第五章

数据拟合和函数逼近(1 – 2)

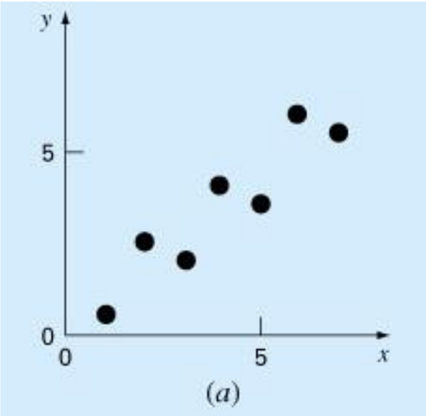
第一节 引言

第二节 最小二乘法

第1节 引言

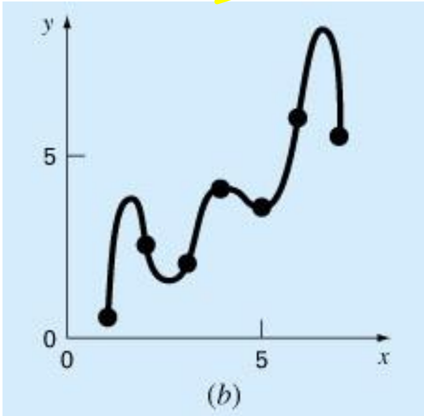
已知一组数据$(x_i, f(x_i))_{i=1}^{n}$，求一个函数$\varphi(x)$，使得$\varphi(x)$能反映数据点的变化趋势。

容易出现Runge现象



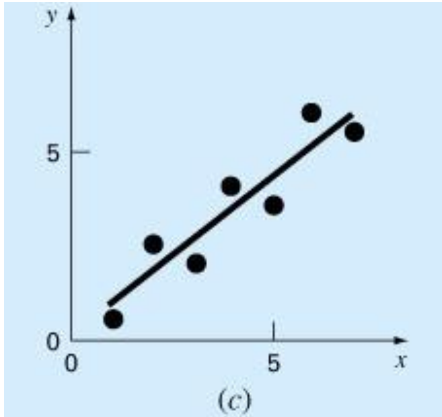
(a)

数据点



(b)

插值方法



(c)

拟合或逼近

拟合的标准？

定义1. 设$(x_i, y_i)(i=1,2,\cdots,n)$为一组已知数据，$\varphi(x)$为一个近似函数，称$\delta_i = y_i - \varphi(x_i)$为数据残差。

求逼近函数$\varphi(x)$的常用规则：

1. $\sum\limits_{i=1}^{n}|\delta_i| = min$

2. $\sum\limits_{i=1}^{n}\delta_i^2 = min$ 最佳平方逼近（拟合：最小二乘法）

3. $\max\limits_{1\le i\le n}|\delta_i| = min$ 最佳一致逼近

第2节 最小二乘法

➢ 2.1 函数组的线性相关性

定义1 函数组$\{\varphi_0(x), \varphi_1(x), \cdots, \varphi_n(x)\}$满足条件：对任意$x \in [a, b]$，线性组合$a_0\varphi_0(x) + a_1\varphi_1(x) + \cdots + a_n\varphi_n(x) = 0$，当且仅当$a_0 = a_1 = \cdots = a_n = 0$时成立，则称$\{\varphi_i(x)\}$在$[a, b]$上是**线性无关**的。

例如：$1, x, x^2, \cdots, x^n$就是$[a, b]$上线性无关函数组.

定义2 设$\{\varphi_0(x),\varphi_1(x),\cdots,\varphi_n(x)\}$线性无关，函数空间$\Phi$满足：

任意$\varphi(x)\in\Phi$均有:$\varphi(x)=c_0\varphi_0(x)+c_1\varphi_1(x)+\cdots+c_n\varphi_n(x)$

则称Φ为由$\varphi_0(x),\varphi_1(x),\cdots,\varphi_n(x)$张成的**线性空间**，记作

$$\Phi=\mathrm{span}\{\varphi_0(x),\varphi_1(x),\cdots,\varphi_n(x)\}$$

称$\varphi_0(x),\varphi_1(x),\cdots,\varphi_n(x)$ 为Φ的一组基.

例如：$1,x,x^2,\cdots,x^n$是多项式函数$p_n(x)$空间的一组基.

➢ 2.2 最小二乘法

问题：已知数据组$(x_i, y_i)(i = 1, 2, \cdots, n)$，

选定近似函数空间Φ, 并求$\varphi(x) \in \Phi$， 使得

$$\sum_{i=1}^{n}[y_i - \varphi^*(x_i)]^2 = \min_{\varphi\in\Phi} \sum_{i=1}^{n}[y_i - \varphi(x_i)]^2$$

$\varphi(x)$称为数据组$(x_i, y_i)(i = 1, 2, \cdots n)$的**最小二乘函数**。

这种确定近似函数的方法称为数据拟合的**最小二乘法**。

近似函数空间Φ主要有：**多项式空间**，指数空间，三角函数空间等

多项式拟合

设Φ为至多$m(m \ll n)$次多项式全体构成的集合；

若$\varphi_m^*(x) = \sum_{k=0}^{m} a_k^* x^k$，$\varphi_m(x) = \sum_{k=0}^{m} a_k x^k$ 满足

$$\sum_{i=1}^{n} [y_i - \varphi_m^*(x_i)]^2 = \min_{\varphi \in \Phi} \sum_{i=1}^{n} [y_i - \varphi_m(x_i)]^2$$

则称$\varphi_m^*(x)$为数据组$\{(x_i, y_i)\}_{i=1}^{n}$的m次最小二乘拟合多项式.

令$Q(a_0, a_1, \cdots, a_m) = \sum_{i=1}^{n} \left[y_i - (a_0 + a_1 x_i + a_2 x_i^2 + \cdots + a_m x_i^m)\right]^2$

求一组系数$a_0, a_1, \cdots, a_m$，使得Q取最小值。

$$Q(a_0,a_1,\cdots,a_m)=\sum_{i=1}^{n}\left[y_i-\left(a_0+a_1x_i+a_2x_i^2+\cdots+a_mx_i^m\right)\right]^2$$

根据极值的必要条件：

若Q在$a_0,a_1,\cdots,a_m$处取极值，则$\frac{\partial Q}{\partial a_i}=0, i=0,1,\cdots,m.$

得，

$$\begin{cases}2\sum_{i=1}^{n}\left[y_i-\left(a_0+a_1x_i+a_2x_i^2+\cdots+a_mx_i^m\right)\right](-1)=0\\2\sum_{i=1}^{n}\left[y_i-\left(a_0+a_1x_i+a_2x_i^2+\cdots+a_mx_i^m\right)\right](-x_i)=0\\2\sum_{i=1}^{n}\left[y_i-\left(a_0+a_1x_i+a_2x_i^2+\cdots+a_mx_i^m\right)\right](-x_i^2)=0\\\cdots\cdots\\2\sum_{i=1}^{n}\left[y_i-\left(a_0+a_1x_i+a_2x_i^2+\cdots+a_mx_i^m\right)\right](-x_i^m)=0\end{cases}$$

将方程组整理成关于$a_0, a_1, \cdots, a_m$的方程组，可得$m+1$个方程，

$$
\begin{cases}
na_0 + a_1\sum\limits_{i=1}^{n} x_i + a_2\sum\limits_{i=1}^{n} x_i^2 + \cdots + a_m\sum\limits_{i=1}^{n} x_i^m = \sum\limits_{i=1}^{n} y_i \\
a_0\sum\limits_{i=1}^{n} x_i + a_1\sum\limits_{i=1}^{n} x_i^2 + a_2\sum\limits_{i=1}^{n} x_i^3 + \cdots + a_m\sum\limits_{i=1}^{n} x_i^{m+1} = \sum\limits_{i=1}^{n} x_i y_i \\
a_0\sum\limits_{i=1}^{n} x_i^2 + a_1\sum\limits_{i=1}^{n} x_i^3 + a_2\sum\limits_{i=1}^{n} x_i^4 + \cdots + a_m\sum\limits_{i=1}^{n} x_i^{m+2} = \sum\limits_{i=1}^{n} x_i^2 y_i \\
\cdots\cdots \\
a_0\sum\limits_{i=1}^{n} x_i^m + a_1\sum\limits_{i=1}^{n} x_i^{m+1} + a_2\sum\limits_{i=1}^{n} x_i^{m+2} + \cdots + a_m\sum\limits_{i=1}^{n} x_i^{2m} = \sum\limits_{i=1}^{n} x_i^m y_i
\end{cases}
$$

写成矩阵形式

$$\begin{pmatrix} n & \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 & \cdots & \sum_{i=1}^n x_i^m \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^3 & \cdots & \sum_{i=1}^n x_i^{m+1} \\ \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^3 & \sum_{i=1}^n x_i^4 & \cdots & \sum_{i=1}^n x_i^{m+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n x_i^m & \sum_{i=1}^n x_i^{m+1} & \sum_{i=1}^n x_i^{m+2} & \cdots & \sum_{i=1}^n x_i^{2m} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_i y_i \\ \sum_{i=1}^n x_i^2 y_i \\ \vdots \\ \sum_{i=1}^n x_i^m y_i \end{pmatrix}$$

记$A = \begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^m \\ 1 & x_2 & x_2^2 & \cdots & x_2^m \\ 1 & x_3 & x_3^2 & \cdots & x_3^m \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^m \end{pmatrix}$，$\alpha = \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix}$，$y = \begin{pmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$

称为正则方程组

则原方程组可以表示为$\boxed{A^T A\alpha = A^T y}$

定理1. 设$A_{n\times m}=\begin{pmatrix}1 & x_1 & \cdots & x_1^m\\ 1 & x_2 & \cdots & x_2^m\\ \vdots & \vdots & \ddots & \vdots\\ 1 & x_n & \cdots & x_n^m\end{pmatrix}(n>m)$，若$x_1,x_2,\cdots,x_n$互异，则$A^TAx=A^Ty$有唯一解.

证明：$r(A)\le m$，又A中存在m阶子式$D_m=\begin{vmatrix}1 & x_1 & \cdots & x_1^m\\ 1 & x_2 & \cdots & x_2^m\\ \vdots & \vdots & \ddots & \vdots\\ 1 & x_m & \cdots & x_m^m\end{vmatrix}\neq 0,$

所以$r(A)\ge m$, 故$r(A)=m$.

另外$r(A)=r\left(A^TA\right)\le r\left(A^TA,A^Ty\right)=r\left[A^T\cdot(A,y)\right]\le r\left(A^T\right)=r(A)$，

所以$r\left(A^TA\right)=r\left(A^TA,A^Ty\right)=m$，

故方程组$A^TAx=A^Ty$有唯一解.

例1：已知$f(0)=1, f(1)=2, f(3)=4, f(5)=8$，求直线拟合.

解 设直线方程为$y=a+bx$,

建立方程组$\begin{pmatrix} 4 & \sum_{i=1}^{4} x_i \\ \sum_{i=1}^{4} x_i & \sum_{i=1}^{4} x_i^2 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix}=\begin{pmatrix} \sum_{i=1}^{4} y_i \\ \sum_{i=1}^{4} x_i y_i \end{pmatrix}$，即$\begin{pmatrix} 4 & 9 \\ 9 & 35 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix}=\begin{pmatrix} 15 \\ 45 \end{pmatrix}$

解得$a=\dfrac{39}{59}$，$b=\dfrac{81}{59}$，所以所求拟合直线为：$y=\dfrac{81}{59}x+\dfrac{39}{59}$.

计算各点的函数值分别为$y(0)=\dfrac{39}{59}, y(1)=\dfrac{120}{59}, y(3)=\dfrac{282}{59}, f(5)=\dfrac{444}{59}$

$$\sum_{i=1}^{4}[f(x_i)-y_i)]^2=0.9492.$$

例2：已知函数$y=f(x)$的观测数据为：

x	1	2	3	4
$f(x)$	0	-5	-6	3

用最小二乘法求形如$y=a+bx+cx^2$的数据拟合。

解：建立方程组$\begin{pmatrix} 4 & \sum_{i=1}^4 x_i & \sum_{i=1}^4 x_i^2 \\ \sum_{i=1}^4 x_i & \sum_{i=1}^4 x_i^2 & \sum_{i=1}^4 x_i^3 \\ \sum_{i=1}^4 x_i^2 & \sum_{i=1}^4 x_i^3 & \sum_{i=1}^4 x_i^4 \end{pmatrix}\begin{pmatrix} a \\ b \\ c \end{pmatrix}=\begin{pmatrix} \sum_{i=1}^4 y_i \\ \sum_{i=1}^4 x_i y_i \\ \sum_{i=1}^4 x_i^2 y_i \end{pmatrix}$,

即$\begin{pmatrix} 4 & 10 & 30 \\ 10 & 30 & 100 \\ 30 & 100 & 354 \end{pmatrix}\begin{pmatrix} a \\ b \\ c \end{pmatrix}=\begin{pmatrix} -8 \\ -16 \\ -26 \end{pmatrix}$

解得$a=13.5, b=-16.7, c=3.5$

所以拟合曲线为：$y=13.5-16.7x+3.5x^2$.