科学计算(Scientific Computing)

第三章　数据拟合法

任课老师: Xiao-liang Cheng(程晓良)

Department of Mathematics, Zhejiang University, Hangzhou 310027, P.R. China.

2023. 3 .27

数据拟合法

1. 问题的提出及最小二乘原理

数据拟合法

1. 问题的提出及最小二乘原理
2. 多变量的数据拟合

数据拟合法

数据拟合法

数据拟合

测定一组数据$(x_i, y_i), i = 1, 2, \cdots, n$，求一个近似简单的解析表达式

$$y = \phi(x),$$

使得数据在此曲线附近.

插值的区别: 不要求曲线经过所有点(x_i, y_i), 反映给定数据的一般趋势.

为什么叫最小二乘?

已知点(x_i, y_i)，求$y = \phi(x) = a + bx$使得这些点尽量靠近这条直线? 即

$$\delta_i = y_i - \phi(x_i) = y_i - a - bx_i,$$

最小二乘法: 求(a, b)使得

$$\sum_{i=1}^{n} \delta_i^2 \to min!$$

如果这样提问题: 求(a, b)使得

$$\sum_{i=1}^{n} |\delta_i|^p \to min!$$

求(a, b)使得

$$\max_{1 \le i \le n} |\delta_i| \to min!$$

最小二乘: 求(a,b)使得

$$F(a,b)=\sum_{i=1}^{n}\delta_i^2=\sum_{i=1}^{n}(y_i-a-bx_i)^2\rightarrow min!$$

则求得解

$$a=\frac{1}{n}\sum_{i=1}^{n}(y_i-bx_i)$$

$$b=\frac{\sum_{i=1}^{n}x_iy_i-\frac{1}{n}(\sum_{i=1}^{n}x_i)(\sum_{i=1}^{n}y_i)}{\sum_{i=1}^{n}x_i^2-\frac{1}{n}(\sum_{i=1}^{n}x_i)^2}$$

记

$$\mathbf{x}_1 = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \ \mathbf{x}_2 = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \ \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}.$$

将向量$\mathbf{y}$表示成$\mathbf{x}_1, \mathbf{x}_2$的线性组合

$$a\mathbf{x}_1 + b\mathbf{x}_2 \approx \mathbf{y}.$$

几何意义: 残量$\mathbf{r} = \mathbf{y} - a\mathbf{x}_1 - b\mathbf{x}_2$和由向量$\mathbf{x}_1, \mathbf{x}_2$张成的平面垂直,

$$\mathbf{x}_1^T\mathbf{r} = 0, \ \ \mathbf{x}_2^T\mathbf{r} = 0.$$

$$\mathbf{X} = (\mathbf{x}_1, \ \mathbf{x}_2), \quad \alpha = (a, b)^T$$

$$\mathbf{X}\alpha = \mathbf{y}, \quad \mathbf{X}^T\mathbf{X}\alpha = \mathbf{X}^T\mathbf{y}.$$

解的存在唯一性: $\mathbf{x}_1, \mathbf{x}_2$线性无关. 最后一个方程称为正规方程!

例: 用最小二乘法求下列超定方程组的解:

$$\begin{cases} 2x_1 - x_2 = 1 \\ 8x_1 + 4x_2 = 0 \\ 2x_1 + x_2 = 1 \\ 7x_1 - x_2 = 8 \\ 4x_1 = 3 \end{cases}$$

解析. 用最小二乘求极小值.

$$f(x_1, x_2) = (2x_1 - x_2 - 1)^2 + (8x_1 + 4x_2)^2 + (2x_1 + x_2 - 1)^2 + (7x_1 - x_2 - 8)^2 + (4x_1 - 3)^2.$$

或记

$$\mathbf{z}_1 = \begin{pmatrix} 2 \\ 8 \\ 2 \\ 7 \\ 4 \end{pmatrix}, \ \mathbf{z}_2 = \begin{pmatrix} -1 \\ 4 \\ 1 \\ -1 \\ 0 \end{pmatrix}, \ \mathbf{y} = \begin{pmatrix} 1 \\ 0 \\ 1 \\ 8 \\ 3 \end{pmatrix}.$$

则求x_1, x_2满足(残量和$\mathbf{z}_1, \mathbf{z}_2$正交)

$$x_1\mathbf{z}_1 + x_2\mathbf{z}_2 \approx \mathbf{y}.$$

已知数据$(x_{1,i}, x_{2,i}, \cdots, x_{k,i}, y_i)$, $i = 1, 2, \cdots, n$.

一般来说，$n >> k$, 如果选择近似方程组为

$$y^* = a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_k x_k.$$

要求

$$\phi(a_0, a_1, \cdots, a_k) = \sum_{i=1}^{N} (y_i - y_i*)^2$$

达到最小.

记

$$\mathbf{x}_0 = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \ \mathbf{x}_1 = \begin{pmatrix} x_{1,1} \\ x_{1,2} \\ \vdots \\ x_{1,n} \end{pmatrix}, \ \cdots, \mathbf{x}_k = \begin{pmatrix} x_{k,1} \\ x_{k,2} \\ \vdots \\ x_{k,n} \end{pmatrix}, \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}.$$

求超定方程组

$$a_0\mathbf{x}_0 + a_1\mathbf{x}_1 + \cdots + a_k\mathbf{x}_k \approx \mathbf{y}.$$

残量$\mathbf{r} = \mathbf{y} - a_0\mathbf{x}_0 - \cdots - a_k\mathbf{x}_k$和由向量$\mathbf{x}_0, \mathbf{x}_1, \cdots, \mathbf{x}_k$张成的平面垂直,

$$\mathbf{x}_0^T\mathbf{r} = 0, \ \mathbf{x}_1^T\mathbf{r} = 0, \cdots, \ \mathbf{x}_k^T\mathbf{r} = 0.$$

$$\mathbf{X} = (\mathbf{x}_0, \mathbf{x}_1, \cdots, \mathbf{x}_k), \quad \alpha = (a_0, \cdots, a_k)^T$$

$$\mathbf{X}\alpha = \mathbf{y}, \quad \mathbf{X}^T\mathbf{X}\alpha = \mathbf{X}^T\mathbf{y}.$$

最后一个为正规方程!

高次多项式

特别地，如果选择近似方程组

$$y^* = a_0 + a_1 x + a_2 x^2 + \cdots + a_k x^k,$$

则为k次多项式数据拟合.
记

$$\mathbf{z}_0 = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \ \mathbf{z}_1 = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \ \cdots, \mathbf{z}_k = \begin{pmatrix} x_1^k \\ x_2^k \\ \vdots \\ x_n^k \end{pmatrix}, \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}.$$

求超定方程组

$$a_0 \mathbf{z}_0 + a_1 \mathbf{z}_1 + \cdots + a_k \mathbf{z}_k \approx \mathbf{y}.$$

非线性曲线的数据拟合

原来的函数关系是

$$f(y') = a + bg(x').$$

令$y = f(y'), x = g(x')$, 则经过变换后即得

$$y = a + bx$$

的形式. 于是许多非线性的问题就转化成线性问题而得到解决.

♠ 比如数据(x_i, y_i), 求拟合

$$\frac{1}{y} = a + b\frac{1}{x}$$

令$y' = \frac{1}{y}, x' = \frac{1}{x}$, 则由数据$(x_i', y_i') = (1/x_i, 1/y_i)$拟合

$$y' = a + bx'$$

♠ 又比如数据(x_i, y_i), 求拟合

$$y = a\,e^{bx}.$$

令$y' = ln(y), x' = x$, 则由数据$(x_i', y_i') = (x_i, ln(y_i))$拟合

$$y' = A + bx',\ A = ln(a).$$

♠ 又比如数据(x_i, y_i), 求拟合

$$y = \frac{1+x}{a+bx}.$$

令$y' = \frac{1}{y}, x' = \frac{1}{x+1}$, 则由数据$(x_i', y_i') = (\frac{1}{x_i+1}, \frac{1}{y_i})$拟合

$$y' = b + (a-b)x'.$$

用多项式拟合

$$P(x) = \sum_{j=0}^{m} a_j x^j,$$

正规方程可得唯一解，但多项式次数较高时出现病态.

考虑更一般的情形, 即用

$$y^* = \psi_m(x) = \alpha_0 P_0(x) + \alpha_1 P_1(x) + \cdots + \alpha_m P_m(x) = \sum_{k=0}^{m} \alpha_k P_k(x).$$

最小二乘，加上权重. 则

$$\phi(\alpha_0, \alpha_1, \cdots, \alpha_m) = \sum_{i=1}^{N} w_i \delta_i^2 = \sum_{i=1}^{N} w_i (y_i - \sum_{k=0}^{m} \alpha_k P_k(x_i))^2$$

达到最小.

可得正规方程组

$$\sum_{k=0}^{m} c_{jk} \alpha_k - c_j = 0, \ j = 0, 1, \cdots, m.$$

如果能找到多项式$P_k(x)$ $(k=0,1,\cdots,m)$满足下面关系:

$$\begin{cases} c_{jk}=(P_k,P_j)=\sum_{i=1}^{N}w_iP_j(x_i)P_k(x_i)=0,\ \ j\neq k \\ c_{jj}=(P_j,P_j)=\sum_{i=1}^{N}w_iP_j^2(x_i)>0,\ \ j,k=0,1,\cdots,m. \end{cases}$$

则正规方程是一个对角矩阵,直接写出解.

这样的多项式$P_k(x)$称为对某组x_i值和与之对应的权数w_i值的正交多项式. 怎样得到$P_k(x)$?

等距节点

设已给定一组$n+1$个等距节点$\xi_i\ (i=0,1,\cdots,n)$, 它们的步长为h, 令权函数$w_i=1$, 引进变换

$$x=\frac{\xi-\xi_0}{h}$$

则以上各点将变为$x=0,1,2,\cdots,n$的$n+1$个整数点. 可以证明

$$P_{m,n}(x)=\sum_{k=0}^{m}(-1)^k C_m^k C_{m+k}^k \frac{x^{(k)}}{n^{(k)}}$$

是满足正交条件,

$$\sum_{x=0}^{n}(x+k)^{(k)}P_{m,n}(x)=0,\quad k=0,1,\cdots,m-1$$

这里$x^{(m)}=x(x-1)\cdots(x-m+1)$.

它的前几项多项式如下:

$$P_{0,n}(x) = 1$$

$$P_{1,n}(x) = 1 - 2\frac{x}{n}$$

$$P_{2,n}(x) = 1 - 6\frac{x}{n} + 6\frac{x(x-1)}{n(n-1)}$$

$$P_{3,n}(x) = 1 - 12\frac{x}{n} + 30\frac{x(x-1)}{n(n-1)} - 20\frac{x(x-1)(x-2)}{n(n-1)(n-2)}$$

最佳平方逼近问题

设$\mathbb{P}_n(x)$表示n次多项式空间，定义内积及范数

$$(f,g)=\int_a^b \rho(x)f(x)\,g(x)\,dx,\qquad \|f\|=(f,f)^{\frac{1}{2}}.$$

函数的最佳平方逼近问题: 求$p_n(x)\in\mathbb{P}_n(x)$满足

$$\|f-p_n\|=\inf_{q_n\in\mathbb{P}_n}\|f-q_n\|.$$

则$p_n(x)=a_0+a_1x+\cdots+a_nx^n$满足线性方程组

$$\sum_{j=0}^{n}h_{i,j}a_j=f_i,\ \ i=0,1,\cdots,n.$$

$$h_{i,j}=\int_a^b\rho(x)x^ix^j\,dx,\qquad f_i=\int_a^b\rho(x)f(x)x^i.$$

特别的，若$\rho(x)=1, a=0, b=1$, 则

$$h_{i,j}=\frac{1}{i+j+1},\ \ i,j=0,1,2,\cdots,n.$$

记矩阵$H=(h_{i,j})_{n\times n}$, (上式中$i,j\leftarrow i+1, j+1$, n阶), 则

$$h_{i,j}=\frac{1}{i+j-1},\ \ i,j=1,2,\cdots,n.$$

这是Hilbert矩阵，是一个高度病态的矩阵. 在解线性方程组时作为试验问题的例子.

如果

$$p_n(x) = a_0P_0(x) + a_1P_1(x) + \cdots + a_nP_n(x)$$

是最佳逼近的解，则

$$\sum_{j=0}^{n} \widetilde{h}_{i,j}a_j = \widetilde{f}_i, \ i = 0, 1, \cdots, n.$$

$$\widetilde{h}_{i,j} = \int_a^b \rho(x)P_i(x)P_j(x)\,dx, \qquad \widetilde{f}_i = \int_a^b f(x)P_i(x).$$

如何选取$P_k(x)$使得

$$(P_i, P_j) = \delta_{i,j} = \begin{cases} 0, & i \neq j \\ 1, & i = j. \end{cases}$$

正交多项式系

如果存在多项式$P_k(x)$, $k=0,1,2,\cdots$使得

$$(P_i,P_j)=\delta_{i,j}=\begin{cases}0, & i\neq j\\ 1, & i=j\end{cases}$$

则称为$[a,b]$上关于权函数$\rho(x)$的规格化正交多项式系.

例. Legendre多项式

$$P_n(x)=\frac{1}{2^n n!}\frac{d^n}{dx^n}(x^2-1)^n,\ n=0,1,2,\cdots$$

它是区间$[-1,1]$关于权函数1的正交多项式系.

例. Chebyshev多项式

$$T_n(x)=\cos(n\,\theta),\quad x=\cos(\theta).$$

它是区间$[-1,1]$关于权函数$\frac{1}{\sqrt{1-x^2}}$的正交多项式系.

一般问题，如何求正交多项式？即给定权重函数，给定内积，如何求？

Gram-Schmit正交化过程！

给定基函数$\phi_1, \phi_2, \phi_3, \cdots,$

$$\phi_1^* = \phi_1; \quad \widetilde{\phi}_1 = \frac{1}{(\phi_1^*, \phi_1^*)^{\frac{1}{2}}}\phi_1^*;$$

$$\phi_2^* = \phi_2 - (\widetilde{\phi}_1, \phi_2)\widetilde{\phi}_1; \quad \widetilde{\phi}_2 = \frac{1}{(\phi_2^*, \phi_2^*)^{\frac{1}{2}}}\phi_2^*;$$

$$\phi_3^* = \phi_3 - (\widetilde{\phi}_2, \phi_3)\widetilde{\phi}_2 - (\widetilde{\phi}_1, \phi_3)\widetilde{\phi}_1; \quad \widetilde{\phi}_3 = \frac{1}{(\phi_3^*, \phi_3^*)^{\frac{1}{2}}}\phi_3^*;$$

Legendre多项式性质

Legendre多项式

$$P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n}(x^2-1)^n, \ n = 0, 1, 2, \cdots$$

它是区间$[-1,1]$关于权函数1的正交多项式系.
性质:
(1) $\int_{-1}^{1} P_n(x)\, P_m(x)\, dx = 0, \ n \neq m;$
$\int_{-1}^{1} P_n(x)\, P_m(x)\, dx = \frac{2}{2n+1}, \ n = m;$
(2) 递推关系:

$$\begin{cases} P_0(x) = 1, \qquad P_1(x) = x \\ P_{n+1}(x) = \dfrac{2n+1}{n+1} x P_n(x) - \dfrac{n}{n+1} P_{n-1}(x), \quad n = 1, 2, \cdots \end{cases}$$

(3) $P_n(x)$在$(-1,1)$中有n个互不相同的实根.

证明: (3) $P_n(x)$在$(-1,1)$中有n个互不相同的实根.

Rolle方法: 记$h_n(x)=\frac{1}{2^n n!}(x^2-1)^n$, 则$-1,1$是$h_n(x)$的$n$重根，由Rolle定理，至少存在一个$\xi_1\in(-1,1)$ 是$\frac{d}{dx}h_n(x)=0$的根.

对于$\frac{d}{dx}h_n(x)=0$有三个不同的根$-1,\xi_1,1$, 至少存在$\eta_1\in(-1,\xi_1),\eta_2\in(\xi_1,1)$是$\frac{d^2}{dx^2}h_n(x)=0$的根,

再对于$\frac{d^2}{dx^2}h_n(x)=0$有四个不同的根$-1,\eta_1,\eta_2,1$, 则至少存在$\chi_1\in(-1,\eta_1),\chi_2\in(\eta_1,\eta_2),\chi_3\in(\eta_2,1)$是$\frac{d^3}{dx^3}h_n(x)=0$的根,

依次类推，$P_n(x)=\frac{d^n}{dx^n}h_n(x)=0$在$(-1,1)$上至少有$n$个不同的实根，且只有$n$个不同的实根.

这个证明依赖于Legendre多项式特殊性，一般是如下利用正交性质的证明方法.

利用正交性质证明: $\int_{-1}^{1} P_n(x)\,Q(x)\,dx = 0$对所有$\leq n-1$次多项式$Q(x)$成立.

(1) 若$P_n(x)$在$(-1,1)$内有$s(<n)$个根$x_1,\cdots,x_s$(可以重根)，则

$$P_n(x) = (x-x_1)\cdots(x-x_s)h(x)$$

这里$h(x)$是$n-s$次多项式且在$(-1,1)$内值不变号.

于是取$Q(x) = (x-x_1)\cdots(x-x_s)$, 则

$$\int_{-1}^{1} P_n(x)Q(x)\,dx = \int_{-1}^{1} (x-x_1)^2\cdots(x-x_s)^2 h(x)\,dx = 0$$

不成立. 矛盾.

(2) 若$P_n(x)$在$(-1,1)$内有n个根$x_1,\cdots,x_n$但存在重根，不妨设$x_1 = x_2 = \cdots = x_t$, $t \geq 2$, 则

$$P_n(x) = c(x-x_1)^t(x-x_{t+1})\cdots(x-x_n).$$

取$Q(x) = (x-x_1)^{t-2}(x-x_{t+1})\cdots(x-x_n)$是小于$n$次的多项式，但

$$\int_{-1}^{1} P_n(x)Q(x)\,dx = \int_{-1}^{1} c(x-x_1)^{2t-2}(x-x_{t+1})^2\cdots(x-x_n)^2\,dx = 0$$

不成立. 矛盾.