数据拟合

数据拟合的基本概念

数据拟合的线性模型

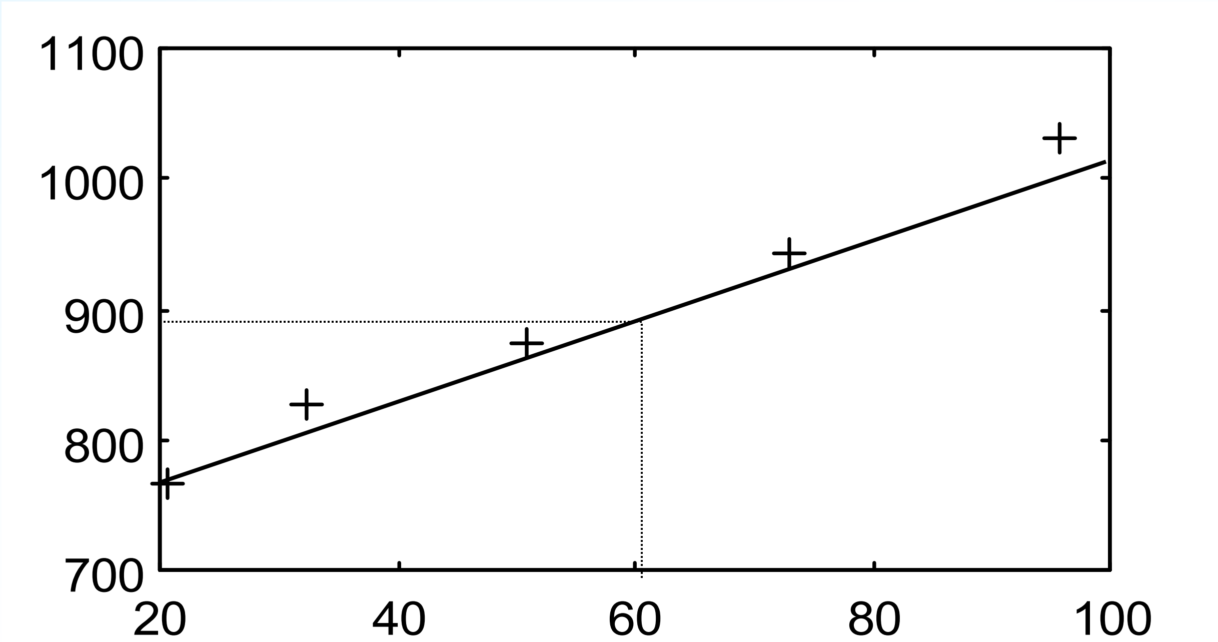
数据拟合的非线性模型

拟 合 问 题 引 例 1

已知热敏电阻数据：

温度t(^{0}C)	20.5	32.7	51.0	73.0	95.7
电阻R(Ω)	765	826	873	942	1032

求60^{0}C时的电阻R。

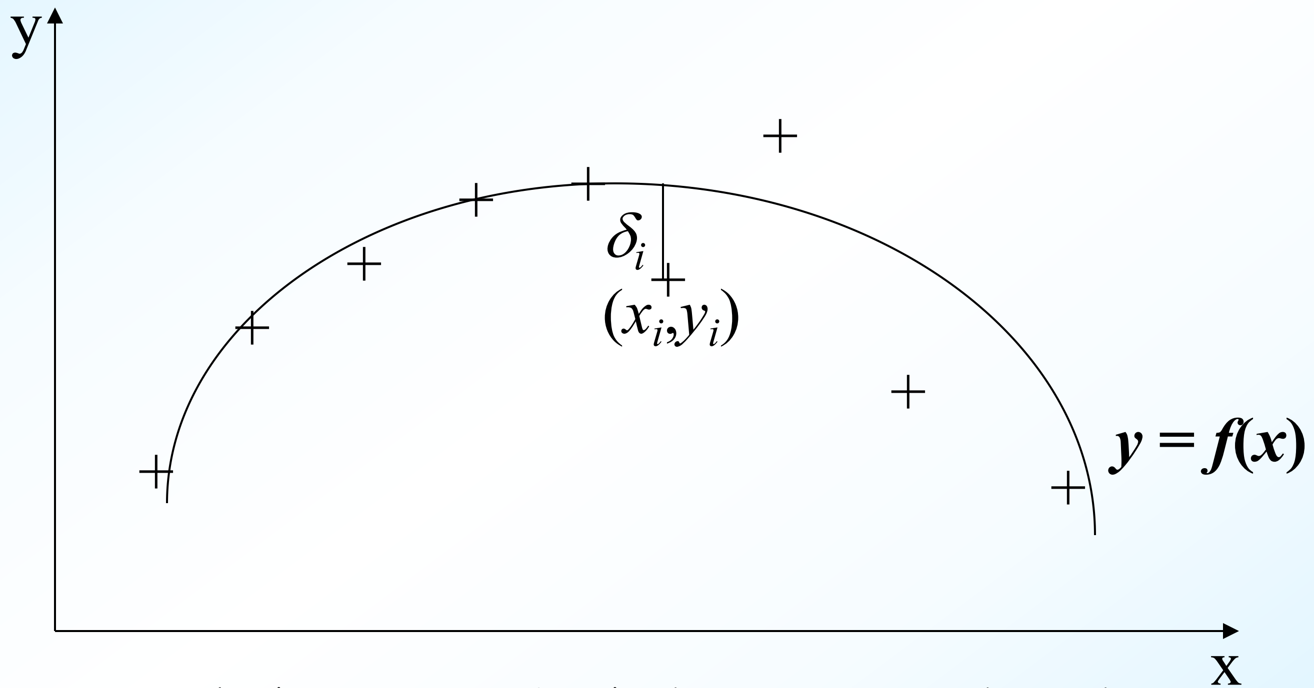


设 $R=at+b$

a,b为待定系数

曲 线 拟 合 问 题 的 提 法

已知一组（二维）数据，即平面上 n个点（x_i,y_i） $i=1,\ldots,n$，寻求一个函数（曲线）$y=f(x)$，使 $f(x)$ 在某种准则下与所有数据点最为接近，即曲线拟合得最好。



δ_i 为点（x_i,y_i) 与曲线 $y=f(x)$ 的距离

拟合与插值的关系

问题：给定一批数据点，需确定满足特定要求的曲线或曲面

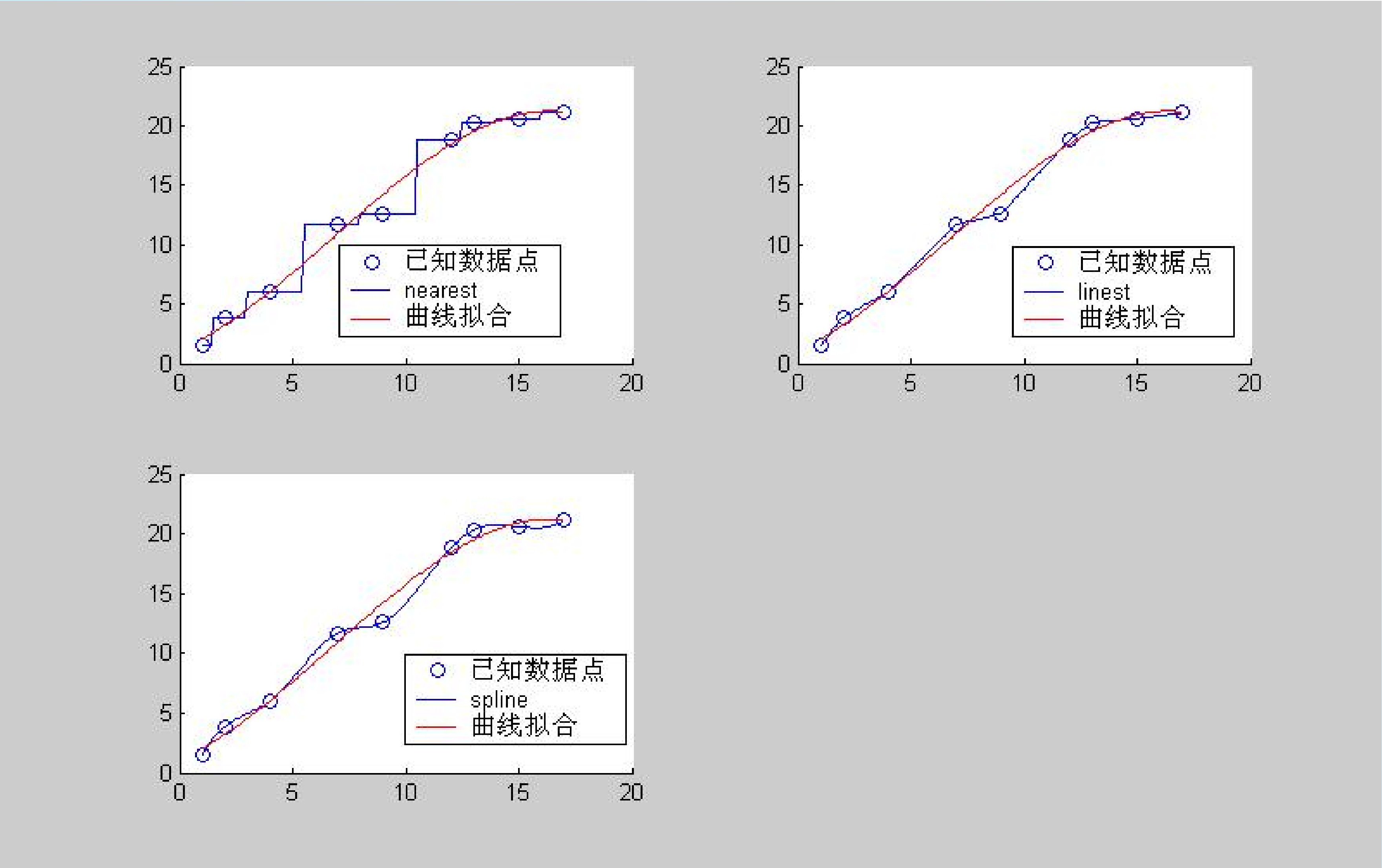
解决方案：

•若要求所求曲线（面）通过所给所有数据点，就是插值问题;

•若不要求曲线（面）通过所有数据点，而是要求它反映对象整体的变化趋势，这就是数据拟合，又称曲线拟合或曲面拟合。

函数插值与曲线拟合都是要根据一组数据构造一个函数作为近似，由于近似的要求不同，二者的数学方法上是完全不同的。

最临近插值、线性插值、样条插值与曲线拟合结果：

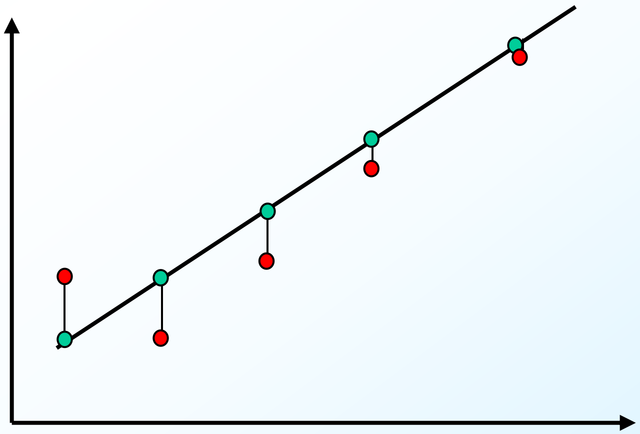


离散数据的线性拟合

已知数据表

x	x_1	x_2	……	x_m
$f(x)$	y_1	y_2	……	y_m

求拟合函数：$\varphi(x) = a + b\,x$



$$\begin{cases} a + b\,x_1 = y_1 \\ a + b\,x_2 = y_2 \\ \cdots\cdots \\ a + b\,x_m = y_m \end{cases} \rightarrow \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} \rightarrow GX=F$$

超定方程组

残差: $r = GX - F$

最小二乘问题 $\min\limits_{X \in R2} \| GX - F \|_2$

$$\| GX - F \|_2^2 = (GX - F, GX - F)$$

$$= (GX, GX) - 2(GX, F) + (F, F)$$

$$= (X, G^T GX) - 2(X, G^T F) + (F, F)$$

$$f(X) = (X, G^T GX) - 2(X, G^T F)$$

求解极值问题:

$$\min_{X \in R^2} f(X) = \min_{X \in R^2} [(X, G^T GX) - 2(X, G^T F)]$$

设 X^* 是函数 $f(X)$ 的极值点,任意 $e\in R^2$

$$f(X^*+te)=(X^*+te, G^TG(X^*+te)) - 2(X^*+te, G^TF)$$

$$g(t)=f(X^*+te) = f(X^*)+2t(e, G^T(GX^*-F))+t^2(e, G^TGe)$$

$$g'(0)=0 \rightarrow (e, G^T(GX^*-F))=0 \rightarrow$$

$$(e, G^T(GX^*-F)) = 0 \rightarrow G^T(GX^*-F) = 0$$

$$G^TGX^*-G^TF = 0 \quad \rightarrow \quad G^TGX^*=G^TF$$

令 $A = G^TG, B = G^TF \rightarrow AX^* =B$ 对称正定方程组

超定方程组最小二乘解 X^* 的几何意义

例 $\begin{cases} 2x+4y=11 \\ 3x-5y=3 \\ x+2y=6 \end{cases}$ $\quad G=\begin{bmatrix} 2 & 4 \\ 3 & -5 \\ 1 & 2 \end{bmatrix} \quad \beta=\begin{bmatrix} 11 \\ 3 \\ 6 \end{bmatrix}$

$$\| \beta - GX^* \| = \min_{X \in R2} \| \beta - GX \|_2$$

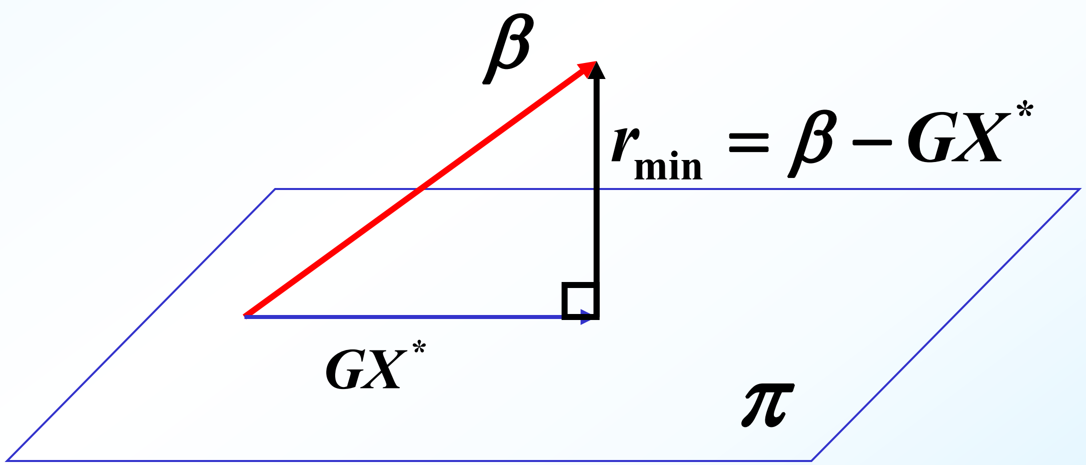
向量组 $G=[\alpha_1 , \alpha_2]$

正规方程 $G^T GX^* = G^T \beta$

平面π $GX = x\alpha_1 + y\alpha_2$

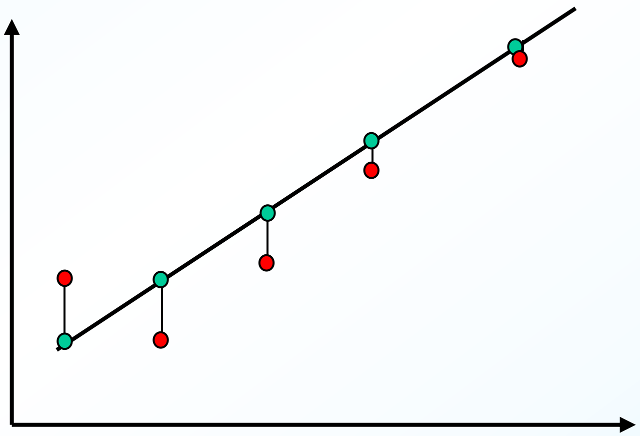
$$G^T(\beta - GX^*) = 0$$

$$G^T r_{\min} = 0$$



$$(GX^*, r_{\min}) = 0$$

$$S(a,b)=\|r\|_2^2=\sum_{k=1}^{m}[(a+bx_k)-y_k]^2$$



求 a, b 使 $S(a, b)=\min$

超定方程组：$GX=F$ $\rightarrow$

正规方程组：$G^TGX=G^TF$

$$G^TG=\begin{bmatrix}1 & 1 & \cdots & 1\\ x_1 & x_2 & \cdots & x_m\end{bmatrix}\begin{bmatrix}1 & x_1\\ 1 & x_2\\ \vdots & \vdots\\ 1 & x_m\end{bmatrix}=\begin{bmatrix}m & \sum_{j=1}^{m}x_j\\ \sum_{j=1}^{m}x_j & \sum_{j=1}^{m}x_j^2\end{bmatrix}$$

例1. 已知实验数据如下, 求线性拟合函数。

x	1	2	3	4	5
$f(x)$	4	4.5	6	8	9

解: 设拟合曲线方程为 $\varphi(x)= a + b\,x$

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \\ 1 & 5 \end{bmatrix}\begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 4 \\ 4.5 \\ 6 \\ 8 \\ 9 \end{bmatrix} \rightarrow$$

$$GX=F \rightarrow G^TGX=G^TF$$

$$\boxed{\begin{aligned} 5a + 15b &= 31.5 \\ 15a + 55b &= 108 \end{aligned}}$$

$$a = 2.25, \quad b = 1.35$$

求数据的二次拟合函数 $P(x)=a_0+a_1x+a_2x^2$

x	1	2	3	4	5
$f(x)$	4	4.5	6	8	9

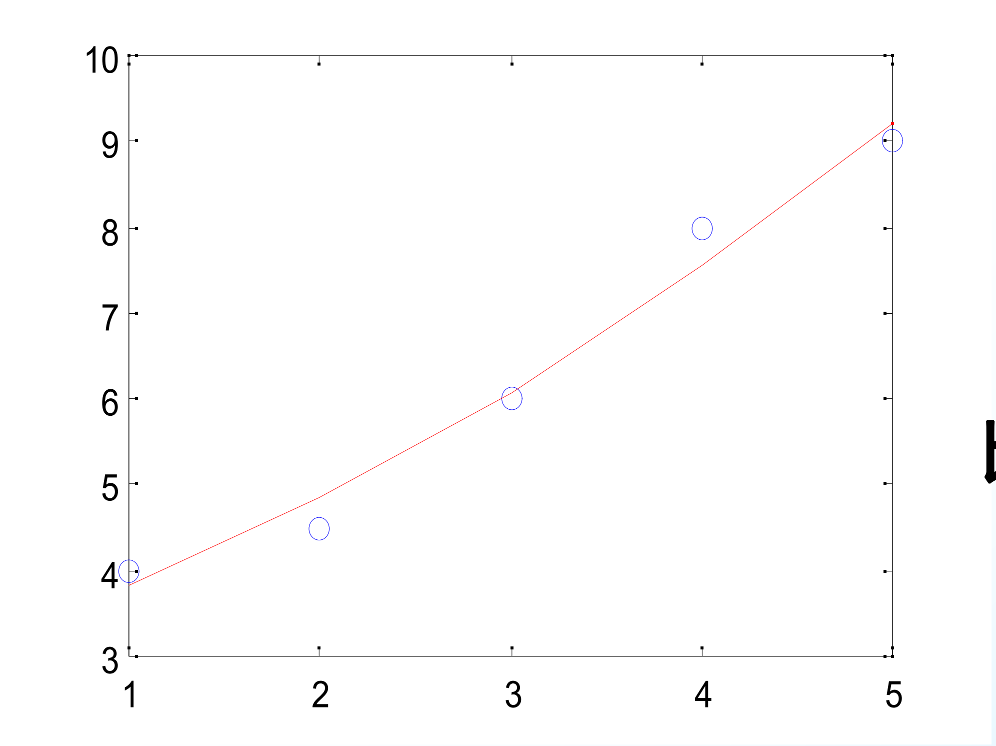
解:将数据点代入，得

$$\begin{cases} a_0 + a_1 + a_2 = 4 \\ a_0 + 2a_1 + 4a_2 = 4.5 \\ \qquad \cdots\cdots \\ a_0 + 5a_1 + 25a_2 = 9 \end{cases} \Rightarrow \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \\ 1 & 5 & 25 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 4 \\ 4.5 \\ 6 \\ 8 \\ 9 \end{bmatrix}$$

$$\begin{bmatrix} 5 & 15 & 55 \\ 15 & 55 & 225 \\ 55 & 225 & 979 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 31.5 \\ 108 \\ 429 \end{bmatrix}$$

$a_0=3$, $a_1=0.7071$,
$a_2=0.1071$

得 $P(x)=3+0.7071x + 0.1071x^2$



二次拟合误差:

$$\| r \|_2 = 0.6437$$

比较线性拟合误差:

$$\| r \|_2 = 0.7583$$

数据拟合的线性模型

给定数据表

x	x_1	x_2	…………	x_m
$f(x)$	y_1	y_2	…………	y_m

多项式拟合函数:

$$\varphi(x)=a_0+a_1x+a_2x^2+\cdots\cdots+a_nx^n$$

数据拟合的线性模型

$$\varphi(x)=a_0\varphi_0(x)+a_1\varphi_1(x)+\cdots\cdots+a_n\varphi_n(x)$$

例如:

$$[\varphi_0(x),\varphi_1(x),\cdots,\varphi_n(x)]=[1,x,\cdots,x^n]$$

$$[\varphi_0(x),\varphi_1(x),\cdots,\varphi_n(x)]=[1,\cos x,\cdots,\cos nx]$$

拟合函数: $\varphi(x)=\sum_{j=0}^{n} a_j\varphi_j(x)$

拟合数据: $f(x_j)=y_j$, $(j=1,2,3,\cdots\cdots,m)$

超定方程组: $GX=F \rightarrow G^TGX=G^TF$

$$\begin{cases}\varphi(x_1)=y_1\\ \varphi(x_2)=y_2\\ \varphi(x_3)=y_3\\ \cdots\cdots\cdots\cdots\\ \varphi(x_m)=y_m\end{cases}$$

$$\begin{bmatrix}\varphi_0(x_1) & \varphi_1(x_1) & \cdots & \varphi_n(x_1)\\ \varphi_0(x_2) & \varphi_1(x_2) & \cdots & \varphi_n(x_2)\\ \cdots & \cdots & \cdots & \cdots\\ \varphi_0(x_m) & \varphi_1(x_m) & \cdots & \varphi_n(x_m)\end{bmatrix}\begin{bmatrix}a_0\\ a_1\\ \vdots\\ a_n\end{bmatrix}=\begin{bmatrix}y_1\\ y_2\\ \vdots\\ y_m\end{bmatrix}$$

$m>n+1$ 超定方程组

系数矩阵按列分块 $G = [\vec{\varphi}_0 \quad \vec{\varphi}_1 \quad \cdots \quad \vec{\varphi}_n]$

$$\vec{\varphi}_0 = \begin{bmatrix} \varphi_0(x_1) \\ \varphi_0(x_2) \\ \vdots \\ \varphi_0(x_m) \end{bmatrix} \quad \vec{\varphi}_1 = \begin{bmatrix} \varphi_1(x_1) \\ \varphi_1(x_2) \\ \vdots \\ \varphi_1(x_m) \end{bmatrix} \quad \cdots\cdot \quad \vec{\varphi}_n = \begin{bmatrix} \varphi_n(x_1) \\ \varphi_n(x_2) \\ \vdots \\ \varphi_n(x_m) \end{bmatrix} \quad F = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

$GX=F \rightarrow G^TGX=G^TF$

$$\begin{bmatrix} (\vec{\varphi}_0,\vec{\varphi}_0) & (\vec{\varphi}_0,\vec{\varphi}_1) & \cdots & (\vec{\varphi}_0,\vec{\varphi}_n) \\ (\vec{\varphi}_1,\vec{\varphi}_0) & (\vec{\varphi}_1,\vec{\varphi}_1) & \cdots & (\vec{\varphi}_1,\vec{\varphi}_n) \\ \cdots & \cdots & \cdots & \cdots \\ (\vec{\varphi}_n,\vec{\varphi}_0) & (\vec{\varphi}_n,\vec{\varphi}_1) & \cdots & (\vec{\varphi}_n,\vec{\varphi}_n) \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} (\vec{\varphi}_0,\vec{y}) \\ (\vec{\varphi}_1,\vec{y}) \\ \vdots \\ (\vec{\varphi}_n,\vec{y}) \end{bmatrix}$$

正规方程组的解称为超定方程组的最小二乘解

给定数表

x	x_1 x_2 ………… x_m
$f(x)$	y_1 y_2 ………… y_m

求拟合函数：$\varphi(x)=a_0\varphi_0(x)+a_1\varphi_1(x)$

取

$$\varphi_0(x)=1,\ \varphi_1(x)=x-\frac{1}{m}\sum_{j=1}^{m}x_j$$

$$\begin{bmatrix}\varphi_0(x_1) & \varphi_1(x_1)\\ \varphi_0(x_2) & \varphi_1(x_2)\\ \vdots & \vdots\\ \vdots & \vdots\\ \varphi_0(x_m) & \varphi_1(x_m)\end{bmatrix}\begin{bmatrix}a_0\\ a_1\end{bmatrix}=\begin{bmatrix}y_1\\ y_2\\ \vdots\\ y_m\end{bmatrix}\iff[\vec{\varphi}_0\quad\vec{\varphi}_1]\vec{a}=\vec{y}$$

$$\begin{bmatrix} (\vec{\varphi}_0,\vec{\varphi}_0) & (\vec{\varphi}_0,\ \vec{\varphi}_1) \\ (\vec{\varphi}_1,\ \vec{\varphi}_0) & (\vec{\varphi}_1,\ \vec{\varphi}_1) \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} (\vec{\varphi}_0,\ \vec{y}) \\ (\vec{\varphi}_1,\ \vec{y}) \end{bmatrix}$$

$$\begin{bmatrix} (\vec{\varphi}_0,\vec{\varphi}_0) & \\ & (\vec{\varphi}_1,\ \vec{\varphi}_1) \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} (\vec{\varphi}_0,\ \vec{y}) \\ (\vec{\varphi}_1,\ \vec{y}) \end{bmatrix}$$

$$a_0 = \frac{(\vec{\varphi}_0,\ \vec{y})}{(\vec{\varphi}_0,\ \vec{\varphi}_0)} \qquad a_1 = \frac{(\vec{\varphi}_1,\ \vec{y})}{(\vec{\varphi}_1,\ \vec{\varphi}_1)}$$

$$\varphi(x) = \frac{(\vec{\varphi}_0,\vec{y})}{(\vec{\varphi}_0,\vec{\varphi}_0)}\varphi_0(x) + \frac{(\vec{\varphi}_1,\vec{y})}{(\vec{\varphi}_1,\vec{\varphi}_1)}\varphi_1(x)$$

数据拟合的非线性模型

观测数据

x	x_1	x_2	$\cdots\cdots\cdots\cdots$	x_m
f	y_1	y_2	$\cdots\cdots\cdots\cdots$	y_m

求拟合函数 $f(x, c_0, c_1, \cdots, c_n)$满足

$$\sum_{i=1}^{m}[f(x_i, c_0, c_1, \cdots, c_n) - y_i]^2 = \min$$

例1. 已知人口统计数据

年	1991	1992	1993	1994	1995	1996
数量	11.58	11.72	11.85	11.98	12.11	12.24

利用最小二乘法求指数拟合 $y = c\,e^{ax}$

$$S(a,c) = \sum_{j=1}^{6}[c\exp(ax_j) - y_j]^2 = \min$$

指数函数拟合人口统计数据(单位：亿)

t	1991	1992	1993	1994	1995	1996
$\ln N$	2.45	2.46	2.47	2.48	2.49	2.50

设 $N = e^{a+bt}$ → $\ln N = a + bt$ 令 $y = \ln N$，有

$$y(t) = a + bt$$

(1)计算对数值 $y_k = \ln N_k$ （$k = 1,2,\ldots,6$）

(2) 列出未知数 a 、b 的超定方程组

$$a + b\,t_k = y_k \quad (k = 1，2，\ldots\ldots，6)$$

(3)求超定方程组的最小二乘解,得 $N = e^{a+bx}$，并预测

$N(2000)$，$N(2008)$

N2000=12.7971，N2008=13.9783

线性最小二乘拟合 $f(x)=a_1r_1(x)+\ldots+a_mr_m(x)$中函数$\{r_1(x),\cdots r_m(x)\}$的选取

1. 通过机理分析建立数学模型来确定 $f(x)$；

2. 将数据 (x_i,y_i) $i=1,\ldots n$ 作图，通过直观判断确定 $f(x)$: