《数值分析》15

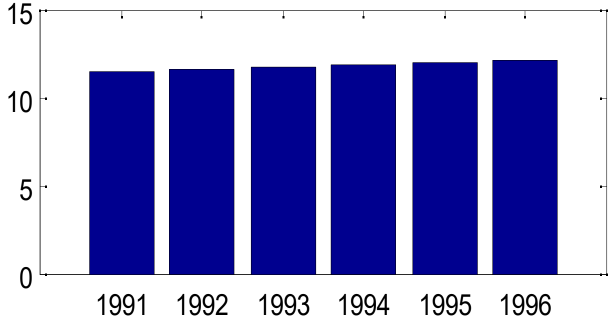
主要内容：

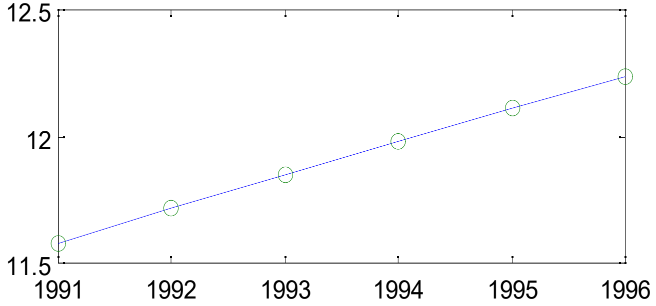
- 离散数据的最小二乘逼近
- 线性拟合与二次拟合
- 数据拟合的线性模型
- 一次多项式拟合公式

离散数据的最小二乘逼近

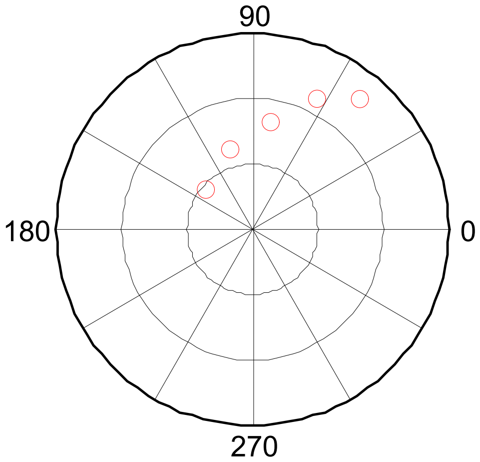
引例1. 我国上世纪90年代初人口数量(单位:亿)

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





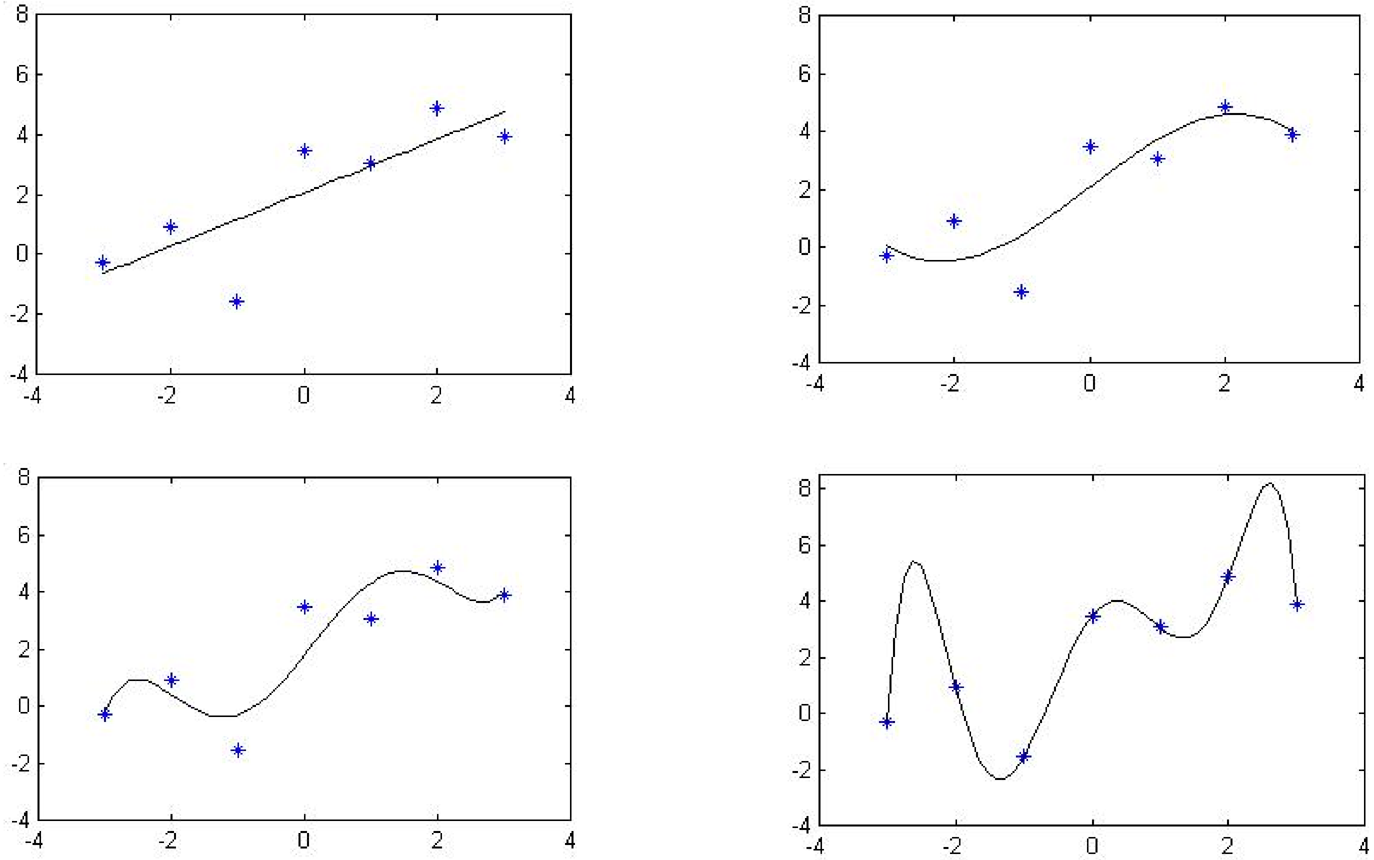
引例2. 极坐标数据拟合慧星轨道方程



2.70	2.00	1.61	1.20	1.02
48°	67°	83°	108°	126°

引例3. 离散数据的多项式(1、3、5、6)拟合实验：

-3	-2	-1	0	1	2	3
-0.277	0.895	-1.565	3.456	3.060	4.856	3.898



离散数据的线性拟合

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

求拟合函数： $\varphi(x) = c_1 + c_2 x$

$$\begin{cases} c_1 + c_2 x_1 = y_1 \\ c_1 + c_2 x_2 = y_2 \\ \cdots\cdots\cdots\cdots \\ c_1 + c_2 x_m = y_m \end{cases} \rightarrow \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

$\rightarrow GX=F$

超定方程组

方程组 $GX=F$ 的残差向量： $r=F-GX$

最小二乘问题 $\min\limits_{X\in R2} \|GX-F\|_2^2$

$$\|GX-F\|_2^2 = (GX-F\,,\,GX-F)$$
$$=(GX, GX)-2(GX, F)+(F, F)$$
$$=(X, G^TGX)-2(X, G^TF)+(F, F)$$

$f(X)=(X, G^TGX)-2(X, G^TF)$ ➔

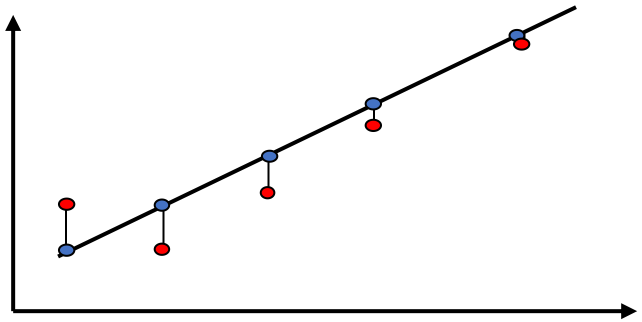
$$\boxed{\begin{gathered}\min_{X\in R^2} f(X)=\min_{X\in R^2}[(X,G^TGX)-2(X,G^TF)] \\ \Leftrightarrow \quad (G^TG)X=(G^TF)\end{gathered}}$$

线性拟合与二次拟合

$$S(c_1, c_2) = \| r \|_2^2 = \sum_{k=1}^{m} [(c_1 + c_2 x_k) - y_k]^2$$

超定方程组：$GX=F$ ➔

正规方程组：$G^T GX=G^T F$



$$G^T G = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_m \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_m \end{bmatrix} = \begin{bmatrix} m & \sum_{j=1}^{m} x_j \\ \sum_{j=1}^{m} x_j & \sum_{j=1}^{m} x_j^2 \end{bmatrix} \quad G^T F = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_m \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{m} y_j \\ \sum_{j=1}^{m} x_j y_j \end{bmatrix}$$

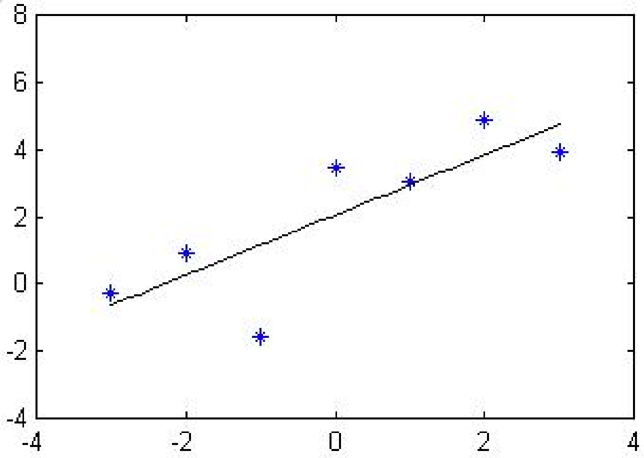
引例4 实验数据线性拟合。

-3	-2	-1	0	1	2	3
-0.277	0.895	-1.565	3.456	3.060	4.856	3.898

解: 设拟合曲线方程为 $\varphi(x) = c_1 + c_2 x$

$$\begin{bmatrix} 1 & -3 \\ 1 & -2 \\ 1 & -1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} -0.277 \\ 0.895 \\ -1.565 \\ 3.456 \\ 3.060 \\ 4.856 \\ 3.898 \end{bmatrix} \Rightarrow$$

$$c_1 = 2.0464, \; c_2 = 0.8955$$



残差2范数: $\| r \|_2 = 3.4142$

引例5 实验数据3次多项式拟合。

-3	-2	-1	0	1	2	3
-0.277	0.895	-1.565	3.456	3.060	4.856	3.898

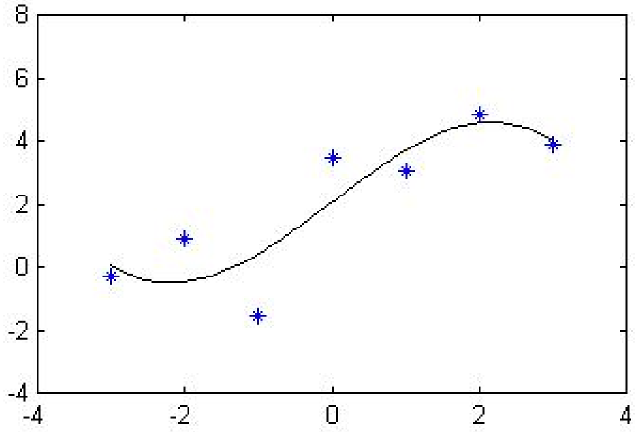
解: 设 $\varphi(x)=c_1+c_2x+c_2x^2+c_3x^3$

$$c_1=2.0563,\ c_2=1.7531$$

$$c_3=-0.0025,\ c_4=-0.1225$$

$$\begin{bmatrix}1 & -3 & 9 & -27\\ 1 & -2 & 4 & -8\\ 1 & -1 & 1 & -1\\ 1 & 0 & 0 & 0\\ 1 & 1 & 1 & 1\\ 1 & 2 & 4 & 8\\ 1 & 3 & 9 & 27\end{bmatrix}\begin{bmatrix}c_1\\ c_2\\ c_3\\ c_4\end{bmatrix}=\begin{bmatrix}-0.277\\ 0.895\\ -1.565\\ 3.456\\ 3.060\\ 4.856\\ 3.898\end{bmatrix}$$

$\Rightarrow$

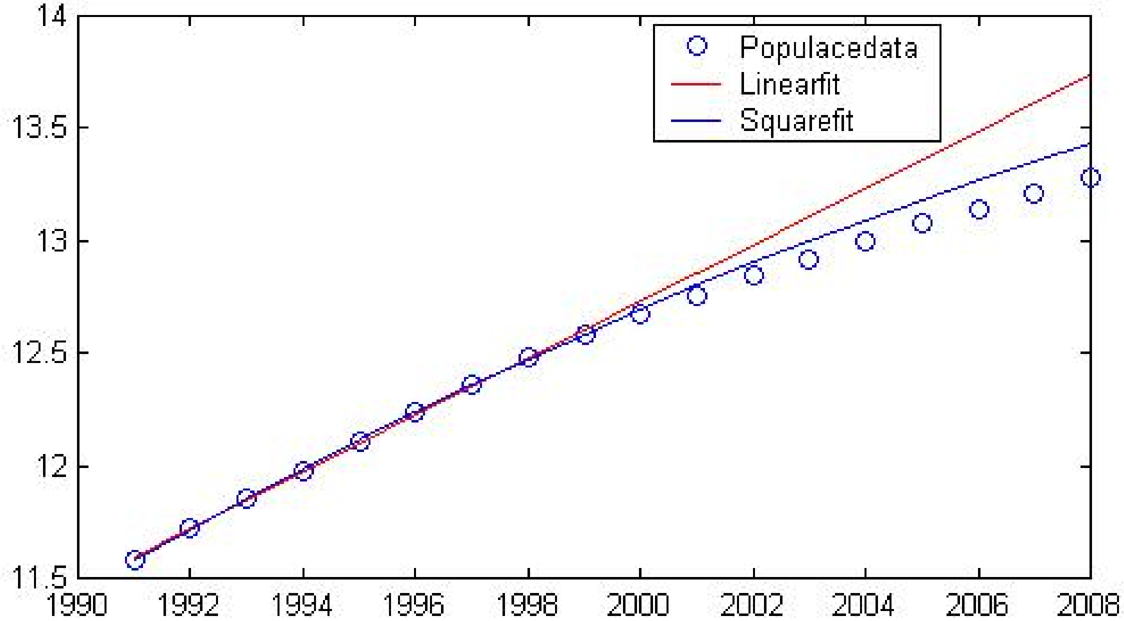


残差2范数: $\|r\|_2=2.9007$

线性拟合与二次拟合

中国人口数据1991--2008(单位:亿)

年份	1991	1992	1993	1994	1995	1996	1997	1998	1999
人数	11.5	11.7	11.8	11.9	12.1	12.2	12.3	12.4	12.5
年份	2000	2001	2002	2003	2004	2005	2006	2007	2008
人数	12.6	12.7	12.8	12.9	13.0	13.0	13.1	13.2	13.2



离散数据表

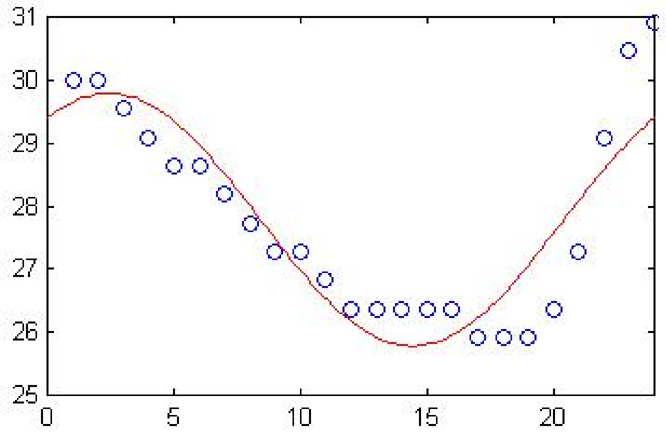
x	x_1	x_2	…………	x_m
$f(x)$	y_1	y_2	…………	y_m

线性模型

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

三角函数 $$\varphi(x) = a_0 + a_1\cos(\frac{\pi}{12}x) + a_2\sin(\frac{\pi}{12}x)$$

收集24小时室外温度数据(每小时记录一次),以三角函数作数据拟合



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

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

$$\begin{cases}\varphi(x_1)=y_1\\ \varphi(x_2)=y_2\\ \varphi(x_3)=y_3\\ \cdots\cdots\cdots\cdots\\ \varphi(x_m)=y_m\end{cases}\qquad \begin{bmatrix}\varphi_0(x_1) & \varphi_1(x_1) & \cdots & \varphi_n(x_1)\\ \varphi_0(x_2) & \varphi_1(x_2) & \cdots & \varphi_n(x_2)\\ \cdots & \cdots & \cdots & \cdots\\ \varphi_0(x_m) & \varphi_1(x_m) & \cdots & \varphi_n(x_m)\end{bmatrix}\begin{bmatrix}a_0\\ a_1\\ \vdots\\ a_n\end{bmatrix}=\begin{bmatrix}y_1\\ y_2\\ \vdots\\ y_m\end{bmatrix}$$

$m>n+1$ 超定方程组

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

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

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

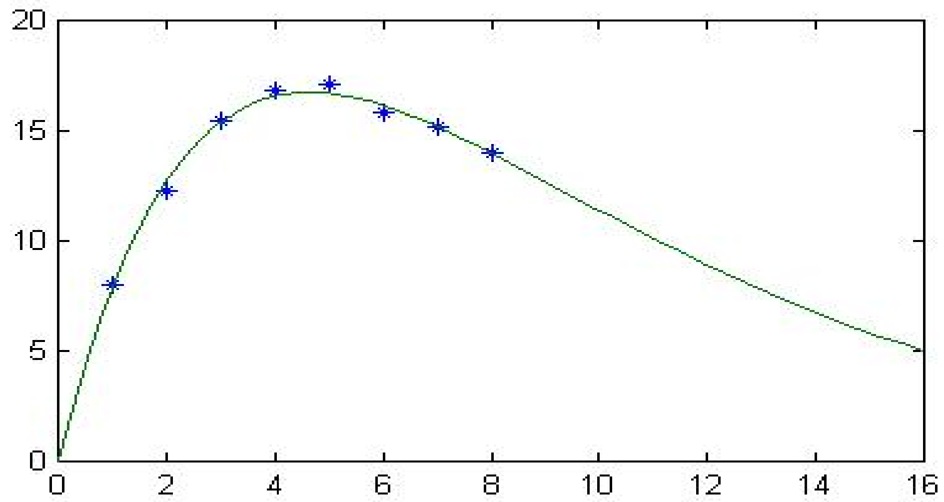
$$\begin{bmatrix}(\vec{\varphi}_0,\vec{\varphi}_0) & (\vec{\varphi}_0,\vec{\varphi}_1) & \cdots & (\vec{\varphi}_0,\vec{\varphi}_n)\\ (\vec{\varphi}_1,\vec{\varphi}_0) & (\vec{\varphi}_1,\vec{\varphi}_1) & \cdots & (\vec{\varphi}_1,\vec{\varphi}_n)\\ \cdots & \cdots & \cdots & \cdots\\ (\vec{\varphi}_n,\vec{\varphi}_0)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

正规方程组的解称为超定方程组的最小二乘解

药物浓度模型：考虑药品针剂注射后，血液中药物浓度的衰减曲线。模型为

$$u(t) = ct\exp(-bt)$$

T小时	1	2	3	4	5	6	7	8
u浓度	8.0	12.3	15.5	16.8	17.1	15.8	15.2	14.0



拟合函数：$u(t) = 9.79t\exp(-0.215t)$

拟合函数 $$u(t) = ct\exp(-bt)$$

对数变换 $$\ln u(t) = \ln c + \ln t - bt$$

线性模型 $$\ln c - bt = \ln\frac{u(t)}{t}$$

令 $\ln c = a$ 即 $c = \exp(a)$

由数据,得超定方程组 $$a - bt_k = \ln\frac{u_k}{t_k}$$

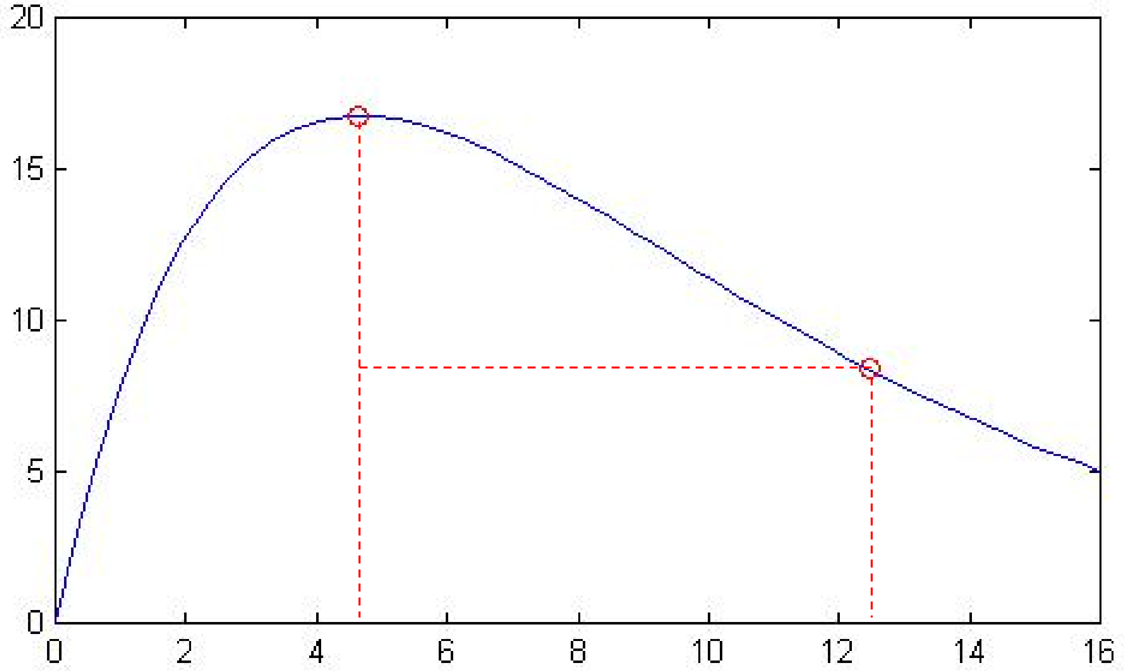
高矩阵 $$G = \begin{bmatrix} 1 & -t_1 \\ 1 & -t_2 \\ \vdots & \vdots \\ 1 & -t_m \end{bmatrix}$$

$$u(t) = 9.79t\exp(-0.215t)$$

$$u'(t) = 9.79(1-0.215t)\exp(-0.215t)$$

最大值点: t0=1/0.215=4.6512, u(t0)=16.7513

求u(t1)=16.7513/2; 半衰期: t1= 12.4574,



数据拟合的线性模型

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

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

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

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

一次多项式拟合公式

MATLAB中的多项式拟合命令

$$P = \text{polyfit}(X,Y,n)$$

求出(最小二乘意义下)n次拟合多项式

$$P(x)=a_0x^n + a_1x^{n-1} + \cdots\cdots + a_{n-1}x + a_n$$

计算结果为系数: $P=[\ a_0, a_1, \cdots\cdots, a_{n-1}, a_n\ ]$

多项式求值命令

$$y1=\text{polyval}(P, x)$$

其中,P是n次多项式的系数, x是自变量的值, $y1$是多项式在x处的值

思考与练习

1. 收集中国人口数据和世界人口数据，利用数据拟合方法分析人口变化规律。
2. 收集我国粮食生产数据，分析总产量或亩产量变化规律。
3. 收集近四年我国高考人数数据，分析并预测今后几年内参加高考人数变化规律。
4. 收集血液中酒精含量测试数据，分析人体内酒精浓度变化规律。

学到了什么?

- 离散数据的最小二乘逼近
- 线性拟合与二次拟合
- 数据拟合的线性模型
- 一次多项式拟合公式