IE 6318 Data Mining and Analytics

Data Exploration

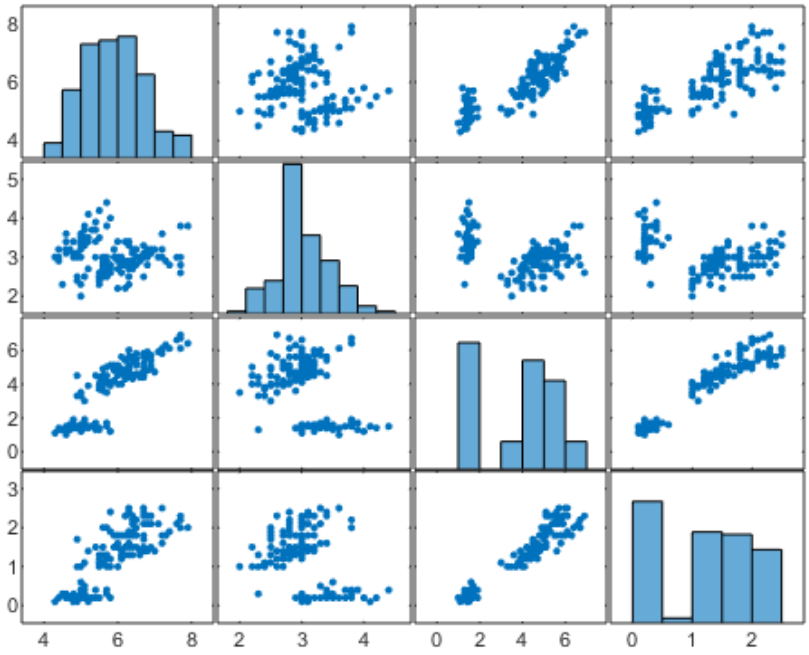
1. Raw data processing

```
% Load Data %
data = importdata('iris.txt');
features = data(:,1:4);
class = data(:,5);
```

2. Exploring Iris dataset

2.1 2D scatter plots

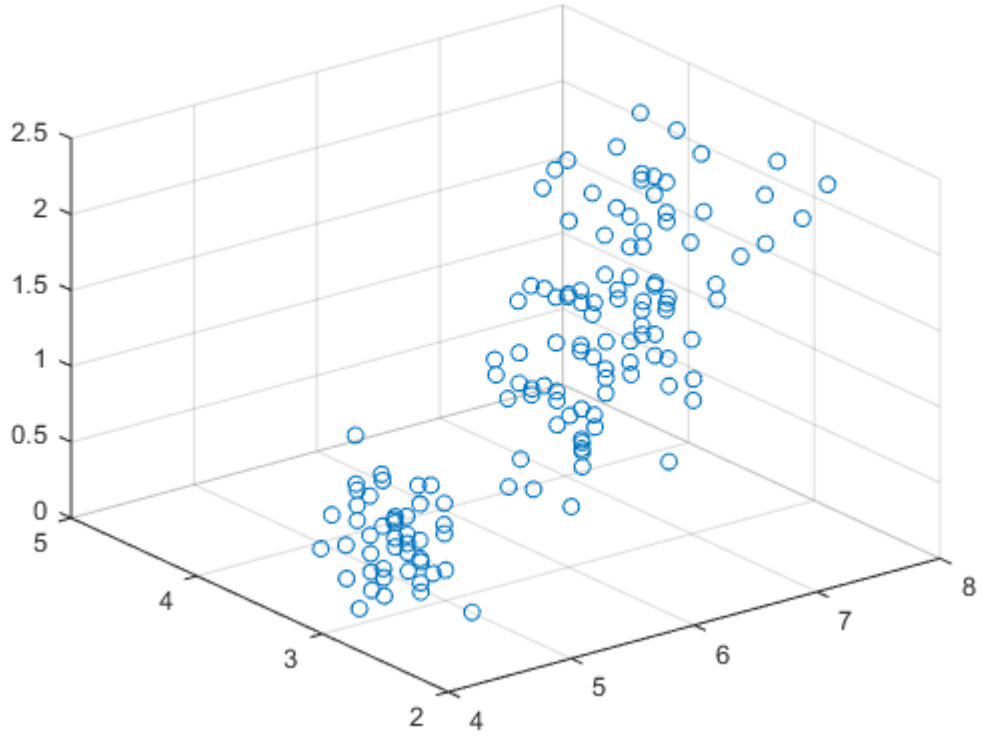
```
% 2.1 - 2D Scatter plot matrix %
figure(1);
plotmatrix(features)
```



2D scatter plot of attributes

2.2 3D scatter plot

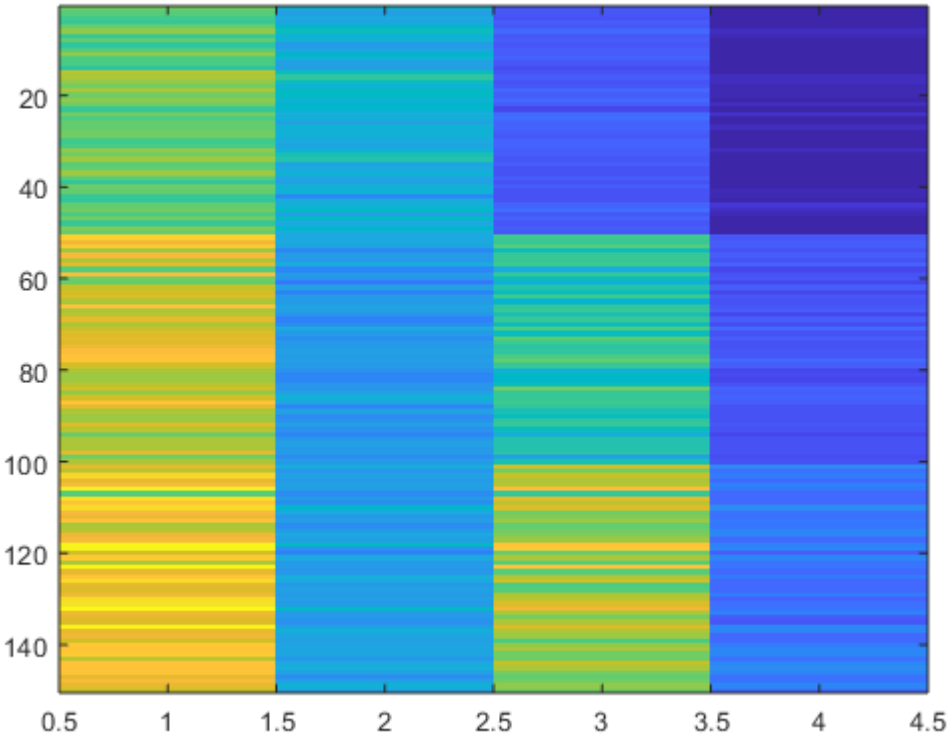
```
% 2.2 - 3D Scatter plot of 3 attributes %
Sepal_length = features(:,1);
Sepal_width = features(:,2);
Petal_width = features(:,4);
figure(2);
scatter3(Sepal_length, Sepal_width, Petal_width)
```



3D scatter plot of three attributes

2.3 Visualization of the feature matrix

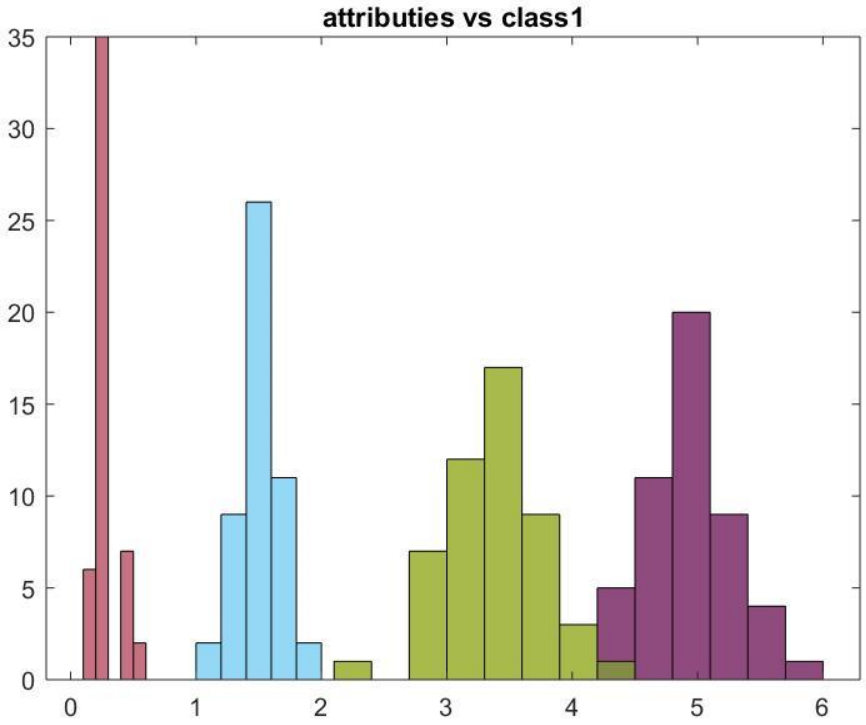
```
% 2.3 - Visualization of the feature matrix (4 columns) %
figure(3);
imagesc(features)
```

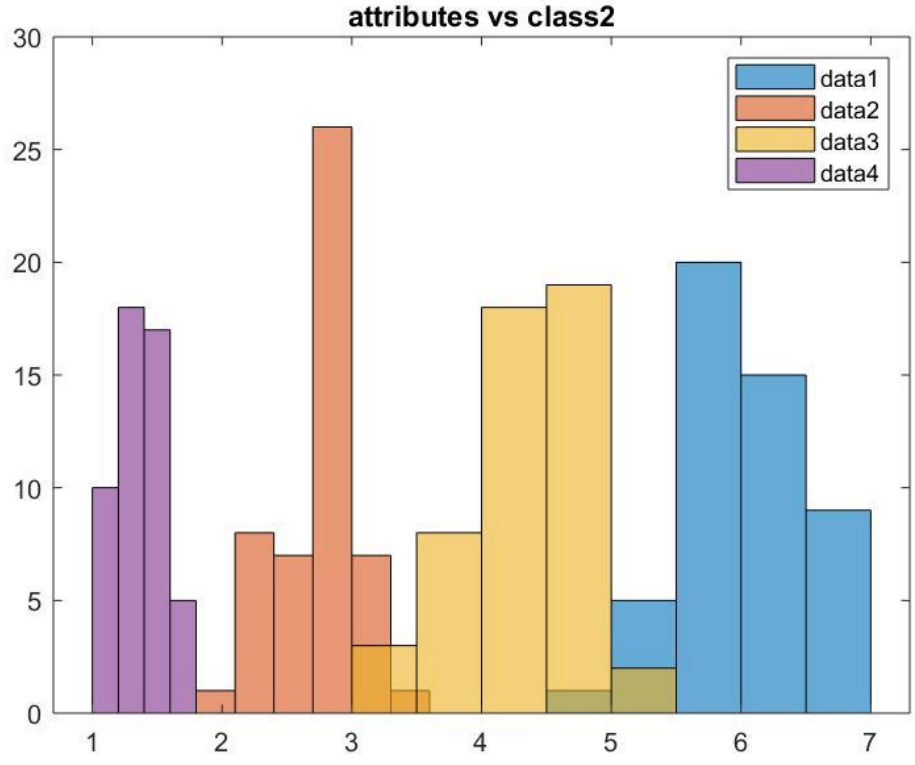


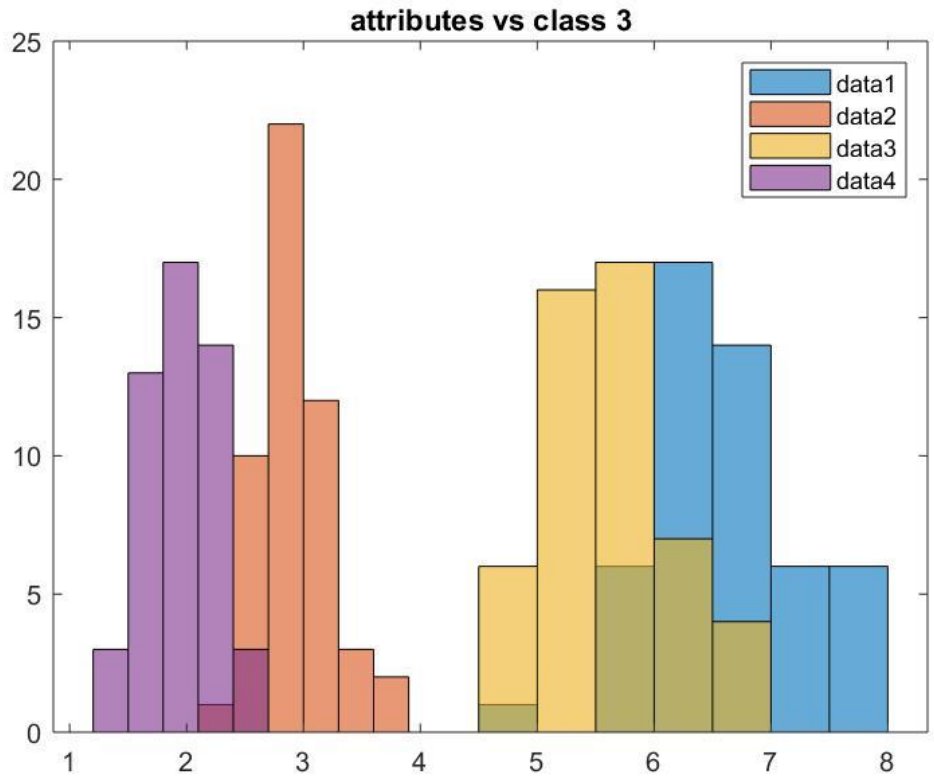
2.3 – Visualization of feature matrix

2.4 Histogram

```
% 2.4 - Histogram of four attribues of 3 classes %
figure;
histogram(data1);
hold on;
histogram(data2);
hold on;
histogram(data3);
hold on;
histogram(data4);
```

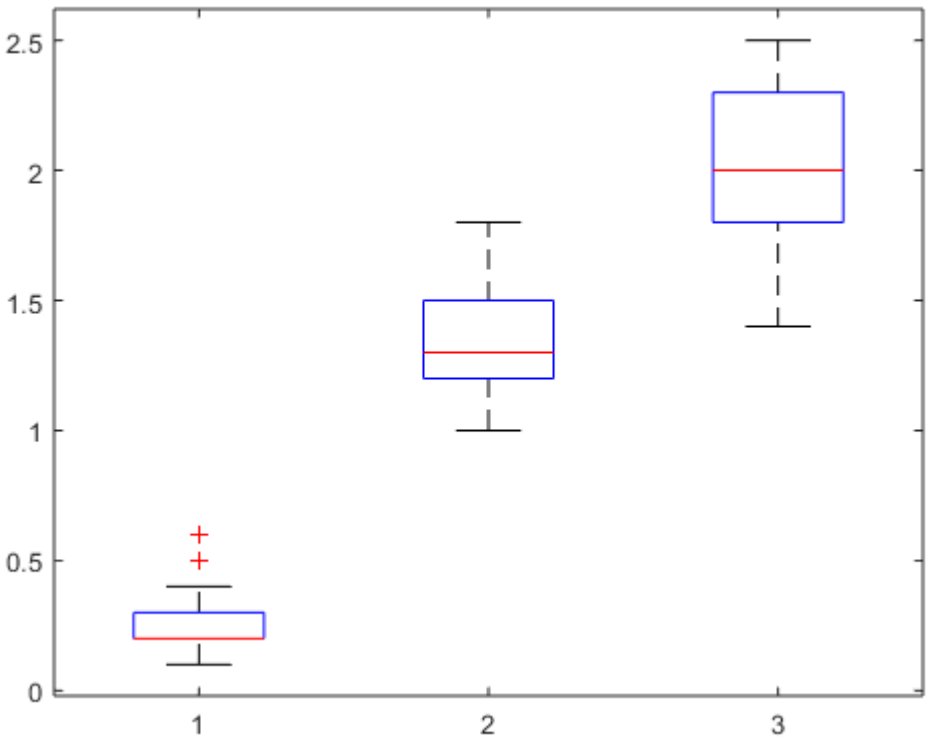

attributies vs class1
35
30
25
20
15
10
5
0
0
1
2
3
4
5
6

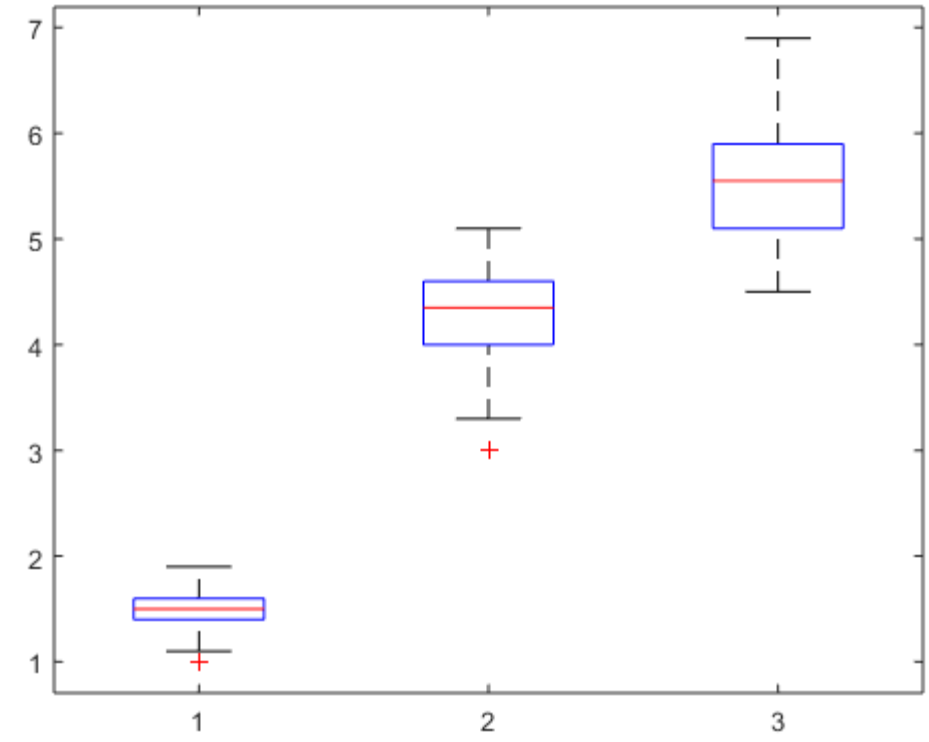

attributes vs class2
data1
data2
data3
data4
30
25
20
15
10
5
0
1
2
3
4
5
6
7

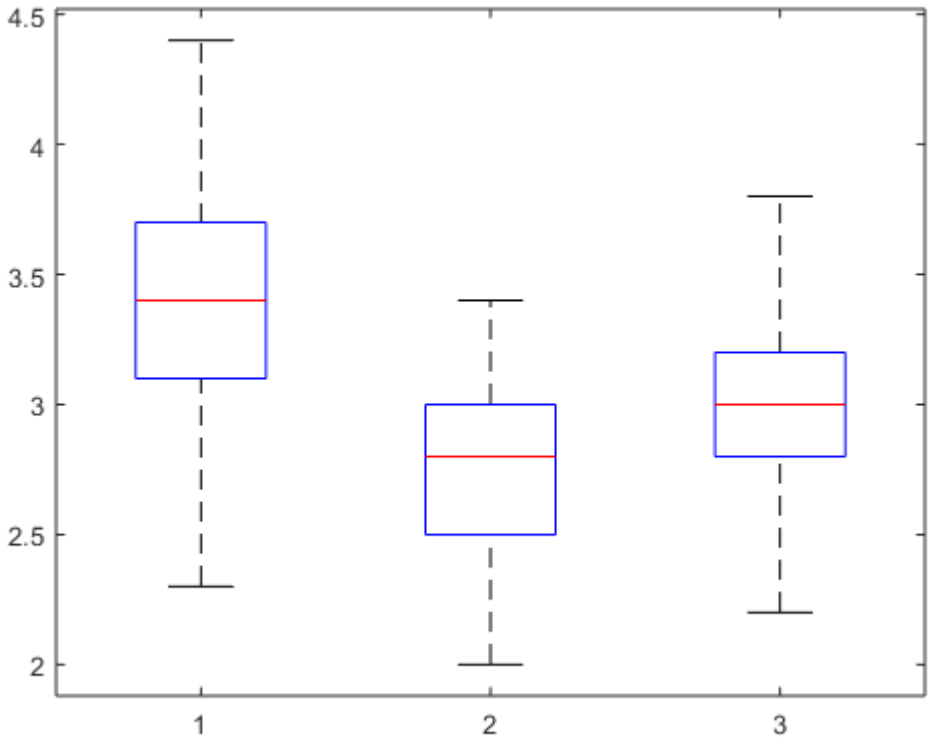


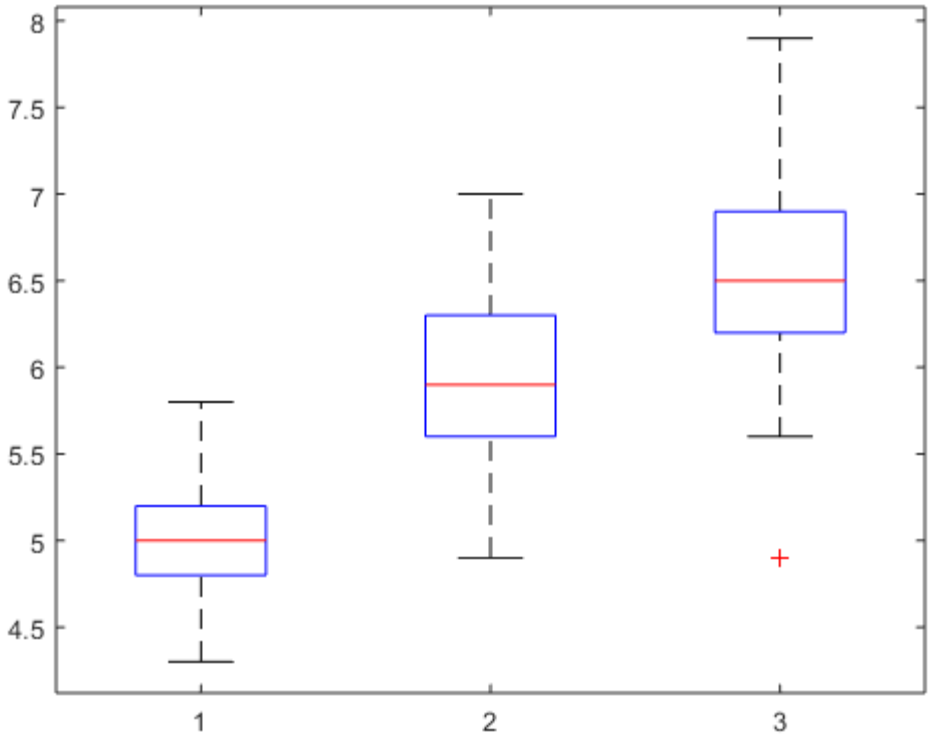
2.5 Boxplots

```
% 2.5 - Boxplots %
figure(4)
boxplot(features(:,1),class);
figure(5)
boxplot(features(:,2),class);
figure(6)
boxplot(features(:,3),class);
figure(7)
boxplot(features(:,4),class);
```


2.5
2
1.5
1
0.5
0
1
2
3


7
6
5
4
3
2
1
1
2
3

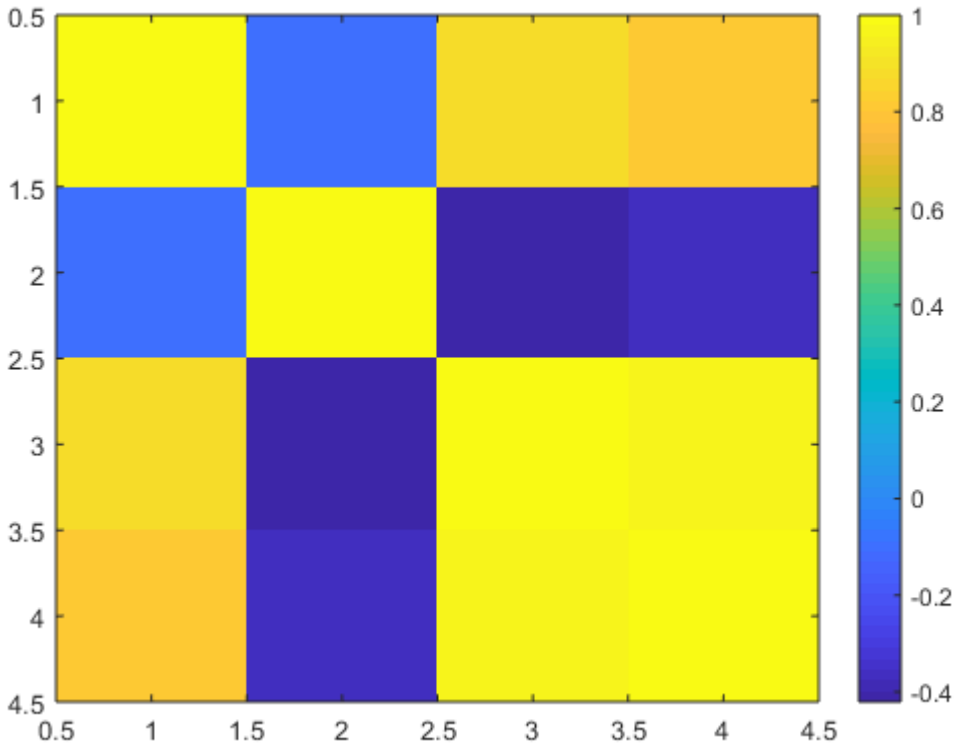

4.5
4
3.5
3
2.5
2
1
2
3


8
7.5
7
6.5
6
5.5
5
4.5
1
2
3

2.6 Correlation matrix and plot

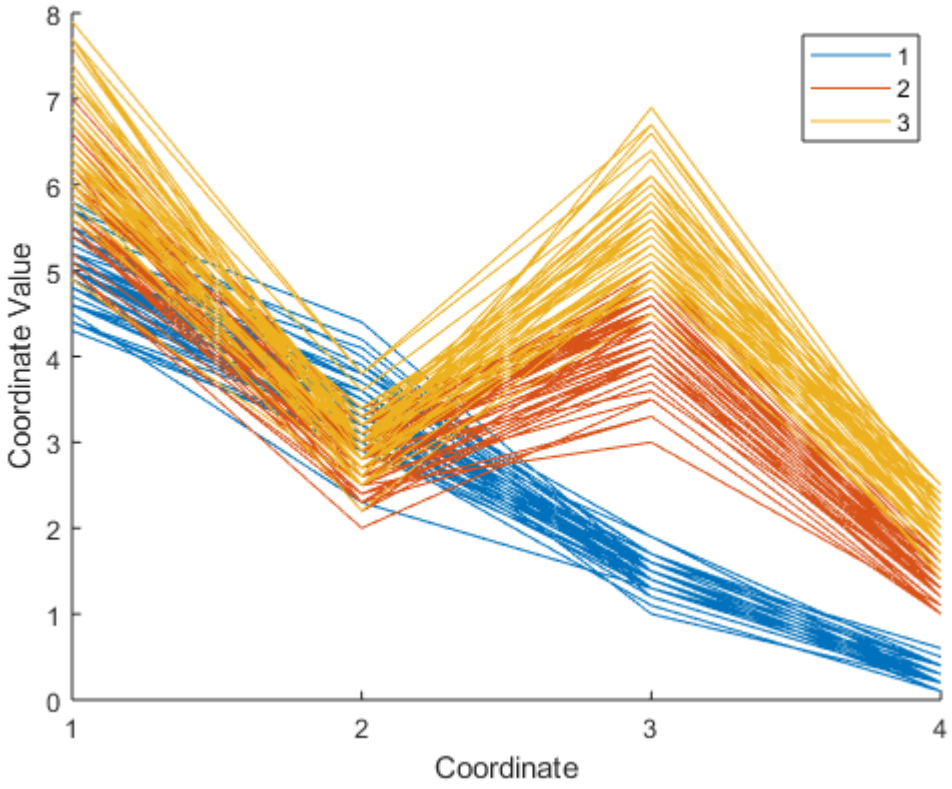
```
% 2.6 - Correlation matrix and plot %
figure(8)
C = corr(features);
imagesc(C)
colorbar;
```

```
      C =

    1.0000   -0.1094    0.8718    0.8180
   -0.1094    1.0000   -0.4205   -0.3565
    0.8718   -0.4205    1.0000    0.9628
    0.8180   -0.3565    0.9628    1.0000
```



2.7 Parallel coordinates plot

```
% 2.7 - Parallel coordinates plot %
figure(9)
parallelcoords(features,'group', class)
```



3. Data Distance Measures

3.1 Minkowski Distance

```
function [mink] = minko(A,B,r)

for i=1:150
minko(i)=(sum((abs(A(i,:)-B)).^r).^(1/r));
end
```

3.2 T-statistics Distance

```
function k =t_dist(X,Y)

Ex =mean(X);
Ey =mean(Y);
c =Ex-Ey;
d =X-Y;
k =(abs(c))/(std(d));
end
```

3.3 Mahalanobis Distance

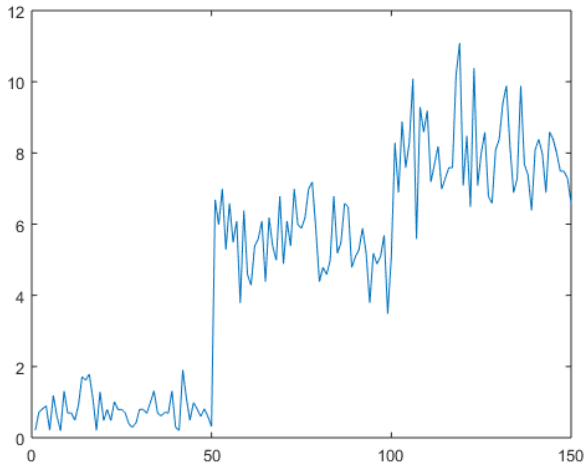
```
function [d] =mahanalobis(A,b,r)
[dist, iidx] = pdist2(X, Y-mean(X),'mahalanobis',r);

num_samples_per_class = 50;
matching_class = ceil(iidx/ num_samples_per_class);
```

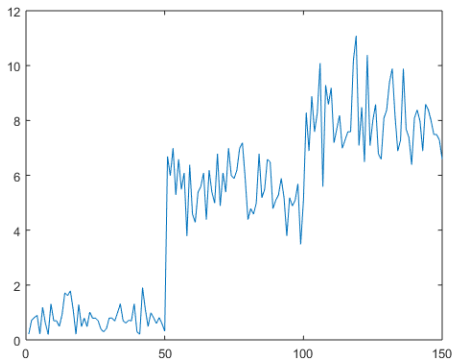
4. Calculate the distances

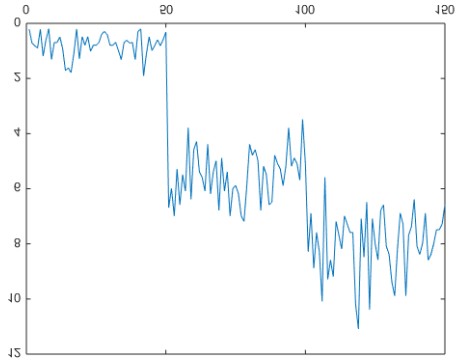
R = 1	R = 2	R = 100
-0.1140	0.0807	0.0600
-0.8140	0.5445	0.5000
-0.9140	0.5277	0.4000
-0.9140	0.6438	0.5000
-0.1140	0.1628	0.1007
1.0860	0.5735	0.4000
-0.6140	0.5155	0.5000
-0.2140	0.1566	0.1007
-1.4140	0.9255	0.7000
-0.7140	0.4747	0.4000
0.4860	0.3668	0.3000
-0.3140	0.3500	0.3000
-1.0140	0.6061	0.5000
-1.8140	1.0214	0.8000
0.8860	0.9003	0.7000
1.6860	1.0922	0.9000
0.6860	0.5449	0.4000
-0.0140	0.0756	0.0600
1.1860	0.7139	0.6000
0.3860	0.3061	0.3000
0.3860	0.4006	0.3000
0.3860	0.2508	0.2000
-0.9140	0.6889	0.5000
0.2860	0.3976	0.2462
-0.0140	0.5445	0.4400
-0.5140	0.5315	0.5000
0.0860	0.2468	0.1460
0.0860	0.1205	0.1000
-0.1140	0.1628	0.1007
-0.6140	0.5220	0.4000
-0.6140	0.5220	0.4000
0.3860	0.3506	0.3000
0.5860	0.6287	0.6000
0.9860	0.8103	0.7000
-0.7140	0.4747	0.4000
-0.7140	0.4129	0.3000
0.1860	0.4342	0.4000
-0.7140	0.4747	0.4000
-1.4140	0.8767	0.7000
-0.1140	0.1205	0.1000
-0.2140	0.1942	0.1600
-1.9140	1.3519	1.2000
-1.2140	0.7801	0.7000
0.3860	0.3864	0.3460

0.8860	0.5522	0.4400
-0.8140	0.5880	0.5000
0.3860	0.3354	0.3000
-0.9140	0.5887	0.5000
0.3860	0.2907	0.2014
-0.4140	0.2377	0.2000
5.9860	3.9384	3.2400
5.2860	3.5460	3.0400
6.0860	4.0971	3.4400
2.7860	3.0242	2.5400
5.0860	3.7232	3.1400
3.9860	3.3445	3.0400
5.5860	3.7134	3.2400
1.2860	2.2786	1.8400
5.0860	3.6829	3.1400
2.8860	2.8137	2.4400
1.1860	2.6416	2.0400
4.2860	3.1544	2.7400
2.8860	3.0835	2.5400
4.7860	3.6292	3.2400
3.0860	2.5067	2.1400
5.2860	3.5605	2.9400
4.2860	3.3607	3.0400
3.2860	2.9421	2.6400
4.0860	3.7006	3.0400
2.7860	2.8141	2.4400
5.3860	3.7783	3.3400
3.8860	3.0059	2.5400
4.8860	3.9782	3.4400
4.4860	3.5892	3.2400
4.5860	3.3481	2.8400
5.0860	3.5294	2.9400
5.4860	3.9811	3.3400
6.0860	4.1752	3.5400
4.5860	3.4589	3.0400
2.4860	2.4265	2.0400
2.4860	2.7498	2.3400
2.2860	2.6352	2.2400
3.2860	2.8246	2.4400
5.0860	4.0634	3.6400
4.0860	3.3368	3.0400
5.1860	3.4458	3.0400
5.6860	3.8432	3.2400
3.9860	3.5521	2.9400
3.6860	2.9264	2.6400
2.9860	2.9505	2.5400
3.3860	3.2417	2.9400
4.7860	3.5246	3.1400
3.2860	2.9405	2.5400
1.2860	2.3221	1.8400
3.4860	3.0809	2.7400
3.7860	3.0021	2.7400
3.7860	3.0531	2.7400
4.3860	3.2756	2.8400
1.3860	2.0217	1.5400
3.5860	2.9856	2.6400
7.7860	5.2092	4.5400
5.1860	4.1339	3.6400

7.7860	5.2318	4.4400
6.2860	4.6184	4.1400
7.1860	4.9832	4.3400
8.9860	6.0272	5.1400
3.2860	3.5175	3.0400
7.9860	5.5692	4.8400
6.4860	4.9785	4.3400
9.0860	5.5672	4.6400
6.4860	4.2835	3.6400
5.9860	4.4480	3.8400
7.0860	4.7822	4.0400
4.8860	4.1158	3.5400
5.7860	4.3399	3.6400
6.8860	4.5510	3.8400
6.4860	4.5740	4.0400
10.0860	6.1721	5.2400
9.1860	6.4304	5.4400
4.3860	4.0724	3.5400
7.7860	5.0491	4.2400
4.9860	3.9525	3.4400
8.8860	6.1446	5.2400
5.3860	4.0378	3.4400
7.4860	4.8975	4.2400
7.8860	5.2442	4.5400
5.2860	3.9046	3.3400
5.4860	3.9337	3.4400
6.5860	4.7673	4.1400
7.2860	5.0306	4.3400
7.8860	5.4790	4.6400
9.7860	5.9483	4.9400
6.6860	4.8069	4.1400
5.3860	4.0905	3.6400
5.3860	4.5014	4.1400
8.7860	5.7206	4.6400
7.3860	4.8161	4.1400
6.4860	4.5345	4.0400
5.2860	3.8217	3.3400
7.1860	4.7256	3.9400
7.4860	4.9462	4.1400
7.0860	4.5646	3.6400
5.1860	4.1339	3.6400
7.8860	5.1846	4.4400
7.8860	5.0618	4.2400
6.8860	4.5808	3.7400
5.3860	4.2049	3.5400
6.3860	4.3870	3.7400
6.9860	4.5749	3.9400
5.4860	4.0657	3.6400



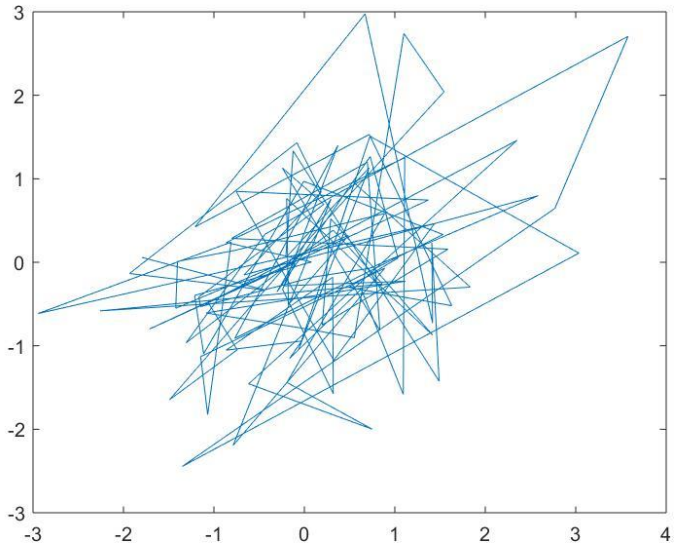




5. Time series data

5.1 Plot of two time series data

```
plot(VarName1,VarName2)
```



5.2 T-statistics distance between the two time series

0.1285

5.3 Correlation of two time series

```
Corr(timeseries)

ans =

          1.0000    0.4030
          0.4030    1.0000
```

5.4 Normalization of feature matrix

```
normalize(a,2);

    1.1700    0.4359   -0.5277   -1.0783
    1.2396    0.3068   -0.4787   -1.0678
    1.1765    0.4255   -0.5257   -1.0763
    1.1766    0.3922   -0.4445   -1.1243
    1.1362    0.4869   -0.5333   -1.0898
    1.1431    0.4707   -0.5155   -1.0982
    1.1233    0.5035   -0.5294   -1.0975
    1.1734    0.4148   -0.4859   -1.1023
    1.1932    0.3703   -0.4526   -1.1109
    1.2084    0.3384   -0.4350   -1.1117
    1.1699    0.4333   -0.5199   -1.0832
    1.1405    0.4463   -0.4463   -1.1405
    1.2180    0.3322   -0.4552   -1.0950
    1.1545    0.4645   -0.5441   -1.0749
    1.1689    0.4676   -0.6234   -1.0131
```

1.0944	0.5675	-0.6080	-1.0539
1.1486	0.4984	-0.6285	-1.0185
1.1778	0.4315	-0.5481	-1.0612
1.1921	0.3903	-0.4958	-1.0866
1.1159	0.5177	-0.5407	-1.0929
1.2175	0.3239	-0.4356	-1.1058
1.1437	0.4834	-0.5542	-1.0729
1.0780	0.5989	-0.6468	-1.0301
1.2276	0.3257	-0.4760	-1.0772
1.1254	0.4173	-0.3414	-1.2012
1.2448	0.2685	-0.4149	-1.0983
1.1882	0.3961	-0.4951	-1.0892
1.1810	0.4088	-0.4997	-1.0902
1.2017	0.3854	-0.5215	-1.0656
1.1667	0.3974	-0.4231	-1.1410
1.2012	0.3414	-0.4173	-1.1254
1.2392	0.3297	-0.5343	-1.0345
1.0583	0.5880	-0.5238	-1.1225
1.0931	0.5618	-0.5823	-1.0726
1.2084	0.3384	-0.4350	-1.1117
1.2176	0.3746	-0.5620	-1.0302
1.2197	0.3712	-0.5621	-1.0288
1.2084	0.3384	-0.4350	-1.1117
1.1745	0.4185	-0.4995	-1.0935
1.1868	0.3956	-0.4887	-1.0937
1.1655	0.4592	-0.5769	-1.0478
1.3361	0.1113	-0.4454	-1.0021
1.1290	0.4915	-0.5180	-1.1025
1.1851	0.4205	-0.5479	-1.0576
1.1109	0.4830	-0.4347	-1.1592
1.2371	0.3188	-0.4974	-1.0585
1.1058	0.5130	-0.4902	-1.1286
1.1588	0.4378	-0.4893	-1.1073
1.1568	0.4517	-0.5178	-1.0907
1.1953	0.3905	-0.5089	-1.0769
1.2336	-0.3690	0.2636	-1.1281
1.2075	-0.3381	0.2898	-1.1592
1.2034	-0.4298	0.3438	-1.1175
1.1992	-0.5255	0.3908	-1.0645
1.2176	-0.4825	0.3446	-1.0798
1.1024	-0.4020	0.4799	-1.1802
1.1612	-0.3371	0.3621	-1.1861
1.2232	-0.3058	0.2446	-1.1621
1.2087	-0.4175	0.3296	-1.1208
1.1679	-0.3688	0.3688	-1.1679
1.2143	-0.5000	0.3571	-1.0714
1.2079	-0.3489	0.2953	-1.1542
1.2375	-0.5042	0.3208	-1.0542
1.1315	-0.4258	0.4501	-1.1558
1.2624	-0.2525	0.1403	-1.1502
1.2530	-0.3580	0.2238	-1.1188
1.0918	-0.3639	0.4759	-1.2038
1.1758	-0.3429	0.3429	-1.1758
1.2062	-0.6495	0.4175	-0.9742
1.2072	-0.4024	0.3245	-1.1293
1.0979	-0.4030	0.4864	-1.1813
1.2578	-0.3699	0.2220	-1.1098
1.1395	-0.5925	0.5014	-1.0483

1.1182	-0.4193	0.4659	-1.1648
1.2362	-0.3813	0.2657	-1.1207
1.2470	-0.3854	0.2494	-1.1110
1.2090	-0.4879	0.3606	-1.0818
1.1810	-0.4997	0.4088	-1.0902
1.1667	-0.4231	0.3974	-1.1410
1.2747	-0.3059	0.1530	-1.1217
1.2179	-0.4236	0.3177	-1.1120
1.2267	-0.3915	0.2871	-1.1223
1.2344	-0.3600	0.2572	-1.1316
1.0503	-0.5618	0.6107	-1.0992
1.0498	-0.3499	0.5249	-1.2247
1.1464	-0.2563	0.3372	-1.2273
1.2140	-0.4047	0.3147	-1.1240
1.2224	-0.5720	0.3701	-1.0206
1.1584	-0.2758	0.3310	-1.2135
1.1932	-0.4526	0.3703	-1.1109
1.0892	-0.4331	0.5118	-1.1680
1.1470	-0.3823	0.4070	-1.1716
1.2205	-0.4068	0.3051	-1.1188
1.2447	-0.3556	0.2371	-1.1261
1.1564	-0.4034	0.4034	-1.1564
1.1428	-0.2758	0.3546	-1.2216
1.1611	-0.3336	0.3603	-1.1878
1.2126	-0.3722	0.3002	-1.1406
1.3118	-0.2563	0.0452	-1.1007
1.1878	-0.3603	0.3336	-1.1611
0.9300	-0.6419	0.7728	-1.0610
1.0296	-0.6284	0.6552	-1.0563
1.0905	-0.6458	0.5823	-1.0270
1.0021	-0.5826	0.6758	-1.0953
1.0143	-0.6563	0.6802	-1.0382
1.0340	-0.6800	0.6614	-1.0154
0.9709	-0.5826	0.7120	-1.1004
1.0322	-0.6345	0.6534	-1.0512
1.0363	-0.7047	0.6632	-0.9948
1.0813	-0.5752	0.5752	-1.0813
1.1529	-0.5013	0.4511	-1.1028
1.0949	-0.6475	0.5769	-1.0243
1.1257	-0.6203	0.5284	-1.0338
1.0417	-0.7128	0.6579	-0.9869
1.0578	-0.7300	0.6406	-0.9684
1.1161	-0.5846	0.5315	-1.0630
1.0579	-0.5520	0.5980	-1.1039
1.0219	-0.5110	0.6289	-1.1398
1.0011	-0.8062	0.7176	-0.9125
1.0737	-0.6812	0.6119	-1.0044
1.1103	-0.6194	0.5493	-1.0401
1.0431	-0.6023	0.6317	-1.0724
1.0283	-0.7092	0.6737	-0.9928
1.1585	-0.5976	0.4756	-1.0366
1.0618	-0.5427	0.5899	-1.1090
1.0668	-0.5435	0.5837	-1.1071
1.1637	-0.5565	0.4553	-1.0625
1.1203	-0.4950	0.4950	-1.1203
1.0382	-0.6802	0.6563	-1.0143
1.0954	-0.5477	0.5477	-1.0954
1.0873	-0.6677	0.5914	-1.0110

1.0917	-0.4651	0.5221	-1.1486
1.0429	-0.7034	0.6548	-0.9944
1.0929	-0.5177	0.5407	-1.1159
0.9517	-0.5798	0.7330	-1.1049
1.1447	-0.6947	0.5185	-0.9686
1.0245	-0.5601	0.6421	-1.1065
1.0371	-0.5185	0.6128	-1.1314
1.1257	-0.4825	0.4825	-1.1257
1.1595	-0.5855	0.4707	-1.0447
1.1063	-0.6638	0.5654	-1.0079
1.2331	-0.6044	0.3627	-0.9913
1.0296	-0.6284	0.6552	-1.0563
1.0502	-0.6301	0.6301	-1.0502
1.0882	-0.6327	0.5821	-1.0376
1.1872	-0.6431	0.4452	-0.9893
1.1441	-0.6865	0.5179	-0.9755
1.1359	-0.5741	0.5008	-1.0627
1.0467	-0.5164	0.6001	-1.1304
1.0350	-0.5042	0.6104	-1.1411