回顾

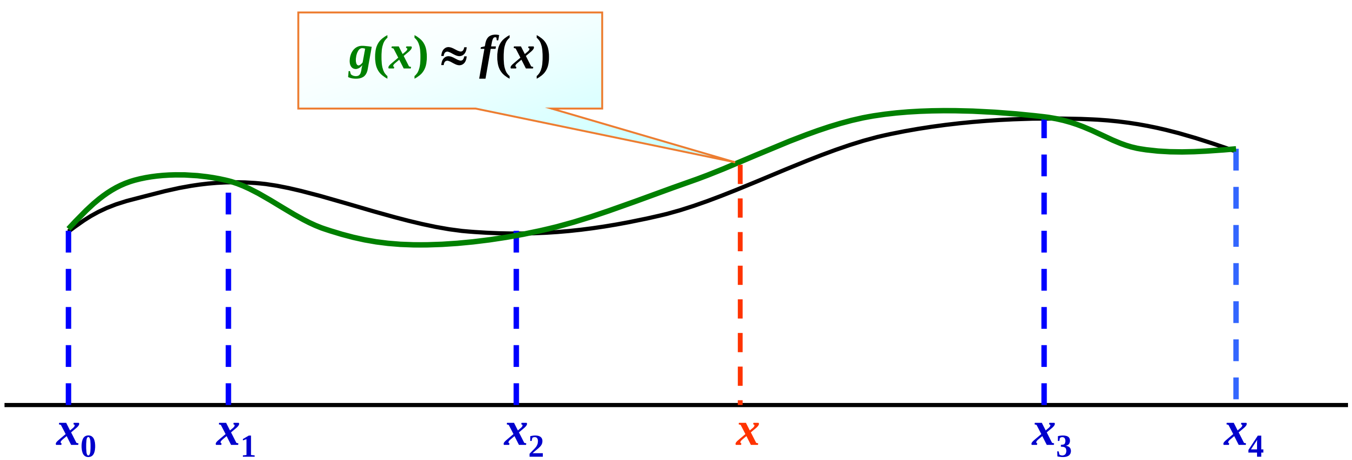
第四章 插值法

4.1 引言
4.2 多项式插值
4.3 拉格朗日插值
4.4 牛顿插值
4.5 埃尔米特插值
4.6 分段插值
4.7 三次样条插值(课本在第五章)

回顾

插值问题定义

- 当函数 $y = f(x)$ 非常复杂或未知时，在一系列节点 $x_0, x_1 \cdots, x_n$ 处测得函数值 $y_0 = f(x_0)$，$\cdots$，$y_n = f(x_n)$，由此构造一个简单易算的近似函数 $g(x) \approx f(x)$，满足条件 $g(x_i) = f(x_i)$ $(i = 0, \cdots n)$。称 $g(x)$ 为 $f(x)$ 的插值函数。



4.2 多项式插值

回顾

取 $\Phi = P_n := \mathrm{span}\left\{1, x, x^2, \cdots, x^n\right\}$，即

$$P_n = \left\{\varphi(x) | \varphi(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n,\ a_i \in \mathbf{R},\ 0 \le i \le n\right\}$$

插值区间　　插值节点

定义4.1　设 $y=f(x)$ 在区间[a,b]上有定义，且已知它在 $n+1$ 个互异点　$a \le x_0 < x_1 \ldots < x_n \le b$ 上的函数值 y_0, $y_1, \ldots, y_n$，若存在一个次数不超过 n 次的多项式

$$p(x) = a_0 + a_1 x + \ldots + a_n x^n$$

满足条件

插值条件

$$p(x_i) = y_i \quad (i = 0, \ldots n)$$

则称 $p(x)$ 为 $f(x)$ 的 n 次插值多项式。

回顾

- 给定n+1个互异点，如何计算Lagrange插值?
- 给定n+1个互异点，如何计算Newton插值?
- Lagrange插值与Newton插值的差别
- 如何应对采样点过多的情况（Runge现象）

第五章 曲线拟合

5.1 引言

5.2 最小二乘拟合曲线

5.3 其他曲线拟合方法

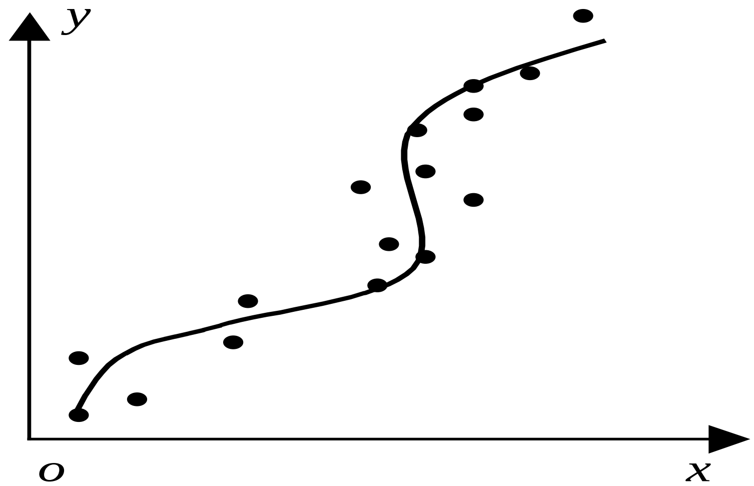
5.1 引言

重要

如果已知函数*f(x)*在若干点$x_i(i=1,2,\cdots,n)$处的值y_i，便可根据插值原理来建立插值多项式作为f(x)的近似，这里的横坐标$\{x_k\}$是明确的。

但在科学实验和生产实践中，往往会遇到这样一种情况，即节点上的函数值并不是很精确的，这些函数值是由实验或观测得到的数据，不可避免地带有测量误差，如果要求所得的近似函数曲线精确无误地通过所有的点(x_i, y_i)，就会使曲线保留着一切测试误差。当个别数据的误差较大时，插值效果显然是不理想的。

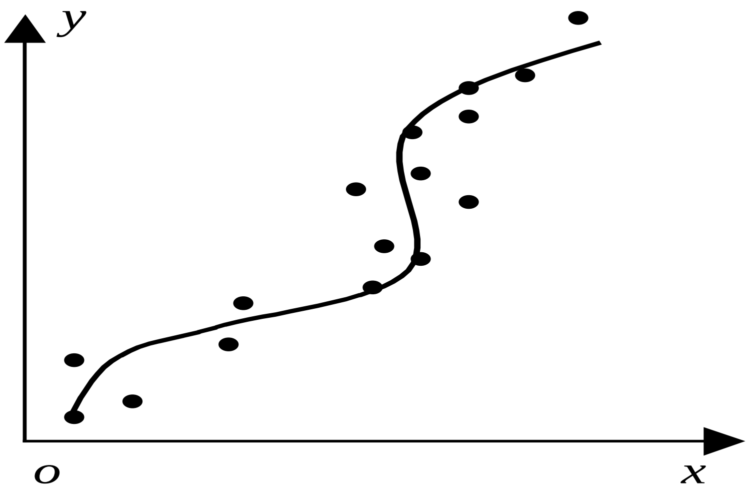
此外，由实验或观测提供的数据个数往往很多，如果用插值法，势必得到次数较高的插值多项式，这样计算起来很烦琐。



为此，就希望由给定的数据(xi, yi)，构造出一个近似函数φ(x)，不要求φ(x)完全通过所有的数据点，只要求所得的近似曲线能反映出数据的基本趋势，如上图。

5.1 引言

实际上，就是求出一条曲线，使数据点均在离此曲线的上方或下方不远处，所求的曲线称为拟合曲线，它既能反映数据的总体分布，又不至于出现局部较大的波动，更能反映被逼近函数的特性，使求得的逼近函数与已知函数从总体上来说，其偏差按某种方法度量达到最小。



非常重要

反映数据的基本关系，更具有实用价值。

拟合与插值的区别

- 插值：过点；(适合精确数据)
- 拟合：不过点，整体近似；(适合有经验公式或有误差的数据)

5.2 最小二乘拟合曲线

通过观测、测量或试验得到如右数值表

x	x_1	x_2	…	x_n
y	y_1	y_2	…	y_n

函数插值是插值函数P(x)与被插函数f(x)在节点处函数值相同，即$P(x_i)=f(x_i)(i=0,1,\ldots,n)$，而曲线拟合函数$\varphi(x)$不要求严格地通过所有数据点$(x_i, y_i)$，也就是说拟合函数$(x_i, y_i)$在$x_i$处的偏差(亦称残差)

$$\varepsilon_i = \varphi(x_i)-f(x_i) \quad i=0,1,\ldots,n$$

不都严格地等于零。

非常重要

但是，为了使近似曲线能尽量反映出所给数据点的变化趋势，要求$|\varepsilon_i|$按某种度量标准最小。若记向量$\varepsilon=[\varepsilon_0, \varepsilon_1,\cdots,\varepsilon_n]$，即要求向量$\varepsilon$的某种范数$\|\varepsilon\|$最小。

向量范数

5.2 最小二乘拟合曲线

常见做法:

$$\|e\|_1 = \sum_{i=0}^{n} |\varepsilon_i| = \sum_{i=0}^{n} |\varphi(x_i) - f(x_i)| \quad \text{最小}$$

太复杂☹

$$\|e\|_\infty = \max_i |\varepsilon_i| = \max_i |\varphi(x_i) - f(x_i)| \quad \text{最小}$$

不可导，
求解困难☹

$$\|e\|_2 = \left(\sum_{i=0}^{n} \varepsilon_i^2 \right)^{\frac{1}{2}} = \left\{ \sum_{i=0}^{n} [\varphi(x_i) - f(x_i)]^2 \right\}^{\frac{1}{2}} \quad \text{最小}$$

即 $$\|e\|_2^2 = \sum_{i=0}^{n} \varepsilon_i^2 = \sum_{i=0}^{n} [\varphi(x_i) - f(x_i)]^2 \quad \text{最小}$$

最小二乘法

非常非常重要！

5.2 最小二乘拟合曲线

(1) 直线拟合

问题 对于给定的数据点(x_i, y_i) $(i=1,2,\ldots,m)$, 求拟合直线

$$y(x)=a_0+a_1x$$

使总误差为最小，即在二元函数式中

$$F(a_0,a_1)=\sum_{i=1}^{m}(a_0+a_1x_i-y_i)^2 \text{ 最小}$$

这里F是关于未知数a_0和a_1的二元函数，这一问题就是要确定a_0和a_1取何值时，二元函数$F(a_0,a_1)$的值最小?

5.2 最小二乘拟合曲线

非常重要

故a_0、a_1应满足下列条件：

$$\begin{cases} \dfrac{\partial F(a_0,a_1)}{\partial a_0} = 2\sum_{i=1}^{m}(a_0 + a_1 x_i - y_i) = 0 \\ \dfrac{\partial F(a_0,a_1)}{\partial a_1} = 2\sum_{i=1}^{m}(a_0 + a_1 x_i - y_i)x_i = 0 \end{cases}$$

即得如下正规方程组

$$\begin{cases} a_0 m + a_1 \sum_{i=1}^{m} x_i = \sum_{i=1}^{m} y_i \\ a_1 \sum_{i=1}^{m} x_i^2 + a_0 \sum_{i=1}^{m} x_i = \sum_{i=1}^{m} x_i y_i \end{cases}$$

要求：会推导或者背过这个公式

求解此正规方程组，得a_0和a_1，即可求出直线拟合方程：$y(x)=a_0+a_1x$

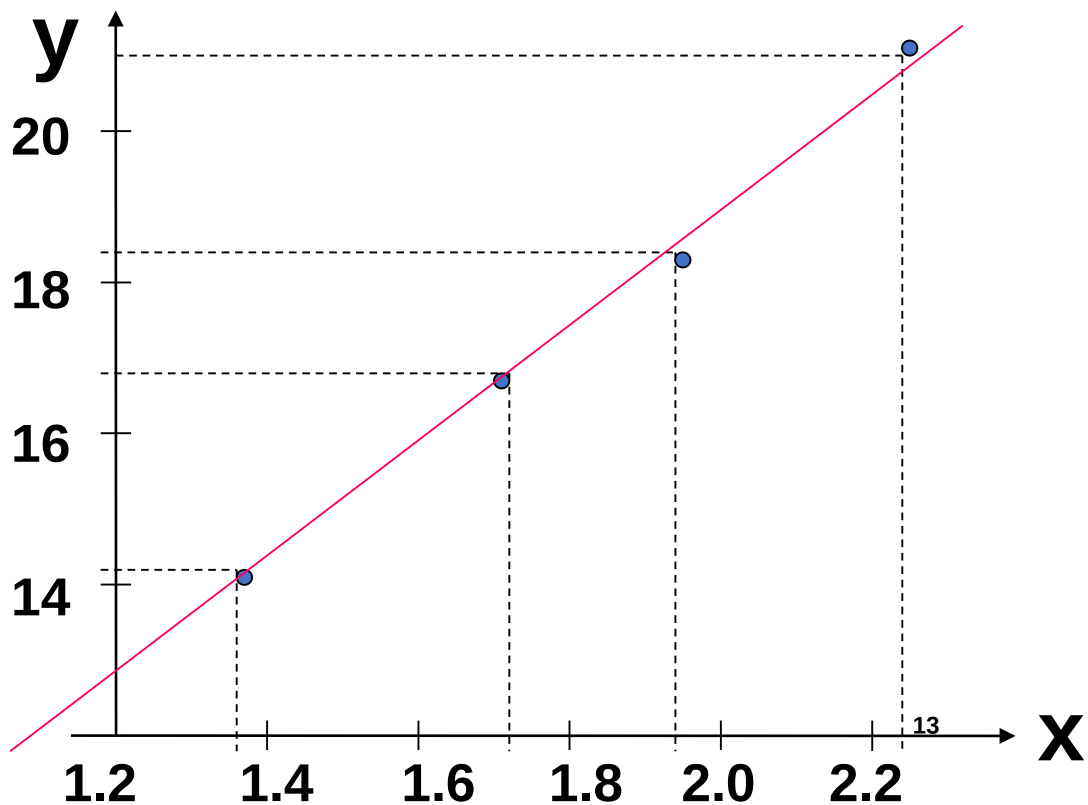
例5.1 设有某实验数据如下：

非常重要

i	1	2	3	4
x_i	1.36	1.73	1.95	2.28
y_i	14.094	16.844	18.475	20.963

用最小二乘法求以上数据的拟合函数。

解：　把表中所给数据画在坐标纸上,将会看到数据点的分布可以用一条直线来近似地描述。



设所求的拟合直线为 $y(x)=a_0+a_1x$，由（x_i, y_i）（$i=1, 2, 3, 4$），则正规方程组为

$$\begin{cases} 4a_0 + a_1\sum_{i=1}^{4} x_i = \sum_{i=1}^{4} y_i \\ a_0\sum_{i=1}^{4} x_i + a_1\sum_{i=1}^{4} x_i^2 = \sum_{i=1}^{4} x_i y_i \end{cases}$$

其中 $\sum_{i=1}^{4} x_i = 7.32$　$\sum_{i=1}^{4} x_i^2 = 13.8434$

$\sum_{i=1}^{4} y_i = 70.376$　$\sum_{i=1}^{4} x_i y_i = 132.12985$

将以上数据代入上式正规方程组, 得

$$\begin{cases} 4a_0 + 7.32a_1 = 70.376 \\ 7.32a_0 + 13.8434a_1 = 132.12985 \end{cases}$$

非常重要

解之，得　$a_0=3.9374, a_1=7.4626$

即得拟合曲线 $y = 3.9374 + 7.4626x$

(2) 多项式拟合

有时所给数据点的分布并不一定近似地呈一条直线，这时仍用直线拟合显然是不合适的，可用多项式拟合。对于给定的一组数据$(x_i,y_i)(i=1,2,\cdots,n)$，寻求次数不超过$m(m<<n)$的多项式$y(x)=a_0+a_1x+a_2x^2+\cdots+a_mx^m$，来拟合所给定的数据，与线性拟合类似，使偏差的平方和

$$Q=\sum_{i=1}^{N}(y_i-\sum_{j=0}^{m}a_jx_i^j)^2$$ **最小**

(2) 多项式拟合

由于Q可以看作是关于(j=0,1,2,···, m)的多元函数, 故上述拟合多项式的构造问题可归结为多元函数的极值问题。令

$$\frac{\partial Q}{\partial a_k}=0 \qquad k=0,1,2,\cdots,m$$

非常重要

得 $$\sum_{i=1}^{N}(y_i-\sum_{j=0}^{m}a_j x_i^j)x_i^k=0 \qquad k=0,1,\cdots,m$$

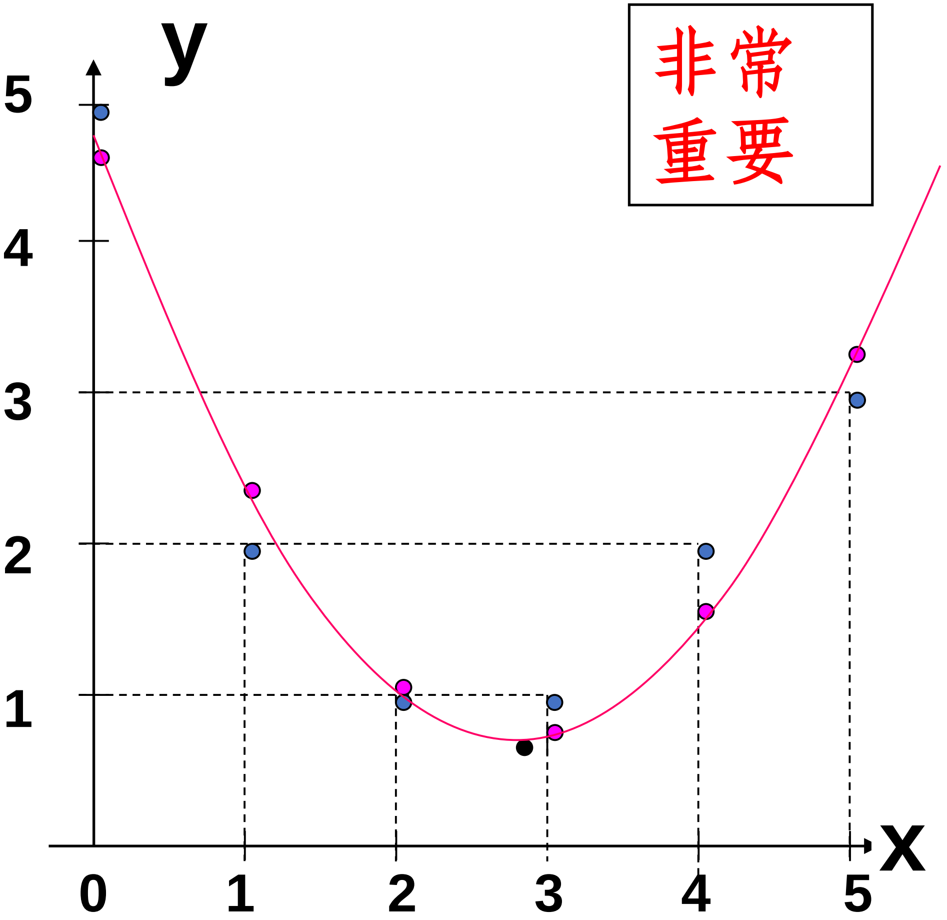
即有

$$\begin{cases} a_0 N + a_1\sum x_i + \cdots + a_m\sum x_i^m = \sum y_i \\ a_0\sum x_i + a_1\sum x_i^2 + \cdots + a_m\sum x_i^{m+1} = \sum x_i y_i \\ \cdots \\ a_0\sum x_i^m + a_1\sum x_i^{m+1} + \cdots + a_m\sum x_i^{2m} = \sum x_i^m y_i \end{cases} \qquad (5.46)$$

这是关于系数a_j的线性方程组，通常称为正规方程组。可以证明，正规方程组有唯一解。

例5.2 设某实验数据为：

i	1	2	3	4	5	6
x_i	0	1	2	3	4	5
y_i	5	2	1	1	2	3



非常
重要

用最小二乘法求一个多项式拟合这组数据 。

解 将已给数据点描在坐标系中，可以看出6个点接近一条抛物线，因此，N=6，设所求的多项式为

$$y(x)=a_0+a_1x+a_2x^2 \qquad \text{即}m=2$$

由N=6，m=2，
其正规方程组为

$$\begin{cases} a_0 * 6 + a_1 \sum x_i + a_2 \sum x_i^2 = \sum y_i \\ a_0 \sum x_i + a_1 \sum x_i^2 + a_2 \sum x_i^3 = \sum x_i y_i \\ a_0 \sum x_i^m + a_1 \sum x_i^3 + a_2 \sum x_i^4 = \sum x_i^2 y_i \end{cases}$$

其中 $\sum_{i=1}^{6} x_i = 15, \sum_{i=1}^{6} x_i^2 = 55, \sum_{i=1}^{6} x_i^3 = 225, \sum_{i=1}^{6} x_i^4 = 797$

$\sum_{i=1}^{6} y_i = 14, \sum_{i=1}^{6} x_i y_i = 30, \sum_{i=1}^{6} x_i^2 y_i = 122$

i	1	2	3	4	5	6
x_i	0	1	2	3	4	5
y_i	5	2	1	1	2	3

代入，则其正规方程组变为

$$\begin{cases} 6a_0 + 15a_1 + 55a_2 = 14 \\ 15a_0 + 55a_1 + 225a_2 = 30 \\ 55a_0 + 225a_1 + 979a_2 = 122 \end{cases}$$

非常重要

解之，得：$a_0=4.7143, a_1=2.7857, a_2=0.5000$

则所求多项式为：$y(x)=4.7143-2.7857x+0.5000x^2$

课堂作业

经调查15个人，他们的体重与身高的数据如下：

身高x（米）	0.75	0.86	0.96	1.08	1.12
体重y (千克)	10	12	15	17	20
身高x（米）	1.26	1.35	1.51	1.55	1.60
体重y (千克)	27	35	41	48	50
身高x（米）	1.63	1.67	1.71	1.78	1.85
体重y (千克)	51	54	59	66	75

试用数据建模的方法建立体重（y）与身高（x）的关系。

（请画出相应的散点图和拟合曲线图）