

统计

- 描述性统计(Descriptive Statistics)
 - 中心趋势(Central Tendency)
 - 变异(Variation)
 - 离散程度
 - 偏度(Skewness)
 - 峰度(Kurtosis)
- 统计推断(Inferential Statistics)

拟合

- 多项式拟合
 - 一元多项式拟合: `polyfit()`
 - 多元线性拟合: `regress()`
- 非线性拟合

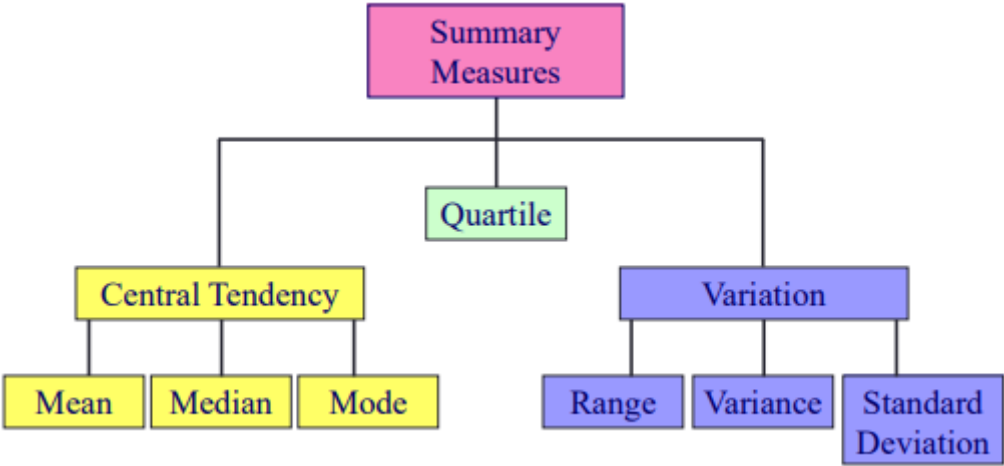
插值

- 一维插值
- 二维插值

统计

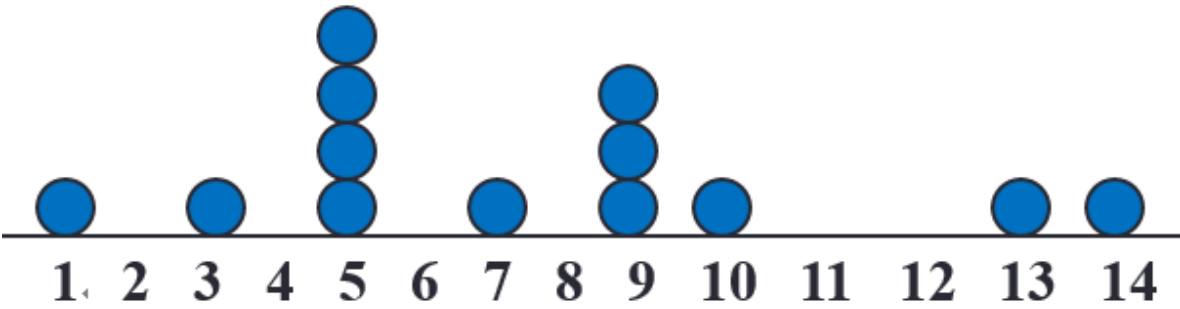
描述性统计(Descriptive Statistics)

描述性统计主要研究数据的**中心趋势**(Central Tendency)和**变异**(Variation).



中心趋势(Central Tendency)

函数	作用
<code>mean()</code>	计算平均值
<code>median()</code>	计算中位数
<code>mode()</code>	计算众数
<code>prctile()</code>	计算百分位数
<code>max()</code>	计算最大值
<code>min()</code>	计算最小值

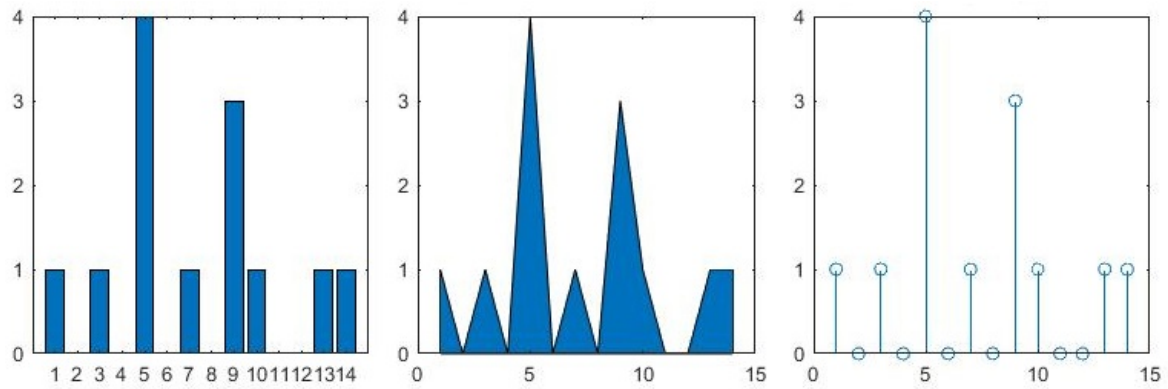


```
1 x = [1 3 5 5 5 5 7 9 9 9 10 13 14];
2
3 mean(x);           % 数据的平均值为 7.3077
4 median(x);         % 数据的中位数为 7
5 mode(x);           % 数据的众数为 5
6 prctile(x, 0);     % 数据的0%分位数为 0
7 prctile(x, 50);    % 数据的50%分位数为 7
8 prctile(x, 100);   % 数据的100%分位数为 14
9 prctile(x, 12.6);  % 数据的12.6%分位数为 3.2760
```

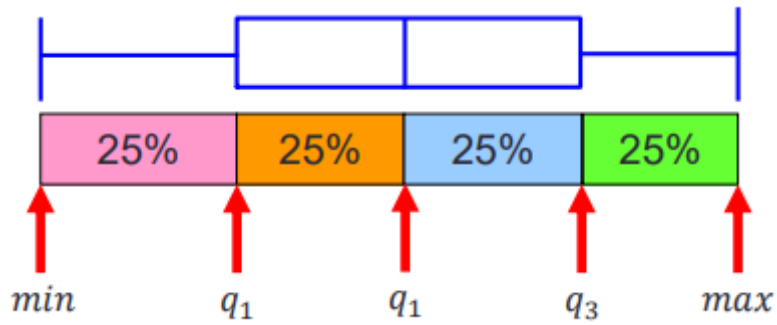
下列函数绘制统计图表:

函数	作用
<code>bar()</code>	绘制条形图
<code>stem()</code>	绘制针状图
<code>area()</code>	绘制填充图
<code>boxplot()</code>	绘制箱线图

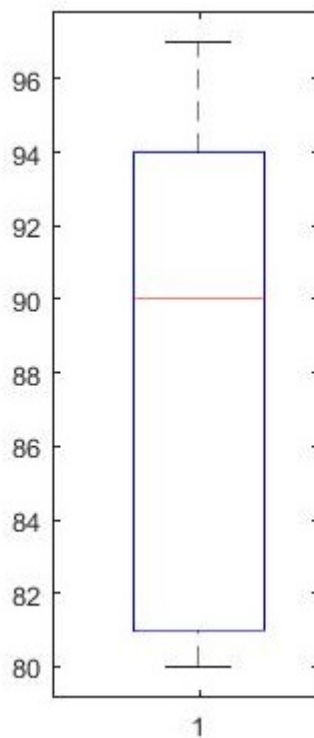
```
1 x = 1:14;
2 freqy = [1 0 1 0 4 0 1 0 3 1 0 0 1 1];
3 subplot(1,3,1); bar(x,freqy); xlim([0 15]);
4 subplot(1,3,2); area(x,freqy); xlim([0 15]);
5 subplot(1,3,3); stem(x,freqy); xlim([0 15]);
```



箱线图可以突出显示数据的四分位点.



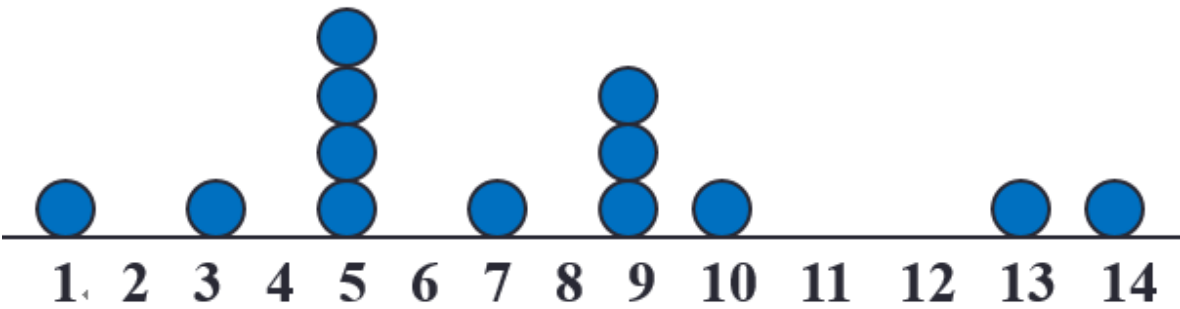
```
1 marks = [80 81 81 84 88 92 92 94 96 97];
2 boxplot(marks)
3 prctile(marks, [25 50 75]) % 得到 [81 90 94]
```



变异(Variation)

离散程度

函数	作用
<code>std()</code>	计算数据的标准差
<code>var()</code>	计算数据的方差



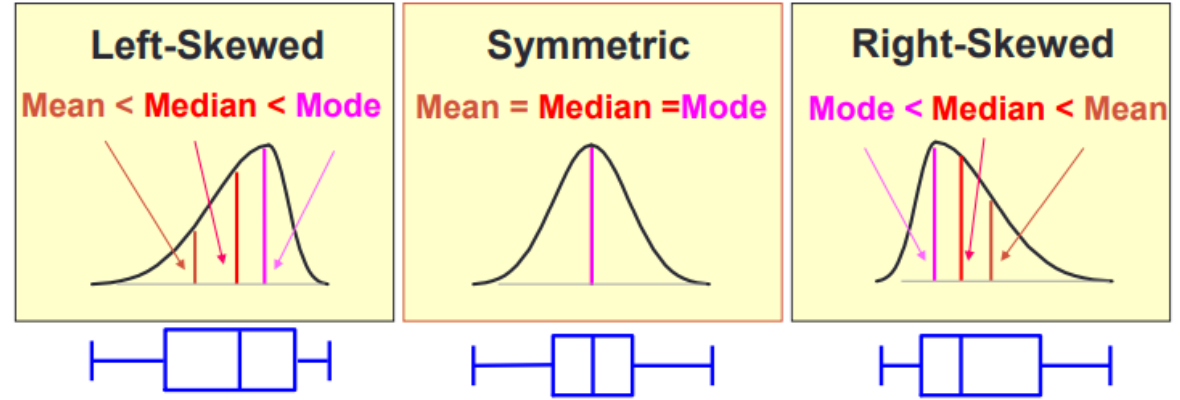
```

1 | x = [1 3 5 5 5 5 7 9 9 9 10 13 14];
2 |
3 | std(x);           % 得到 3.7944
4 | var(x);           % 得到 14.3974

```

偏度(Skewness)

函数	作用
<code>skewness()</code>	计算数据的偏度



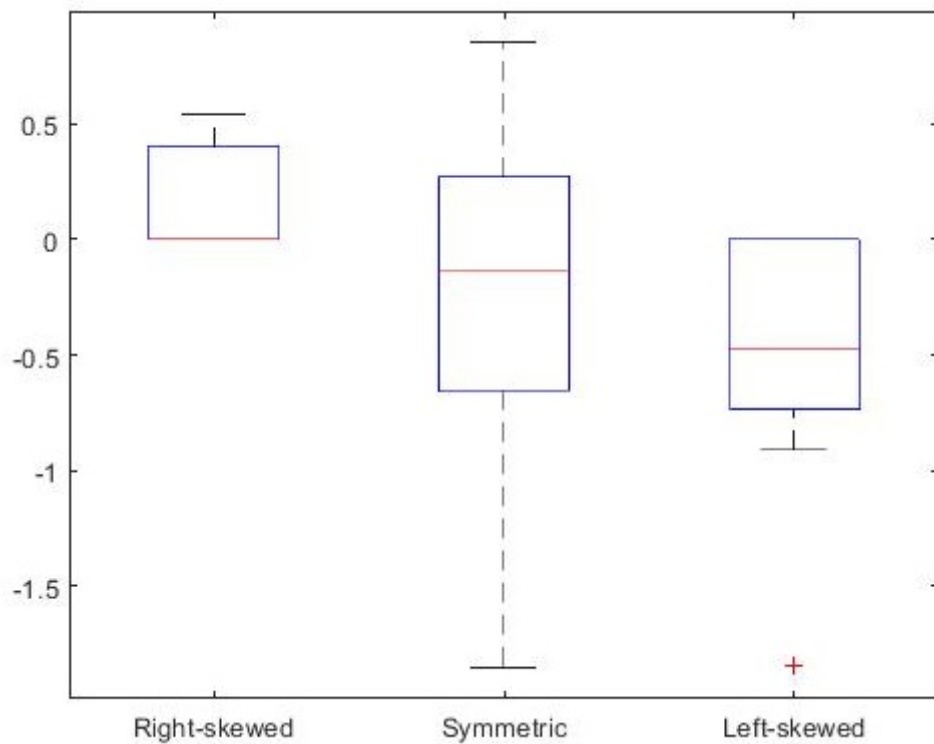
偏度反映数据的对称程度

- 数据左偏时,其偏度小于0.
- 数据完全对称时,其偏度等于0.
- 数据右偏时,其偏度大于0.

```

1 | x = randn([10 3]);           % 构造10*3的矩阵
2 | x(x(:,1)<0, 1) = 0;         % 将第一列数据右偏
3 | x(x(:,3)>0, 3) = 0;         % 将第二列数据左偏
4 | boxplot(x, {'Right-skewed', 'Symmetric', 'Left-skewed'});
5 | skewness(x);               % 得到 [0.5162 -0.7539 -1.1234]

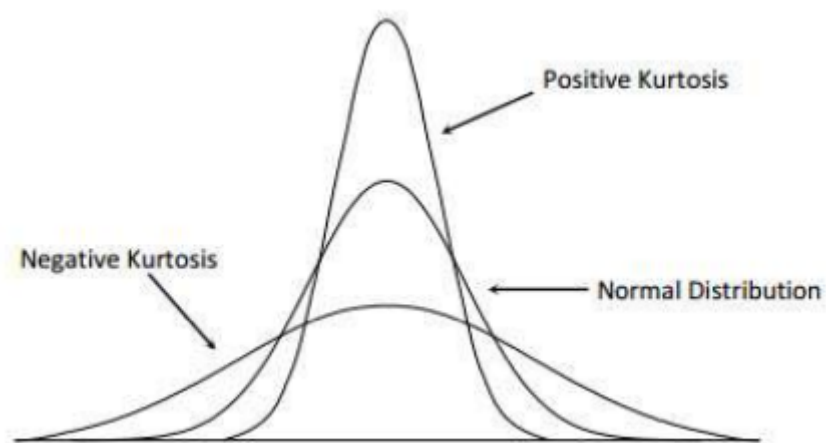
```



峰度(Kurtosis)

函数	作用
<code>kurtosis()</code>	计算数据的峰度

峰度(Kurtosis)表征概率密度分布曲线在平均值处峰值的高低.直观来看,峰度反映了峰部的尖度.



统计推断(Inferential Statistics)

推论统计的核心即为假设检验.下列函数用于进行假设检验.

函数	作用
<code>ttest()</code>	进行T检验
<code>ztest()</code>	进行Z检验
<code>ranksum()</code>	进行秩和检验
<code>signrank()</code>	进行符号秩检验

```

1 load examgrades
2 x = grades(:,1);
3 y = grades(:,2);
4 [h,p] = ttest(x,y);

```

执行上述程序,得到 `[h p] = [0 0.9805]`,表示在默认显著性水平(5%)下我们没有理由拒绝 `x` 与 `y` 同分布.

拟合

多项式拟合

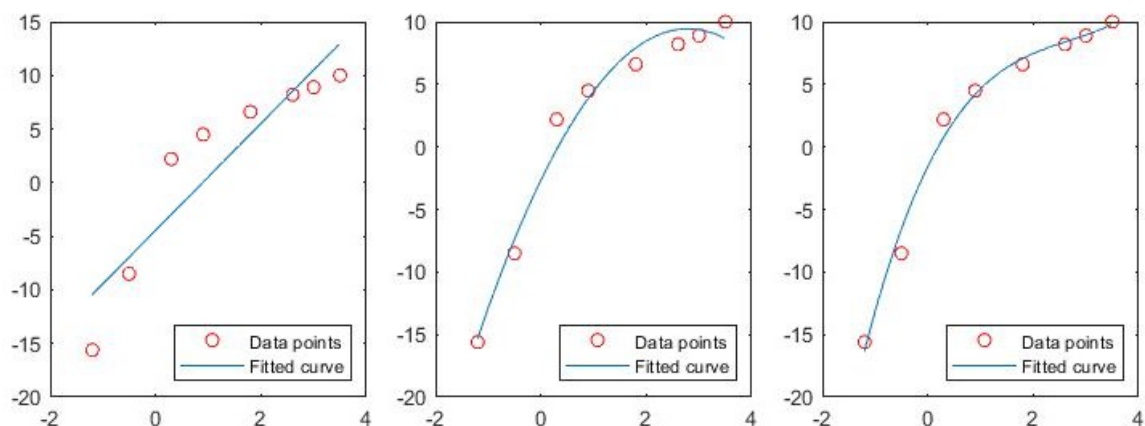
一元多项式拟合: `polyfit()`

使用 `polyfit(x, y, n)` 函数对数据 `x` 和 `y` 进行 `n` 次多项式拟合.

```

1 x = [-1.2 -0.5 0.3 0.9 1.8 2.6 3.0 3.5];
2 y = [-15.6 -8.5 2.2 4.5 6.6 8.2 8.9 10.0];
3 for i=1:3
4     p = polyfit(x,y,i); % 分别进行一次,二次,三次拟合
5     xfit = x(1):0.1:x(end); yfit = polyval(p,xfit);
6     subplot(1,3,i); plot(x,y,'ro',xfit,yfit);
7     legend('Data points','Fitted curve','Location','southeast');
8 end

```



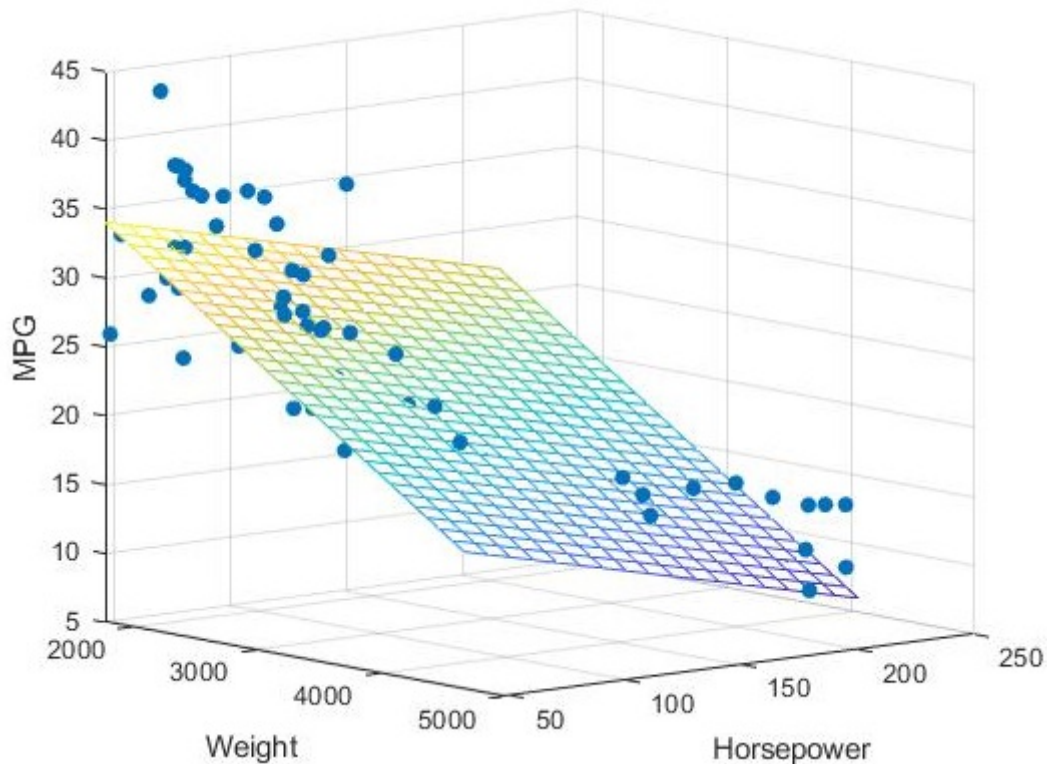
多元线性拟合: `regress()`

使用 `regress(y, x)` 函数对数据 `x` 和 `y` 进行多元线性回归.

```

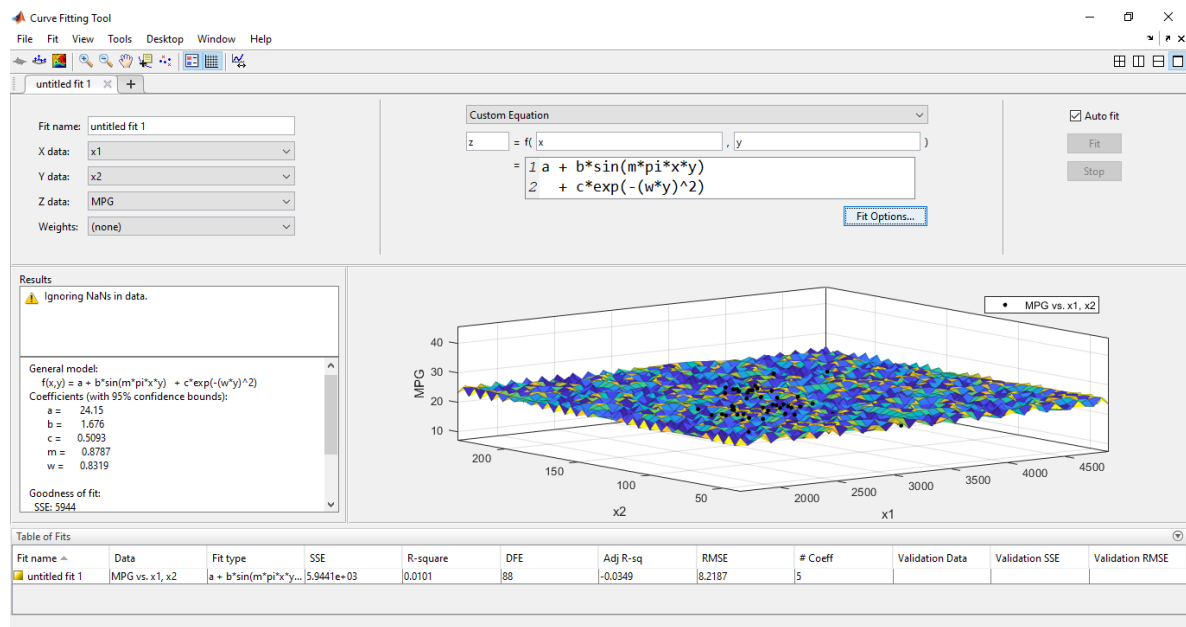
1 load carsmall;
2 y = MPG; x1 = weight; x2 = Horsepower; % 导入数据集
3 X = [ones(length(x1),1) x1 x2]; % 构建增广x矩阵
4 b = regress(y,X); % 进行线性回归
5
6 % 下面是绘图语句
7 x1fit = min(x1):100:max(x1); x2fit = min(x2):10:max(x2);
8 [X1FIT,X2FIT] = meshgrid(x1fit,x2fit);
9 YFIT = b(1)+b(2)*X1FIT+b(3)*X2FIT;
10 scatter3(x1,x2,y,'filled'); hold on;
11 mesh(X1FIT,X2FIT,YFIT); hold off;
12 xlabel('weight'); ylabel('Horsepower'); zlabel('MPG'); view(50,10);

```



非线性拟合

对于非线性拟合,需要使用曲线拟合工具箱.在命令窗口输入 `cftool()` 打开曲线拟合工具箱.



插值

一维插值

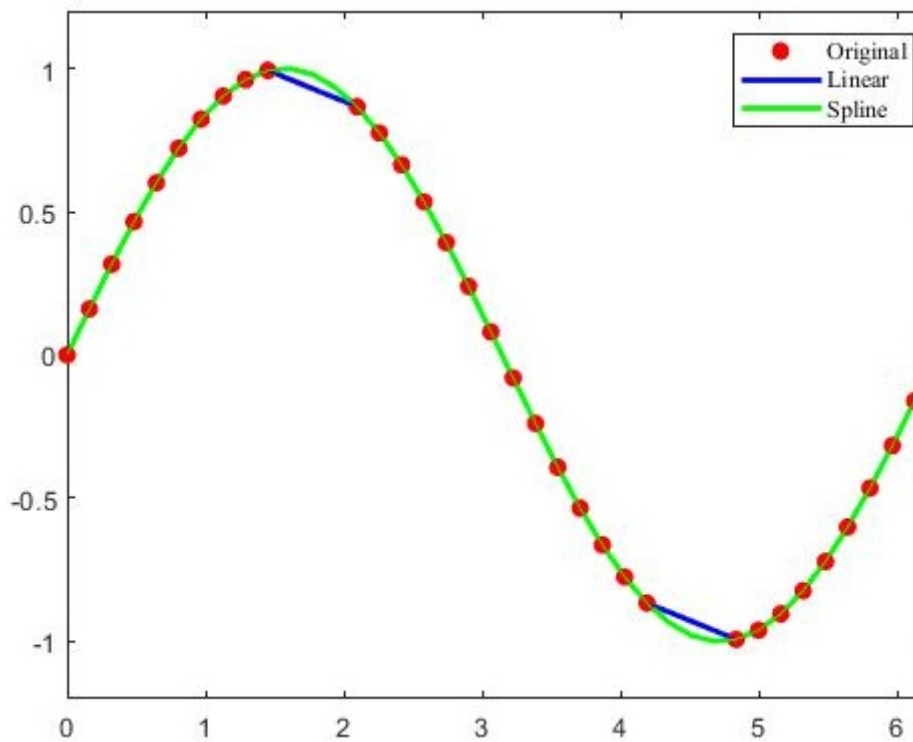
下列函数均与一维插值有关:

函数	作用
<code>interp1(x,v)</code> 或 <code>interp1(x,v,xq)</code>	线性插值
<code>spline(x,v)</code> 或 <code>spline(x,v,xq)</code>	三次样条插值
<code>pchip(x,v)</code> 或 <code>pchip(x,v,xq)</code>	三次Hermite插值
<code>mkpp(breaks,coefs)</code>	生成分段多项式
<code>ppval(pp,xq)</code>	计算分段多项式的插值结果

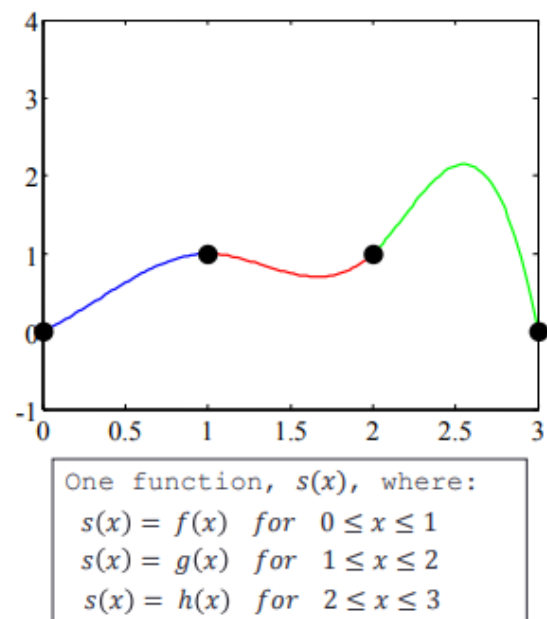
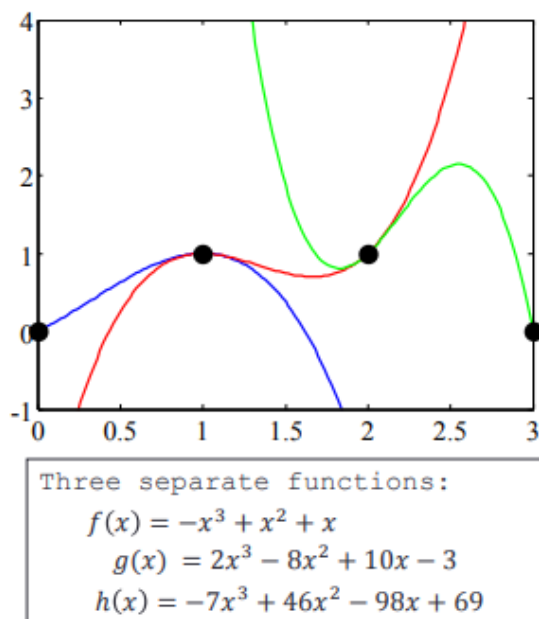
下面例子演示使用 `interp1(x, v, xq)` 进行线性插值和使用 `spline(x, v, xq)` 进行三次样条插值.各参数意义如下:

- `x,v`: 待插值样本点.
- `xq`: 查询点,函数返回在这些点处的插值结果.

```
1 % 构造数据
2 x = linspace(0, 2*pi, 40); x_m = x; x_m([11:13, 28:30]) = NaN;
3 y_m = sin(x_m);
4 plot(x_m, y_m, 'ro', 'MarkerFaceColor', 'r'); hold on;
5
6 % 对数据进行线性插值
7 m_i = ~isnan(x_m);
8 y_i = interp1(x_m(m_i), y_m(m_i), x);
9 plot(x,y_i, '-b'); hold on;
10
11 % 对数据进行三次样条插值
12 m_i = ~isnan(x_m);
13 y_i = spline(x_m(m_i), y_m(m_i), x);
14 plot(x,y_i, '-g');
15
16 legend('Original', 'Linear', 'Spline');
```

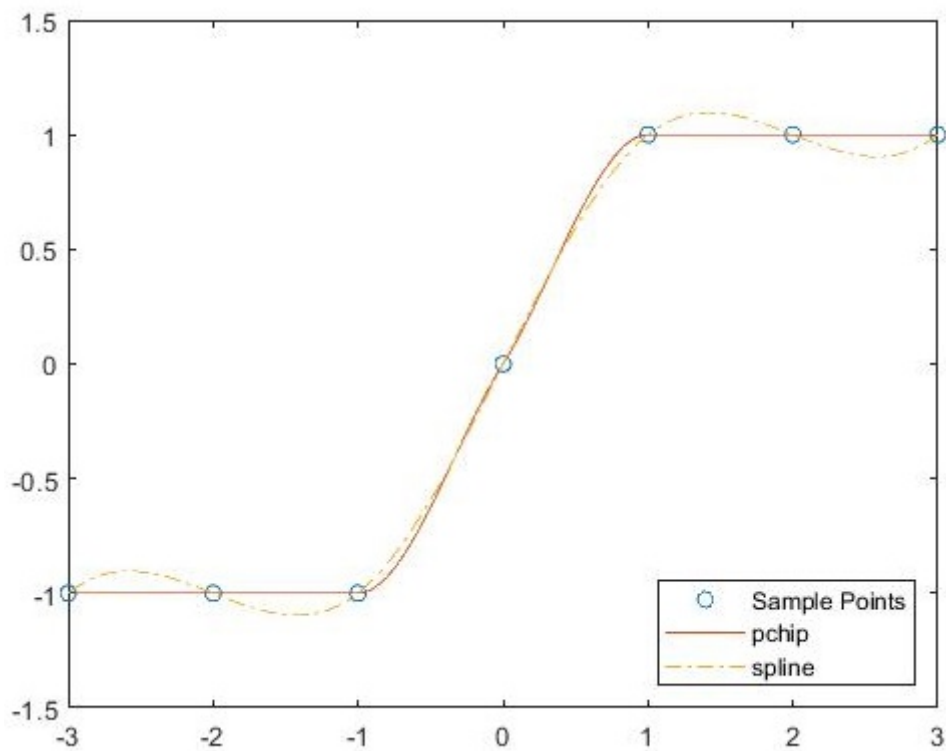



三次样条插值的原理在每两个样本点之间用两两相切的三次函数曲线来插值.若 `spline()` 方法不指定查询点,则返回一个结构体,其中封装了插值三次函数的系数.使用 `mkpp(breaks, coefs)` 可以创建一个类似的结构体,使用 `ppval(pp, xq)` 可以计算查询点的插值结果.



使用 `pchip(x, y, xq)` 函数可以进行三次Hermite插值,该算法同样以三次函数进行插值,但得到的曲线更平缓.

```
1 x = -3:3; y = [-1 -1 -1 0 1 1 1]; xq1 = -3:.01:3;
2 p = pchip(x,y,xq1);
3 s = spline(x,y,xq1);
4 plot(x,y,'o',xq1,p,'-',xq1,s,'-.-')
5 legend('Sample Points','pchip','spline','Location','SouthEast')
```



二维插值

使用 `interp2()` 可以进行二维插值,向其 `method` 参数传入字符串可以指定插值算法.

方法	说明	连续性
'linear'	(默认)在查询点插入的值基于各维中邻点网格点处数值的线性插值.	C0
'spline'	在查询点插入的值基于各维中邻点网格点处数值的三次插值.插值基于使用非结终止条件的三次样条.	C2
'nearest'	在查询点插入的值是距样本网格点最近的值.	不连续
'cubic'	在查询点插入的值基于各维中邻点网格点处数值的三次插值.插值基于三次卷积.	C1
'makima'	修改后的Akima三次Hermite插值.在查询点插入的值基于次数最大为3的多项式的分段函数,使用各维中相邻网格点的值进行计算.为防过冲,已改进Akima 公式.	C1

```

1 % 构建样本点
2 xx = -2:.5:2; yy = -2:.5:3; [x,y] = meshgrid(xx,yy);
3 xx_i = -2:.1:2; yy_i = -2:.1:3; [x_i,y_i] = meshgrid(xx_i,yy_i);
4 z = x.*exp(-x.^2-y.^2);
5
6 % 线性插值
7 subplot(1, 2, 1);
8 z_i = interp2(xx,yy,z,x_i,y_i);
9 surf(x_i,y_i,z_i); hold on;

```

```

10 plot3(x,y,z+0.01,'ok','MarkerFaceColor','r'); hold on;
11
12 % 三次插值
13 subplot(1, 2, 2);
14 z_ic = interp2(xx,yy,z,x_i,y_i, 'spline');
15 surf(x_i,y_i,z_ic); hold on;
16 plot3(x,y,z+0.01,'ok','MarkerFaceColor','r'); hold on;

```

