دانشکده ی مهندسی برق

دانشگاه صنعتی امیرکبیر

مقدمه ای بر

آمار و احتمالات در MatLab

گردآوری

دانیال خشابی ، مصطفی علیزاده، سید فریبرز زارعی

{ danyal.khashabi, mostafa_ali_kk, zareeseyyedfariborz }@yahoo.com

اردیبهشت ۱۳۸۸

اولین نکته ای که در برنامه نویسی **MatLab** باید به خاطر داشته باشید این است که تمامی نوع داده ها در این محیط بصورت آرایه هستند. بنابرین تمامی داده ها برای آنالیز باید در آرایه قرار گیرند.

۱.۱ مقدار دهی به ماتریس اعداد

مقدار دهی به یک ماتریس شامل اعداد، مطابق نمونه، بصورت زیر انجام می گیرد:

```
>> A = [1 2 3 4; 1 10 8 5; 9 8 7 0; 0 0 0 1]
```

نمایش مقدار دهی به ماتریس توسط **MatLab** بصورت زیر است :

```
A =

     1     2     3     4
     1    10     8     5
     9     8     7     0
     0     0     0     1
```

در واقع به ازای هر سطر، در مقابل اعداد، سمی کالن(**;**) قرار خواهیم داد. برای دسترسی به هریک از خانه های ماتریس بصورت **MatrixName(row,column)** خانه ی مورد نظر را فراخوانی می کنیم. برای مثال:

```
>> A(2,1)
```

خروجی:

```
ans =

-1
```

۱.۱ شاخص های مرکزی

۱.۱.۱ میانگین ها

Mean(matrixName)	میانگین حسابی
Geomean(matrixName)	میانگین هندسی
Harmmean(matrixName)	میانگین هارمونیک

برای مثال داریم :

```
>> mean(A)

ans =

    2.7500    5.0000    4.5000    2.5000

>> geomean(A)

ans =
```

```
     0      0      0      0
>> harmmean(A)
ans =
     0      0      0      0
```

مشاهده می شود که این توابع تنها میانگین ستونی ماتریس را بدست آورده اند. در واقع در اصل تابع به صورت **y=mean(x,dim)** می باشد که مقدار **dim** مشخص خواهد کرد که میانگین به چه شیوه ای بر روی ماتریس اعمال شود. اگر **dime=1** باشد تابع مانند **y=mean(x)** عمل می کند واگر **dime=2** باشد میانگین از هر سطرماتریس **x** گرفته می شود واگر **dim=3** باشد در بعد سوم ماتریس میانگین گرفته می شود در واقع چون در اینجا از لحاظ بعد سوم، هر خانه تنها یک عضو دارد لذا **y=mean(x,3)=x** خواهد بود.

۲.۱.۱ میانه:

برای مثال:

```
>> median(A)
ans =
0.5000  5.0000  5.0000   2.5000
```

مشاهده می شود تابع **median** مانند تابع میانگین برای حالات پیشفرض روی ستون ها عمل می کند. مانند توابع میانگین می توان پارا متر دومی را برای مشخص کردن نحوه ی محاسبه ی میانه بکار برد.

```
>> median(A, 2)
ans =
    2.5000
    6.5000
    7.5000
         0
```

۳.۱.۱ مُد:

برای مثال:

```
>> mode(A)
ans =
    -1      0      0      0
```

مانند توابع بالا می توان پرامتر سومی را برای معین کردن نحوه ی مد گیری برای داده ها معین کرد.

۲.۱ شاخص های پراکندگی

۱.۲.۱ دامنه

بصورت بدیهی برابر است با اختلاف مثبت بزرگترین داده و کوچکترین داده:

```
>> range(A)

ans =

     9    10     8     5
```

توابع بزرگترین داده و کوچکترین داده نیز در محاسبات آماری پرکاربردند:

```
>> min(A)
ans =

-1 0 0 0

>> max(A)

ans =
```

9 10 8 5

Rang(x,dime) **دامنه تغییرات را روی بعد** dime **ماتریس** x **مشخص می کند.**

۲.۲.۱ واریانس

```
>> var(A)

ans =

   17.5833   22.6667   13.6667    5.6667
```

باید دقت کنیم که تابع بالا واریانس را از طریق فرمول زیر حساب می کند:

در صورتی که بخواهیم واریانس را بصورت $\frac{1}{n}\sum_{k=1}^{n}(\mathbf{x_i}-\overline{\mathbf{x}})^2$ حساب کنیم، باید پارامتر دوم تابع را ۱ قرار دهیم:

```
>> var(A, 1)

ans =

   13.1875   17.0000   10.2500    4.2500
```

در واقع **var(x)** برابر است با **var(x, 0)** .

گاهی لازم می شود تا واریانس برای داده ها بصورت وزن دار حساب شود.در اینصورت تابع **var(x,w)** واریانس را با استفاده از ماتریس وزن بردار**w** حساب کند. بدیهی است که طول **w** با طول بعدی که واریانس روی آن عمل می کند

باید برابر باشد و عناصر **w** باید نامنفی باشند. تابع خود مقادیر **w** را به گونه ای تنظیم می کند که حاصل جمع آنها برابر با ۱ باشد.

در حالت کلی داریم **var(x,w,dime)**

```
>> w=[2 2 2 2 ]
w =
     1     1     1     1


>> var(A, 1)
ans =
   13.1875   17.0000   10.2500    4.2500


>> var(A, w , 1)
ans =
   13.1875   17.0000   10.2500    4.2500


>> w=[2 2 2 2 ]
w =
     2     2     2     2


>> var(A, w, 1)
ans =
   13.1875   17.0000   10.2500    4.2500


>> w=[1 2 3 4]
w =
     1     2     3     4

>> var(A, w)
ans =
   15.6000   18.4400   12.4000    3.7600
```

۳.۲.۱ **انحراف معیار** (Standard Deviation)

```
>> std(A)

ans =

4.5735  4.7610  3.6968  2.3805
```

بدیهی است با توجه به رابطه ی رادیکالی واریانس و انحراف معیار، پارامتر دومی برای تعیین نوع انحراف معیار گیری، تعریف شده باشد بطوریکه :

$$std(x,0) = std(x) = \left(\frac{1}{n-1}\sum_{k=1}^{n}(\mathrm{x_i}-\overline{\mathrm{x}})^2\right)^{\frac{1}{2}}$$

```
>> std(A,1)
ans =
```

```
    3.6315    4.1231    3.2016    2.0616

>> std(A,0)
ans =
    4.1932    4.7610    3.6968    2.3805
```

پارامتر سومی نیز برای تعیین بعدی که واریانس در آن عمل می کند تعریف می شود. بنابرین در حالت کلی داریم :
s=var(x,flag,dime) بطوریکه **flag=0** یا **flag=1** باشد انحراف معیار به ترتیب از نوع اول و دوم حساب می شود و اگر **dime=1** باشد انحراف معیاربرای هر ستون ماتریس **x** حساب می شود و اگر **dime=2** باشد انحراف معیاربرای هرردیف ماتریس **x** حساب می شود .

۱.۲.٤ کواریانس(Covariance)

Cov(x) مقدار کواریانس را به ازای یک تابع ورودی برابر با واریانس محاسبه می کند.

```
>> cov(A)
ans =
   17.5833    9.6667    8.1667   -6.1667
    9.6667   22.6667   17.3333    2.6667
    8.1667   17.3333   13.6667    2.3333
   -6.1667    2.6667    2.3333    5.6667

>> diag(cov(A))

ans =

   17.5833
   22.6667
   13.6667
    5.6667

>> var(A)

ans =

   17.5833   22.6667   13.6667    5.6667
```

y=cov(x, y, flag) در این تابع **flag** نوع تابع را مشخص می کند. اگر **flag=0** باشد به جای اینکه تعداد داده ها را در معادله ی کواریانس **n** قرار دهد **n-1** را در معادله قرار داده و جواب را می دهد (یعنی کوراریانس از نوع اول) و اگر **flag=1** باشد کواریانس از نوع دوم خواهد بود.

۳.۱ برازش نمودار ها روی داده ها و بدست آوردن خطا

۱.۳.۱ برازش بر روی توابع چند جمله ای :

با استفاده از تابع **polyfit(x, y, degree)** می توان داده ها را بر روی یک تابع چند جمله ای از درجه ی **degree** برازش نمود.

به عنوان مثال:

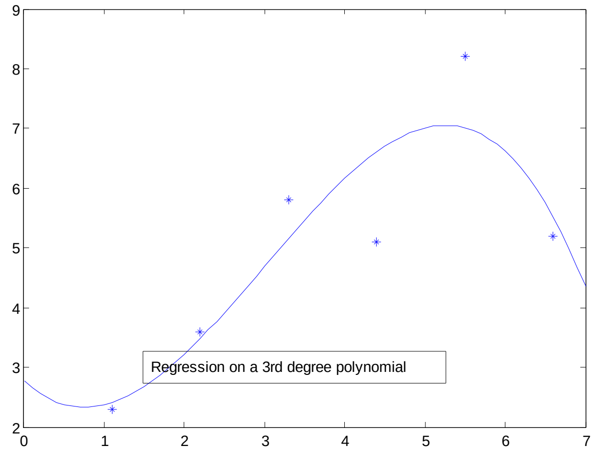
```
>> x=[1.100  2.200  3.300  4.400  5.500  6.600]';
>> y=[2.300  3.600  5.800  5.100  8.200  5.200]';


>> polyfit(x, y, 1)
ans =
    0.7169    2.2733


>> polyfit(x, y, 2)
ans =
   -0.2642    2.7510   -0.7100


>> polyfit(x, y, 3)
ans =
   -0.1037    0.9330   -1.2249    2.7667
```

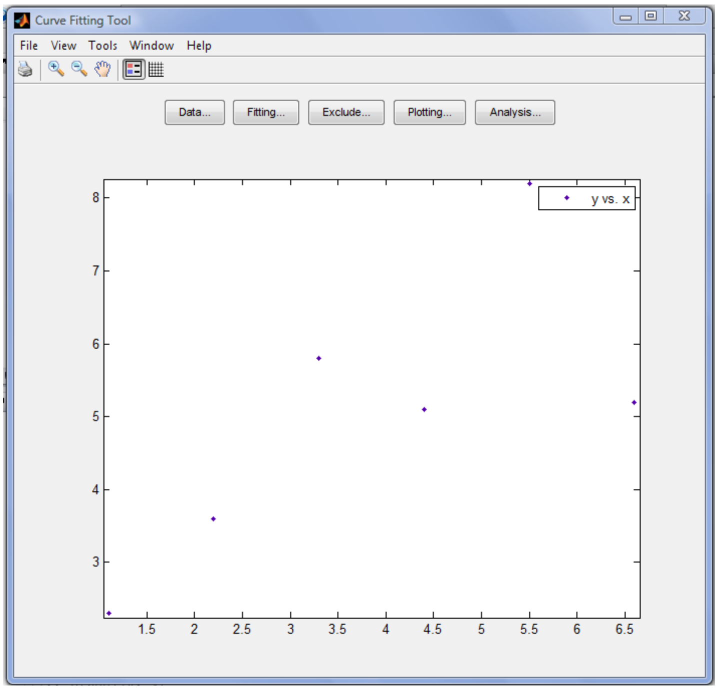
به عنوان نتایج نموداری برای داده های بالا داریم:



۲.۳.۱ برازش بر روی انواع توابع مختلف به کمک ابزار برازش(Curve Fitting Tool)

برای استفاده از این ابزار از دستور **Cftool(x, y)** استفاده می کنیم. برای مثال:

```
>> x=[1.100  2.200  3.300  4.400  5.500  6.600]';
>> y=[2.300  3.600  5.800  5.100  8.200  5.200]';
>> cftool(x,y)
```

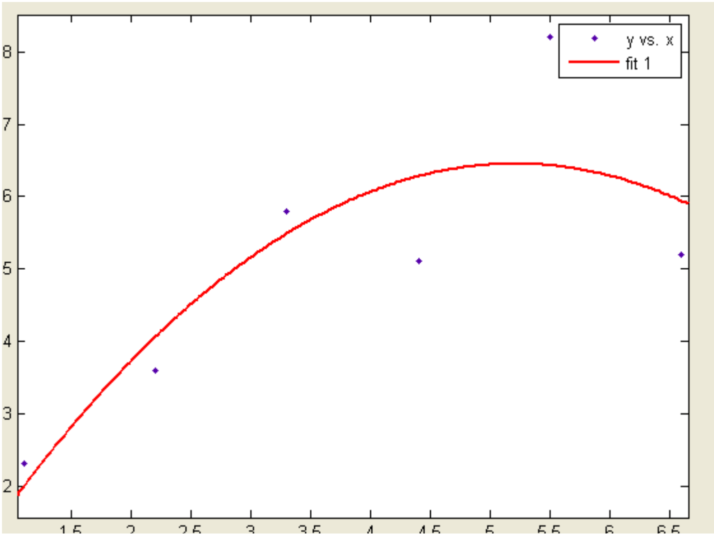


از پنجره ی باز شده می توانیم:

- داده ها را بر روی انواع منحنی های چند چمله ای خطی، چند جمله ای درجه ۲ ، چند جمله ای درجه ۳(**cubic polynomial**) و ... ، منحنی گاوسی(**Gaussian Curve**) ، منحنی های توانی، توابع مثلثاتی و مجموع آنها ، برازش کرد.
- می توان برای داده ها وزن تعریف کرد و برازش را به صورت وزن دار انجام داد.
- برخی نقاط خاص را از محدوده ی برازش خارج کرد.
- نتایج داده ها را آنالیز کرد.

و ...

برای برازش دلخواه داریم:



٤.١ چند تابع دیگر

١.٤.١ گشتاور حول مبدا

اگر **x** یک بردار باشد، **Moment(x, r)** گشتاور مرتبه ی **r** حول مبدا برای تمام اعضای بردار **x** حساب می کند و اگر **x** یک ماتریس باشد این تابع گشتاور مرتبه ی **r** را حول مبدا برای هر ستون ماتریس **x** حساب می کند.

٢.٤.١ چولگی

می دانیم تابع چولگی همان B_3 یا گشتاور استاندارد مرتبه ۳ می باشد. **skewness(x)** دستوری است که چولگی داده ها را حساب می کند، وقتی که **x** یک بردار باشد(یعنی داده ها فقط در یک ردیف باشند) و اگر **x** ماتریس باشد چولگی در هر ستون را حساب می کند.

٣.٤.١ قدر مطلق انحراف از میانگین/ میانه

تابع **mad(x)** (**Mean/Median**) قدر مطلق انجراف ها را از میانگین/میانه را می هد .

اگر **x** یک بردار باشد دستوری است که قدر مطلق انحرافات از میانگین را برای تمام درایه های **x** حساب می کندو اگر **x**یک ماتریس باشد **y** قدر مطلق انحرافات از میانگین برای هر ستون **x** می باشد.

mad(x,flag) در این دستور اگر **flag=0** باشد مانند **mad(x)** کند(یعنی به **صورت** $\overline{|x-\bar{x}|}$)

اگر **flag** هر عددی دیگر باشد **y** برای یک بردار، میانه ی قدر مطلق انحرافات از میانه را حساب می کند و برای یک ماتریس میانه ی قدرمطلق انحرافات از میانه ی هر ستون آن ماتریس را بدست می آورد.(یعنی به صورت : **(median(abs(x-median(x)))**

y =mad(x,flag,dim) اگر **flag=0** باشد **y** برابر است با میانگین قدر مطلق انحرافات از میانگین روی بعد **dime** برای ماتریس **x** و اگر **flag** هر عددی به جز این باشد **y** برابر است با قدرمطلق انحرافات از میانه روی بعد **dime** برای ماتریس **x** .

٤.٤.١ تابع ضریب همبستگی (R_{xy}):

corrcoef(x, y) ضریب همبستگی برای برای داده های ستونی هر ماتریس حساب می کند.