第3章　最佳逼近和最小二乘法

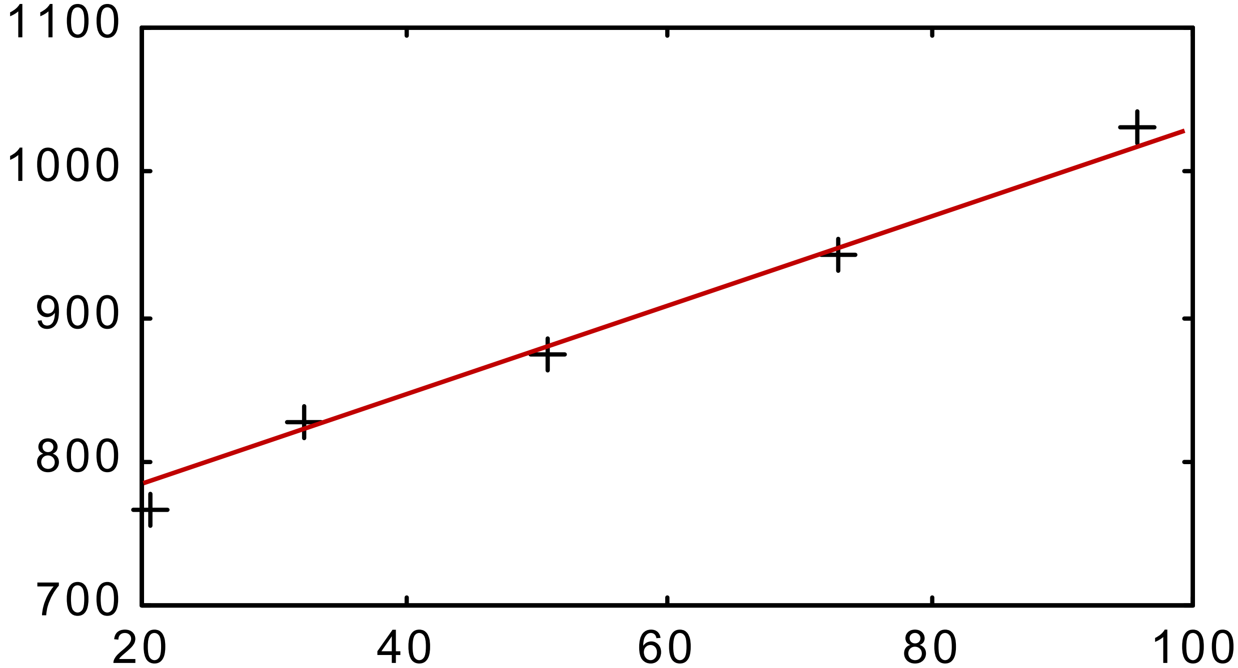
3.5 曲线拟合的最小二乘法

在工程计算和统计分析中，常常遇到大量带有误差的实验数据，需要将这些数据拟合成一条函数曲线. 从而总结、验证变量之间满足的函数关系，称这种问题为曲线拟合. 它要求构造一个不必严格满足所有离散数据的近似函数（曲线），但使逼近的总体误差达到最小. 根据最佳平方逼近准则(2-范数度量误差)进行的曲线拟合方法就称为最小二乘法.

例1 **已知热敏电阻数据：**

温度 $t(^0C)$	20.5	32.7	51.0	73.0	95.7
电阻 $R(\Omega)$	765	826	873	942	1032

求60^0C时的电阻R。



根据数据描绘散点图，数据点近似分布在一条直线上，故设拟合曲线为 $S= a+bt$ ，

a, b为待定系数

求出待定常数a和b，使得计算值S_i与实测值R_i之间的整体绝对误差最小.

比如　误差 $E_1 = \sum_{i=1}^{5} \left| (a + bt_i) - R_i \right|$

误差 $E_2 = \sqrt{\sum_{i=1}^{5} [(a + bt_i) - R_i]^2}$

最小

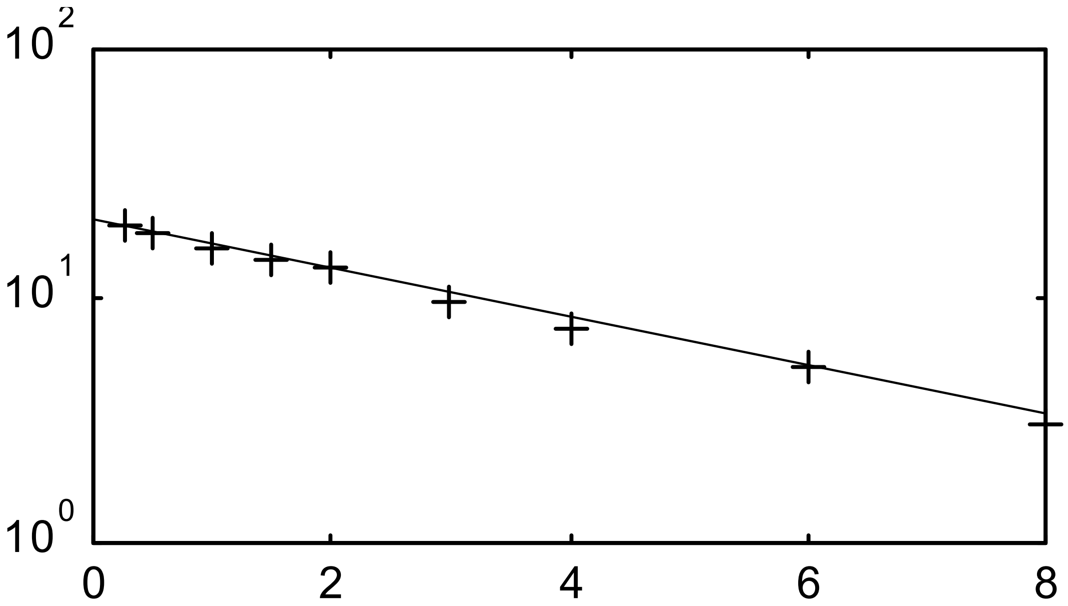
用E_2度量方法求出拟合曲线 $S=\boldsymbol{a}+b\,\boldsymbol{t}$ 称为一次最小二乘拟合或一次最佳平方逼近.

例2 **一种药物快速静脉注射下的血药浓度数据(t =0注射300mg)**

t (h)	0.25	0.5	1	1.5	2	3	4	6	8
c (μg/ml)	19.21	18.15	15.36	14.10	12.89	9.32	7.45	5.24	3.01

求血药浓度随时间的变化规律 $c(t)$.

作半对数坐标系(semilogy)下的图形



根据数据描绘散点图，设拟合曲线为

$$c(t) = c_0 e^{-kt}$$

(c_0, k为待定系数)

$$\Leftrightarrow s(t) = \ln c(t) = \ln c_0 - kt$$

即 $s(t) = a + bt$

问题的描述：已知离散数据表$(x_i, f(x_i))_{i=1\sim m}$，选取逼近表格

函数$f(x)$的函数类ϕ，求拟合函数$S^*(x)\in\phi$，使得

$$\left\|\left\{f(x_i)-S^*(x_i)\right\}_{i=1\sim m}\right\| = \min_{S(x)\in\phi}\left\|\left\{f(x_i)-S(x_i)\right\}_{i=1\sim m}\right\|$$

由于 $\delta=\{\delta_i\}\triangleq\{f(x_i)-S(x_i)\}\in R^m$,

$$\delta^*=\{\delta_i^*\}\triangleq\{f(x_i)-S^*(x_i)\}\in R^m,$$

问题即是在$\boldsymbol{m}$维向量空间R^m中的最佳逼近问题.

取2-范数时，该方法称为最小二乘法(最佳平方逼近)；

取∞-范数时，该方法称为最佳一致逼近法.

离散数据的最小二乘拟合

定义函数的离散加权内积

$$(f,\ g)=\sum_{i=1}^{m}\omega_i f(x_i)g(x_i)$$

其中ω_i为$f(x)$在点x_i处的权，$\omega_i\geq 0$，$i=1,2,\cdots,m.$

内积诱导的2-范数 $\|f\|_2=\sqrt{\sum_{i=1}^{m}\omega_i f^2(x_i)}$

注：ω_i可以表示在$(x_i,f(x_i))$处重复观测的次数.

定义 设$\{\varphi_1(x),\varphi_2(x),\cdots,\varphi_n(x)\}$是$C[a,b]$中的$n$个线性无关的函数，子空间$M=\mathrm{span}\{\varphi_1,\varphi_2,\cdots,\varphi_n\}$， $S(x)=\sum_{i=1}^{n}\alpha_i\varphi_i(x)\in M$.

已知$f(x)\in C[a,b]$在m个互异点$x_i(i=1\sim m)$测试数据$(x_i,f(x_i))_{i=1\sim m}$，

求$S^*(x)=\sum_{i=1}^{n}\alpha_i^*\varphi_i(x)\in M$，

s.t. $$\sqrt{\sum_{i=1}^{m}\omega_i[f(x_i)-S^*(x_i)]^2}=\min_{S(x)\in M}\sqrt{\sum_{i=1}^{m}\omega_i[f(x_i)-S(x_i)]^2}$$

即 $$\sum_{i=1}^{m}\omega_i[f(x_i)-\sum_{j=1}^{n}\alpha_j^*\varphi_j(x_i)]^2=\min_{\{\alpha_i\}\in K^m}\sum_{i=1}^{m}\omega_i[f(x_i)-\sum_{j=1}^{n}\alpha_j\varphi_j(x_i)]^2$$

称$s^*(x)$为$f(x)$在M中的加权最小二乘拟合函数，

求$s^*(x)$的方法称为最小二乘法.

由法方程

$$\begin{bmatrix} (\varphi_1,\varphi_1) & (\varphi_2,\varphi_1) & \cdots & (\varphi_n,\varphi_1) \\ (\varphi_1,\varphi_2) & (\varphi_2,\varphi_2) & \cdots & (\varphi_n,\varphi_2) \\ \vdots & \vdots & \ddots & \vdots \\ (\varphi_1,\varphi_n) & (\varphi_2,\varphi_n) & \cdots & (\varphi_n,\varphi_n) \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} = \begin{bmatrix} (f,\varphi_1) \\ (f,\varphi_2) \\ \vdots \\ (f,\varphi_n) \end{bmatrix}$$

求出 $(\alpha_1^*,\cdots,\alpha_j^*,\cdots,\alpha_n^*)^T \Rightarrow s^*(x) = \sum_{j=1}^{n} \alpha_j^* \varphi_j(x)$

当 $f(x) \approx s^*(x)$时，

均方误差 $\|\delta\|_2 = \sqrt{\sum_{i=1}^{m} \omega(x_i)[f(x_i) - s^*(x_i)]^2}$

$$= \sqrt{(f，f) - \sum_{i=1}^{n} \alpha_i^* (f,\varphi_i)}$$

例 3. 已知数据组

x_i	0.2	0.5	0.7	0.85	1
y_i	1.221	1.649	2.014	2.340	2.718

，试用最小二乘法求 $f(x)$ 的二次拟合多项式 $P_2(x)=a_0+a_1x+a_2x^2$.

解 权 $\omega_i=1$, 取 $M=\mathrm{span}\{1,x,x^2\}\triangleq\mathrm{span}\{\varphi_0,\varphi_1,\varphi_2\}$，计算

$$(\varphi_0,\varphi_0)=\sum_{i=1}^{5}1\times 1=5,\qquad (\varphi_1,\varphi_0)=\sum_{i=1}^{5}x_i\times 1=3.250,$$

$$(\varphi_2,\varphi_0)=\sum_{i=1}^{5}x_i^2\times 1=2.503,\qquad (\varphi_1,\varphi_1)=\sum_{i=1}^{5}x_i\times x_i=2.503,$$

$$(\varphi_2,\varphi_1)=\sum_{i=1}^{5}x_i^2\times x_i=2.090,\qquad (\varphi_2,\varphi_2)=\sum_{i=1}^{5}x_i^2\times x_i^2=1.826,$$

$$(f,\varphi_0)=\sum_{i=1}^{5}y_i\times 1=9.942,\qquad (f,\varphi_1)=\sum_{i=1}^{5}y_i\times x_i=7.185,$$

$$(f,\varphi_2)=\sum_{i=1}^{5}y_i\times x_i^2=5.857.$$

$$法方程\begin{bmatrix} 5 & 3.250 & 2.503 \\ 3.250 & 2.503 & 2.090 \\ 2.503 & 2.090 & 1.826 \end{bmatrix}\begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 9.942 \\ 7.185 \\ 5.857 \end{bmatrix}.$$

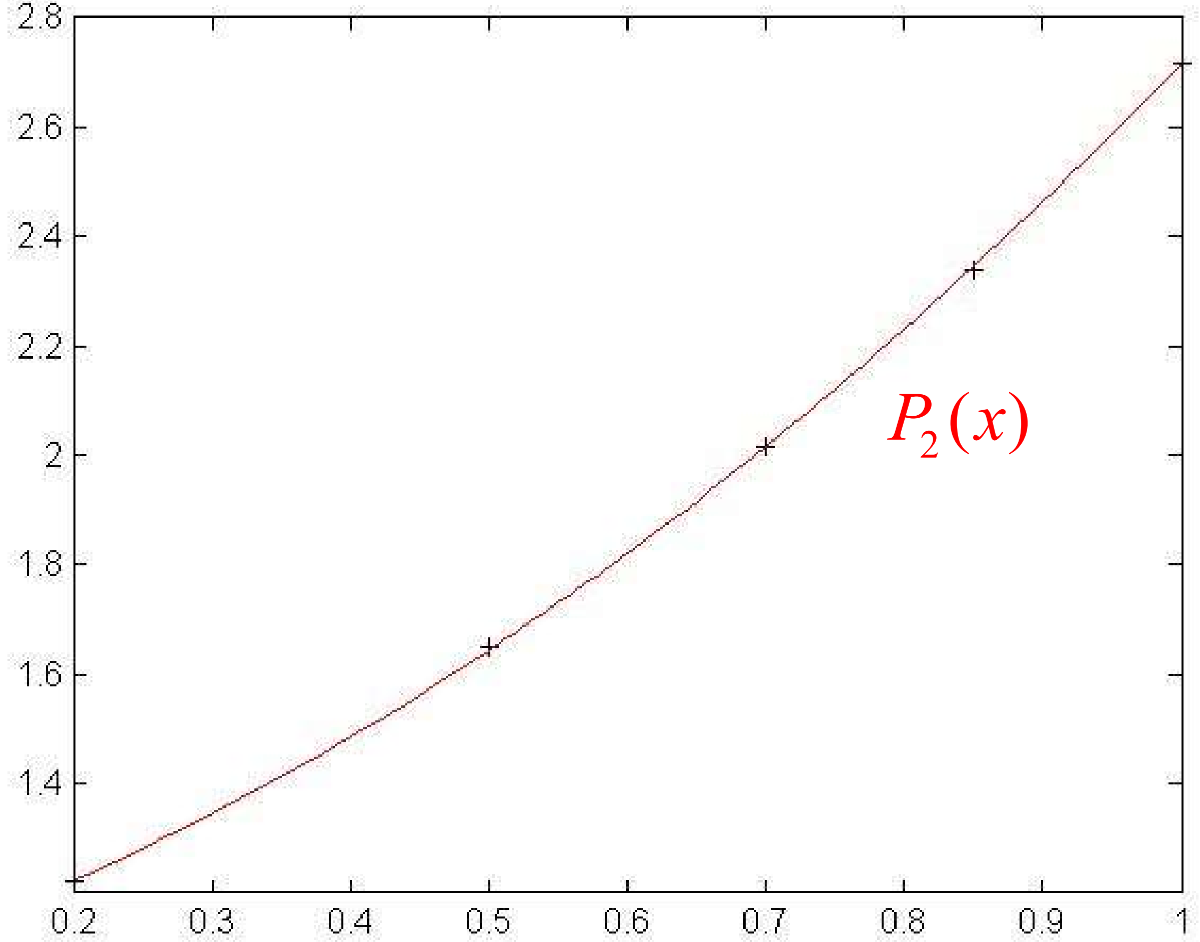
求得 $a_0 = 1.036,\ a_1 = 0.751,\ a_2 = 0.928.$

故 $f(x) \approx P_2(x) = 1.036 + 0.751x + 0.928x^2$

均方误差 $\|\delta\|_2 = \sqrt{\sum_{i=1}^{5}[f(x_i) - s^*(x_i)]^2} \approx 0.0086$

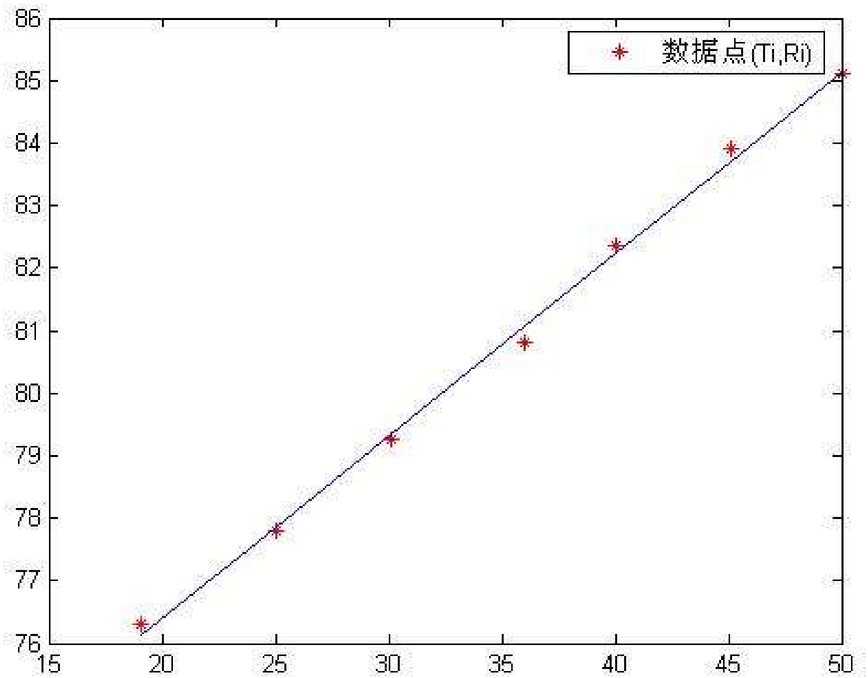
均匀误差 $\|\delta\|_\infty = \max_{1\le i\le 5}\left|f(x_i) - s^*(x_i)\right| \approx 0.0055$

$$P_2(x) = 1.036 + 0.751x + 0.928x^2$$



例4 已知测得铜导线在温度t_i时的电阻R_i,求电阻R与温度t的关系.

i	1	2	3	4	5	6	7
T_i	19.1	25.0	30.1	36.0	40.0	45.1	50.0
R_i	76.30	77.80	79.25	80.80	82.35	83.90	85.10



设各点的权 $\omega_i=\dfrac{1}{7}\,(i=1\sim7)$

根据数据描绘散点图,
拟合曲线近似为

$$S=a+bt$$

i	1	2	3	4	5	6	7
t_i	19.1	25.0	30.1	36.0	40.0	45.1	50.0
R_i	76.30	77.80	79.25	80.80	82.35	83.90	85.10

解 令 $\varphi_0=1$，$\varphi_1=t$，权 $\omega_i=\dfrac{1}{7}\,(i=1,2,\cdots,7)$

$$(\varphi_0,\varphi_0)=\frac{1}{7}\sum_{i=1}^{7}1\times 1=1,\ (\varphi_1,\varphi_0)=\frac{1}{7}\sum_{i=1}^{7}t_i\times 1=\frac{245.3}{7},\quad (\varphi_1,\varphi_1)=\frac{1}{7}\sum_{i=1}^{7}t_i\times t_i=\frac{9325.83}{7},$$

$$(R,\varphi_0)=\frac{1}{7}\sum_{i=1}^{7}R_i=\frac{566.5}{7},\quad (R,\varphi_1)=\frac{1}{7}\sum_{i=1}^{7}R_i\times t_i=\frac{20029.445}{7}$$

法方程 $\begin{cases}7a+245.3b=566.5\\ 245.3a+9325.83b=20029.445\end{cases}\Rightarrow R\approx 70.572+0.291t$

问： 若权 $\omega_i=1\,(i=1,2,\cdots,7)$，计算结果如何？

用最小二乘法解决实际问题的步骤：

(1) 由实验数据表，画出表格函数的粗略图形——散点图；

(2) 根据离散点的图形，选取合适的拟合函数类型 Φ，

(3) 用最佳逼近方法求出拟合函数 $\varphi(x)\in\Phi$；

(4) 通过实验结果验证拟合曲线的吻合性.

例5 已知一组实验数据(t_i, y_i)如表，用适当的函数对它们进行拟合.

t_i	1.00	1.25	1.50	1.75	2.00
y_i	5.10	5.79	6.53	7.45	8.46
$\tilde{y}_i = \ln y_i$	1.6292	1.7561	1.8764	2.0082	2.1353

解 根据数据点分布的散点图，选取拟合函数为指数函数.

设 $y \approx Ae^{bt} \Leftrightarrow \ln y \approx \ln A + bt$，记 $\tilde{y} = \ln y,\ a = \ln A$，

则只需求出拟合函数 $\tilde{y} = a + bt$.

应用最小二乘法，解相应的法方程，求得

$a = 1.1225$，$b = 0.5057 \Rightarrow A = e^a = 3.0725.$

所求数据的拟合曲线为 $y \approx 3.0725e^{0.5057t}$

练习.

已知一组试验数据，其中ρ_i为各数据点出现的次数，试用最小二乘法求这些数据的拟合曲线.

t_i	1	2	3	4	5
f_i	4	4.5	6	8	8.5
ρ_i	2	1	3	1	1

答案： 所求拟合曲线为 $S^*(t)=2.5648+1.2037\,t$

数值分析作业 4：

此次作业使用 python 语言编写代码完成，计算了普通最小二乘和加权最小二乘并作图加以比较。

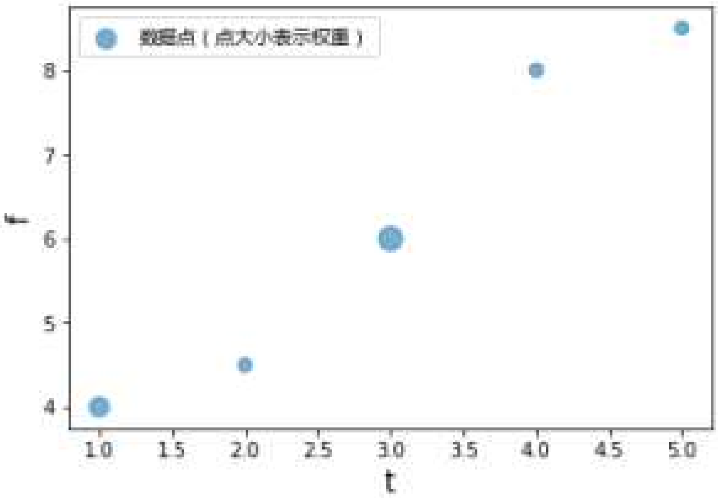
代码展示：

```python
import numpy as np
import matplotlib.pyplot as plt
import matplotlib.font_manager as fm
myfont = fm.FontProperties(fname='C:/Windows/Fonts/msyh.ttc') ##解决图乱码
%matplotlib inline
from sympy import *

```

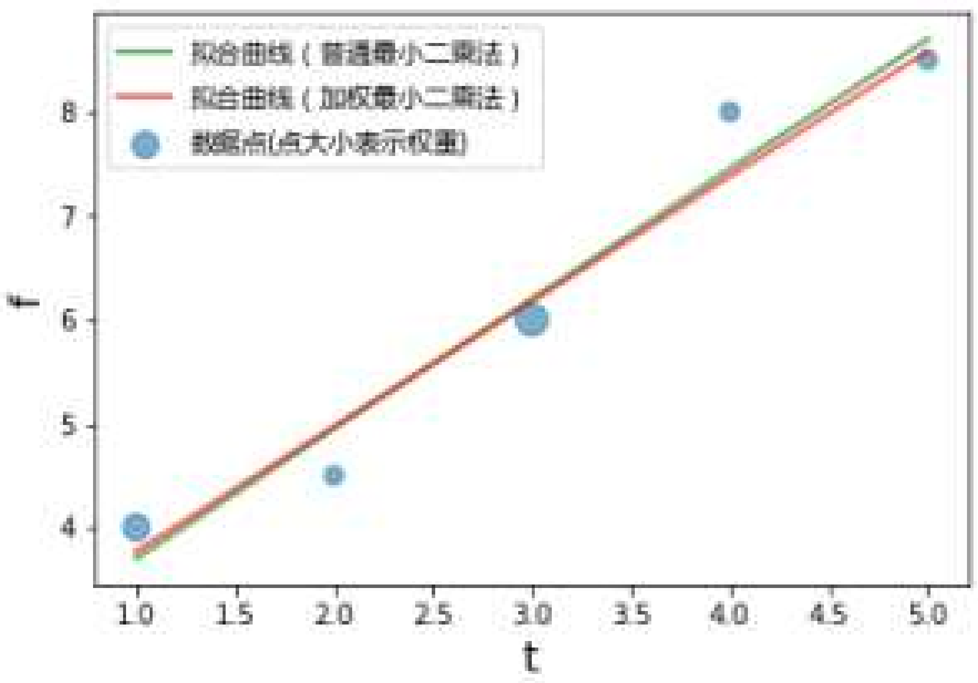
结果展示：

画出表格函数的粗略图形：



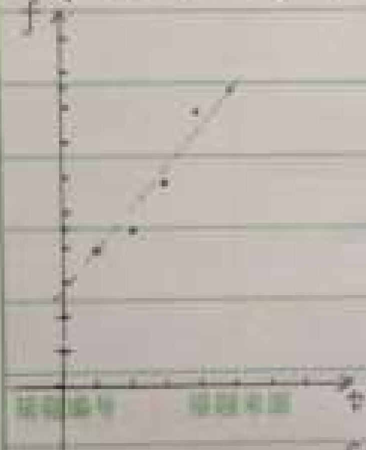
```
-------------------------------------
普通最小二乘法求出的参数：
a_noweight=2.45000000000000
b_noweight=1.25000000000000
普通最小二乘法求出的拟合曲线为：
s(t)=2.4500+1.2500*t
-------------------------------------
加权最小二乘法求出的参数：
a_weight=2.56481481481481
b_weight=1.20370370370370
加权最小二乘法求出的拟合曲线为：
s(t)=2.5648+1.2037*t
-------------------------------------
```

数据点与两条拟合曲线对比图如下：



(4). 已知一组试验数据，ρ_i为各数据点的权数
试用最小二乘法求这些数据的拟合曲线

t_i	1	2	3	4	5
f_i	4	4.5	6	8	8.5
ρ_i	2	1	3	1	1



解：根据数据作出散点图

假拟合曲线为 $S=a+bt$

令 $\varphi_0=1$，$\varphi_1=t$，权为 ρ_i $(i=1,2,3,4,5)$

则有 $(\varphi_0,\varphi_0)=\sum_{i=1}^{5}\rho_i=8$

$(\varphi_0,\varphi_1)=\sum_{i=1}^{5}\rho_i t_i=2+2+9+4+5=22$

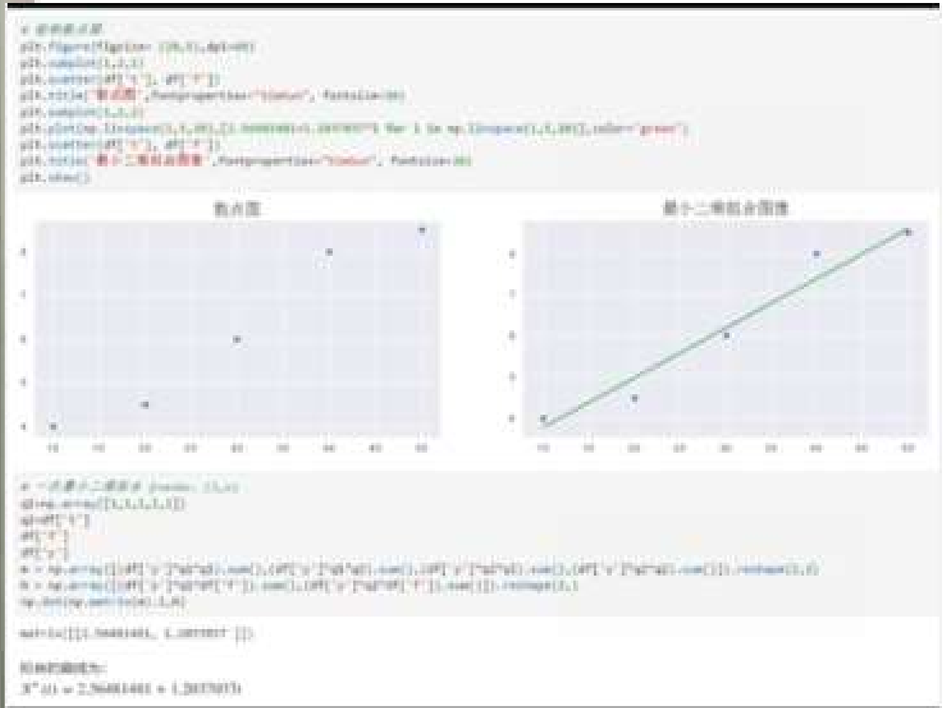
$(\varphi_1,\varphi_1)=\sum_{i=1}^{5}\rho_i t_i^2=2+4+27+16+5=74$

$(f,\varphi_0)=\sum_{i=1}^{5}\rho_i f_i=8+4.5+18+8+8.5=47$

$(f,\varphi_1)=\sum_{i=1}^{5}\rho_i f_i t_i=8+9+54+32+42.5=145.5$

法方程 $\begin{bmatrix}8 & 22\\ 22 & 74\end{bmatrix}\begin{bmatrix}\alpha_0\\ \alpha_1\end{bmatrix}=\begin{bmatrix}47\\ 145.5\end{bmatrix}$ 解得 $\begin{cases}\alpha_1=1.20370\\ \alpha_0=2.56481\end{cases}$

故得所求拟合曲线为 $S=2.56481+1.20370t$



由观察可知，拟合曲线近似认为直线（一次）.

设 $S^*(t)=\alpha_0^*+\alpha_1^* t$.　　$\varphi_0=1$,　$\varphi_1=t$.

权分别为：$\frac{1}{4}$, $\frac{1}{8}$, $\frac{3}{8}$, $\frac{1}{8}$, $\frac{1}{8}$

$(\varphi_0,\varphi_0)=1$　　$(\varphi_0,\varphi_1)=(\varphi_1,\varphi_0)=2.75$

$(\varphi_1,\varphi_1)=9.25$.　$(f,\varphi_0)=5.875$　$(f,\varphi_1)=18.1875$.

$$\therefore \begin{pmatrix} 1 & 2.75 \\ 2.75 & 9.25 \end{pmatrix} \begin{pmatrix} \alpha_0^* \\ \alpha_1^* \end{pmatrix} = \begin{pmatrix} 5.875 \\ 18.1875 \end{pmatrix}$$

解得：
$$\begin{pmatrix} \alpha_0^* \\ \alpha_1^* \end{pmatrix} = \begin{pmatrix} 2.5648 \\ 1.2037 \end{pmatrix}$$

$$\therefore S^*(t)=2.5648+1.2037t$$

拟合曲线类型的选择依据

- 专业知识和经验；
- 离散点图的分布形状及特点；
- 反复计算比较误差、修正类型；
- 现有的自动选择数学模型的程序.

法方程的病态问题

- 在最小二乘拟合中，若取基函数为 $\varphi_j(x)=x^j\,(j=0,1,\cdots,n)$.

当 n 较大时，法方程组是病态的.

通常采用离散点的正交多项式可避免法方程出现病态.

- 实际计算中也常采用低次多项式分段拟合.

例如

若子空间$M=\operatorname{span}\{\varphi_0,\varphi_1,\cdots,\varphi_n\}$中的基$\varphi_i$ 关于数据点

(x_i,y_i) $(i=0,1,2,\cdots,m)$加权$\omega(x_i)$正交，即

$$(\varphi_j,\varphi_k)=\sum_{i=0}^{m}\omega(x_i)\varphi_j(x_i)\varphi_k(x_i)=\begin{cases}0, & j\neq k\\ A_k>0, & j=k\end{cases}$$

则法方程的解

$$a_k^*=\frac{(f,\varphi_k)}{(\varphi_k,\varphi_k)}=\frac{\sum_{i=0}^{m}\omega(x_i)f(x_i)\varphi_k(x_i)}{\sum_{i=0}^{m}\omega(x_i)\varphi_k^2(x_i)}\quad(k=0,1,\cdots,n)$$

均方误差 $\|\delta\|_2=\|f-s^*\|_2=\sqrt{\|f\|_2^2-\sum_{k=0}^{n}A_k(a_k^*)^2}$

构造离散点的正交多项式方法

根据给定的节点$x_0,x_1,\cdots,x_m$ 及权函数$\omega(x)\geq 0$，构造出带权$\omega(x)$的正交多项式 $P_0(x),\cdots,P_n(x)$ $(n\leq m)$.

（由G.-S正交化方法得到）

$$\begin{cases} P_0(x)=1, \\ P_1(x)=(x-\alpha_1)P_0(x), \\ P_{k+1}(x)=(x-\alpha_{k+1})P_k(x)-\beta_k P_{k-1}(x) \quad (k=1,2,\cdots,n-1). \end{cases}$$

$$\begin{cases} \alpha_{k+1}=\dfrac{\sum\limits_{i=0}^{m}\omega(x_i)x_iP_k^2(x_i)}{\sum\limits_{i=0}^{m}\omega(x_i)P_k^2(x_i)}=\dfrac{(xP_k,P_k)}{(P_k,P_k)}, \\ \beta_k=\dfrac{\sum\limits_{i=0}^{m}\omega(x_i)P_k^2(x_i)}{\sum\limits_{i=0}^{m}\omega(x_i)P_{k-1}^2(x_i)}=\dfrac{(P_k,P_k)}{(P_{k-1},P_{k-1})} \end{cases} \quad (k=0,1,\cdots,n-1)$$

若取$M=\operatorname{span}\{P_0,P_1,\cdots,P_n\}$，则 $f(x)$在M中最小二乘拟合函数为

$$s^*(x)=\sum_{k=0}^{n}\alpha_k^*P_k(x)$$

其中 $$\alpha_k^*=\frac{(f,P_k)}{(P_k,P_k)}=\frac{\sum_{i=0}^{m}\omega(x_i)f(x_i)P_k(x_i)}{\sum_{i=0}^{m}\omega(x_i)P_k^2(x_i)}.$$

此方法求最小二乘函数不用解方程组，计算简单、稳定.

2.6 $C[a,b]$中的最佳一致逼近

以2－范数作为度量误差的标准，研究了函数空间$C[a,\mathrm{b}]$中的最佳逼近及离散情况下的最小二乘法，由此得到的最佳逼近函数$S^*(x)$，误差$\|\delta\|_2$反应了$f(x)$的平均误差，但可能出现在某些点上的误差大于平均误差的情况。

为了解决这类问题，考虑用∞－范数作为度量误差的标准.

定义 设 $f(x)\in C[a,b]$, $M_n=\mathrm{span}\left\{1,x,x^2,\cdots x^n\right\}$,

若存在多项式函数$P_n^*(x)\in M_n$,使得

$$\left\|f(x)-P_n^*(x)\right\|_\infty=\min_{P_n(x)\in M_n}\left\|f(x)-P_n(x)\right\|_\infty$$

即 $$\max_{x\in[a,b]}\left|f(x)-P_n^*(x)\right|=\min_{P_n(x)\in M_n}\max_{x\in[a,b]}\left|f(x)-P_n(x)\right|$$

则称$P_n^*(x)$是$f(x)$在$[a,b]$上的**最佳一致逼近多项式**

或切比雪夫逼近多项式.

连续函数的一次最佳逼近多项式

设 $f(x)\in C^2[a,b]$，且 $f''(x)$在(a,b)内不变号, $M_1=\mathrm{span}\{1,x\}$，则$f(x)$在M_1中的最佳一致逼近多项式$P_1(x)=a_0+a_1x$，即是连续函数$f(x)$的一次最佳一致逼近多项式.

记 $x_1=a$, $x_3=b$, $f'(x)-a_1=0$的解记作x_2. 解方程租

$$\begin{cases} a_0+a_1a-f(a)=a_0+a_1b-f(b) \\ a_0+a_1a-f(a)=f(x_2)-(a_0+a_1x_2) \end{cases}$$

$$\Rightarrow \quad a_1=\frac{f(b)-f(a)}{b-a}=f'(x_2) \quad \text{由微分中值定理得}$$

$$a_0=\frac{1}{2}[f(a)+f(x_2)]-\frac{f(b)-f(a)}{b-a}\cdot\frac{a+x_2}{2}.$$

故$f(x)\approx P_1(x)=a_0+a_1x$，均匀误差 $R(x)=\max\limits_{x\in[a,b]}\left|f(x)-P_1(x)\right|\le ?$

内容小结

内积空间中的最佳逼近

1. 在$C[a,b]$中 $\begin{cases} M = \mathrm{span}\{1, x, x^2, \cdots, x^n\} \\ M = \mathrm{span}\{\text{正交多项式}\} \end{cases}$

勒让德多项式
切比雪夫多项式

最佳平方逼近

$$\|\delta\|_2^2 = \left\|f(x) - S^*(x)\right\|_2^2 = \min_{S(x)\in M} \int_a^b p(x)[f(x) - S(x)]^2 dx$$

2. 在R^m中 $M = \mathrm{span}\{x_1, x_2, \cdots, x_n\}$

最小二乘法 $\|\delta\|_2^2 = \min_{S(x)\in M} \sum_{i=1}^{n} [f(x_i) - S(x_i)]^2$

3. 在 C[a,b] 中，最佳一致逼近

$$\|\delta\|_\infty = \left\|f(x) - P_n^*(x)\right\|_\infty = \min_{P_n(x)\in M_n} \left\|f(x) - P_n(x)\right\|_\infty$$

作业　　**P85　习题3**

1, 2, 3, 5, 7, 8, 10, 11, 12, 13