石家莊鐵道大學
SHIJIAZHUANG TIEDAO UNIVERSITY

在线开放课程

MATLAB数据统计分析

数据插值与拟合

主讲 ： 卞建鹏

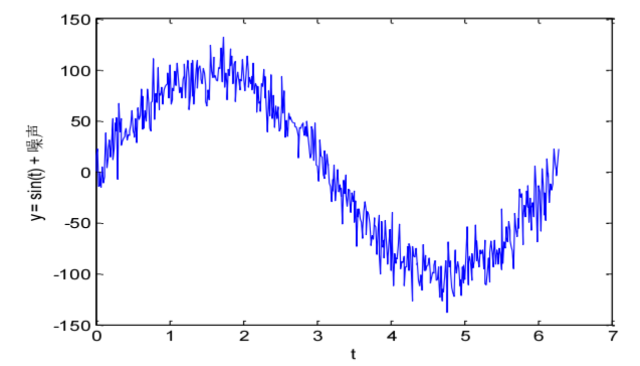
1、数据的平滑处理


```
yy = smooth(y)

yy = smooth(y,span)    %span: 滤波器的窗宽

yy = smooth(y,method)

yy = smooth(y,span,method)

yy = smooth(y,'sgolay',degree)

yy = smooth(y,span,'sgolay',degree)

yy = smooth(x,y,…)
```

1、数据的平滑处理

产生一列正弦波信号，加入噪声信号，然后调用smooth函数对加入噪声的正弦波进行滤波（平滑处理）

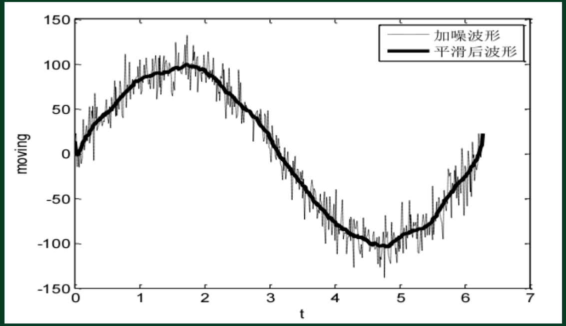
```
>> t = linspace(0,2*pi,500)';
>> y = 100*sin(t);
>> noise = normrnd(0,15,500,1);
>> y = y + noise;
>> plot(t,y);
>> xlabel('t');
>> ylabel('y = sin(t) + 噪声');
```



1、数据的平滑处理

（1）移动平均法

```
>> yy1 = smooth(y,30);
>> figure;
>> plot(t,y,'k:');
>> hold on;
>> plot(t,yy1,'k','linewidth',3);
>> xlabel('t');
>> ylabel('moving');
>> legend('加噪波形','平滑后波形');
```

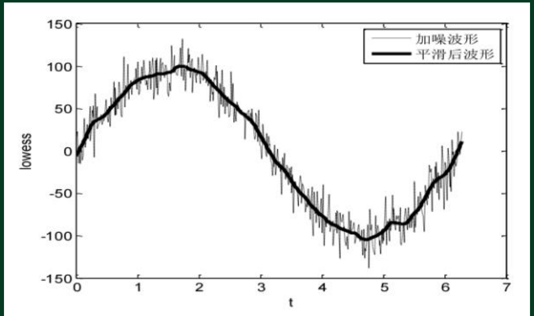


1、数据的平滑处理

在线开放课程

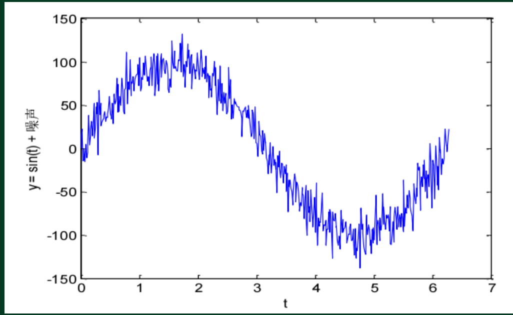
（2）lowess方法

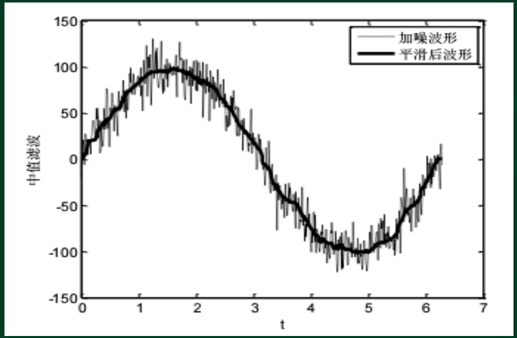
```
>> yy2 = smooth(y,30,'lowess');
>> figure;
>> plot(t,y,'k:');
>> hold on;
>> plot(t,yy2,'k','linewidth',3);
>> xlabel('t');
>> ylabel('lowess');
>> legend('加噪波形','平滑后波形');
```



1、数据的平滑处理


```
>> yy = medfilt1(y,30);
>> figure;
>> plot(t,y,'k:');
>> hold on
>> plot(t,yy,'k','LineWidth',3);
>> xlabel('t');
>> ylabel('中值滤波');
>> legend('加噪波形','平滑后波形');
```





2、数据的标准化变换

Z = zscore(X)

[Z,mu,sigma] = zscore(X)

[…] = zscore(X,1)

[…] = zscore(X,flag,dim)

标准化变换公式

$$\mathbf{X}=\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}, \qquad \mathbf{X}^*=\begin{pmatrix} x_{11}^* & x_{12}^* & \cdots & x_{1p}^* \\ x_{21}^* & x_{22}^* & \cdots & x_{2p}^* \\ \vdots & \vdots & \cdots & \vdots \\ x_{n1}^* & x_{n2}^* & \cdots & x_{np}^* \end{pmatrix},$$

$$x_{ij}^*=\frac{x_{ij}-\bar{x}_j}{\sqrt{s_{jj}}},\quad i=1,2,\cdots,n,\quad j=1,2,\cdots,p,$$

$$\bar{x}_j=\frac{1}{n}\sum_{i=1}^{n}x_{ij},\quad \sqrt{s_{jj}}=\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2},\quad j=1,2,3,\cdots,p.$$

2、数据的标准化变换

在线开放课程

调用rand函数产生随机矩阵，然后调用zscore函数将其按列标准化。

```
>> x = [rand(3,1), 5*rand(3,1), 10*rand(3,1)]

>> [xz,mu,sigma] = zscore(x)

>> s=std(xz)      %标准差
```

```
x =

    0.6948    0.1722    7.6552
    0.3171    2.1937    7.9520
    0.9502    1.9078    1.8687
```

```
xz =

    0.1280   -1.1448    0.5335
   -1.0578    0.7031    0.6201
    0.9298    0.4417   -1.1536
```

```
s =

    1.0000    1.0000    1.0000
```

2、数据的标准化变换

在线开放课程

数据的极差归一化变换

```
>> data = [1 2 3;4 2 4;5 8 6];
[m,n]=size(data);
X = max(data); Y = min(data);
yy = repmat(Y, [m,1]) ;
x = (data-yy)./(X-Y)
```

```
data =

     1     2     3
     4     2     4
     5     8     6
```

```
x =

         0         0         0
    0.7500         0    0.3333
    1.0000    1.0000    1.0000
```

$$\mathbf{X}=\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \cdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{pmatrix}, \qquad \mathbf{X}^R=\begin{pmatrix} x_{11}^R & x_{12}^R & \cdots & x_{1p}^R \\ x_{21}^R & x_{22}^R & \cdots & x_{2p}^R \\ \vdots & \vdots & \cdots & \vdots \\ x_{n1}^R & x_{n2}^R & \cdots & x_{np}^R \end{pmatrix},$$

$$x_{ij}^R=\frac{x_{ij}-\min\limits_{1\le k\le n} x_{kj}}{\max\limits_{1\le k\le n} x_{kj}-\min\limits_{1\le k\le n} x_{kj}}, \quad i=1,2,\cdots,n, \ \ j=1,2,\cdots,p,$$

2、数据的标准化变换

在线开放课程

读取Excel数据文件，去除有缺失数据的行

>> data = xlsread('Excel数据文件.xlsx')

>> A= **isnan(data)**　　% data中有非数（NaNs，比如0/0），

则取1，否则取0；

```
命令行窗口
>> data = xlsread('Excel数据文件.xlsx')

data =

    1.0000  168.4000   74.2000  686.0000
    2.0000  162.3000       NaN  275.0000
    3.0000       NaN   63.8000  867.0000
    4.0000  169.8000   48.7000  327.0000
    5.0000  174.0000   71.5000       NaN
    6.0000  161.9000   52.1000  625.0000
```

```
>> isnan(data)

ans =

  6×4 logical 数组

   0   0   0   0
   0   0   1   0
   0   1   0   0
   0   0   0   0
   0   0   0   1
   0   0   0   0
```

	A	B	C	D
1	序号	身高	体重	肺活量
2	1	168.4	74.2	686
3	2	162.3		275
4	3		63.8	867
5	4	169.8	48.7	327
6	5	174	71.5	
7	6	161.9	52.1	625
8				

读取Excel数据文件，去除有缺失数据的行

>> id =any(A, 2)　　%对A的行向量判断，如果有非零数，则取1，否则取0。

>> D = data(~id,:)

	A	B	C	D
1	序号	身高	体重	肺活量
2	1	168.4	74.2	686
3	2	162.3		275
4	3		63.8	867
5	4	169.8	48.7	327
6	5	174	71.5	
7	6	161.9	52.1	625
8				

```
>> isnan(data)

ans =

  6×4 logical 数组

   0   0   0   0
   0   0   1   0
   0   1   0   0
   0   0   0   0
   0   0   0   1
   0   0   0   0
```

```
>> id =any(A, 2)

id =

  6×1 logical 数组

   0
   1
   1
   0
   1
   0
```

```
>> D = data(~id, :)

D =

    1.0000  168.4000   74.2000  686.0000
    4.0000  169.8000   48.7000  327.0000
    6.0000  161.9000   52.1000  625.0000
```

3、数据插值

通过y=f(x)的数据点（x1,y1),(x2,y2),…,(xn,yn)构造通过这些数据点，并且光滑的函数y=g(x)。数据点越密集，g(x) 与f(x)越接近。

根据自变量个数，插值分为：一维、二维、多维插值。

根据插值函数不同，插值分为：线性插值、最近点插值、多项式插值、三次样条插值等。

Matlab提供了一维、二维、N维插值函数interp1，interp2，interpN以及三次样条插值函数spline。

3、数据插值

在线开放课程

一维数据插值（单变量函数）

Y1=interp1(X,Y,X1,'method')

函数根据X, Y的值，计算函数在X1处的值。X, Y是两个等长的已知向量，分别描述采样点和样本值，X1是一个向量或标量，描述欲插值的点，Y1是一个与X1等长的插值结果。

method是插值方法，允许的取值有linear（线性插值, 为默认的插值方法）、nearest（最近点插值）、cubic（三次多项式插值）、spline（三次样条插值）。

一维数据插值（单变量函数）

例：用不同的插值方法计算y=sin(x)在π/2点的值。

```
x=0:0.2:pi;
y=sin(x);
s1=interp1(x,y,pi/2)
s2=interp1(x,y,pi/2,'linear')
s3=interp1(x,y,pi/2,'nearest')
s4=interp1(x,y,pi/2,'cubic')
s5=interp1(x,y,pi/2,'spline')
```

```
s1 =
    0.9975
s2 =
    0.9975
s3 =
    0.9996
s4 =
    0.9992
s5 =
    1.0000
```

在线开放课程

例:某观测站测得某日6:00时至18:00时之间每隔2小时的室内外温度(℃)，用3次样条插值分别求得该日室内外6:30至17:30时之间每隔2小时各点的近似温度(℃)。

设时间变量h为一行向量，温度变量t为一个两列矩阵，其中第一列存放室内温度，第二列储存室外温度。

```
h =6:2:18;
t=[18,20,22,25,30,28,24 ; 15,19,24,28,34,32,30]';
XI =6.5:2:17.5
YI=interp1(h,t,XI,'spline')
```

```
YI =

   18.5020   15.6553
   20.4986   20.3355
   22.5193   24.9089
   26.3775   29.6383
   30.2051   34.2568
   26.8178   30.9594
```

3、数据插值

二维数据插值

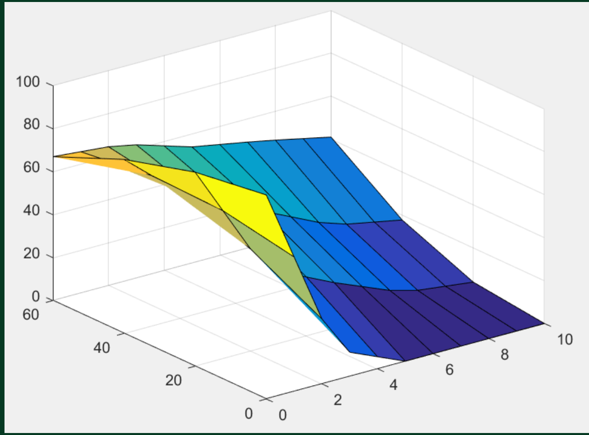
Z1=interp2(X,Y,Z,X1,Y1,'method')

其中X,Y是两个向量，分别描述两个参数的采样点，Z是与参数采样点对应的函数值，**X1,Y1是两个向量或标量，描述欲插值的点**。Z1是根据相应的插值方法得到的插值结果。 method的取值与一维插值函数相同。X,Y,Z也可以是矩阵形式。

同样，X1,Y1的取值范围不能超出X,Y的给定范围，否则，会给出“NaN”错误。

例：某实验对一根长10米的钢轨进行热源的温度传播测试。用x表示测量点0:2.5:10(米)，用h表示测量时间0:30:60(秒)，用T表示测试所得各点的温度(℃)。试用线性插值求出在一分钟内每隔20秒、钢轨每隔1米处的温度TI。


```
x=0:2.5:10; h=[0:30:60]';
T=[95,14,0,0,0; 88,48,32,12,6; 67,64,54,48,41];
xi=[0:10]; hi=[0:20:60]';
TI=interp2(x,h,T, xi, hi)
surf(xi,hi,TI)
```



4、曲线拟合

通过polyfit函数来求得最小二乘拟合多项式的系数，再用polyval函数按所得的多项式计算所给出的点上的函数近似值。

[p,s]=polyfit(x,y,m)

函数根据采样点x和采样点函数值y，产生一个m次多项式p及其在采样点的误差向量s。其中x, y是两个等长的向量，p是一个长度为m+1的向量，p的元素为多项式系数。

polyval(p,x)函数的功能是按多项式的系数计算x点多项式的值。

石家庄铁道大学 SHIJIAZHUANG TIEDAO UNIVERSITY
在线开放课程

例: 已知数据表[t,y]，试求2次拟合多项式p(t)，然后求ti=1,1.5,2, 2.5, …,9.5,10 各点的函数近似值。

t=1:10; y=[9.6 4.1 1.3 0.4 0.05 0.1 0.7 1.8 3.8 9.0];

p=polyfit(t,y,2)

结果：p =0.4561   -5.0412   13.2533

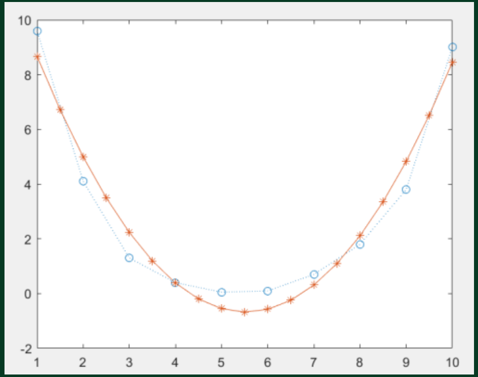
相当于 $y= 0.4561x^2 -5.0412 x+13.2533$

然后计算各点近似值并绘制拟合曲线。

ti=1:0.5:10

yi=polyval(p,ti)

plot(t,y,':o',ti,yi,'-*')



小结

1. 数据的平滑处理
2. 数据的标准化变换
3. 数据插值
4. 曲线拟合