矛盾方程组。对于 rank$A = n$（A的秩为n）的矛盾方程组（$N>n$），我们寻求其最小二乘意义下的解。

二、用最小二乘法求解矛盾方程组

1.最小二乘原则

由于矛盾方程组的精确解不存在，我们转而寻求其某种意义下，即最小二乘意义下的解。

令 $$\delta_i = \sum_{j=1}^{n} a_{ij}x_j - b_i \qquad (i = 1,2,\cdots,N)$$

称 δ_i 为**偏差**。

工程实际中的许多问题都可以归结为矛盾方程组，实际中需要寻求矛盾方程组的一组解，以使得偏差的绝对值之和 $\sum_{i=1}^{N}|\delta_i|$ 尽可能地小。为了便于分析计算和应用，常采用使偏差的平方和

$$Q = \sum_{i=1}^{N}\delta_i^2 = \sum_{i=1}^{N}\left(\sum_{j=1}^{n} a_{ij}x_j - b_i\right)^2$$

达到最小值，这一条件称为**最小二乘原则**。

按照最小二乘原则来选择未知数$x_1,x_2,\ldots,x_n$的一组取值的方法称为求解矛盾方程组的**最小二乘法**。符合条件的一组取值称为矛盾方程组的**最小二乘解**。

把Q看成是n个自变量$x_1,x_2,\ldots,x_n$的二次函数，记为$Q=f(x_1,x_2,\ldots,x_n)$，因此，求矛盾方程组的最小二乘解就是求二次函数$Q=f(x_1,x_2,\ldots,x_n)$的最小值点。

问题：二次函数$Q=f(x_1,x_2,\ldots,x_n)$是否存在最小值?若最小值存在，如何求出该最小值点?

2.最小二乘解的存在唯一性

引理1：设n元实函数$f(x_1,x_2,\ldots,x_n)$在点$P_0(a_1,a_2,\ldots,a_n)$的某个邻域内连续，且有一阶及二阶连续的偏导数，如果

(1) $$\left.\frac{\partial f}{\partial x_k}\right|_{P_0}=0 \qquad (k=1,2,\cdots,n)$$

(2)矩阵

$$M=\begin{bmatrix} \left.\frac{\partial^2 f}{\partial x_1^2}\right|_{P_0} & \left.\frac{\partial^2 f}{\partial x_1\partial x_2}\right|_{P_0} & \cdots & \left.\frac{\partial^2 f}{\partial x_1\partial x_n}\right|_{P_0} \\ \left.\frac{\partial^2 f}{\partial x_2\partial x_1}\right|_{P_0} & \left.\frac{\partial^2 f}{\partial x_2^2}\right|_{P_0} & \cdots & \left.\frac{\partial^2 f}{\partial x_2\partial x_n}\right|_{P_0} \\ \vdots & \vdots & \ddots & \vdots \\ \left.\frac{\partial^2 f}{\partial x_n\partial x_1}\right|_{P_0} & \left.\frac{\partial^2 f}{\partial x_n\partial x_2}\right|_{P_0} & \cdots & \left.\frac{\partial^2 f}{\partial x_n^2}\right|_{P_0} \end{bmatrix}$$

是正（负）定矩阵，则$f(a_1,a_2,\ldots,a_n)$是n元实函数$f(x_1,x_2,\ldots,x_n)$的极小（大）值。

引理2：设非齐次线性方程组$A\vec{x}=\vec{b}$的系数矩阵$A=(a_{ij})_{N\times n}$，若rank$A=n$，则

(1)矩阵A^TA是对称正定矩阵；

(2)n阶线性方程组$A^TA\vec{x}=A^T\vec{b}$有唯一的解。

证明：（1）矩阵A^TA显然是对称矩阵。

设齐次线性方程组 $A\vec{x}=\vec{0}$

因为rank$A=n$，故齐次方程组有唯一零解。

因此，对于任意的$\vec{x}\neq\vec{0}$，有$A\vec{x}\neq\vec{0}$，从而

$$(A\vec{x})^T(A\vec{x})=\vec{x}^T(A^TA)\vec{x}>0$$

故矩阵A^TA是对称正定矩阵。

(2)因为矩阵A^TA是正定矩阵，故rank$(A^TA)=n$，从而线性方程组 $A^TA\vec{x}=A^T\vec{b}$ 有唯一的解。**证毕**

引理2说明,在条件RankA=n下,无论线性方程组Ax=b是否有解,构造的n阶方程组$A^TAx=A^Tb$一定有唯一解。

定理：设矛盾方程组的系数矩阵的秩为n，则二次函数

$$Q=f(x_1,x_2,\cdots,x_n)=\sum_{i=1}^{N}\left(\sum_{j=1}^{n}a_{ij}x_j-b_i\right)^2$$

一定存在最小值。

证明：因为Q是$x_1,x_2,\ldots,x_n$的二次函数，故Q不仅是连续函数，且有连续的一阶及二阶偏导数。

因为
$$\frac{\partial Q}{\partial x_k}=2a_{1k}(\sum_{j=1}^{n}a_{1j}x_j-b_1)+$$

$$2a_{2k}(\sum_{j=1}^{n}a_{2j}x_j-b_2)+\cdots+2a_{Nk}(\sum_{j=1}^{n}a_{Nj}x_j-b_N)$$

$$= 2\begin{pmatrix} a_{1k} & a_{2k} & \cdots & a_{Nk} \end{pmatrix} \begin{pmatrix} \sum_{j=1}^{n} a_{1j}x_j - b_1 \\ \sum_{j=1}^{n} a_{2j}x_j - b_2 \\ \vdots \\ \sum_{j=1}^{n} a_{Nj}x_j - b_N \end{pmatrix}$$

$$= 2\begin{pmatrix} a_{1k} & a_{2k} & \cdots & a_{Nk} \end{pmatrix} (A\vec{x} - \vec{b})$$

故 $\begin{pmatrix} \dfrac{\partial Q}{\partial x_1} \\ \dfrac{\partial Q}{\partial x_2} \\ \vdots \\ \dfrac{\partial Q}{\partial x_n} \end{pmatrix} = 2A^T(A\vec{x}-\vec{b}) = 2(A^TA\vec{x}-A^T\vec{b})$

令 $\dfrac{\partial Q}{\partial x_k} = 0 \qquad (k=1,2,\cdots,n)$

即 $A^TA\vec{x} = A^T\vec{b} \qquad (*)$

因为rank$A=n$，故由引理2知，上式有唯一解。设解为$x_1=a_1$，$x_2=a_2$,…，$x_n=a_n$，记为点$P_0(a_1,a_2,\ldots,a_n)$，即二元函数Q存在点P_0，使 $\left.\dfrac{\partial f}{\partial x_k}\right|_{P_0}=0 \qquad (k=1,2,\cdots,n)$。故满足引理1的条件（1）。

因为 $$\frac{\partial^2 Q}{\partial x_k \partial x_t} = 2(a_{1k}a_{1t} + a_{2k}a_{2t} + \cdots + a_{Nk}a_{Nt})$$

$$= 2\sum_{i=1}^{N} a_{ik}a_{it} \qquad (k,t = 1,2,\cdots,n)$$

故

$$M = 2\begin{bmatrix} \sum_{i=1}^{N} a_{i1}^2 & \sum_{i=1}^{N} a_{i1}a_{i2} & \sum_{i=1}^{N} a_{i1}a_{i3} & \cdots & \sum_{i=1}^{N} a_{i1}a_{in} \\ \sum_{i=1}^{N} a_{i1}a_{i2} & \sum_{i=1}^{N} a_{i2}^2 & \sum_{i=1}^{N} a_{i2}a_{i3} & \cdots & \sum_{i=1}^{N} a_{i2}a_{in} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{N} a_{i1}a_{in} & \sum_{i=1}^{N} a_{i2}a_{in} & \sum_{i=1}^{N} a_{i3}a_{in} & \cdots & \sum_{i=1}^{N} a_{in}^2 \end{bmatrix} = 2A^T A$$

由引理2知，当rank$A=n$时，矩阵M是对称正定阵，M满足引理1的条件（2），故由引理1知，二次函数Q存在极小值。

又因方程组（*）式有唯一解，故Q存在的极小值就是最小值，线性方程组（*）式的解就是最小值点。

证毕

Remark1：线性方程组（*）式称为正则方程组。

Remark2：该定理说明，只要矛盾方程组的系数矩阵A的秩rank$A=n$，则

（1）矛盾方程组的最小二乘解存在；

（2）正则方程组有唯一解，此解就是矛盾方程组的最小二乘解。

3.最小二乘法解矛盾方程组

计算步骤：

（1）判断方程组的秩是否满足rank$A=n$?

（2）写出正则方程组；

（3）求解正则方程组，其解就是矛盾方程组的最小二乘解。

§5.2 多项式拟合

一、曲线拟合模型

定义：依据某种标准选择一条“最好”的简单曲线作为一组离散数据$\{(x_i, y_i)\}_{i=0}^{N}$的连续模型。

确定曲线的类型：一般选取简单的低次多项式。

求一个次数不高于N - 1次的多项式：

$$y=\varphi(x)=a_0+a_1x+a_2x^2+\cdots+a_mx^m \qquad (m<N-1)$$

（其中$a_0,a_1,\ldots,a_m$待定），使其“最好”的拟合这组数据。“最好”的标准是：使得$\varphi(x)$在x_i的偏差

$$\delta_i=\varphi(x_i)-y_i \quad (i=1,2,\cdots,N)$$

的平方和 $$Q=\sum_{i=1}^{N}\delta_i^2=\sum_{i=1}^{N}[\varphi(x_i)-y_i]^2$$

达到最小。

由于拟合曲线$y=\varphi(x)$不一定过点(x_i,y_i)，因此，把点(x_i,y_i)带入$y=\varphi(x)$ ，便得到以$a_0,a_1,\ldots,a_m$为未知量的矛盾方程组

$$\left.\begin{array}{l} a_0+a_1x_1+a_2x_1^2+\cdots+a_mx_1^m=y_1 \\ a_0+a_1x_2+a_2x_2^2+\cdots+a_mx_2^m=y_2 \\ \vdots \\ a_0+a_1x_N+a_2x_N^2+\cdots+a_mx_N^m=y_N \end{array}\right\}$$

其矩阵形式为

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

其中

$$A=\begin{bmatrix}1 & x_1 & x_1^2 & \cdots & x_1^m\\ 1 & x_2 & x_2^2 & \cdots & x_2^m\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 1 & x_N & x_N^2 & \cdots & x_N^m\end{bmatrix},\vec{x}=\begin{bmatrix}a_0\\ a_1\\ \vdots\\ a_m\end{bmatrix},\vec{b}=\begin{bmatrix}y_1\\ y_2\\ \vdots\\ y_N\end{bmatrix}$$

$\varphi(x)$在x_i的偏差就是矛盾方程组各方程的偏差。曲线拟合的条件就是确定$a_0,a_1,\ldots,a_m$，使得偏差的平方和Q达到最小值。

据此可知，$a_0,a_1,\ldots,a_m$就是矛盾方程组的最小二乘解，也就是正则方程组 $A^TA\vec{x}=A^T\vec{b}$ 的解。

二、曲线拟合的最小二乘解法

$$A^T A=\begin{bmatrix} N & \sum_{i=1}^{N} x_i & \sum_{i=1}^{N} x_i^2 & \cdots & \sum_{i=1}^{N} x_i^m \\ \sum_{i=1}^{N} x_i & \sum_{i=1}^{N} x_i^2 & \sum_{i=1}^{N} x_i^3 & \cdots & \sum_{i=1}^{N} x_i^{m+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{N} x_i^m & \sum_{i=1}^{N} x_i^{m+1} & \sum_{i=1}^{N} x_i^{m+2} & \cdots & \sum_{i=1}^{N} x_i^{2m} \end{bmatrix}, A^T\vec{b}=\begin{bmatrix} \sum_{i=1}^{N} y_i \\ \sum_{i=1}^{N} x_i y_i \\ \vdots \\ \sum_{i=1}^{N} x_i^m y_i \end{bmatrix}$$

正则方程组为：

$$\left.\begin{aligned} & a_0 N + a_1 \sum_{i=1}^{N} x_i + a_2 \sum_{i=1}^{N} x_i^2 + \cdots + a_m \sum_{i=1}^{N} x_i^m = \sum_{i=1}^{N} y_i \\ & a_0 \sum_{i=1}^{N} x_i + a_1 \sum_{i=1}^{N} x_i^2 + a_2 \sum_{i=1}^{N} x_i^3 + \cdots + a_m \sum_{i=1}^{N} x_i^{m+1} = \sum_{i=1}^{N} x_i y_i \\ & \vdots \\ & a_0 \sum_{i=1}^{N} x_i^m + a_1 \sum_{i=1}^{N} x_i^{m+1} + a_2 \sum_{i=1}^{N} x_i^{m+2} + \cdots + a_m \sum_{i=1}^{N} x_i^{2m} = \sum_{i=1}^{N} x_i^m y_i \end{aligned}\right\}$$

三、解的存在唯一性

定理：设$x_1,x_2,\ldots,x_N$互异，且$N>m+1$，则上面的正则方程组有唯一的解。

证明：只需证明矛盾方程组的系数矩阵A的秩rank$A=m+$

矛盾方程组的系数矩阵A是$N\times(m+1)$的矩阵，记A的前$m+1$行构成$m+1$阶子矩阵

$$A_1=\begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^m \\ 1 & x_2 & x_2^2 & \cdots & x_2^m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{m+1} & x_{m+1}^2 & \cdots & x_{m+1}^m \end{bmatrix}$$

该矩阵是范德蒙矩阵，由$x_1,x_2,\ldots,x_N$互异知行列式不为零，从而有rank$A=m+1$。由引理2知，正则方程组有唯一解。

证毕

四、最小二乘法拟合曲线的步骤

1..通过观察、分析得到拟合曲线的数学模型，或根据经验公式确定数学模型。

2.将拟合曲线的数学模型转换为多项式。

3.写出矛盾方程组。

4.写出正则方程组。（可由多项式模型直接得到）

5.求解正则方程组，得到拟合曲线的待定系数。

6.将正则方程组的解带回到数学模型中，得到拟合曲线。

Remark

1.同一问题可以有不同的拟合曲线，通常根据均方误差 $\sqrt{\sum_{i=1}^{N}[\varphi(x_i)-y_i]^2}$ 和最大偏差 $\max\limits_{1\le i\le N}|\varphi(x_i)-y_i|$ 的大小来衡量拟合曲线的优劣。均方误差和最大偏差较小的拟合曲线为较优的拟合曲线。

2.在解决实际问题时，有时通过观察选择多个函数类型进行计算、分析、比较，最终获得较好的数学模型；有时把经验公式作为数学模型，只是用最小二乘法来确定公式中的待定常数。

Remark

3.当拟合曲线$\varphi(x)$中的待定常数是线性形式时，可直接根据矛盾方程组得到正则方程组而求解。当待定常数不是线性形式时，则应该先将待定常数线性化，再根据矛盾方程组写出正则方程组而求解。

例1： $y = ae^{bx} \Rightarrow \ln y = \ln a + bx \Rightarrow$

$u = \ln y, A = \ln a, B = b \Rightarrow u = A + Bx$

例2： $y = \dfrac{1}{a+bx} \Rightarrow \dfrac{1}{y} = a + bx \Rightarrow$

$u = \dfrac{1}{y} \Rightarrow u = a + bx$

曲线拟合应用实例:

例1: 试用最小二乘法求一个形如 $y = ae^{bx}$ (*a,b*为常数)的经验公式,使它与下列数据相拟合(取四位小数)

x_i	1	2	3	4	5	6	7	8
y_i	15.3	20.5	27.4	36.6	49.1	65.6	87.8	117.6

解:由于经验公式中待定常数*a,b*是非线性形式,故做如下变形: $\ln y = \ln a + bx$

令: $u = \ln y, A = \ln a, B = b$

则有: $u = A + Bx$

将*x,u*带入得到关于*A,B*的矛盾方程组，进而得正规方程组并求出*A,B*，由*A,B*得到*a,b*即可。

(具体计算数据见书P141页例6.3)

例2. 对彗星1968Tentax的移动在某极坐标系下有如下表所示的观察数据，假设忽略来自行星的干扰,坐标应满足：$r=\dfrac{p}{1-e\cos p}$ 其中：*p*为参数，*e*为偏心率，试用最小二乘法拟合*p*和*e*。

r	2.70	2.00	1.61	1.20	1.02
φ	48^0	67^0	83^0	108^0	126^0

解:变形为: $\dfrac{1}{r}=\dfrac{1}{p}-\dfrac{e}{p}\cos\varphi,$ 则有如下数据

$y=\dfrac{1}{r}$	0.370370	0.50000	0.621118	0.83333	0.980392
$t=\cos\varphi$	0.669131	0.390731	0.121869	-0.309017	-0.587785

记 $a=\dfrac{1}{p}, b=-\dfrac{e}{p}$，得拟合模型：$a+bt=y$

则矛盾方程组为：

$$\begin{pmatrix} 1 & 0.669131 \\ 1 & 0.390731 \\ 1 & 0.121869 \\ 1 & -0.309017 \\ 1 & -0.587785 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix}=\begin{pmatrix} 0.370370 \\ 0.500000 \\ 0.621118 \\ 0.833333 \\ 0.980392 \end{pmatrix}$$

得正则方程组为：

$$\begin{pmatrix} 5.0 & 0.284929 \\ 0.284929 & 1.056242 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix}=\begin{pmatrix} 3.305214 \\ -0.314887 \end{pmatrix}$$

解得： $a = 0.688617 \qquad b = -0.483880$

则： $p = \dfrac{1}{a} = 1.452186 \quad e = -bp = 0.702684$

则拟合方程为： $$r = \frac{1.452186}{1 - 0.702684\cos\varphi}$$