

Õunte suuruste seast anomaalia leidmine

Õunte andmed võeti [veebiaadressilt](http://www.tlu.ee/~jaagup/andmed/muu/ounad/antoonovka_2_imelikega.txt)

(http://www.tlu.ee/~jaagup/andmed/muu/ounad/antoonovka_2_imelikega.txt)

Anomaaliad püüti leida meetoditega

- One-class Vector support machine
- PCA-based anomaly detection

In [1]:

```
from azureml import Workspace
ws = Workspace()
experiment = ws.experiments['66e373b2084d4ffa9395c0e34ce9ccaa.f-id.e773e947bd7d4c68b4da26e992d0122f']
ds = experiment.get_intermediate_dataset(
    node_id='ebd1c86f-2ee7-47f3-b8d9-428740f5e5d8-36086',
    port_name='Results dataset',
    data_type_id='GenericCSV'
)
frame = ds.to_dataframe()
```

In [4]:

```
frame.head()
```

Out[4]:

	august	september	Label	Scored Labels	Scored Probabilities
0	6.0	7.9	1	0	-1.043081e-07
1	4.0	5.7	1	0	-7.145673e-03
2	5.2	6.6	1	0	-5.228825e-03
3	4.1	5.4	1	0	-7.064961e-03
4	5.7	7.9	1	0	-9.832531e-04

In [3]:

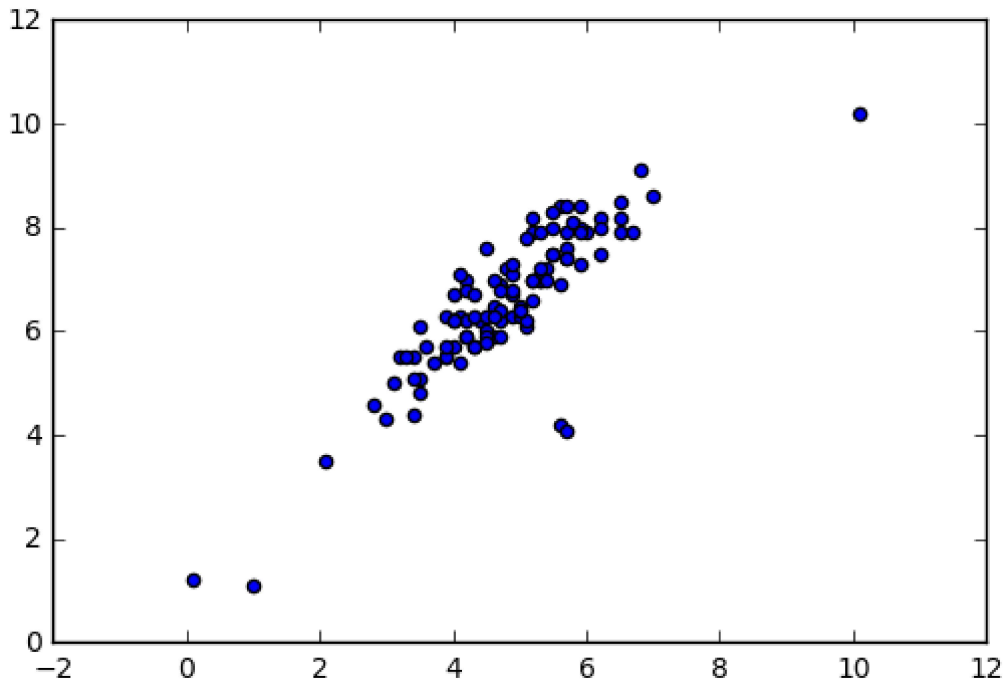
```
frame.tail()
```

Out[3]:

	august	september	Label	Scored Labels	Scored Probabilities
100	5.6	4.2	2	0	-0.002141
101	5.7	4.1	2	0	-0.001486
102	10.1	10.2	2	0	0.032764
103	0.1	1.2	2	0	0.019103
104	1.0	1.1	2	0	0.015089

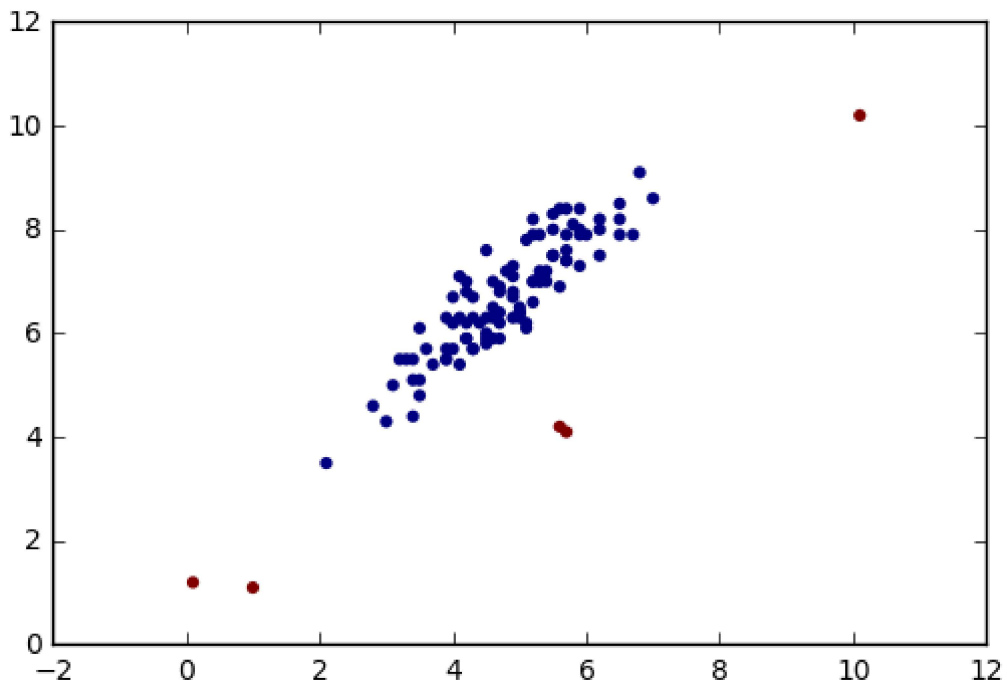
In [6]:

```
import matplotlib.pyplot as plt
plt.scatter(frame.august, frame.september)
plt.show()
```



In [16]:

```
plt.scatter(frame.august, frame.september, c=frame.Label, edgecolors="none")
plt.show()
```



In [36]:

```
frame["Scored Probabilities"]
frame["Scored Probabilities"].mean()
tulp=frame["Scored Probabilities"]
frame.normskoor=(tulp-tulp.min())/(tulp.max()-tulp.min())
frame.normskoor.max()
```

Out[36]:

1.0

In [37]:

```
plt.scatter(frame.august, frame.september, c=frame.Label, s=frame.normskoor*20+0.5, edgecolors="none")
plt.show()
```

