数学建模模型算法精讲课——

非线性回归分析

—— 江北老师

一个胜利者不会放弃，

而一个放弃者永远不会胜利

非线性回归

- 配曲线求解
- 多项式回归求解
- 一般非线性回归求解
- 具体代码

非线性回归模型

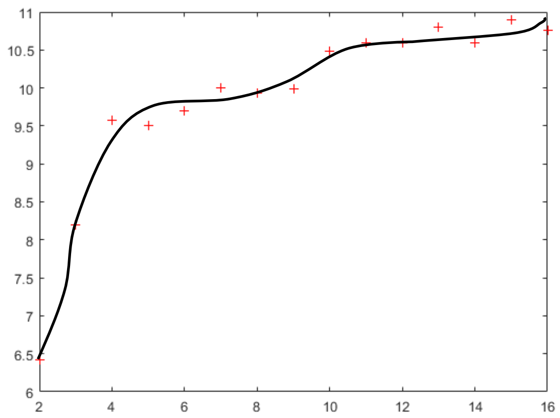
- 出钢时所用的盛钢水的钢包，由于钢水对耐火材料的侵蚀，容积不断增大。我们希望知道使用次数与增大的容积之间的关系，对一钢包作试验，测得的数据列于下表：

使用次数	增大容积	使用次数	增大容积
1	/	9	9.99
2	6.42	10	10.49
3	8.20	11	10.59
4	9.58	12	10.60
5	9.50	13	10.80
6	9.70	14	10.60
7	10.00	15	10.90
8	9.93	16	10.76

非线性回归——配曲线求解

➢ 非线性回归模型

- 画出散点图



- 显然不是线性回归，此即非线性回归或曲线回归问题（需要**配曲线**）
- 配曲线的一般方法是:

  （1）画出散点图;

  （2）根据散点图确定须配曲线的类型;

  （3）估计其中的未知参数.

非线性回归模型

- 由散点图我们可以选配**倒指数曲线**

$$y = ae^{\frac{b}{x}}$$

- 取对数后得

$$lny = lna + \frac{b}{x}lne = lna + \frac{b}{x}$$

- 设$\hat{y}' = lny\,, x' = \frac{1}{x}\,, \hat{b} = lna\,, \hat{A} = b$

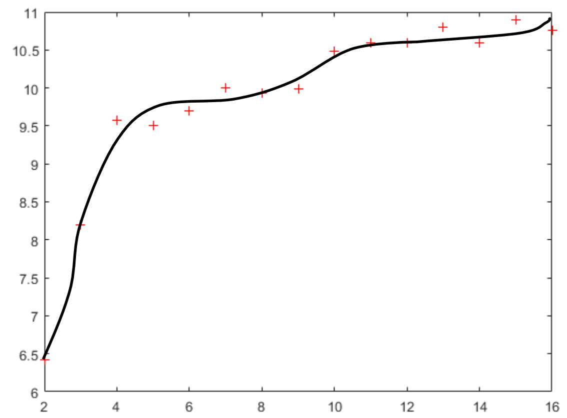
得$\hat{y} = \hat{b} + \hat{A}x'$

- 利用线性回归的方法，算得

$\hat{A} = -1.1107\,, \hat{b} = 2.4587$

$a = e^{\hat{b}} = 11.6789$

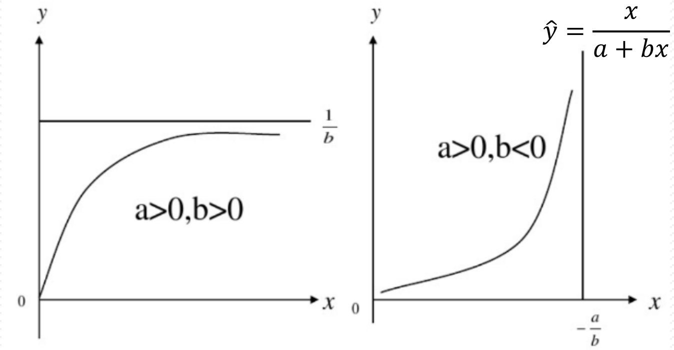
- 所以回归模型为：$y = 11.6789e^{-\frac{1.1107}{x}}$



➢ 常见的六类曲线如下:

（1）**双曲函数曲线:** 变形双曲线

$$\begin{cases} \hat{y} = \dfrac{x}{a+bx} \\ \hat{y} = \dfrac{a+bx}{x} \\ \hat{y} = \dfrac{1}{a+bx} \end{cases}$$



- $\hat{y} = \frac{x}{a+bx}$该曲线通过原点(0,0)，当 $a > 0$、$b > 0$时，y随x的增大而增大，但速率趋小，并向$y = \frac{1}{b}$渐进，是凸曲线；当 $a > 0$、$b < 0$时，y随x的增大而增大，速率趋大，并向$x = -\frac{a}{b}$渐进，是凹曲线
- 变换方式:

$\hat{y} = \frac{x}{a+bx}$，两边取倒数后，令$y' = \frac{x}{\hat{y}}$, 得$y' = ax + b$

$\hat{y} = \frac{a+bx}{x}$, 令$y' = \hat{y}x$, 得$y' = ax + b$；$\hat{y} = \frac{1}{a+bx}$, 两边取倒数后, 令$y' = \frac{1}{y}$, 得$y' = ax + b$

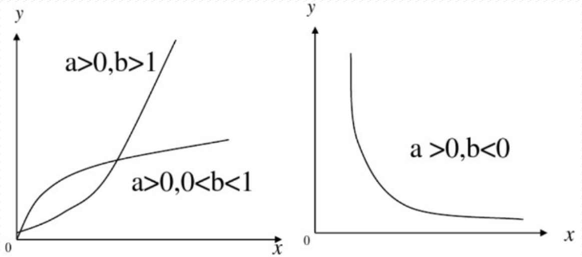
MATH 数模加油站

- **常见的六类曲线如下：**

（2）**幂函数曲线**

- 幂函数（y是x某次幂的函数）方程形式:

$$\hat{y} = ax^b$$



当 $a > 0$、$b > 1$时，y随x的增大而增大（增长），是凹曲线;
当 $a > 0$、$0 < b < 1$时，y随x的增大而增大（增长），但变化缓慢，是凸曲线;
当 $a > 0$、$b < 0$时，y随x的增大而减小，且以x,y轴为渐近线，是凹曲线。

- 变换方式：两边取对数，令$y' = ln\hat{y}, x' = lnx, a' = lna$, 得$y' = a' + bx'$

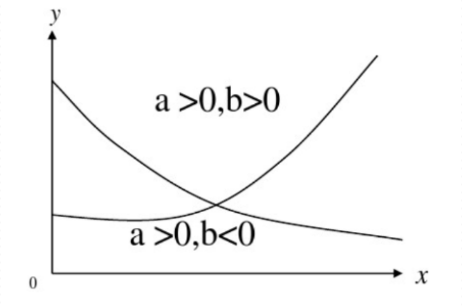
MATH 数模加油站

➢ **常见的六类曲线如下：**

（3）**指数函数曲线**

- 指数函数（x 作为指数出现）方程形式：

$$\begin{cases} \hat{y} = ae^{bx} \\ \hat{y} = ab^{x} \end{cases}$$



- $\hat{y} = ae^{bx}$

  参数b一般用来描述增长或衰减的速度。
  当 $a > 0$、$b > 0$时，y随x的增大而增大（增长），是凹曲线；
  当 $a > 0$、$b < 0$时，y随x的增大而减小（衰减），是凹曲线。

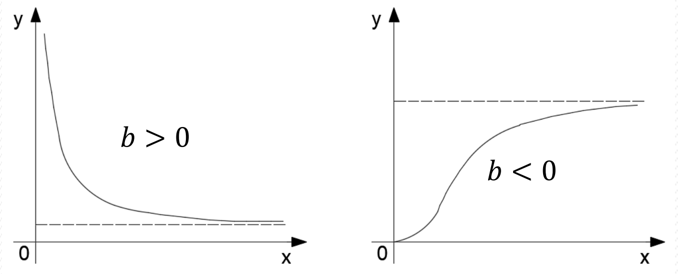
- 变换方式：两边取对数，令$y' = ln\hat{y}, a' = lna$, 得$y' = a' + bx$

- **常见的六类曲线如下:**

（4）**倒指数曲线**

- 指数函数（x 作为指数出现）方程形式:

$$\hat{y} = ae^{\frac{b}{x}}, 其中 a > 0$$



当 $a > 0$、$b > 0$时，y随x的增大而减小（衰减），是凹曲线;
当 $a > 0$、$b < 0$时，y随x的增大而增大（增长），是先凹后凸曲线。

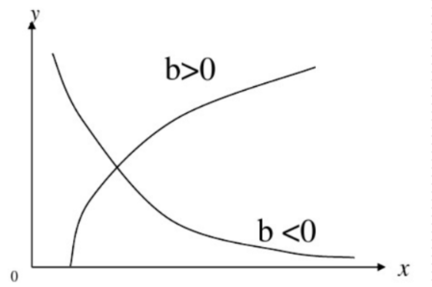
- 变换方式：两边取对数，令$y' = ln\hat{y}, a' = lna, x' = \frac{1}{x}$，得$y' = a' + bx'$

➢ **常见的六类曲线如下：**

（5）**对数函数曲线**

- 对数函数（x 作为自然对数出现）方程形式：

$$\hat{y} = a + blnx(x > 0)$$



- 对数函数表示：x变数的较大变化可引起y变数的较小变化。
  $b > 0$时，y随x的增大而增大，是凸曲线；
  $b < 0$时，y随x的增大而减小，是凹曲线。

- 变换方式：令$x' = lnx$，得$\hat{y} = a + bx'$

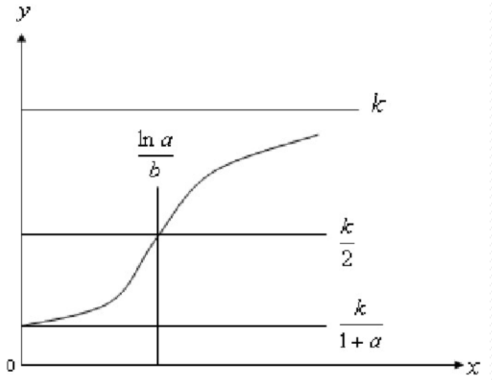
➤ 常见的六类曲线如下：

（6）S型曲线

- 主要描述动、植物的**自然生长过程**，又称生长曲线，也可以描述传染病的**发展趋势**。
- 生长过程的基本特点是开始增长较慢，而在以后的某一范围内迅速增长，达到一定的限度后增长又缓慢下来，曲线呈拉长的‘S’型曲线。著名的‘S’型曲线是$Logistic$生长曲线。

$$\hat{y}=\frac{k}{1+ae^{-bx}}(a、b、k均大于0)$$

$$x=0,\hat{y}=\frac{k}{1+a};\ x\to\infty,\hat{y}=k$$



- 所以时间为0的起始量为$\frac{k}{1+a}$，时间为无限延长的，终极量为k

曲线$x=\frac{lna}{b}$时有一个拐点，这时$\hat{y}=\frac{k}{2}$，恰好是终极量的一半

拐点左侧，为凹曲线，速率由小趋大；拐点右侧，为凸曲线速率由大趋小。

- 变换方式：两边取倒数再取对数后，$y'=ln(\frac{k-\hat{y}}{\hat{y}}),a'=lna,$得$y'=a'+bx$

非线性回归——多项式回归

- **当六类曲线都配不上时，怎么办?**
  - 多项式回归

    设变量x、Y的回归模型为

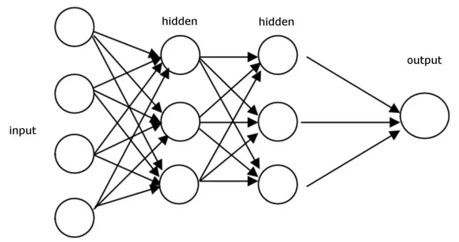
    $$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_p x^p + \epsilon$$

    其中p是已知的，$\beta_i (i = 1,2,\cdots,p)$是未知参数，$\epsilon$服从正态分布$N(0,\sigma)$

    $$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_p x^p$$

    称为回归多项式，上面的回归模型称为多项式回归。
  - 其他可能的方法

    比如神经网络、支持向量机



➢ 多项式回归——一元多项式回归

$$y = a_1x^m + a_2x^{m-1} + \cdots + a_mx + a_{m+1}$$

- 1、回归

1）确定多项式系数的命令：$[p, S] = polyfit(x, y, m)$

其中$x = (x_1, x_2, \ldots, x_n), y = (y_1, y_2, \ldots, y_n)$;

$p = (a_1, a_2, \cdots, a_{m+1})$是多项式$y = a_1x^m + a_2x^{m-1} + \cdots + a_mx + a_{m+1}$的系数;

S是一个结构数据，用来估计预测误差

2）一元多项式回归命令：$polytool(x, y, m)$

- 2、预测和预测误差估计:

1）$Y = polyval(p, x)$ 求$polyfit$所得的回归多项式在x处预测值Y;

2）$[Y, DELTA] = polyconf(p, x, S, alpha)$

求$polyfit$所得的回归多项式在x处的预测值Y及预测值的显著性为$alpha$的置信区间$Y \pm DELTA$，$alpha$缺省时为0.05

➢ **例1、观测物体降落的距离s与时间t的关系，得到数据如下表，求s关于t的回归方程**

$$\hat{s} = a + bt + ct^2$$

$t(s)$	1/30	2/30	3/30	4/30	5/30	6/60	7/30
$s(cm)$	11.86	15.67	20.60	26.69	33.71	41.93	51.13
$t(s)$	8/30	9/30	10/.30	11/30	12/30	13/30	14/30
$s(cm)$	61.49	72.90	85.44	99.08	113.77	129.54	146.48

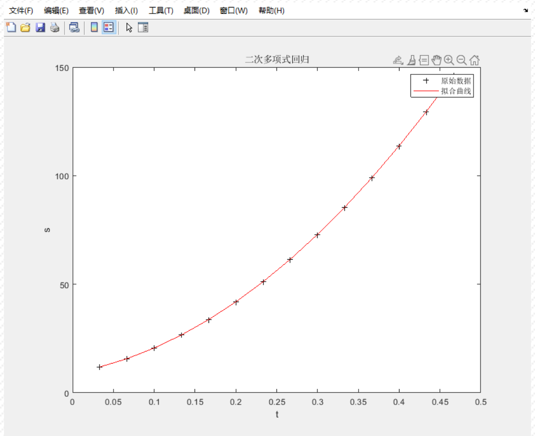
- 1、方法一
  - ✓ 直接作二次多项式回归：

    $t = 1/30:1/30:14/30;$

    $s = [11.86\ 15.67\ \ 20.6\ \ 26.69\ 33.71\ \ 41.93\ 51.13\ 61.49\ 72.9\ 85.44\ 99.08\ \ 113.77\ 129.54\ 146.48]$

    $[p,S] = polyfit(t,s,2)$
  - ✓ 得回归模型为：$\hat{s} = 489.2946t^2 + 65.8896t + 9.1329$
  - ✓ 预测及作图：$Y = polyconf(p,t,S)$

    $plot(t,s,'k+',t,Y,'r')$



非线性回归——多项式回归

- 例1、观测物体降落的距离s与时间t的关系，得到数据如下表，求s关于t的回归方程

$$\hat{s} = a + bt + ct^2$$

$t(s)$	1/30	2/30	3/30	4/30	5/30	6/60	7/30
$s(cm)$	11.86	15.67	20.60	26.69	33.71	41.93	51.13
$t(s)$	8/30	9/30	10/.30	11/30	12/30	13/30	14/30
$s(cm)$	61.49	72.90	85.44	99.08	113.77	129.54	146.48

- 2、方法二
  - ✓ 化为多元线性回归

$t = 1/30{:}\,1/30{:}\,14/30;$

$s = [11.86\ \ 15.67\ \ 20.6\ \ 26.69\ \ 33.71\ \ 41.93\ \ 51.13\ \ 61.49$

$72.9\ \ 85.44\ \ 99.08\ \ 113.77\ \ 129.54\ \ 146.48];$

$T = [\mathrm{ones}(14\,,1)\ \ \mathrm{t}'\ \ (\mathrm{t}.\,\hat{}\,2)'];$

$[b\,,bint\,,r\,,rint\,,stats] = regress(s',T);$

$b, stats$

  - ✓ 得回归模型为：$\hat{s} = 489.2946t^2 + 65.8896t + 9.1329$

数模加油站

多项式回归——多元二项式回归

- 命令：$rstool(x, y, model', alpha)$

x——n(样本个数) $\times$ m(自变量个数)矩阵

y——n (样本个数)维列向量

$alpha$——显著性水平（默认0.05）

$model$——由下列4个模型中选择1个（用字符串输入，缺省时为线性模型）

$linear$(线性)：$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m$

$purequadratic$(纯二次)：$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m + \sum_{j=1}^{n} \beta_{jj} x_j^2$

$interaction$(交叉)：$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m + \sum_{1 \le j \ne k \le m} \beta_{jk} x_j x_k$

$quadratic$(完全二次)：$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m + \sum_{1 \le j, k \le m} \beta_{jk} x_j x_k$

MATH 数模加油站

- 例2、设某商品的需求量与消费者的平均收入、商品价格的统计数据如下，建立回归模型，预测平均收入为1000、价格为6时的商品需求量

需求量	100	75	80	70	50	65	90	100	110	60
收入	1000	600	1200	500	300	400	1300	1100	1300	300
价格	5	7	6	6	8	7	5	4	3	9

选择纯二次模型，即$y=\beta_0+\beta_1x_1+\beta_2x_2+\beta_{11}x_1^2+\beta_{22}x_2^2$

- 1、方法一
  - ✓ 直接用多元二项式回归：

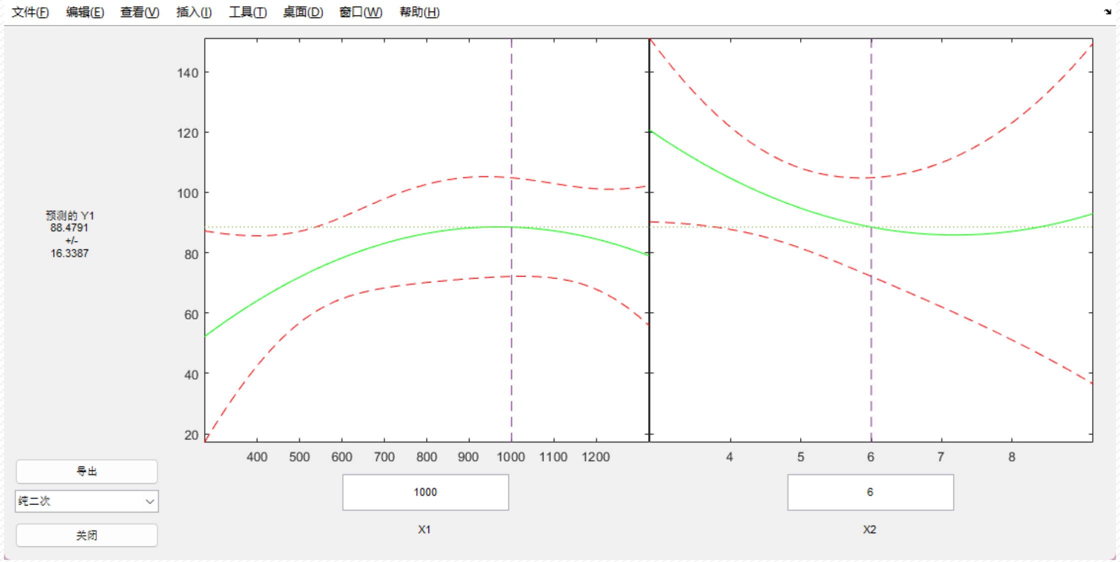
$x1=[1000\ 600\ 1200\ 500\ 300\ 400\ 1300\ 1100\ 1300\ 300];$

$x2=[5\ 7\ 6\ 6\ 8\ 7\ 5\ 4\ 3\ 9];y=[100\ 75\ 80\ 70\ 50\ 65\ 90\ 100\ 110\ 60]';$

$x=[x1'\ \ x2'];$

$rstool(x,y,'purequadratic')$

非线性回归——多项式回归



✓ 在左边图形下方的方框中输入1000，右边图形下方的方框中输入6。

则画面左边的“*Predicted Y*”下方的数据变为88.47981，即预测出平均收入为1000、价格为6时的商品需求量为88.4791。在画面左下方的点击导出，则*beta*、*rmse*和*residuals*都传送到*Matlab*工作区中。

✓ 在Matlab工作区中输入命令：$beta, rmse$

得结果：

$beta = 110.5313 \quad 0.1464 \quad -26.5709 \quad -0.0001 \quad 1.8475$

$rmse = 4.5362$

✓ 故回归模型为：$y = 110.5313 + 0.1464x_1 - 26.5709x_2 - 0.0001x_1^2 + 1.8475x_2^2$

✓ 剩余标准差仅为4.5362，说明此回归模型的预测效果较好

- 2、方法二

✓ 将$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^2 + \beta_{22} x_2^2$，化为多元线性回归：

$X = [ones(10,1) \quad x1' \quad x2' \quad (x1.\hat{}2)' \quad (x2.\hat{}2)'];$

$[b, bint, r, rint, stats] = regress(y, X);$

$b, stats$

结果为：

$b = 110.5313 \quad 0.1464 \quad -26.5709 \quad -0.0001 \quad 1.8475$

$stats = 0.9702 \quad 40.6656 \quad 0.0005$

➢ 一般非线性回归的MATLAB实现

- 1、回归命令

✓ 确定回归系数的命令:

$[beta, r, J] = nlinfit(x, y, 'model', beta0)$

$beta$——估计出的回归系数

r——残差

J——Jacobian矩阵

x, y——输入数据x, y分别为$n \times m$矩阵和n维列向量，对一元非线性回归，x为n维列向量

'model'——是事先用$m-$文件定义的非线性函数

$beta0$——回归系数的初值

✓ 非线性回归命令：$nlintool(x, y, 'model', beta0, alpha)$

- 2、预测和预测误差估计

$[Y, DELTA] = nlpredci('model', x, beta, r, J)$

求$nlinfit$或$nlintool$所得的回归函数在x处的预测值Y及预测值的显著性为$alpha$的置信区间$Y \pm DELTA$

非线性回归——一般求解

一般非线性回归求解

- 例3、出钢时所用的盛钢水的钢包，由于钢水对耐火材料的侵蚀，容积不断增大。我们希望知道使用次数与增大的容积之间的关系，对一钢包作试验，测得的数据列于下表：

使用次数	增大容积	使用次数	增大容积
1	/	9	9.99
2	6.42	10	10.49
3	8.20	11	10.59
4	9.58	12	10.60
5	9.50	13	10.80
6	9.70	14	10.60
7	10.00	15	10.90
8	9.93	16	10.76

- 由散点图我们可以选配倒指数曲线 $y = ae^{\frac{b}{x}}$

非线性回归——一般求解

➢ 一般非线性回归求解

- 1、对将要拟合的非线性模型$y=ae^{\frac{b}{x}}$，建立$m-$文件$volum.m$如下：

  $function\ \ yhat=volum(beta,x)$

  $yhat=beta(1)*\exp(beta(2)./x)$

- 2、输入数据：

  $x=2:16;$

  $y=[6.42\ \ 8.20\ \ 9.58\ \ 9.5\ \ 9.7\ \ 10\ \ 9.93\ \ 9.99\ \ 10.49\ \ 10.59\ \ 10.60\ \ 10.80\ \ 10.60\ \ 10.90\ \ 10.76]';$

  $beta0=[8\ \ 2]';$

- 3、求回归系数：

  $[beta,r,J]=nlinfit(x',y,'volum',beta0);$

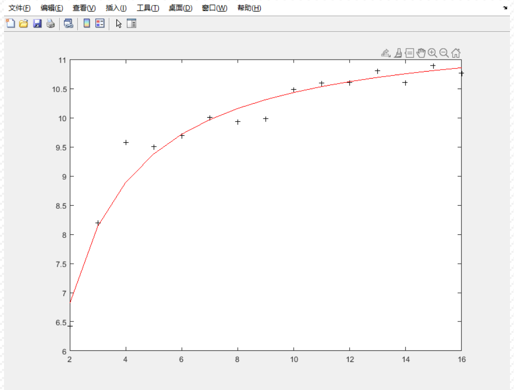
  $beta$

  得结果：$beta=11.6036\quad -1.0641$

  即得回归模型为：$y=11.6036e^{-\frac{1.10641}{x}}$

- 4、预测及作图：$[YY,delta]=nlpredci('volum',x',beta,r,J);$

  $plot(x,y,'k+',x,YY,'r')$



- **例1代码：**

```
% 1、二次多项式回归
t = 1/30:1/30:14/30; % 自变量 t
s = [11.86 15.67 20.6 26.69 33.71 41.93 51.13 61.49 72.9 85.44 99.08 113.77 129.54 146.48]; % 因变量 s
[p, S] = polyfit(t, s, 2); % 进行二次多项式拟合

% 预测及作图
Y = polyconf(p, t, S); % 计算拟合曲线的预测值
plot(t, s, 'k+', t, Y, 'r'); % 绘制原始数据点和拟合曲线
xlabel('t'); % 设置 x 轴标签
ylabel('s'); % 设置 y 轴标签
title('二次多项式回归'); % 设置图表标题
legend('原始数据', '拟合曲线'); % 设置图例

% 2、化为多元线性回归
t = 1/30:1/30:14/30; % 自变量 t
s = [11.86 15.67 20.6 26.69 33.71 41.93 51.13 61.49 72.9 85.44 99.08 113.77 129.54 146.48]; % 因变量 s
% 令x1=t, x2=t^2
T = [ones(14, 1), t', (t.^2)']; % 构建多元线性回归的自变量矩阵
[b, bint, r, rint, stats] = regress(s', T); % 进行多元线性回归分析
```

MATH 数模加油站

➤ **例2代码:**

```
% 1、直接用多元二项式回归
x1=[1000 600 1200 500 300 400 1300 1100 1300 300] ;
x2=[5 7 6 6 8 7 5 4 3 9];y=[100 75 80 70 50 65 90 100 110 60]';
x=[x1' x2' ];
rstool(x ,y,'purequadratic') % 使用多元二项式回归工具箱进行回归分析
beta, rmse % 回归系数和剩余标准差
% beta=110.5313 0.1464 -26.5709 -0.0001 1.8475
% rmse=4.5362
% 故回归模型为: y=110.5313+0.1464x_1-26.5709x_2-0.0001x_1^2+1.8475x_2^2
% 剩余标准差仅为4.5362，说明此回归模型的预测效果较好

% 2、化为多元线性回归
% 令x3=x1^2 x4=x2^2
X=[ones(10,1) x1' x2' (x1.^2)' (x2.^2)']; % 多元线性回归的自变量矩阵
[b,bint,r,rint,stats]=regress(y,X); % 进行多元线性回归分析
b,stats % b为回归系数，stats为统计信息
% 结果为:
% b=110.5313 0.1464 -26.5709 -0.0001 1.8475
% stats=0.9702 40.6656 0.0005
```

- **例3代码：**

```
% 1) 输入数据
x=2:16;
y=[6.42 8.20 9.58 9.5 9.7 10 9.93 9.99 10.49 10.59 10.60 10.80 10.60 10.90 10.76];
beta0=[8 2]'; %设定初始值，根据经验选择
% 2) 求回归系数
[beta,r,J]=nlinfit(x',y','volum',beta0);
% 使用非线性最小二乘法进行回归拟合，'volum'为定义曲线函数的函数名
beta  % 回归系数

%3) 预测及作图
[YY,delta]=nlpredci('volum',x',beta,r,J); % 预测因变量的值及置信区间
plot(x,y,'k+',x,YY,'r')   % 绘制原始数据点和拟合曲线
```

欢迎关注数模加油站

THANKS