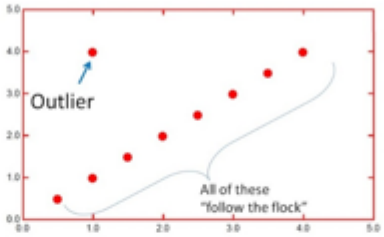

Never mind what the axes mean...

Dr. Bilal Wajid
Assistant Professor
University of Management and Technology
bilal.awan@umt.edu.pk

Umer Saeed (MS Data Science, BSc Telecommunication Engineering)
Sr. RF Optimization & Planning Engineer
f2017313017@umt.edu.pk

Detecting Outliers using Elliptic Envelope

Import required Libraries

In [1]:
```
import pandas as pd
from sklearn.covariance import EllipticEnvelope
```

Import DataSet

In [2]:
```
dset= pd.read_csv('ENB2012_data.csv')
dset
```

Out[2]:

	X1	X2	X3	X4	X5	X6	X7	X8	Y1	Y2
0	0.98	514.5	294.0	110.25	7.0	2	0.0	0	15.55	21.33
1	0.98	514.5	294.0	110.25	7.0	3	0.0	0	15.55	21.33
2	0.98	514.5	294.0	110.25	7.0	4	0.0	0	15.55	21.33
3	0.98	514.5	294.0	110.25	7.0	5	0.0	0	15.55	21.33
4	0.90	563.5	318.5	122.50	7.0	2	0.0	0	20.84	28.28
5	0.90	563.5	318.5	122.50	7.0	3	0.0	0	21.46	25.38
6	0.90	563.5	318.5	122.50	7.0	4	0.0	0	20.71	25.16
7	0.90	563.5	318.5	122.50	7.0	5	0.0	0	19.68	29.60
8	0.86	588.0	294.0	147.00	7.0	2	0.0	0	19.50	27.30
9	0.86	588.0	294.0	147.00	7.0	3	0.0	0	19.95	21.97
10	0.86	588.0	294.0	147.00	7.0	4	0.0	0	19.34	23.49
11	0.86	588.0	294.0	147.00	7.0	5	0.0	0	18.31	27.87
12	0.82	612.5	318.5	147.00	7.0	2	0.0	0	17.05	23.77
13	0.82	612.5	318.5	147.00	7.0	3	0.0	0	17.41	21.46
14	0.82	612.5	318.5	147.00	7.0	4	0.0	0	16.95	21.16
15	0.82	612.5	318.5	147.00	7.0	5	0.0	0	15.98	24.93
16	0.79	637.0	343.0	147.00	7.0	2	0.0	0	28.52	37.73
17	0.79	637.0	343.0	147.00	7.0	3	0.0	0	29.90	31.27
18	0.79	637.0	343.0	147.00	7.0	4	0.0	0	29.63	30.93
19	0.79	637.0	343.0	147.00	7.0	5	0.0	0	28.75	39.44
20	0.76	661.5	416.5	122.50	7.0	2	0.0	0	24.77	29.79
21	0.76	661.5	416.5	122.50	7.0	3	0.0	0	23.93	29.68
22	0.76	661.5	416.5	122.50	7.0	4	0.0	0	24.77	29.79
23	0.76	661.5	416.5	122.50	7.0	5	0.0	0	23.93	29.40
24	0.74	686.0	245.0	220.50	3.5	2	0.0	0	6.07	10.90
25	0.74	686.0	245.0	220.50	3.5	3	0.0	0	6.05	11.19
26	0.74	686.0	245.0	220.50	3.5	4	0.0	0	6.01	10.94
27	0.74	686.0	245.0	220.50	3.5	5	0.0	0	6.04	11.17
28	0.71	710.5	269.5	220.50	3.5	2	0.0	0	6.37	11.27
29	0.71	710.5	269.5	220.50	3.5	3	0.0	0	6.40	11.72
...	...	...	...	...	...	...	...	...	...	...
738	0.79	637.0	343.0	147.00	7.0	4	0.4	5	41.09	47.01
739	0.79	637.0	343.0	147.00	7.0	5	0.4	5	40.79	44.87

	X1	X2	X3	X4	X5	X6	X7	X8	Y1	Y2
740	0.76	661.5	416.5	122.50	7.0	2	0.4	5	38.82	39.37
741	0.76	661.5	416.5	122.50	7.0	3	0.4	5	39.72	39.80
742	0.76	661.5	416.5	122.50	7.0	4	0.4	5	39.31	37.79
743	0.76	661.5	416.5	122.50	7.0	5	0.4	5	39.86	38.18
744	0.74	686.0	245.0	220.50	3.5	2	0.4	5	14.41	16.69
745	0.74	686.0	245.0	220.50	3.5	3	0.4	5	14.19	16.62
746	0.74	686.0	245.0	220.50	3.5	4	0.4	5	14.17	16.94
747	0.74	686.0	245.0	220.50	3.5	5	0.4	5	14.39	16.70
748	0.71	710.5	269.5	220.50	3.5	2	0.4	5	12.43	15.59
749	0.71	710.5	269.5	220.50	3.5	3	0.4	5	12.63	14.58
750	0.71	710.5	269.5	220.50	3.5	4	0.4	5	12.76	15.33
751	0.71	710.5	269.5	220.50	3.5	5	0.4	5	12.42	15.31
752	0.69	735.0	294.0	220.50	3.5	2	0.4	5	14.12	16.63
753	0.69	735.0	294.0	220.50	3.5	3	0.4	5	14.28	15.87
754	0.69	735.0	294.0	220.50	3.5	4	0.4	5	14.37	16.54
755	0.69	735.0	294.0	220.50	3.5	5	0.4	5	14.21	16.74
756	0.66	759.5	318.5	220.50	3.5	2	0.4	5	14.96	17.64
757	0.66	759.5	318.5	220.50	3.5	3	0.4	5	14.92	17.79
758	0.66	759.5	318.5	220.50	3.5	4	0.4	5	14.92	17.55
759	0.66	759.5	318.5	220.50	3.5	5	0.4	5	15.16	18.06
760	0.64	784.0	343.0	220.50	3.5	2	0.4	5	17.69	20.82
761	0.64	784.0	343.0	220.50	3.5	3	0.4	5	18.19	20.21
762	0.64	784.0	343.0	220.50	3.5	4	0.4	5	18.16	20.71
763	0.64	784.0	343.0	220.50	3.5	5	0.4	5	17.88	21.40
764	0.62	808.5	367.5	220.50	3.5	2	0.4	5	16.54	16.88
765	0.62	808.5	367.5	220.50	3.5	3	0.4	5	16.44	17.11
766	0.62	808.5	367.5	220.50	3.5	4	0.4	5	16.48	16.61
767	0.62	808.5	367.5	220.50	3.5	5	0.4	5	16.64	16.03

768 rows × 10 columns

Splitting the dataset into the features set

In [3]:
```
features = dset.iloc[:, 0:8]
```

Apply sklearn.covariance.EllipticEnvelope

In [4]:
```
outlier=EllipticEnvelope(contamination=0.1)
outlier.fit(features)
```

```
C:\Users\uwx161178.CHINA\AppData\Local\Continuum\anaconda3\lib\site-packages\sk
learn\covariance\robust_covariance.py:622: UserWarning: The covariance matrix a
ssociated to your dataset is not full rank
  warnings.warn("The covariance matrix associated to your dataset "
C:\Users\uwx161178.CHINA\AppData\Local\Continuum\anaconda3\lib\site-packages\sk
learn\covariance\robust_covariance.py:165: RuntimeWarning: Warning! det > previ
ous_det (-32.295632627830258 > -32.935337015896074)
  % (det, previous_det), RuntimeWarning)
C:\Users\uwx161178.CHINA\AppData\Local\Continuum\anaconda3\lib\site-packages\sk
learn\covariance\robust_covariance.py:165: RuntimeWarning: Warning! det > previ
ous_det (-31.132234375043936 > -56.843294175691824)
  % (det, previous_det), RuntimeWarning)
C:\Users\uwx161178.CHINA\AppData\Local\Continuum\anaconda3\lib\site-packages\sk
learn\covariance\robust_covariance.py:165: RuntimeWarning: Warning! det > previ
ous_det (-32.692726673441420 > -32.903089904383826)
  % (det, previous_det), RuntimeWarning)
C:\Users\uwx161178.CHINA\AppData\Local\Continuum\anaconda3\lib\site-packages\sk
learn\covariance\robust_covariance.py:165: RuntimeWarning: Warning! det > previ
ous_det (-32.447146904825232 > -33.002603506738986)
  % (det, previous_det), RuntimeWarning)
C:\Users\uwx161178.CHINA\AppData\Local\Continuum\anaconda3\lib\site-packages\sk
learn\covariance\robust_covariance.py:165: RuntimeWarning: Warning! det > previ
ous_det (-32.028055811904842 > -32.669467368408107)
  % (det, previous_det), RuntimeWarning)
C:\Users\uwx161178.CHINA\AppData\Local\Continuum\anaconda3\lib\site-packages\sk
learn\covariance\robust_covariance.py:165: RuntimeWarning: Warning! det > previ
ous_det (-31.239421042492335 > -63.208709683315988)
  % (det, previous_det), RuntimeWarning)
C:\Users\uwx161178.CHINA\AppData\Local\Continuum\anaconda3\lib\site-packages\sk
learn\covariance\robust_covariance.py:165: RuntimeWarning: Warning! det > previ
ous_det (-32.121023735345361 > -32.144038159051540)
  % (det, previous_det), RuntimeWarning)
C:\Users\uwx161178.CHINA\AppData\Local\Continuum\anaconda3\lib\site-packages\sk
learn\covariance\robust_covariance.py:165: RuntimeWarning: Warning! det > previ
ous_det (-31.216489320829933 > -31.927008432034885)
  % (det, previous_det), RuntimeWarning)
C:\Users\uwx161178.CHINA\AppData\Local\Continuum\anaconda3\lib\site-packages\sk
learn\covariance\robust_covariance.py:165: RuntimeWarning: Warning! det > previ
ous_det (-31.434515210006403 > -33.014125872545101)
  % (det, previous_det), RuntimeWarning)
C:\Users\uwx161178.CHINA\AppData\Local\Continuum\anaconda3\lib\site-packages\sk
learn\covariance\robust_covariance.py:165: RuntimeWarning: Warning! det > previ
ous_det (-32.024085232037571 > -32.412114671741847)
  % (det, previous_det), RuntimeWarning)
C:\Users\uwx161178.CHINA\AppData\Local\Continuum\anaconda3\lib\site-packages\sk
learn\covariance\robust_covariance.py:165: RuntimeWarning: Warning! det > previ
ous_det (-31.160376493233429 > -31.602742960199702)
  % (det, previous_det), RuntimeWarning)
C:\Users\uwx161178.CHINA\AppData\Local\Continuum\anaconda3\lib\site-packages\sk
learn\covariance\robust_covariance.py:165: RuntimeWarning: Warning! det > previ
ous_det (-30.431539074525386 > -56.092816378061841)
  % (det, previous_det), RuntimeWarning)
C:\Users\uwx161178.CHINA\AppData\Local\Continuum\anaconda3\lib\site-packages\sk
learn\covariance\robust_covariance.py:165: RuntimeWarning: Warning! det > previ
```

```
ous_det (-31.544405354720446 > -32.361765894000719)
  % (det, previous_det), RuntimeWarning)
```

Out[4]:
```
EllipticEnvelope(assume_centered=False, contamination=0.1, random_state=None,
         store_precision=True, support_fraction=None)
```

In [5]:
```
dset['forecasting']=outlier.predict(features)
```

Filter Forecasting =-1

In [6]:
```
dset[dset['forecasting']==-1]
```

Out[6]:

	X1	X2	X3	X4	X5	X6	X7	X8	Y1	Y2	forecasting
0	0.98	514.5	294.0	110.25	7.0	2	0.00	0	15.55	21.33	-1
1	0.98	514.5	294.0	110.25	7.0	3	0.00	0	15.55	21.33	-1
2	0.98	514.5	294.0	110.25	7.0	4	0.00	0	15.55	21.33	-1
3	0.98	514.5	294.0	110.25	7.0	5	0.00	0	15.55	21.33	-1
8	0.86	588.0	294.0	147.00	7.0	2	0.00	0	19.50	27.30	-1
9	0.86	588.0	294.0	147.00	7.0	3	0.00	0	19.95	21.97	-1
10	0.86	588.0	294.0	147.00	7.0	4	0.00	0	19.34	23.49	-1
11	0.86	588.0	294.0	147.00	7.0	5	0.00	0	18.31	27.87	-1
20	0.76	661.5	416.5	122.50	7.0	2	0.00	0	24.77	29.79	-1
21	0.76	661.5	416.5	122.50	7.0	3	0.00	0	23.93	29.68	-1
22	0.76	661.5	416.5	122.50	7.0	4	0.00	0	24.77	29.79	-1
23	0.76	661.5	416.5	122.50	7.0	5	0.00	0	23.93	29.40	-1
68	0.76	661.5	416.5	122.50	7.0	2	0.10	1	32.96	33.87	-1
69	0.76	661.5	416.5	122.50	7.0	3	0.10	1	32.12	34.07	-1
70	0.76	661.5	416.5	122.50	7.0	4	0.10	1	32.94	34.14	-1
71	0.76	661.5	416.5	122.50	7.0	5	0.10	1	32.21	33.67	-1
116	0.76	661.5	416.5	122.50	7.0	2	0.10	2	33.16	33.91	-1
117	0.76	661.5	416.5	122.50	7.0	3	0.10	2	32.40	34.07	-1
118	0.76	661.5	416.5	122.50	7.0	4	0.10	2	33.12	34.17	-1
119	0.76	661.5	416.5	122.50	7.0	5	0.10	2	32.41	33.78	-1
164	0.76	661.5	416.5	122.50	7.0	2	0.10	3	32.52	33.85	-1
165	0.76	661.5	416.5	122.50	7.0	3	0.10	3	33.28	34.11	-1
166	0.76	661.5	416.5	122.50	7.0	4	0.10	3	32.33	34.48	-1
167	0.76	661.5	416.5	122.50	7.0	5	0.10	3	33.24	34.50	-1
212	0.76	661.5	416.5	122.50	7.0	2	0.10	4	33.08	34.11	-1
213	0.76	661.5	416.5	122.50	7.0	3	0.10	4	32.38	33.62	-1
214	0.76	661.5	416.5	122.50	7.0	4	0.10	4	33.09	33.89	-1
215	0.76	661.5	416.5	122.50	7.0	5	0.10	4	32.31	34.05	-1
248	0.86	588.0	294.0	147.00	7.0	2	0.10	5	27.03	25.82	-1
251	0.86	588.0	294.0	147.00	7.0	5	0.10	5	26.45	31.28	-1
...	...	...	...	...	...	...	...	...	...	...	...
453	0.76	661.5	416.5	122.50	7.0	3	0.25	4	36.45	36.76	-1
454	0.76	661.5	416.5	122.50	7.0	4	0.25	4	36.81	37.05	-1
455	0.76	661.5	416.5	122.50	7.0	5	0.25	4	36.26	37.51	-1

	X1	X2	X3	X4	X5	X6	X7	X8	Y1	Y2	forecasting
488	0.86	588.0	294.0	147.00	7.0	2	0.25	5	29.71	28.02	-1
500	0.76	661.5	416.5	122.50	7.0	2	0.25	5	35.69	36.93	-1
501	0.76	661.5	416.5	122.50	7.0	3	0.25	5	36.64	37.01	-1
502	0.76	661.5	416.5	122.50	7.0	4	0.25	5	36.06	35.73	-1
503	0.76	661.5	416.5	122.50	7.0	5	0.25	5	36.70	36.15	-1
536	0.86	588.0	294.0	147.00	7.0	2	0.40	1	31.89	35.99	-1
548	0.76	661.5	416.5	122.50	7.0	2	0.40	1	40.78	39.55	-1
549	0.76	661.5	416.5	122.50	7.0	3	0.40	1	39.97	40.85	-1
550	0.76	661.5	416.5	122.50	7.0	4	0.40	1	40.71	40.63	-1
551	0.76	661.5	416.5	122.50	7.0	5	0.40	1	40.43	39.48	-1
596	0.76	661.5	416.5	122.50	7.0	2	0.40	2	40.78	39.48	-1
597	0.76	661.5	416.5	122.50	7.0	3	0.40	2	40.15	40.40	-1
598	0.76	661.5	416.5	122.50	7.0	4	0.40	2	40.57	40.47	-1
599	0.76	661.5	416.5	122.50	7.0	5	0.40	2	40.42	39.70	-1
644	0.76	661.5	416.5	122.50	7.0	2	0.40	3	39.32	38.17	-1
645	0.76	661.5	416.5	122.50	7.0	3	0.40	3	39.84	38.48	-1
646	0.76	661.5	416.5	122.50	7.0	4	0.40	3	38.89	39.66	-1
647	0.76	661.5	416.5	122.50	7.0	5	0.40	3	39.68	40.10	-1
692	0.76	661.5	416.5	122.50	7.0	2	0.40	4	40.68	40.36	-1
693	0.76	661.5	416.5	122.50	7.0	3	0.40	4	40.40	39.67	-1
694	0.76	661.5	416.5	122.50	7.0	4	0.40	4	40.60	39.85	-1
695	0.76	661.5	416.5	122.50	7.0	5	0.40	4	40.11	40.77	-1
728	0.86	588.0	294.0	147.00	7.0	2	0.40	5	32.31	29.69	-1
740	0.76	661.5	416.5	122.50	7.0	2	0.40	5	38.82	39.37	-1
741	0.76	661.5	416.5	122.50	7.0	3	0.40	5	39.72	39.80	-1
742	0.76	661.5	416.5	122.50	7.0	4	0.40	5	39.31	37.79	-1
743	0.76	661.5	416.5	122.50	7.0	5	0.40	5	39.86	38.18	-1

77 rows × 11 columns

Filter Forecasting =1

In [7]:
```
dset[dset['forecasting']!=-1]
```

Out[7]:

	X1	X2	X3	X4	X5	X6	X7	X8	Y1	Y2	forecasting
4	0.90	563.5	318.5	122.5	7.0	2	0.0	0	20.84	28.28	1
5	0.90	563.5	318.5	122.5	7.0	3	0.0	0	21.46	25.38	1
6	0.90	563.5	318.5	122.5	7.0	4	0.0	0	20.71	25.16	1
7	0.90	563.5	318.5	122.5	7.0	5	0.0	0	19.68	29.60	1
12	0.82	612.5	318.5	147.0	7.0	2	0.0	0	17.05	23.77	1
13	0.82	612.5	318.5	147.0	7.0	3	0.0	0	17.41	21.46	1
14	0.82	612.5	318.5	147.0	7.0	4	0.0	0	16.95	21.16	1
15	0.82	612.5	318.5	147.0	7.0	5	0.0	0	15.98	24.93	1
16	0.79	637.0	343.0	147.0	7.0	2	0.0	0	28.52	37.73	1
17	0.79	637.0	343.0	147.0	7.0	3	0.0	0	29.90	31.27	1
18	0.79	637.0	343.0	147.0	7.0	4	0.0	0	29.63	30.93	1
19	0.79	637.0	343.0	147.0	7.0	5	0.0	0	28.75	39.44	1
24	0.74	686.0	245.0	220.5	3.5	2	0.0	0	6.07	10.90	1
25	0.74	686.0	245.0	220.5	3.5	3	0.0	0	6.05	11.19	1
26	0.74	686.0	245.0	220.5	3.5	4	0.0	0	6.01	10.94	1
27	0.74	686.0	245.0	220.5	3.5	5	0.0	0	6.04	11.17	1
28	0.71	710.5	269.5	220.5	3.5	2	0.0	0	6.37	11.27	1
29	0.71	710.5	269.5	220.5	3.5	3	0.0	0	6.40	11.72	1
30	0.71	710.5	269.5	220.5	3.5	4	0.0	0	6.37	11.29	1
31	0.71	710.5	269.5	220.5	3.5	5	0.0	0	6.40	11.67	1
32	0.69	735.0	294.0	220.5	3.5	2	0.0	0	6.85	11.74	1
33	0.69	735.0	294.0	220.5	3.5	3	0.0	0	6.79	12.05	1
34	0.69	735.0	294.0	220.5	3.5	4	0.0	0	6.77	11.73	1
35	0.69	735.0	294.0	220.5	3.5	5	0.0	0	6.81	11.93	1
36	0.66	759.5	318.5	220.5	3.5	2	0.0	0	7.18	12.40	1
37	0.66	759.5	318.5	220.5	3.5	3	0.0	0	7.10	12.23	1
38	0.66	759.5	318.5	220.5	3.5	4	0.0	0	7.10	12.40	1
39	0.66	759.5	318.5	220.5	3.5	5	0.0	0	7.10	12.14	1
40	0.64	784.0	343.0	220.5	3.5	2	0.0	0	10.85	16.78	1
41	0.64	784.0	343.0	220.5	3.5	3	0.0	0	10.54	16.80	1
...	...	...	...	...	...	...	...	...	...	...	...
734	0.82	612.5	318.5	147.0	7.0	4	0.4	5	29.06	33.84	1
735	0.82	612.5	318.5	147.0	7.0	5	0.4	5	29.92	32.54	1
736	0.79	637.0	343.0	147.0	7.0	2	0.4	5	42.11	38.56	1

	X1	X2	X3	X4	X5	X6	X7	X8	Y1	Y2	forecasting
737	0.79	637.0	343.0	147.0	7.0	3	0.4	5	41.96	37.70	1
738	0.79	637.0	343.0	147.0	7.0	4	0.4	5	41.09	47.01	1
739	0.79	637.0	343.0	147.0	7.0	5	0.4	5	40.79	44.87	1
744	0.74	686.0	245.0	220.5	3.5	2	0.4	5	14.41	16.69	1
745	0.74	686.0	245.0	220.5	3.5	3	0.4	5	14.19	16.62	1
746	0.74	686.0	245.0	220.5	3.5	4	0.4	5	14.17	16.94	1
747	0.74	686.0	245.0	220.5	3.5	5	0.4	5	14.39	16.70	1
748	0.71	710.5	269.5	220.5	3.5	2	0.4	5	12.43	15.59	1
749	0.71	710.5	269.5	220.5	3.5	3	0.4	5	12.63	14.58	1
750	0.71	710.5	269.5	220.5	3.5	4	0.4	5	12.76	15.33	1
751	0.71	710.5	269.5	220.5	3.5	5	0.4	5	12.42	15.31	1
752	0.69	735.0	294.0	220.5	3.5	2	0.4	5	14.12	16.63	1
753	0.69	735.0	294.0	220.5	3.5	3	0.4	5	14.28	15.87	1
754	0.69	735.0	294.0	220.5	3.5	4	0.4	5	14.37	16.54	1
755	0.69	735.0	294.0	220.5	3.5	5	0.4	5	14.21	16.74	1
756	0.66	759.5	318.5	220.5	3.5	2	0.4	5	14.96	17.64	1
757	0.66	759.5	318.5	220.5	3.5	3	0.4	5	14.92	17.79	1
758	0.66	759.5	318.5	220.5	3.5	4	0.4	5	14.92	17.55	1
759	0.66	759.5	318.5	220.5	3.5	5	0.4	5	15.16	18.06	1
760	0.64	784.0	343.0	220.5	3.5	2	0.4	5	17.69	20.82	1
761	0.64	784.0	343.0	220.5	3.5	3	0.4	5	18.19	20.21	1
762	0.64	784.0	343.0	220.5	3.5	4	0.4	5	18.16	20.71	1
763	0.64	784.0	343.0	220.5	3.5	5	0.4	5	17.88	21.40	1
764	0.62	808.5	367.5	220.5	3.5	2	0.4	5	16.54	16.88	1
765	0.62	808.5	367.5	220.5	3.5	3	0.4	5	16.44	17.11	1
766	0.62	808.5	367.5	220.5	3.5	4	0.4	5	16.48	16.61	1
767	0.62	808.5	367.5	220.5	3.5	5	0.4	5	16.64	16.03	1

691 rows × 11 columns

Counts

In [8]:
```
print('Count of Forecasting is -ve:')
print(dset[dset['forecasting']==-1].count())
print('Count of Forecasting is +ve:')
print(dset[dset['forecasting']!=-1].count())
```

```
Count of Forecasting is -ve:
X1             77
X2             77
X3             77
X4             77
X5             77
X6             77
X7             77
X8             77
Y1             77
Y2             77
forecasting    77
dtype: int64
Count of Forecasting is +ve:
X1             691
X2             691
X3             691
X4             691
X5             691
X6             691
X7             691
X8             691
Y1             691
Y2             691
forecasting    691
dtype: int64
```