设有实验数据 $(x_i, y_i),(i=1,2,\cdots,n)$,寻找函数 $f(x,\theta)$

使得函数在点 $x_i,(i=1,2,\cdots,n)$ 处的函数值与观测数据偏差的平方和达到最小.即求满足如下条件的函数 $f(x,\hat{\theta})$ 使得

$$min\sum_{i=1}^{n}(f(x_i,\theta)-y_i)^2=\sum_{i=1}^{n}(f(x_i,\hat{\theta})-y_i)^2$$

其中 θ 是待定的参数，而 $\hat{\theta}$ 就是最小二乘法所确定的最佳参数.

解决此类问题有以下几个步骤：（**1**）首先作出散点图，确定函数的类别；（**2**）根据已知数据确定待定参数的初始值，利用**Matlab**软件计算最佳参数；（**3**）根据可决系数，比较拟合效果，计算可决系数的公式为

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

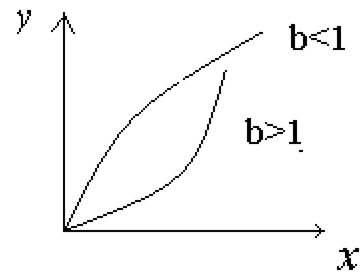
其中 $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ **R^2越趋近于1表明拟合效果越好.**

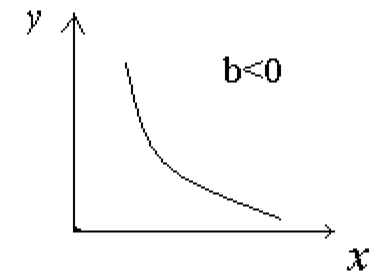
在Matlab中实现可决系数的计算的例子：

```
x=[2:16];
y=[6.42,8.2,9.58,9.5,9.7,10,9.93,9.99,10.49,10.59,10.6,10.8,10.6,10.9,10.76];
y1=x./(0.1152+0.0845*x); % 拟合曲线
R2=1-sum((y-y1).^2)/sum((y-mean(y)).^2)
```

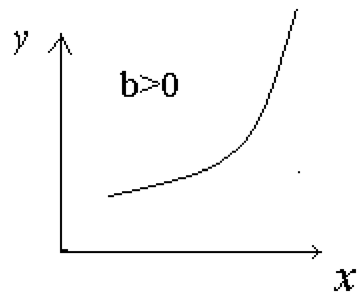
如果是多项式函数，则称为多项式回归，此时的参数即多项式的系数；如果为指数函数、对数函数、幂函数或三角函数等，则称为非线性拟合.下面的图形给出了常见曲线与方程的对应关系：

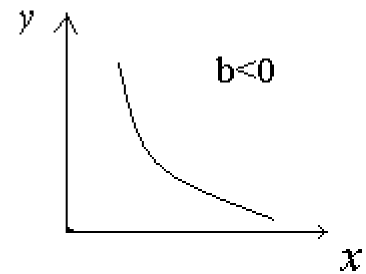
幂函数 $y = ax^{b}$



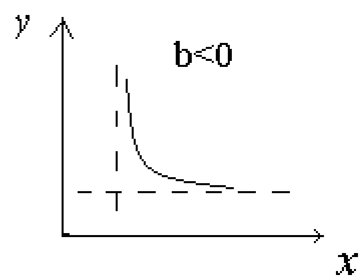


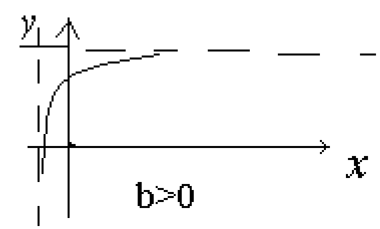
指数函数 $y = ae^{bx}$



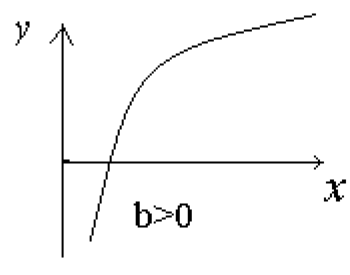


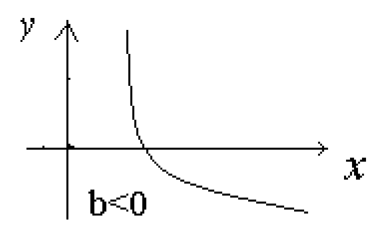
双曲线函数 $y = \dfrac{x}{ax + b}$



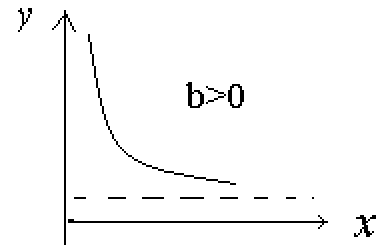


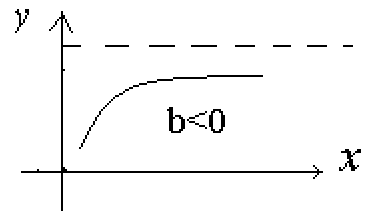
对数函数 $y = a + b\ln x$



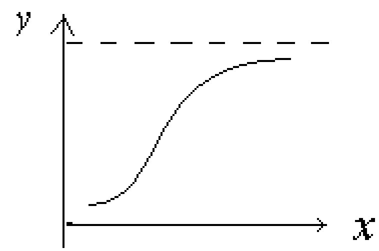


指数函数 $y = ae^{\frac{b}{x}}$





S形曲线 $y = \dfrac{1}{a + be^{-x}}$



具有S形曲线的常见方程有：

罗杰斯蒂（**logistic**）模型： $y = \dfrac{\alpha}{1 + \beta e^{-\gamma x}}$

龚帕兹（**Gomperty**）模型： $y = \alpha \exp(-\beta e^{-kx})$

理查德（**Richards**）模型： $y = \alpha / [1 + \exp(\beta - \gamma x)]^{1/\delta}$

威布尔(**Weibull**)模型： $y = \alpha - \beta \exp(-\gamma t^{\delta})$

为了实现非线性拟合，首先要定义在线函数

1. inline 定义的函数：用于曲线拟合、数值计算

步骤：(1)建立M文件；

(2)fun=inline('f(x) ', '参变量'，'x')

例1. 建立函数：$y = a(1 - be^{-cx})$ a,b,c为待定的参数

fun=inline('b(1)*(1-b(2)*exp(-b(3)*x))','b','x');

此处，将b看成参变量，b(1),b(2),b(3)为其分量.

若计算函数在x=0:0.1:1上的函数值，由于此时x为矩阵，只需将函数表达式中的某些量表示成向量有些*改成.*即可.

在实际问题中，有时散点图作出后未必是多项式的图形，可能像其他的曲线，这时可以猜测曲线类型，然后利用如下命令：

```
[beta,r,J] = nlinfit(x,y,fun,beta0)
```

其中，x,y为原始数据，fun是在M文件中定义的函数，beta0是函数中参数的初始值；beta为参数的最优值，r是各点处的拟合残差，J为雅克比矩阵的数值.

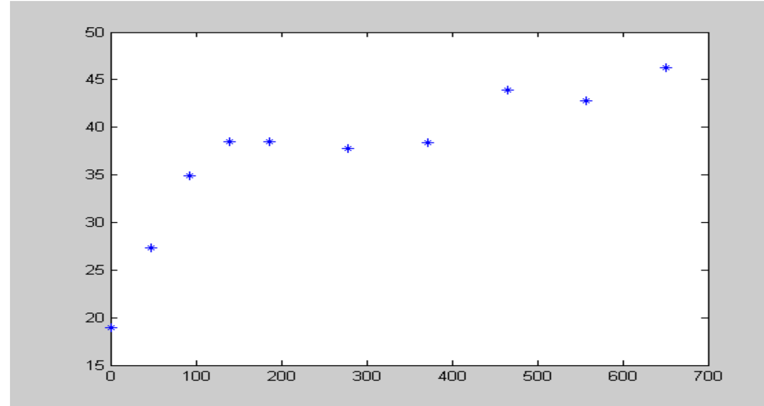
例2. 已知如下数据，求拟合曲线

```
k=[ 0,47,93,140,186,279,372,465,558,651];

y=[18.98,27.35,34.86,38.52,38.44,37.73,38.43,43.87,42.77,46.22];

plot(k,y,'*')
```

根据右图，我们猜测曲线为：

$$y = b_1(1 - b_2 e^{-b_3 k})$$

现在利用最小二乘法确定最佳参数：b1,b2,b3



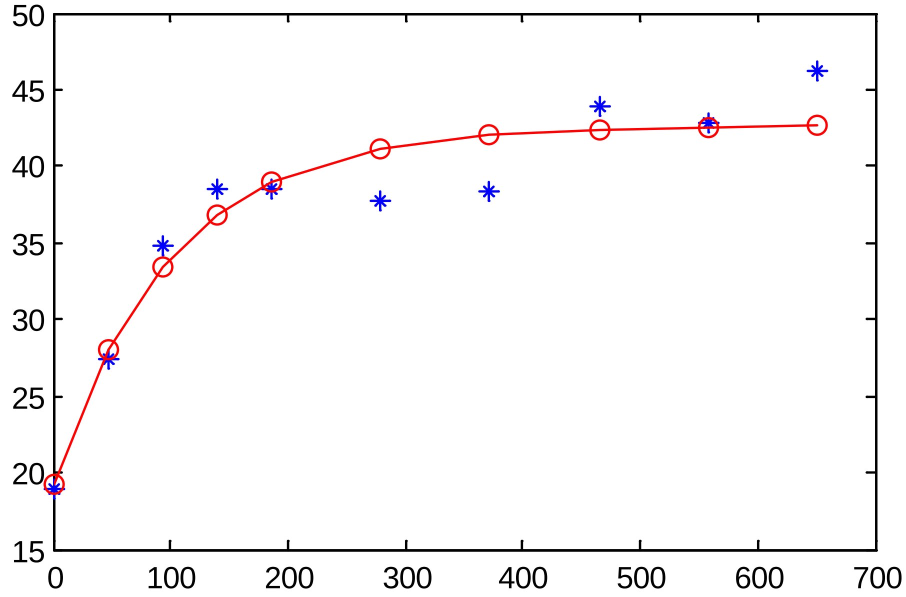
(图6.3)

```
b0=[43,0.6,0.1];                    %初始参数值
fun=inline('b(1)*(1-b(2)*exp(-b(3)*k))','b','k');
[b,r,j]=nlinfit(k,y,fun,b0);
b                                   %最佳参数
R=sum(r.^2)                         %误差平方和
```

b = 42.6643，0.5483，0.0099

即拟合曲线为： $y = 42.6643(1-0.5483e^{-0.0099k})$

拟合结果如右图所示，红色为拟合曲线图形，*为原始散点图.



(图6.4)

作图程序为：

```
 y1=42.6643*(1-0.5483*exp(-0.0099*k));
plot(k,y,'*',k,y1,'-or')
```

练习：计算可决系数

例3.炼钢厂出钢时所用盛钢水的钢包，由于钢水对耐火材料的侵蚀，容积不断增大，我们希望找出使用次数与增大容积之间的函数关系. 实验数据如下：

表4.2 钢包使用次数与增大容积

使用次数	2	3	4	5	6	7	8	9
增大容积	6.42	8.2	9.58	9.5	9.7	10	9.93	9.99
使用次数	10	11	12	13	14	15	16	
增大容积	10.49	10.59	10.6	10.8	10.6	10.9	10.76	

分别选择函数 $y=\dfrac{x}{ax+b}$ $y=a(1+be^{cx})$ $y=ae^{\frac{b}{x}}$

拟合钢包容积与使用次数的关系 ,在同一坐标系内作出函数图形.

下面给出分式函数拟合程序：

```
x1=[2:16];
y1=[6.42,8.2,9.58,9.5,9.7,10,9.93,9.99,10.49,10.59,10.6,10.8
,10.6,10.9,10.76];
b01=[0.1435,0.084];   %初始参数值
fun1=inline('x./(b(1)+b(2)*x)','b','x'); % 定义函数
[b1,r1,j1]=nlinfit(x1,y1,fun1,b01);
y=x1./(0.1152+0.0845*x1); %根据b1写出具体函数
plot(x1,y1,'*',x1,y,'-or');
```

可决系数计算：

初始参数b0的计算， 由于确定两个参数值，因此我们选择已知数据中的两点（2,6.42）和（16,10.76）代入方程，得到方程组：

$$\begin{cases} 6.42 = \dfrac{2}{2a+b} \\ 10.76 = \dfrac{16}{16a+b} \end{cases} \Rightarrow \begin{cases} 6.42(2a+b) = 2 \\ 10.76(16a+b) = 16 \end{cases} \Rightarrow \begin{cases} a = 0.084 \\ b = 0.1435 \end{cases}$$

上述方程组有两种解法：手工，**Matlab,**下面介绍**Matlab** 解方程组的方法

```
[x,y]=solve('6.42*(2*a+b)=2','10.76*(16*a+b)=16')
```

$$y = a(1+be^{cx}) \Leftrightarrow \ln \boldsymbol{b} + \boldsymbol{cx} = \ln(\boldsymbol{y}/\boldsymbol{a}-1)$$

取点：**(2,6.42),(8,9.93),(10,10.49)**代入上述方程

```
[a,b,c]=solve('log(b)+c*2=log(6.42/a-1)','log(b)+c*10=log(10.49/a-1)','log(b)+c*8=log(9.93/a-1)')
```

注意：如果出现复数解，则只取实部