数学建模中的数据处理方法

范筑军

主要内容

* 曲线插值与拟合
* 数值微分与积分
* 微分方程数值解
* 优化问题
* 回归分析
* 判别分析

曲线插值与拟合

* **一维插值**
* **二维插值**
* **曲线拟合**

一维插值

* **对表格给出的函数，求出没有给出的函数值。**
* 在实际工作中，经常会遇到插值问题。
* 下表是待加工零件下轮廓线的一组数据，现需要得到x坐标每改变0.1时所对应的y的坐标.

x	0	3	5	7	9	11	12	13	14	15
y	0	1.2	1.7	2.0	2.1	2.0	1.8	1.2	1.0	1.6

一维插值

- 下面是关于插值的两条命令(专门用来解决这类问题):
- y=interp1(x0,y0,x,’method’)　分段线性插值
- y=spline(x0,y0,x)　三次样条插值
- x0,y0是已知的节点坐标，是同维向量。
- y对应于x处的插值。y与x是同维向量。
- method可选’nearest’(最近邻插值),’linear’(线性插值),’spline’(三次样条插值),’cubic’(三次多项式插值)

一维插值

* 解决上述问题,我们可分两步:
  * 用原始数据绘图作为选用插值方法的参考.
  * 确定插值方法进行插值计算

一维插值(px_lc11.m)

- 对于上述问题,可键入以下的命令:
- x0=[0,3,5,7,9,11,12,13,14,15]';
- y0=[0,1.2,1.7,2.0,2.1,2.0,1.8,1.2,1.0,1.6]'
- plot(x0,y0)   %完成第一步工作
- x=0:0.1:15;
- y=interp1(x0,y0,x');   %用分段线性插值完成第二步工作
- plot(x,y)
- y=spline(x0,y0,x');
- plot(x,y)   %用三次样条插值完成第二步工作

练习

1. 对$y=1/(1+x^2)$,$-5\le x\le 5$，用n（=11）个节点（等分）作上述两种插值，用m（=21）个插值点（等分）作图，比较结果。(see:px_ex_lc1.m)
2. 在某处测得海洋不同深度处水温如下表：求深度为500、1000、1500米处的水温。(see:px_ex_lc2.m)

深度	446	714	950	1422	1634
水温	7.04	4.28	3.40	2.54	2.13

二维插值

* MATLAB中二维插值的命令是：
* z=interp2（x0,y0,z0,x,y,'meth'）

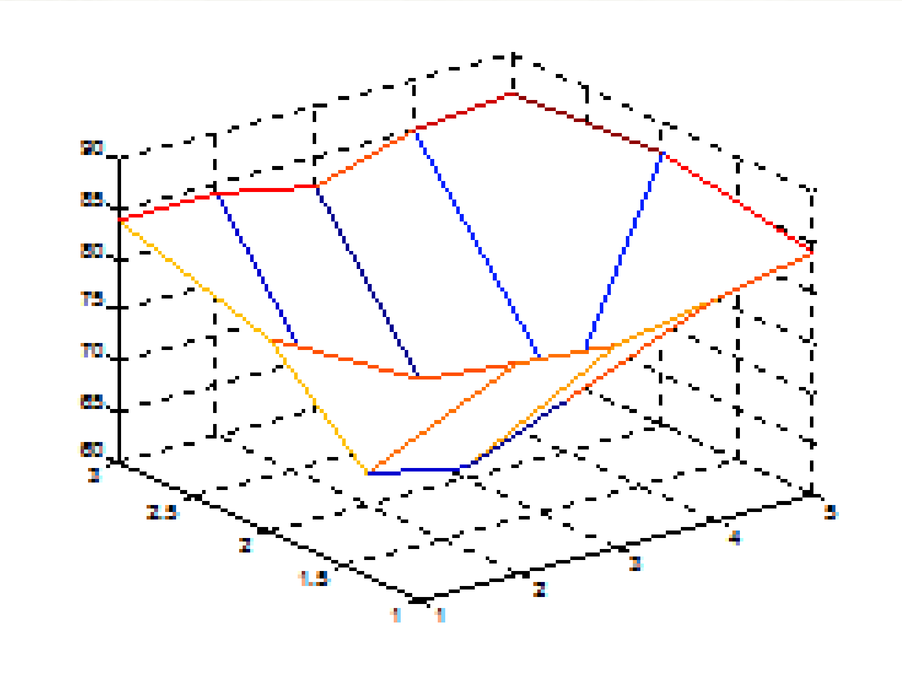
二维插值

* 在一个长为5个单位，宽为3个单位的金属薄片上测得15个点的温度值，试求出此薄片的温度分布，并绘出等温线图。（数据如下表）

y_i \ x_i	1	2	3	4	5
1	82	81	80	82	84
2	79	63	61	65	87
3	84	84	82	85	86

二维插值(px_lc21.m)

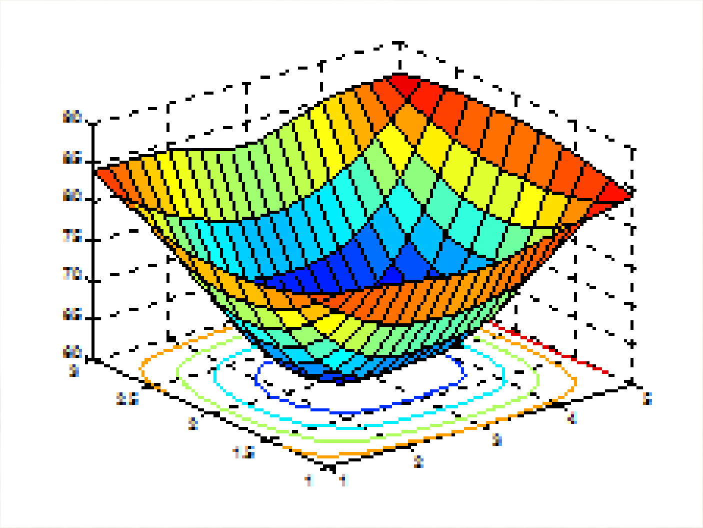
- temps=[82,81,80,82,84;79,63,61,65,87;84,84,82,85,86];
- mesh(temps)　　　%根据原始数据绘出温度分布图，可看到此图的粗造度。

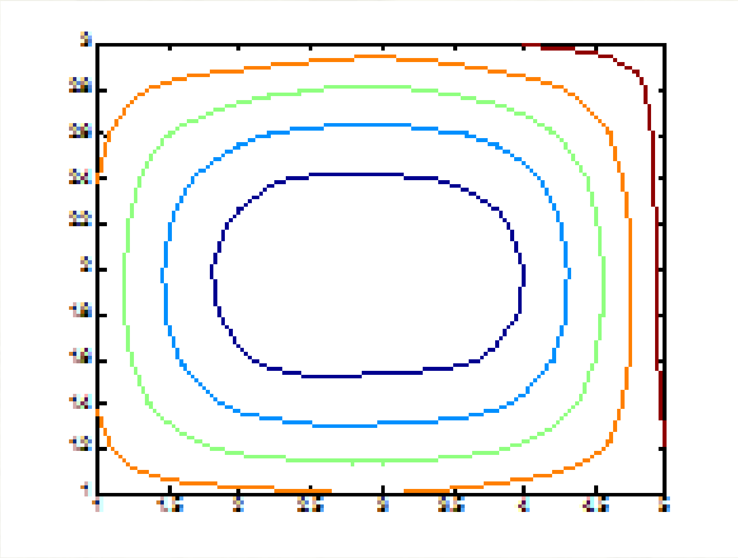


二维插值

- %下面开始进行二维函数的三阶插值。
- width=1:5; depth=1:3; di=1:0.2:3; wi=1:0.2:5;
- [WI,DI]=meshgrid(wi,di);%增加了节点数目
- ZI=interp2(width,depth,temps,WI,DI,'cubic'); % 对数据（width,depth,temps）进
- % 行三阶插值拟合。
- surfc(WI,DI,ZI)
- contour(WI,DI,ZI)

二维插值





曲线拟合

* **假设一函数g(x)是以表格形式给出的，现要求一函数f(x)，使f(x)在某一准则下与表格函数（数据）最为接近。**
* 由于与插值的提法不同，所以在数学上理论根据不同，解决问题的方法也不同。
* 此处，我们总假设f(x)是多项式。

曲线拟合

* 问题：弹簧在力F的作用下伸长x厘米。F和x在一定的范围内服从虎克定律。试根据下列数据确定弹性系数k,并给出不服从虎克定律时的近似公式。

x	1	2	4	7	9	12	13	15	17
F	1.5	3.9	6.6	11.7	15.6	18.8	19.6	20.6	21.1

曲线拟合

- 解题思路：可以用一阶多项式拟合求出k，以及近似公式。
- 在MATLAB中，用以下命令拟合多项式。
- polyfit（x0,y0,n）
- 一般，也需先观察原始数据的图像，然后再确定拟和成什么曲线。

曲线拟合(px_lc31.m)

- 对于上述问题,可键入以下的命令:
- x=[1,2,4,7,9,12,13,15,17]'; F=[1.5,3.9,6.6,11.7,15.6,18.8,19.6,20.6,21.1]';
- plot(x,F,'.')
- 从图像上我们发现:前5个数据应与直线拟合,后5个数据应与二次曲线拟合。于是键入:
- a=polyfit(x(1:5),F(1:5),1);
- a=polyfit(x(5:9),F(5:9),2)

曲线拟合

* 注意：有时，面对一个实际问题，究竟是用插值还是用拟合不好确定，还需大家在实际中仔细区分。同时，大家（包括学过计算方法的同学）注意去掌握相应的理论知识。

数值微分与积分

* 数值积分
* 数值微分

数值积分

- 先看一个例子：
- 现要根据瑞士地图计算其国土面积。于是对地图作如下的测量：以西东方向为横轴，以南北方向为纵轴。（选适当的点为原点）将国土最西到最东边界在x轴上的区间划取足够多的分点x_i，在每个分点处可测出南北边界点的对应坐标y_1，y_2。用这样的方法得到下表
- 根据地图比例知18mm相当于40km，试由上表计算瑞士国土的近似面积。（精确值为41288km^2）。

数值积分

x	7.0	10.5	13.0	17.5	34.0	40.5	44.5	48.0	56.0
y_1	44	45	47	50	50	38	30	30	34
y_2	44	59	70	72	93	100	110	110	110
x	61.0	68.5	76.5	80.5	91.0	96.0	101.0	104.0	106.5
y_1	36	34	41	45	46	43	37	33	28
y_2	117	118	116	118	118	121	124	121	121
x	111.5	118.0	123.5	136.5	142.0	146.0	150.0	157.0	158.0
y_1	32	65	55	54	52	50	66	66	68
y_2	121	122	116	83	81	82	86	85	68

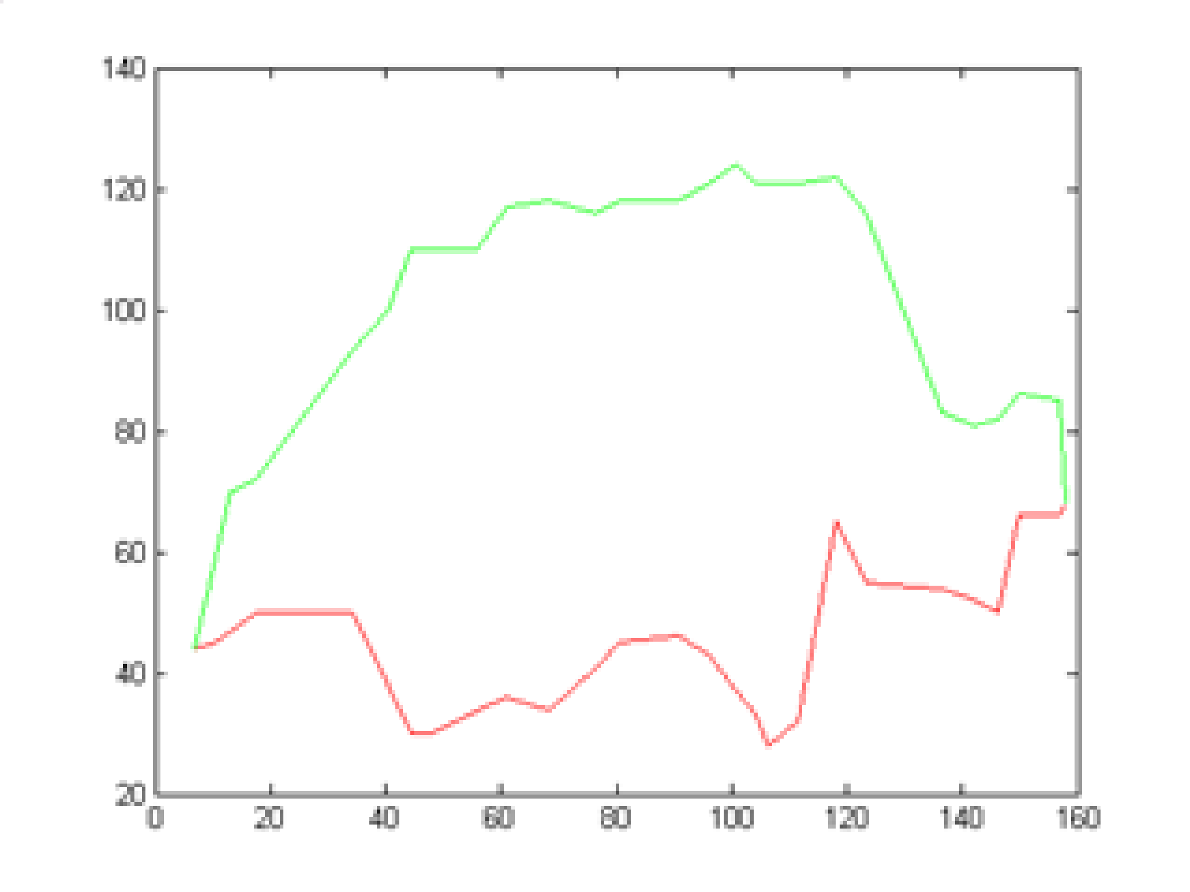
数值积分

* 解题思路：数据实际上表示了两条曲线，实际上我们要求由两曲线所围成的图形的面积。
* 解此问题的方法是数值积分的方法。具体解时我们遇到两个问题：
  * 1。数据如何输入；
  * 2。没有现成的命令可用。

数值积分(px_wj11.m)

- 对于第一个问题，我们可把数据拷贝成M文件（或纯文本文件）。
- 然后，利用数据绘制平面图形。键入
- load mianji.txt
- A=mianji';
- plot(A(:,1),A(:,2),'r',A(:,1),A(:,3),'g')

数值积分



数值积分

- 接下来可以计算面积。键入：
- a1=trapz(A(:,1)*40/18,A(:,2)*40/18);
- a2=trapz(A(:,1)*40/18,A(:,3)*40/18);
- d=a2-a1
- d = 4.2414e+004

数值积分

- 至此，问题可以说得到了解决。
- 之所以说还有问题，是我们觉得误差较大。但计算方法的理论给了我们更精确计算方法。只是MATLAB没有相应的命令。
- 想得到更理想的结果，我们可以自己设计解决问题的方法。（可以编写辛普森数值计算公式的程序，或用拟合的方法求出被积函数，再利用MATLAB的命令quad,quad8）

数值微分

* 已知20世纪美国人口统计数据如下，根据数据计算人口增长率。（其实还可以对于后十年人口进行预测）

年份	1900	1910	1920	1930	1940	1950	1960	1970	1980	1990
人口 $\times 10^6$	76.0	92.0	106.5	123.2	131.7	150.7	179.3	204.0	226.5	251.4

$r(t)=\frac{\dot{x}}{x}$

数值微分

* 解题思路：设人口是时间的函数x(t).于是人口的增长率就是x(t)对t的导数.如果计算出人口的相关变化率$r(t)=\frac{\dot{x}}{x}$。那么人口增长满足$\dot{x}=r(t)x(t)$，它在初始条件x(0)=x0下的解为$x(t)=x0e^{\int_0^t r(u)du}$.（用以检查计算结果的正确性）

数值微分

* 解：此问题的特点是以离散变量给出函数x(t),所以就要用差分来表示函数x(t)的导数.

$$f'(a)\approx\frac{f(a+h)-f(a)}{h},\quad f'(a)\approx\frac{f(a)-f(a-h)}{h},\quad f'(a)\approx\frac{f(a+h)-f(a-h)}{2h}$$

* 常用后一个公式。（因为，它实际上是用二次插值函数来代替曲线x(t)）即常用三点公式来代替函数在各分点的导数值：

$$f'(x_k)\approx\frac{y_{k+1}-y_k}{2h}\quad k=1,2,\cdots n-1$$

$$f'(x_0)\approx\frac{-3y_0+4y_1-y_2}{2h}$$

$$f'(x_n)\approx\frac{y_{n-2}-4y_{n-1}+3y_n}{2h}$$

数值微分

- MATLAB用命令diff按两点公式计算差分;此题自编程序用三点公式计算相关变化率.编程如下(diff3.m):

```
for i=1:length(x)
    if i==1
        r(1)=(-3*x(1)+4*x(1+1)-x(1+2))/(20*x(1));
    elseif i~=length(x)
        r(i)=(x(i+1)-x(i-1))/(20*x(i));
    else
        r(length(x))=(x(length(x)-2)-4*x(length(x)-1)+3*x(length(x)))/(20*x(length(x)));
    end
end
r=r;
```

数值微分

- 保存为diff3.m文件听候调用.再在命令窗内键入
- X=[1900,1910,1920,1930,1940,1950,1960,1970,1980,1990];
- x=[76.0, 92.0, 106.5, 123.2, 131.7, 150.7, 179.3, 204.0, 226.5, 251.4];
- diff3;
- 由于r以离散数据给出,所以要用数值积分计算.键入
- x(1,1)*exp(trapz(X(1,1:9),r(1:9)))
- 数值积分命令：trapz(x),trapz(x,y),quad(‘fun’,a,b)等.

微分方程数值解(单摆问题)

单摆问题的数学模型是

$$l\ddot{\theta} = -g\sin\theta$$

在初始角度不大时,问题可以得到很好地解决,但如果初始角较大,此方程无法求出解析解.现问题是当初始角为10^0和30^0时,求出其解,画出解的图形进行比较。

微分方程数值解(单摆问题)

- 解:若θ_0较小,则原方程可用$l\ddot{\theta}+g\theta=0$ 来近似.其解析解为$\theta(t)=\theta_0\cos\omega t$, $\omega=\sqrt{\frac{l}{g}}$.
- 若不用线性方程来近似,那么有两个模型:

$$\begin{cases}\ddot{\theta}=-\frac{g}{l}\sin\theta\\ \theta(0)=10^0\end{cases}\qquad\qquad\begin{cases}\ddot{\theta}=-\frac{g}{l}\sin\theta\\ \theta(0)=30^0\end{cases}$$

微分方程数值解(单摆问题)

- 取g=9.8,l=25, 10^0=0.1745, 30^0=0.5236.用MATLAB求这两个模型的数值解,先要作如下的处理:令x1=θ,x2=θ',则模型变为

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = -\frac{g}{l}\sin x_1 \\ x_1(0) = 0.1745, x_2(0) = 0 \end{cases} \qquad \begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = -\frac{g}{l}\sin x_1 \\ x_1(0) = 0.5236, x_2(0) = 0 \end{cases}$$

微分方程数值解(单摆问题)

- 再编函数文件(danbai.m)
- function xdot=danbai(t,x)
- xdot=zeros(2,1);
- xdot(1)=x(2);xdot(2)=-9.8/25*sin(x(1));

微分方程数值解(单摆问题)

* 在命令窗口键入()
* [t,x]=ode45(‘danbai’,[0:0.1:20],[0.1745,0]);
* [t,y]=ode45(‘danbai’,[0:0.1:20],[0.5236,0]);
* plot(t,x(:,1),’r’,t,y(:,1),’k’);

优化问题

* 线性规划有约束极小问题
* 非线性规划有约束极小问题
* 非线性无约束极小问题
* 非线性最小二乘问题
* 二次规划

线性规划有约束极小问题

* 模型

$$\min z = cx,$$
$$s.t.\quad Ax \le b,$$
$$A_1 x = b_1,$$
$$LB \le x \le UB$$

* 用命令
* [x, fval]= linprog(f,A,b,A1,b1,lb,ub)

线性规划有约束极小问题

- Find **x** that minimizes
- $f(\mathbf{x})=-5x_1-4x_2-6x_3$
- subject to
- $x_1-x_2+x_3 \leqq 20$
- $3x_1+2x_2+4x_3 \leqq 42$
- $3x_1+2x_2 \leqq 30$
- $0 \leqq x_1,\ 0 \leqq x_2, 0 \leqq x_3$

线性规划有约束极小问题

- First, enter the coefficients:
- f = [-5; -4; -6]
- A = [1 -1 1
- 3 2 4
- 3 2 0];
- b = [20; 42; 30];
- lb = zeros(3,1);
- Next, call a linear programming routine:
- [x,fval,exitflag,output,lambda] = linprog(f,A,b,[],[],lb)；
- Entering x,

fval,lambda.ineqlin, and lambda.lower gets

- x =
- 0.0000
- 15.0000
- 3.0000
- fval =
- -78.0000
- 和其它信息。

线性规划有约束极小问题

- 解问题

$$
\begin{aligned}
\max\ & z = 2x_1 + 3x_2 - 5x_3 \\
s.t.\ & x_1 + x_2 + x_3 = 7, \\
& 2x_1 - 5x_2 + x_3 \geq 10, \\
& x_1, x_2, x_3 \geq 0
\end{aligned}
$$

- 把问题极小化并将约束标准化

$$
\begin{aligned}
\min\ & z = -2x_1 - 3x_2 + 5x_3 \\
s.t.\ & -2x_1 + 5x_2 - x_3 \leq -10, \\
& x_1 + x_2 + x_3 = 7, \\
& 0 \leq x_1, x_2, x_3
\end{aligned}
$$

线性规划有约束极小问题

* 键入c=[-2,-3,5];a=[-2,5,-1];
* b=-10;a1=[1,1,1];b1=7;LB=[0,0,0];
* [x,y]=linprog(c,a,b,a1,b1,LB)
* 得当X=(6.4286,0.5714,0.0000)时,
* z=-14.5714最大.

线性规划有约束极小问题

* 解问题

$$\min\ z = -2x_1 - x_2 + x_3$$

$$\begin{aligned} s.t.\quad & x_1 + x_2 + 2x_3 = 6, \\ & x_1 + 4x_2 - x_3 \le 4, \\ & 2x_1 - 2x_2 + x_3 \le 12, \\ & x_1 \ge 0, x_2 \ge 0, x_3 \le 5. \end{aligned}$$

线性规划有约束极小问题

- 解:键入
- c=[-2,-1,1];a=[1,4,-1;2,-2,1];
- b=[4;12];a1=[1,1,2];b1=6;
- lb=[0;0;-inf];ub=[inf;inf;5];
- [x,z]=linprog(c,a,b,a1,b1,lb,ub)
- 得当X=(4.6667,0.0000,0.6667)时,
- z=-8.6667最小.

非线性规划有约束极小问题

- 模型：

$$
\begin{aligned}
&\min f(x), \\
&s.t.\ Ax \le b, \\
&\quad A_1 x = b_1, \\
&\quad c(x) \le 0, \\
&\quad c_1(x) = 0, \\
&\quad LB \le x \le UB.
\end{aligned}
$$

- MATLAB求解此问题的命令是：
- [x,fval,exitflag,output,lambda,grad,hessian]=fmincon(‘fun’,x0,A,b,A1,b1,LB,UB,’nonlcon’,options,p1,p2,...)
- fun是目标函数的m_文件名.nonlcon是约束函数C(x)和C1(x)的m_文件名.文件输出为[C,C1].

非线性规划有约束极小问题

* 求解最优化问题

$$\min\ f(x)=e^{x_1}(4x_1^2+2x_2^2+4x_1x_2+2x_2+1),$$

$$\begin{aligned} s.t.\quad & x_1+x_2=0, \\ & 1.5+x_1x_2-x_1-x_2\le 0, \\ & -x_1x_2-10\le 0. \end{aligned}$$

非线性规划有约束极小问题

- 第1步: 建立目标函数和非线性约束的m_文件.
- function y=e1511(x)% 目标函数的m_文件
- y=exp(x(1))*(4*x(1)^2+2*x(2)^2+4*x(1)*x(2)+2*x(2)+1);
-
- function [c1,c2]=e1511b(x)% 非线性约束的m_文件
- c1=[1.5+x(1)*x(2)-x(1)-x(2);-x(1)*x(2)-10];
- c2=0;

非线性规划有约束极小问题

- 第2步: 运行程序.键入
- x0=[-1,1];a1=[1,1];b1=0;
- [x,f,exitflab,output]=fmincon('e1511',x0,[],[], a1,b1,[],[],'e1511b')
- 得结果.
- 输出结果的意义:经过4次迭代(iterations:4)收敛到了(exitfag=1)最优解
- x(1)=-1.2247,x(2)=1.2247,
- 目标函数最优值为1.8951.

非线性无约束极小问题

$$\min f(x), x \in R$$

* 用命令x=fmin('f',x0)。
* 或用命令x=fminu('f ',x0)，或用命令x=fmins('f ',x0)。

非线性最小二乘问题

$$\min f^T(x)f(x)$$

* 用命令x=leastsq('f ',x0)，或用命令x=curvefit('f ',x0)。

二次规划

$$\min x^T Hx/2 + c^T x$$

$$s.t. Ax \le b$$

- 用命令x=qp(H,c,A,b)。
- 关于这些命令的详细使用规则和例子，用借助help进行查阅。

回归分析

* 前面我们曾学过拟合。但从统计的观点看，对拟合问题还需作回归分析。例如：有描述问题甲和问题乙的两组数据(x,y)和(x,z)。设x=[1，2，3，4]；y=[1.0，1.3，1.5,2.02.3];z=[0.6,1.95,0.9,2.85,1.8]。如果在平面上画出散点图，那么问题甲的四个点基本在一条直线上而问题乙的四个点却很散乱。如果都用命令polyfit(x,y,1)，polyfit(x,z,1)来拟合，将得到同一条直线。

回归分析

* 自然对问题甲的信任程度会高于对问题乙的信任程度。所以有必要对所得结果作科学的评价分析。回归分析就是解决这种问题的科学方法。
* 下面结合三个具体的例子介绍MATLAB实现回归分析的命令。

回归分析

- 合金强度y与其中含碳量x有密切关系，如下表

x	0.10	0.11	0.12	0.13	0.14	0.15	0.16	0.17	0.18	0.20	0.21	0.23
y	42.0	41.5	45.0	45.5	45.0	47.5	49.0	55.0	50.0	55.0	55.5	60.5

- 根据此表建立y(x)。并对结果作可信度进行检验、判断x对y影响是否显著、检查数据中有无异常点、由x的取值对y作出预测。

回归分析

- 解：
- 在x-y平面上画散点图，直观地知道y与x大致为线性关系。
- 用命令polyfit(x,y,1)可得y=140.6194x+27.0269。
- 作回归分析用命令
- [b,bint,r,rint,ststs]=regress(y,x,alpha) 可用help查阅此命令的具体用法
- 残差及置信区间可以用rcoplot(r,rint)画图

回归分析

- 设回归模型为 $y=\beta_0+\beta_1 x$,
- 在MATLAB命令窗口中键入下列命令进行回归分析(px_reg11.m)
- x=0.1:0.01:0.18;x=[x,0.2,0.21,0.23]';
- y=[42,41.5,45,45.5,45,47.5,49,55,50,55,55.5,60.5]';
- X=[ones(12,1),x];
- [b,bint,r,rint,stats]=regress(y,X,0.05);
- b,bint,stats,rcoplot(r,rint)

回归分析

- 得结果和图
- b =
- 27.0269
- 140.6194
- bint =
- 22.3226   31.7313
- 111.7842 169.4546
- stats =
- 0.9219 118.0670   0.0000   3.1095

回归分析

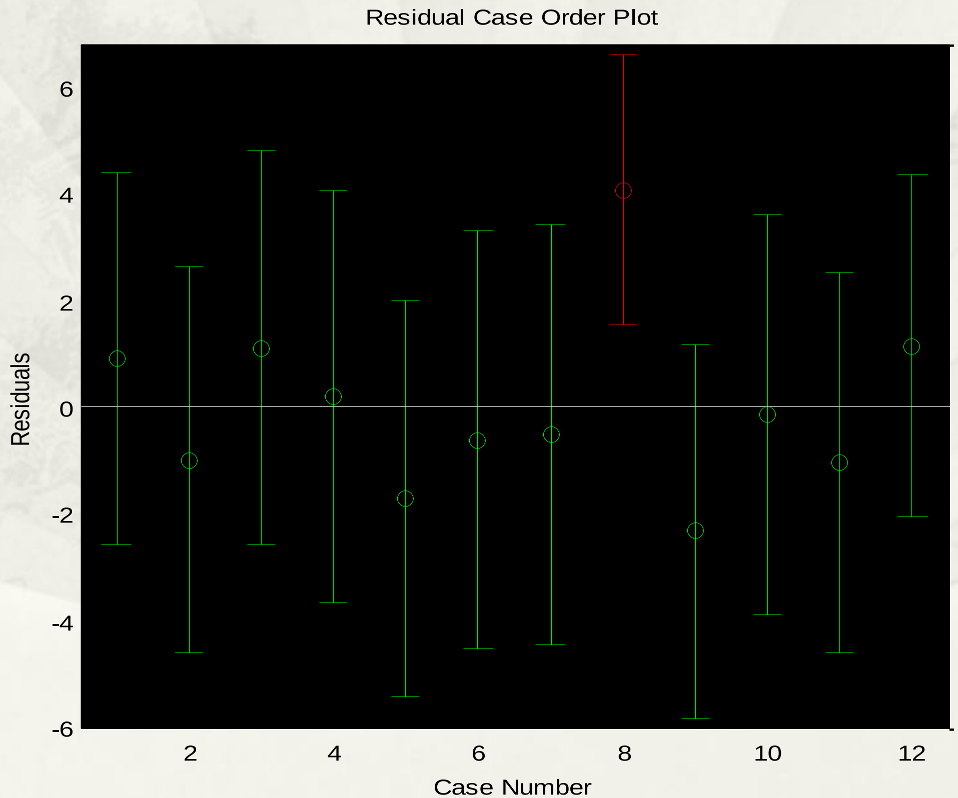
- 结果含义为
- β_0=27.0269 β_1=140.6194
- β_0的置信区间是 [22.3226,31.7313]
- β_1的置信区间是 [111.7842,169.4546]

回归分析

- $R^2=0.9219$ $F=118.0670$，$p<10^{-4}$.
- R是衡量y与x的相关程度的指标，称为相关系数。R越大，x与y关系越密切。通常R大于0.9才认为相关关系成立。
- F是一统计指标
- p是与F对应的概率，当 $p<0.05$时，回归模型成立。
- 此例中 $p=0<10^{-4}<0.05$，所以，所得回归模型成立。

回归分析

观察所得残差分布图，看到第8个数据的残差置信区间不含零点，此点视为异常点，剔除后重新计算。



回归分析

- 此时键入:(px_reg12.m)
- X(8,:)=[];
- y(8)=[];
- [b,bint,r,rint,stats]=regress(y,X);
- b,bint,stats,rcoplot(r,rint)

回归分析

- b =
- 27.0992
- 137.8085
- bint =
- 23.8563 30.3421
- 117.8534 157.7636
- stats =
- 0.9644 244.0571 0.0000 1.4332
- 可以看到：置信区间缩小；R^2、F变大，所以应采用修改后的结果。

回归分析

- 将17至19岁的运动员每两岁一组分为7组，每组两人测量其旋转定向能力，以考察年龄(x)对这种运动能力(y)的影响。现得到一组数据如下表

年龄	17	19	21	23	25	27	29
第一人	20．48	25．13	26．15	30．0	26．1	20．3	19．35
第二人	24．35	28．11	26．3	31．4	26．92	25．7	21．3

- 试建立关系y(x)，并作必要的统计分析。

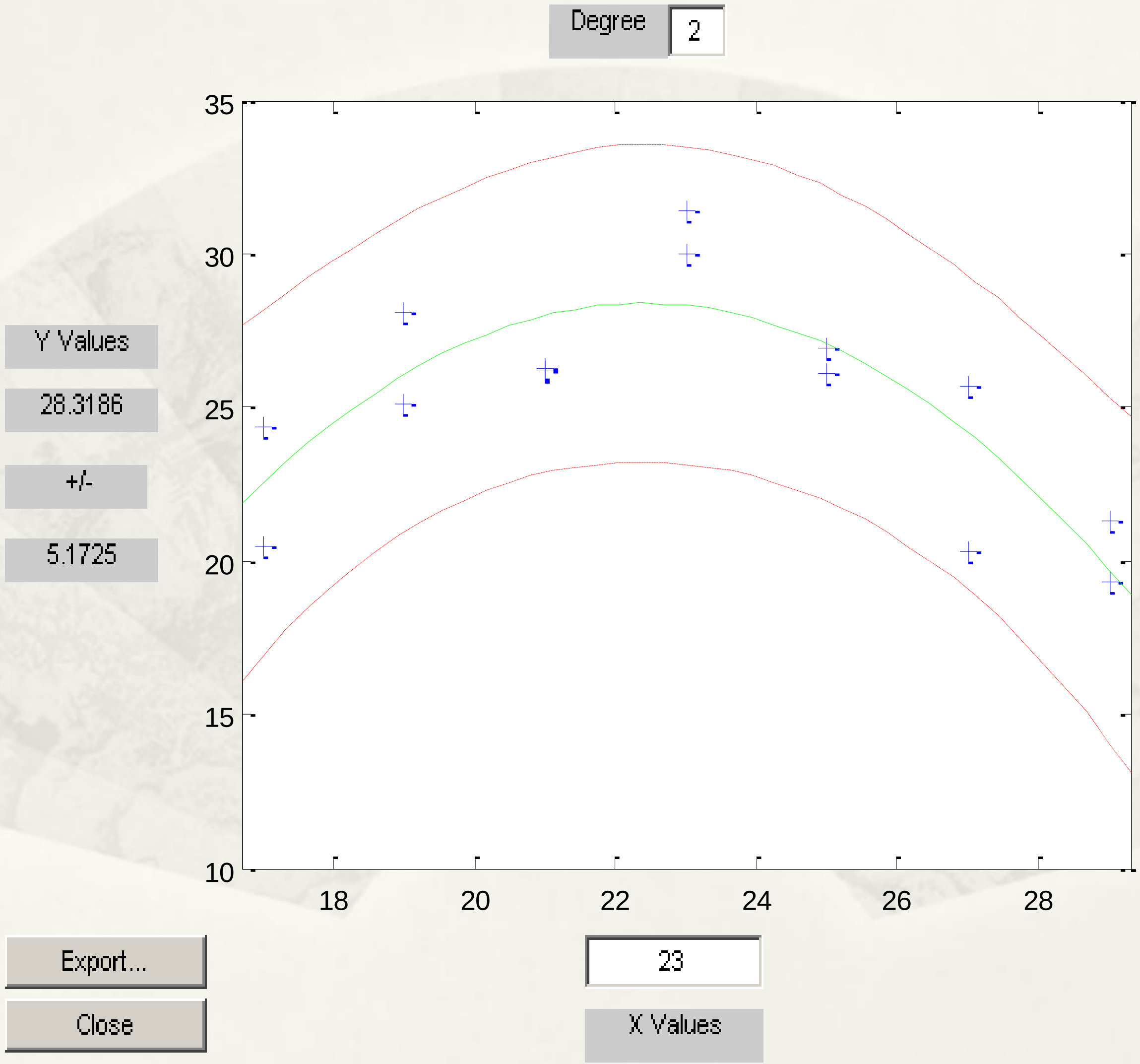
回归分析

- 在x-y平面上画散点图，直观地知道y与x大致为二次函数关系。
- 设模型为$y=a_1x^2+a_2x+a_3$
- 此问题可以利用命令polyfit(x,y,2)来解，也可以像上题一样求解。下面介绍用命令polytool来解。

回归分析

- 首先在命令窗口键入(px_reg21.m)
- x=17:2:29;x=[x,x];
- y=[20.48,25.13,26.15 30,26.1,20.3,19.35,24.35,28.11,26.3,31.4,26.92,25.7,21.3];
- polytool（x,y,2）
- 得到一个交互式窗口

回归分析



Degree
2
35
30
25
20
15
10
18
20
22
24
26
28
Y Values
28.3186
+/-
5.1725
Export...
Close
23
X Values

回归分析

* 窗口中绿线为拟合曲线、红线为y的置信区间、可通过移动鼠标的十字线或通过在窗口下方输入来设定x值，窗口左边则输出与x对应的y值及y的置信区间。通过左下方的Export下拉菜单可输出回归系数等。更详细的解释可通过help查阅。

回归分析

- 某厂生产的某产品的销售量与竞争对手的价格x1和本厂的价格x2有关。下表是该产品在10个城市的销售记录。

x1	120	140	190	130	155	175	125	145	180	150
x2	100	110	90	150	210	150	250	270	300	250
y(个)	102	100	120	77	46	93	26	69	65	85

- 试建立关系y(x1，x2)，对结果进行检验。若某城市本厂产品售价160（元），对手售价170（元），预测此产品在该城市的销售量。

回归分析

- 这是一个多元回归问题。若设回归模型是线性的，即设$y=\beta_0+\beta_1 x_1+\beta_2 x_2$
- 那么依然用regress(y,x,alpha)求回归系数。

回归分析

- 键入(px_reg31.m)
- x1=[120,140,190,130,155,175,125,145,180,150];
- x2=[100,110,90,150,210,150,250,270,300,250];
- y=[102,100,120,77,46,93,26,69,65,85]';
- x=[ones(10,1),x1',x2'];
- [b,bint,r,rint,stats]=regress(y,x);
- b,bint,stats,

回归分析

- b =
- 66.5176
- 0.4139
- -0.2698
- bint =
- -32.5060 165.5411
- -0.2018 1.0296
- -0.4611 -0.0785
- stats =
- 0.6527 6.5786 0.0247 351.0445

回归分析

* p=0.0247,若显著水平取0,01，则模型不能用；R2=0.6527较小;β_0，β_1的置信区间包含零点。因此结果不理想。于是设模型为二次函数。此题设模型为纯二次函数：
* $y=\beta_0+\beta_1 x_1+\beta_2 x_2+\beta_{11} x_1^2+\beta_{22} x_2^2$

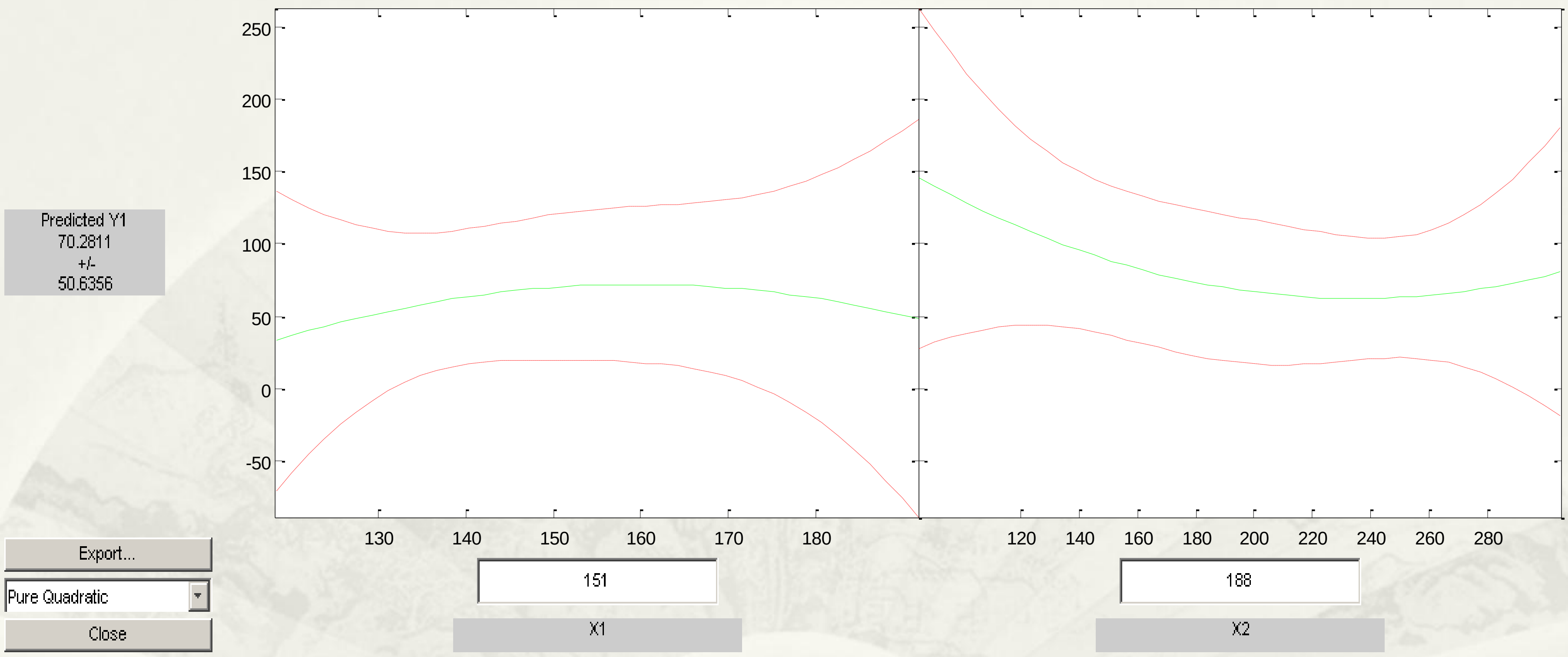
回归分析

- MATLAB提供的多元二项式回归命令为rstool(x,y,model,alpha).其中alpha为显著水平、model在下列模型中选一个：
- Linear(线性)
  $$y=\beta_0+\beta_1x_1+\cdots\beta_mx_m$$
- Purequadratic(纯二次)
  $$y=\beta_0+\beta_1x_1+\cdots\beta_mx_m+\sum_{j=1}^{m}\beta_{jj}x_j^2$$
- Interaction(交叉)
  $$y=\beta_0+\beta_1x_1+\cdots\beta_mx_m+\sum_{1\le j\ne k\le m}\beta_{jk}x_jx_k$$
- Quadratic(完全二次)
  $$y=\beta_0+\beta_1x_1+\cdots\beta_mx_m+\sum_{1\le j,k\le m}\beta_{jk}x_jx_k$$

回归分析

- 对此例，在命令窗中键入
- x(:,1)=[];
- rstool(x,y,'purequadratic')
- 得到一个对话窗：

回归分析



Predicted Y1
70.2811
+/-
50.6356
250
200
150
100
50
0
-50
130
140
150
160
170
180
120
140
160
180
200
220
240
260
280
151
188
X1
X2
Export...
Pure Quadratic
Close

回归分析

* 其意义与前面的对话窗意义类似。若要回答“本厂售价160，对手售价170，预测该市销售量”的问题，只需在下方窗口中分别肩入160和170，就可在左方窗口中读到答案及其置信区间。

回归分析

* 下拉菜单Export向工作窗输出数据具体操作为：
* 弹出菜单，选all，点击确定。此时可到工作窗中读取数据。可读数据包括：beta（回归系数） rmse（剩余标准差） residuals（残差）本题只要键入beta,rmse,residuals

回归分析

- beta =
- -312.5871
- 7.2701
- -1.7337
- -0.0228
- 0.0037
- rmse =
- 16.6436
- residuals =
- 6.6846
- -12.6703
- -0.2013
- 6.4855
- -19.6533
- 7.9989
- -11.4737
- 5.4303
- -4.9932
- 22.3926

判别分析

* 判别分析是判别样品所属类型的一种统计方法，其应用之广泛可与回归分析媲美。
* 判别分析与聚类分析不同。
* 判别分析的分类
  * 距离判别法
  * Fisher 判别法
* [判别分析]

* MATLAB中还包括神经网络工具箱，小波分析工具箱，在网上还可以下载遗传算法工具箱，有兴趣的同学可以借这次机会，结合学习MATLAB，好好学习一下相关理论知识。
* 最后，祝大家学习，竞赛都取得成功。谢谢大家。