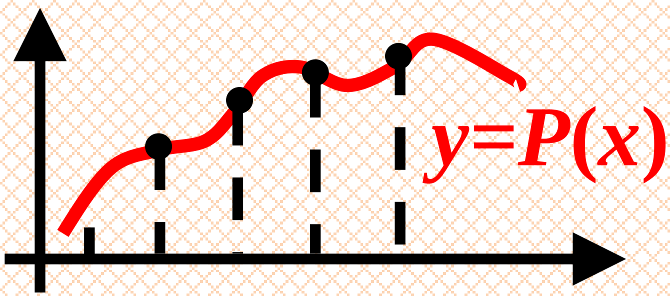
第三章 插值与拟合

（2）拟合方法

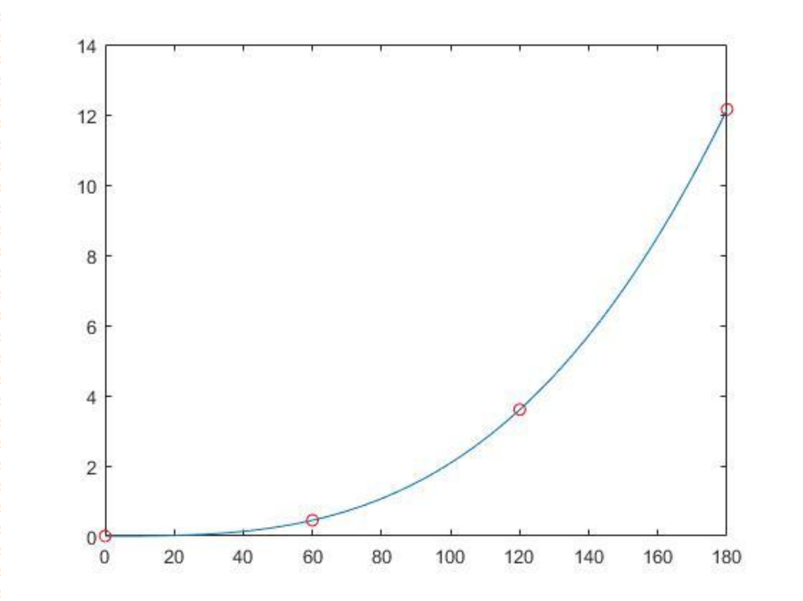


插值的Matlab函数

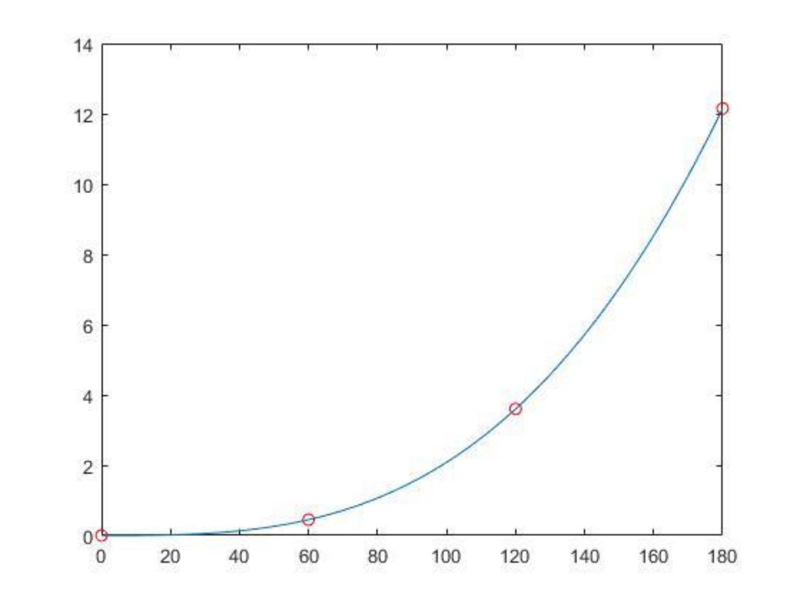
yc=interp1(x,y,cx,’method’)

给定曲线上4个点的坐标为(0,0),(60,0.45),(120,3.6), (180,12.15)，求x=100时y值。

```
x=[0,60,120,180];
y=[0,0.45,3.6,12.15];
y1=interp1(x,y,100,'spline')
```



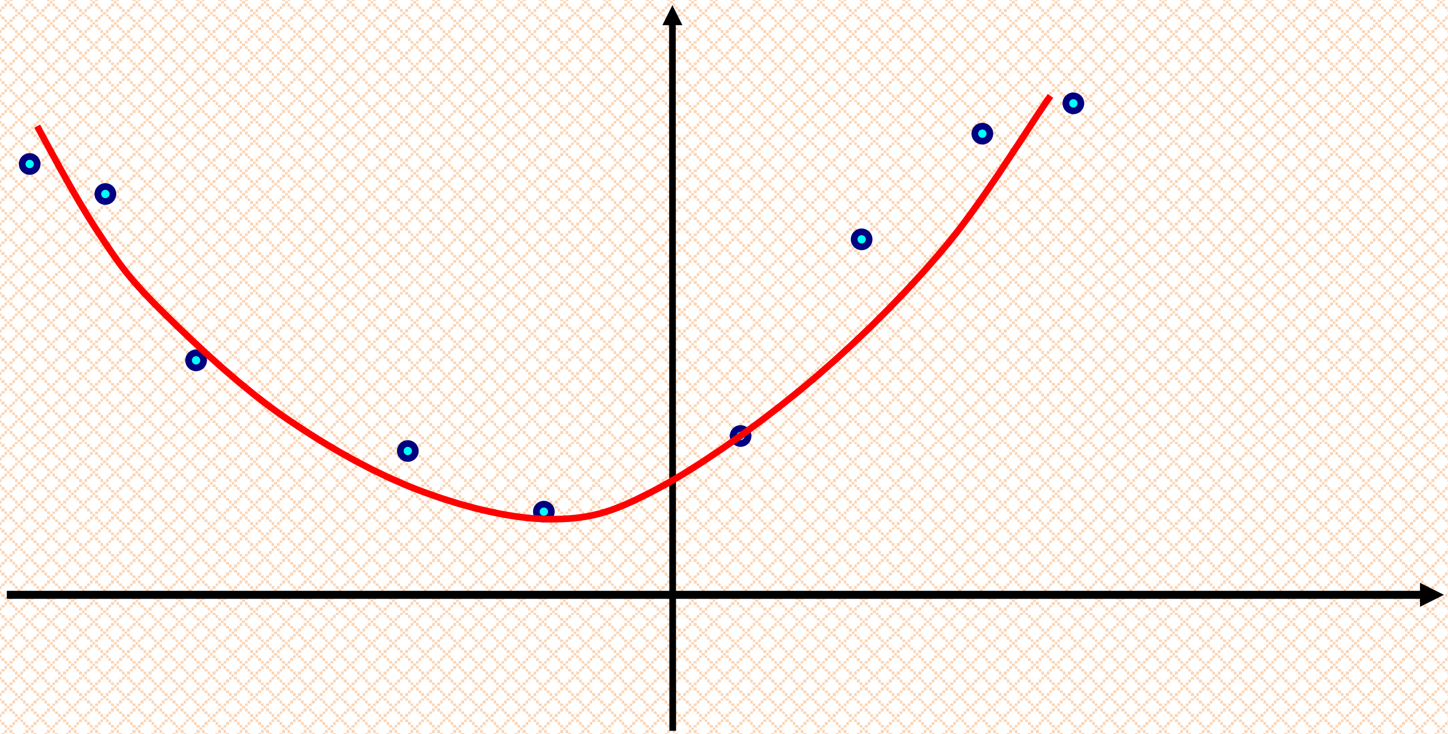
给定曲线上4个点的坐标为(0,0),(60,0.45),(120,3.6),(180,12.15)，求该曲线方程。



?

3.2 曲线拟合

所谓”曲线拟合”,是指根据给定的数据表,寻找一个简单的表达式来”拟合”该组数据,此处的”拟合”的含义为:不要求该表达式对应的近似曲线完全通过所有的数据点,只要求该近似曲线能够反映数据的基本变化趋势.

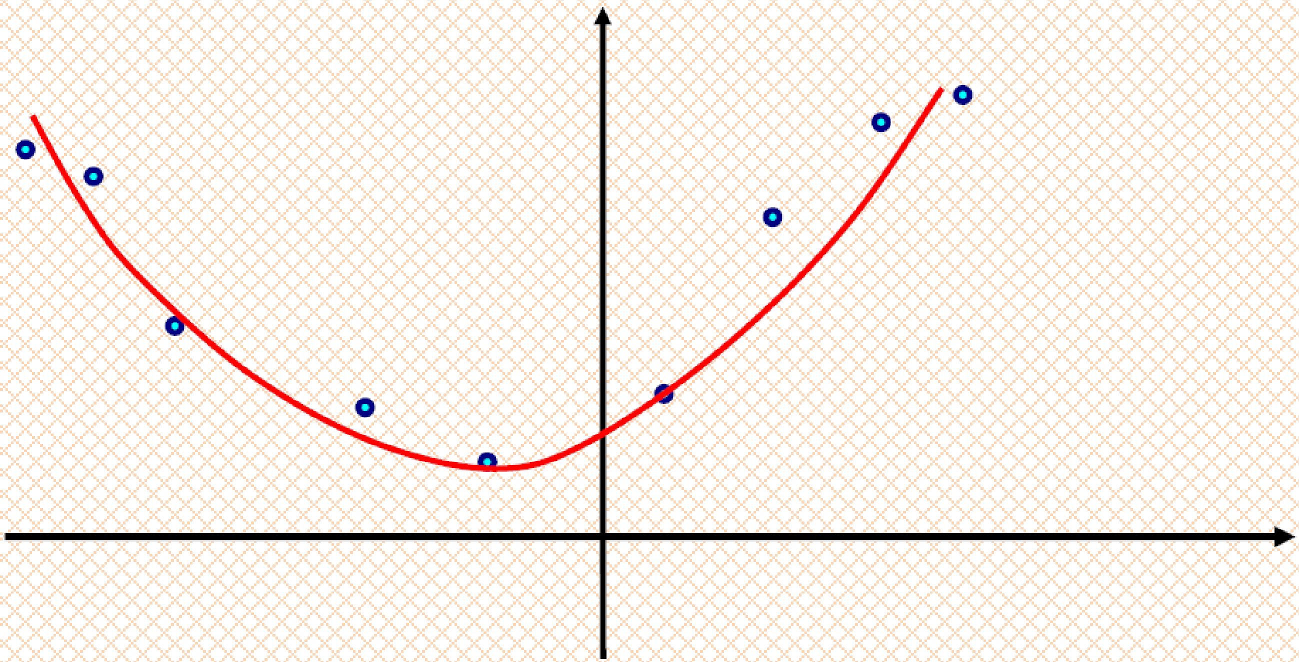


最小二乘拟合

已知数据对$(x_j, y_j)(j=1,2,\ldots,n)$，求多项式

$$P(x)=\sum_{i=0}^{m} a_i x^i \,(m<n)$$

使得 $\Phi(a_0,a_1,\cdots,a_n)=\sum_{j=1}^{n}\left(\sum_{i=0}^{m} a_i x_j^{\;i}-y_j\right)^2$ 为最小



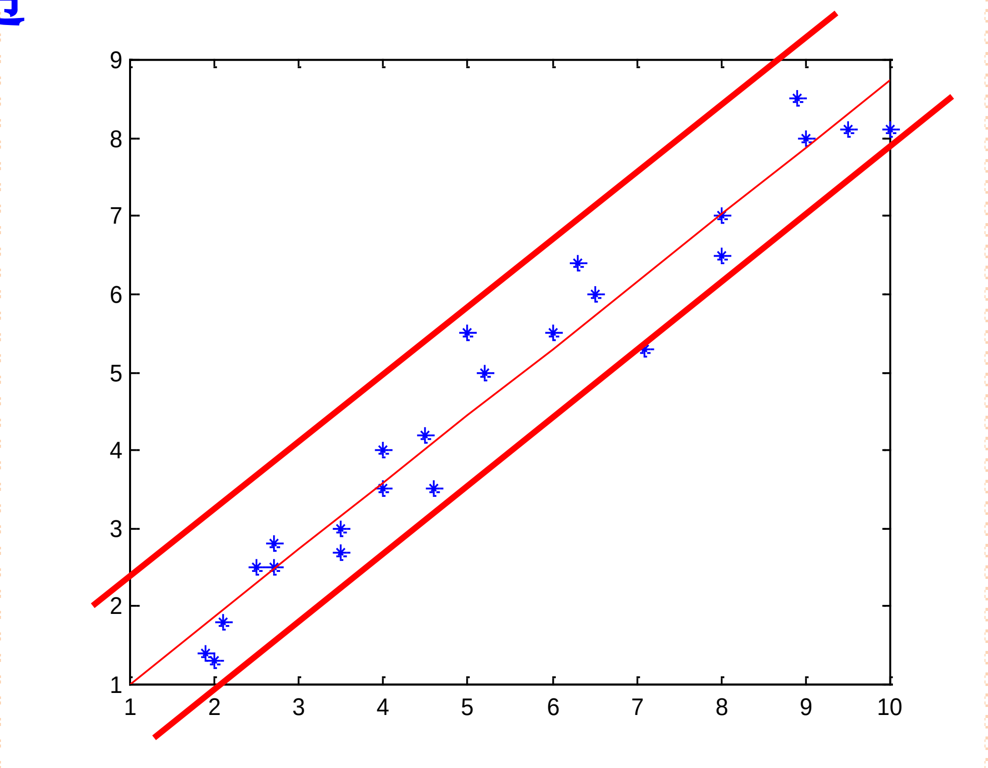
(1) 线性拟合

例：考察某种纤维的强度与其拉伸倍数的关系. 下表是实际测定的24个纤维样品的强度与相应的拉伸倍数的数据记录:

编号	拉伸倍数 x_i	强　　度 y_i	编号	拉伸倍数 x_i	强　度 y_i
1	1.9	1.4	13	5	5.5
2	2	1.3	14	5.2	5
3	2.1	1.8	15	6	5.5
4	2.5	2.5	16	6.3	6.4
5	2.7	2.8	17	6.5	6
6	2.7	2.5	18	7.1	5.3
7	3.5	3	19	8	6.5
8	3.5	2.7	20	8	7
9	4	4	21	8.9	8.5
10	4	3.5	22	9	8
11	4.5	4.2	23	9.5	8.1
12	4.6	3.5	24	10	8.1

可以看出，纤维强度随拉伸倍数增加而增加

并且24个点大致分布在一条直线附近



因此可认为**强度与拉伸倍数**之间的主要关系是**线性关系**

$$y \approx \varphi(x) = a + bx$$

该直线称为这一问题的**数学模型**。

怎样确定a, b, 使得直线能较好地反映所给数据的基本“变化趋势”？

采用最小二乘的思想

令 $$S(a,b)=\sum_{i=1}^{m}(a+bx_i-y_i)^2$$

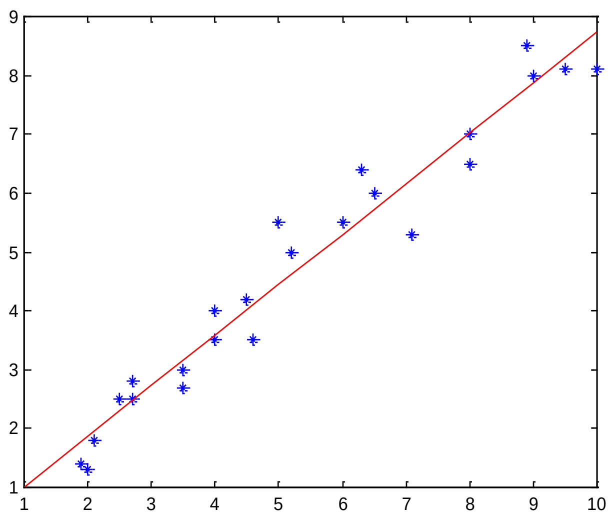
问题转化为求参数 a,b 使 $S(a,b)$ 达到最小值。

$$\frac{\partial S}{\partial a}=\frac{\partial S}{\partial b}=0 \Rightarrow \begin{bmatrix} m & \sum_{i=1}^{m} x_i \\ \sum_{i=1}^{m} x_i & \sum_{i=1}^{m} x_i^2 \end{bmatrix}\begin{bmatrix} a \\ b \end{bmatrix}=\begin{bmatrix} \sum_{i=1}^{m} y_i \\ \sum_{i=1}^{m} x_i y_i \end{bmatrix}$$

$$\begin{pmatrix} 24 & 127.5 \\ 127.5 & 829.61 \end{pmatrix}\begin{pmatrix} a \\ b \end{pmatrix}=\begin{pmatrix} 113.1 \\ 731.6 \end{pmatrix}$$

$$a \approx 0.1505 \qquad b \approx 0.8587$$

$$\varphi(x)=0.1505+0.8587x$$



这种求线性函数y=a+bx的过程称为线性拟合。

(2) 非线性拟合

已知函数 $f(x)$ 在若干个点上的数据表，确定参数 a 和 b，

利用经验函数 $y = ae^{bx}$ 拟合某组数据：

线性化处理：$y = ae^{bx} \Leftrightarrow \ln y = \ln a + bx$

令 $\bar{y} = \ln y; \bar{a} = \ln a$ 则 $\bar{y} = \bar{a} + bx$

x_i	x_1	x_2	$\cdots$	x_m
$\bar{y}_i$	$\ln f_1$	$\ln f_2$	$\cdots$	$\ln f_m$

由线性拟合方法可得到 $\bar{a}$ 和 b，从而得到 a 和 b。

用MATLAB解拟合问题

1、线性最小二乘拟合

2、非线性最小二乘拟合

用MATLAB作线性最小二乘拟合

1. 多项式$f(x)=a_1x^m+ \dots +a_mx+a_{m+1}$拟合, 可利用函数:

a=polyfit(x,y,m)

输出拟合多项式系数
$a=[a_1, \dots a_m, a_{m+1}]$ (数组)）

输入同长度的数组X，Y

拟合多项式次数

2. 多项式在x处的值y可用以下命令计算:

y=polyval（a，x）

例6. 对下面一组数据作二次多项式拟合

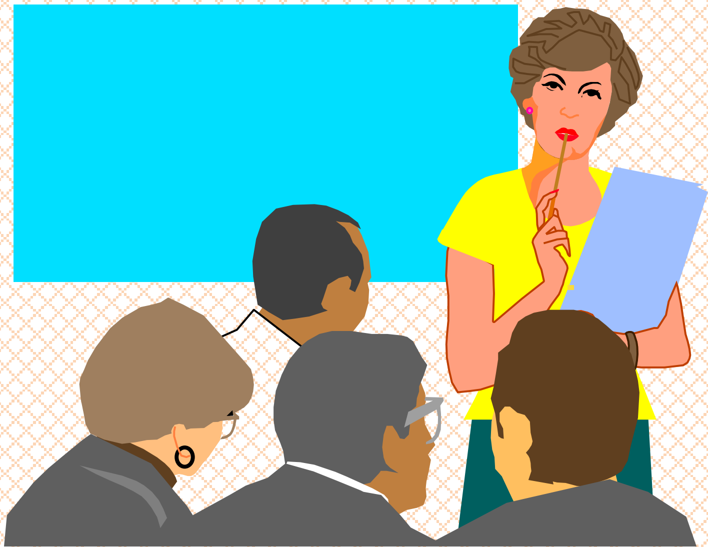
xi	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
yi	-0.447	1.978	3.28	6.16	7.08	7.34	7.66	9.58	9.48	9.30	11.2

即要求 出二次多项式:

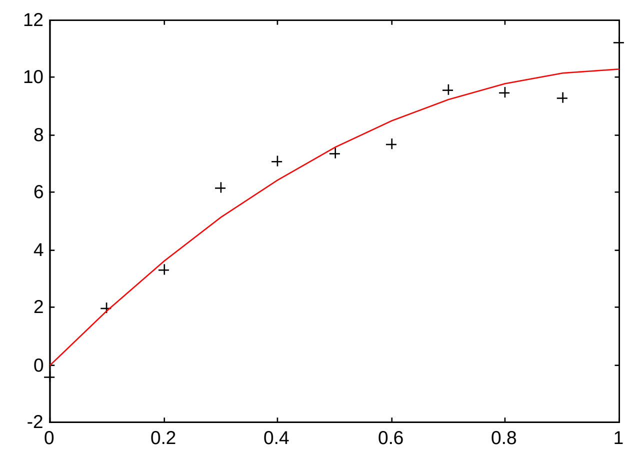
$$f(x) = a_1 x^2 + a_2 x + a_3$$

中 的 $A = (a_1, a_2, a_3)$ **使得:**

$$\sum_{i=1}^{10} [f(x_i) - y_i]^2 \quad \text{最小}$$



用多项式拟合的命令



1）输入以下命令：

```
x=0:0.1:1;
y=[-0.447 1.978 3.28 6.16 7.08 7.34 7.66 9.56 9.48 9.30 11.2];
A=polyfit(x,y,2)
z=polyval(A,x);
plot(x,y,'k+',x,z,'r')      %作出数据点和拟合曲线的图形
```

2）计算结果：　　A = -9.8108　20.1293　-0.0317

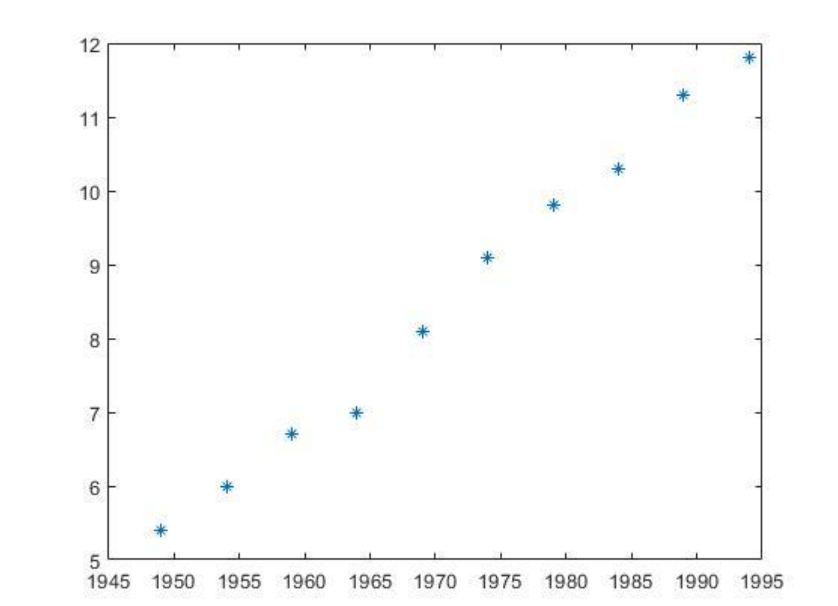
$$f(x) = -9.8108x^2 + 20.1293x - 0.0317$$

例7：1949年—1994年我国人口数据资料如下：

年 份xi	1949	1954	1959	1964	1969	1974	1979	1984	1989	1994
人口数yi	5.4	6.0	6.7	7.0	8.1	9.1	9.8	10.3	11.3	11.8

建模分析我国人口增长的规律，预报1999年我国人口数。

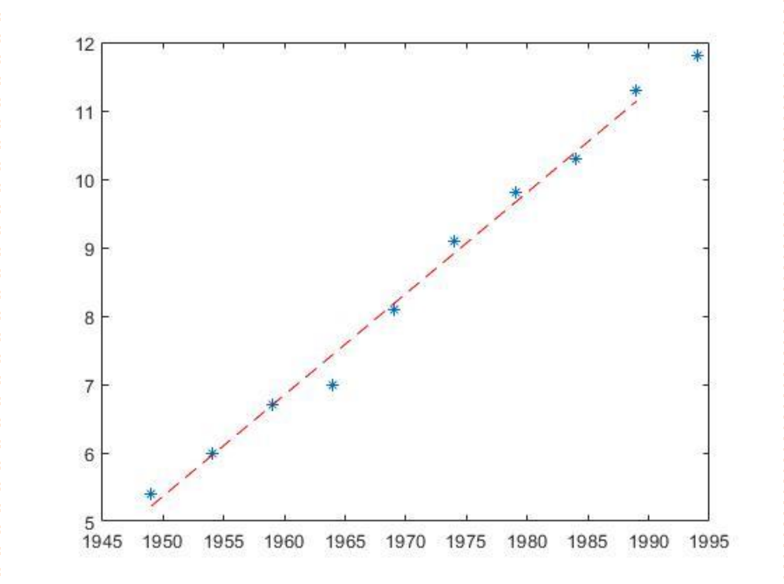
```
x=[1949 1954 1959 1964 1969 1974 1979 1984 1989 1994];
y=[5.4  6.0  6.7  7.0  8.1  9.1  9.8  10.3 11.3  11.8 ];
plot(x,y,'*')
```



假设人口随时间线性地增加

模型： $y = a + bx$

[a,b]=polyfit(x,y,1)



可以算出： a = -283.2320

b=0.1480

模型： y = 0.148 x - 283.232

实验题1

给定数据组 x0, y0，求拟合三阶多项式，并图示拟合情况。

xi	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
yi	-0.447	1.978	3.11	5.25	5.02	4.66	4.01	4.58	3.45	5.35	9.22

```
x0=0:0.1:1;y0=[-0.447,1.978,3.11,5.25,5.02,4.66,4.01,4.58,3.45,5.35,9.22];
n=3;
P=polyfit(x0,y0,n);
xx=0:0.01:1;yy=polyval(P, xx);
plot(xx,yy,'-b',x0,y0,'.r','MarkerSize',20)
legend('拟合曲线','原始数据','Location','SouthEast')
xlabel('x')
```

实验题2

- 给定曲线上4个点的坐标为(0,0),(60,0.45),(120,3.6),(180,12.15)，求该曲线方程。

```
x=[0,60,120,180];
y=[0,0.45,3.6,12.15];
P=polyfit(x,y,3)
xx=0:180;
yy=polyval(P,xx);
plot(x,y,'ro',xx,yy)
```

用MATLAB作非线性最小二乘拟合

Matlab的提供了两个求非线性最小二乘拟合的函数：lsqcurvefit和lsqnonlin。两个命令都要先建立M-文件fun.m，在其中定义函数f(x)，但两者定义f(x)的方式是不同的.

1. lsqcurvefit

已知数据点： xdata=（$\mathbf{xdata_1}$，$\mathbf{xdata_2}$，…，$xdata_n$），

ydata=（$ydata_1$，$ydata_2$，…，$ydata_n$）

lsqcurvefit用以求含参量x（向量）的向量值函数

F(x,xdata)=（F（x，$xdata_1$），…，F（x，$xdata_n$））^{T}

中的参变量x(向量),使得

$$\sum_{i=1}^{n}\left(F(x,xdata_i)-ydata_i\right)^2 \quad 最小$$

输入格式为:

(1) x = lsqcurvefit ('fun',x0,xdata,ydata) ;

(2) [x, options,funval, Jacob] = lsqcurvefit ('fun',x0,xdata,ydata,…);

说明：**x = lsqcurvefit ('fun',x0,xdata,ydata,options);**

fun是一个事先建立的定义函数F(x,xdata) 的M-文件, 自变量为x和xdata

迭代初值

已知数据点

选项见无约束优化

2. lsqnonlin

已知数据点： xdata=（$xdata_1$，$xdata_2$，…，$xdata_n$）

ydata=（$ydata_1$，$ydata_2$，…，$ydata_n$）

lsqnonlin用以求含参量x（向量）的向量值函数

$f(x)=(f_1(x), f_2(x), \ldots, f_n(x))^T$ 中的参量x，使得

$$f^T(x)f(x)=f_1(x)^2+f_2(x)^2+\cdots+f_n(x)^2$$

最小。

其中 $f_i(x)=f(x, xdata_i, ydata_i)$

$=F(x, xdata_i)-ydata_i$

输入格式为：

1） x= lsqnonlin （‘fun’，x0，options）；

2） [x，options，funval]= lsqnonlin （‘fun’，x0，...）；

说明：**x=** lsqnonlin （**‘fun’**，**x0**，**options**）；

fun是一个事先建立的定义函数*f(x)*的M-文件，自变量为x

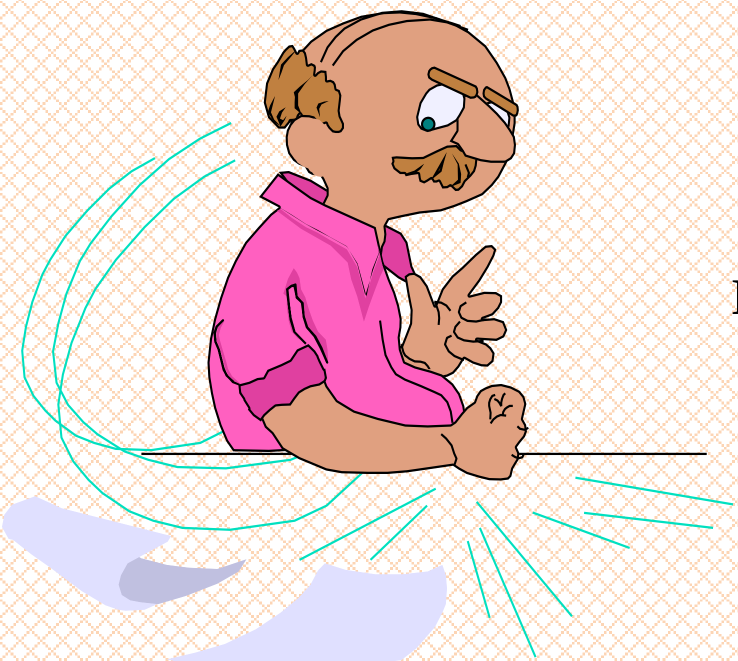
迭代初值

选项见无约束优化

例8. 用下面一组数据拟合 $c(t)=a+be^{0.0.2kt}$

中的参数a，b，k

t_j	100	200	300	400	500	600	700	800	900	1000
$c_j\times 10^3$	4.54	4.99	5.35	5.65	5.90	6.10	6.26	6.39	6.50	6.59



该问题即解最优化问题：

$$\min F(a,b,k)=\sum_{j=1}^{10}[a+be^{-0.02kt_j}-c_j]^2$$

解法1. 用命令lsqcurvefit

F(x, tdata)= $(a+be^{-0.02kt_1},\cdots,a+be^{-0.02kt_{10}})^T$, x=(a, b, k)

1）编写M-文件 curvefun1.m

```
function f=curvefun1(x,tdata)
f=x(1)+x(2)*exp(-0.02*x(3)*tdata)
                    %其中 x(1)=a;  x(2)=b;  x(3)=k;
```

2）输入命令

```
tdata=100:100:1000
cdata=1e-03*[4.54,4.99,5.35,5.65,5.90,6.10,6.26,6.39,
      6.50,6.59];
x0=[0.2,0.05,0.05];
x=lsqcurvefit ('curvefun1',x0,tdata,cdata)
f= curvefun1(x,tdata)
```

3）运算结果为：

f =0.0043 0.0051 0.0056 0.0059 0.0061

0.0062 0.0062 0.0063 0.0063 0.0063

x = 0.0063 -0.0034 0.2542

4）结论：**a=0.0063, b=-0.0034, k=0.2542**

解法 2　用命令lsqnonlin

f(x)=F(x,tdata,ctada)= $(a+be^{-0.02kt_1}-c_1,\cdots,a+be^{-0.02kt_{10}}-c_1)^T$

x=（a，b，k）

函数curvefun2的自变量是x，cdata和tdata是已知参数，故应将cdata tdata的值写在curvefun2.m中

1）编写M-文件 curvefun2.m

```
function f=curvefun2(x)
tdata=100:100:1000;
cdata=1e-03*[4.54,4.99,5.35,5.65,5.90,
            6.10,6.26,6.39,6.50,6.59];
f=x(1)+x(2)*exp(-0.02*x(3)*tdata)- cdata
```

2）输入命令:

```
x0=[0.2,0.05,0.05];
x=lsqnonlin('curvefun2',x0)
f= curvefun2(x)
```

3）运算结果为

f =1.0e-003 *(0.2322 -0.1243 -0.2495 -0.2413 -0.1668 -0.0724 0.0241 0.1159 0.2030 0.2792

x =0.0063 -0.0034 0.2542

4）结论：即拟合得a=0.0063 b=-0.0034 k=0.2542

可以看出,两个命令的计算结果是相同的.

例9. 体重约70kg的某人在短时间内喝下2瓶啤酒后，隔一定时间测量他的血液中酒精含量（mg/100ml），得到数据如表所示。试用所给数据用函数 $\varphi(t)=at^{b}e^{ct}$ 进行拟合，并求出未知常数。

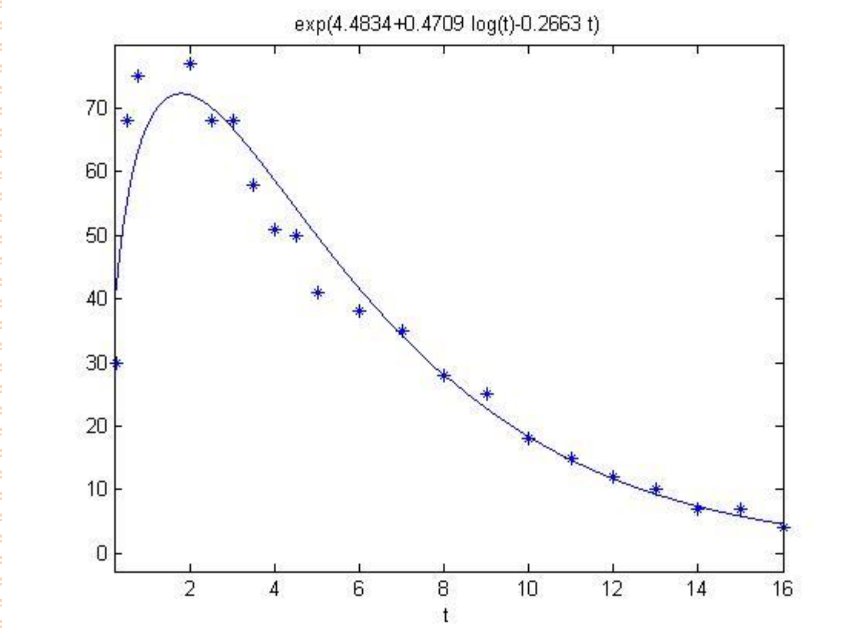
时间t/h	0.25	0.5	0.75	1	1.5	2	2.5	3	3.5	4	4.5	5
酒精含量h mg/100ml	30	68	75	82	82	77	68	68	58	51	50	41
时间t/h	6	7	8	9	10	11	12	13	14	15	16	
酒精含量h mg/100ml	38	35	28	25	18	15	12	10	7	7	4	

```
t=[0.25 0.5 0.75 1 1.5 2 2.5 3 3.5 4 4.5 5 6 7 8 9 10 11 12 13 14 15 16];
h=[30 68 75 82 82 77 68 68 58 51 50 41 38 35 28 25 18 15 12 10 7 7 4];
plot(t,h,'*')
```

```
clear all;
close all;
t=[0.25 0.5 0.75 1 1.5 2 2.5 3 3.5 4 4.5 5 6 7 8 9 10 11 12 13 14 15 16];
h=[30 68 75 82 82 77 68 68 58 51 50 41 38 35 28 25 18 15 12 10 7 7 4];
h1=log(h);  %对数变换
f=inline('a(1)+a(2).*log(t)+a(3).*t', 'a', 't');
[x,r]=lsqcurvefit(f, [1, 0.5, -0.5], t,h1)   %求参数lna,b,c的拟合值
```

输出结果：

```
x =
    4.4834    0.4709   -0.2663
r =
    0.4097
```



实验题3

在一次传染病中，已知t 时刻的人数$i(t)$ 满足$i(t)=\dfrac{1}{a+be^{ct}}$，公共部门每隔2天记录一次传染病的人数，具体数据如表1 所示，求a、b、c的值

表1　数据

天数	0	2	4	6	8	10	12	14	16	18	20	22	24
人数	0.2	0.4	0.5	0.9	1.5	2.4	3.1	3.8	4.1	4.2	4.5	4.4	4.5

解：程序如下

```
tdata=0:2:24;
ydata=[0.2,0.4,0.5,0.9,1.5,2.4,3.1,3.8,4.1,4.2,4.5,4.4,4.5];
f=@(x,t)1./(x(1)+x(2)*exp(x(3)*t));          %定义匿名函数
x=lsqcurvefit(f,[0.5,10,0],tdata,ydata);
plot(tdata,ydata,'ko')
hold on
fplot(@(t)f(x,t),[0,30])
legend({'数据散点图','拟合曲线'},'location','northwest')
title(sprintf('a=%.4f,b=%.4f,c=%.4f',x))
```

实验题4

1、已知数据如表 2 所示，求一个形如 $y=ae^{bx}$ 的经验公式（a、b 为常数）。

表 2 题 1 的数据

x	1	2	3	4	5	6	7	8
y	15.3	20.5	27.4	36.6	49.1	65.5	87.8	117.6

2、近年来我国的电信事业发展迅速，现已成世界第一电信大国。据统计某市在过去 9 年中通信工具的拥有量（单位：万台）如表 3 所示，求该市通信工具的发展规律。

表 3 题 2 的数据

x	1	2	3	4	5	6	7	8	9
y	1.78	2.24	2.74	3.74	4.45	5.31	6.92	8.85	10.97

实验题5

4、许多储油罐在使用一段时间后，由于地基变形等原因，使罐体的位置发生纵向倾斜和横向偏转等变化（以下称为变位），从而导致罐容表发生改变。按照有关规定，需要定期对罐容表进行重新标定。

请根据附件的数据，完成以下问题：

（**1**）为了掌握罐体变位后对罐容表的影响，分别对罐体无变位和倾斜角为α=**4.10**的纵向变位两种情况做了实验，实验数据如附件**3**所示。请分别绘出罐体变位前后油位高度间隔为**1cm**的罐容表数据散点图和曲线拟合图，并尝试用拟合方法给出曲线方程。

（**2**）请利用罐体变位后在进/出油过程中的实际检测数据（附件**4**），绘出罐体变位后油位高度间隔为**10cm**的罐容表标定数据散点图和曲线拟合图，并给出曲线方程。

Q & A

- 有什么问题吗？