IDS - Assignment 3

Bjarke Kingo Iversen - SWD457

March 2018

Introduction

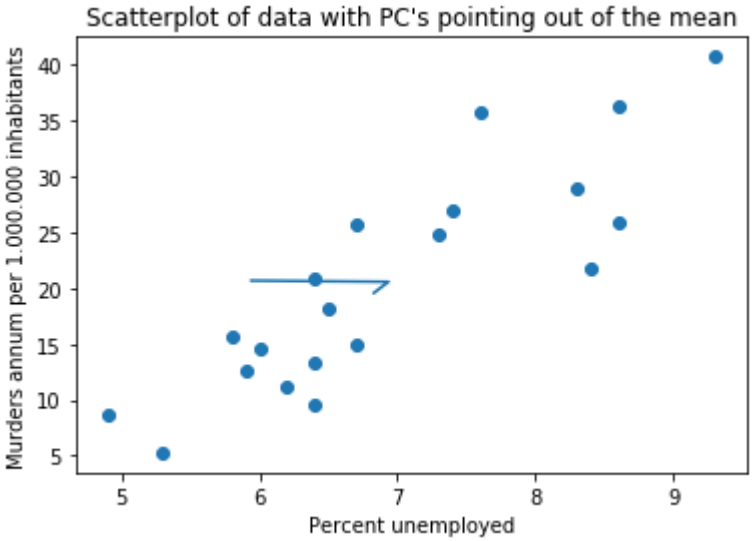
For all my code, please see the Jupyter notebook file ***Assignment 3.ipynb*** in the ***src.zip*** folder.

1 Exercise 1: Performing PCA

1.1 a

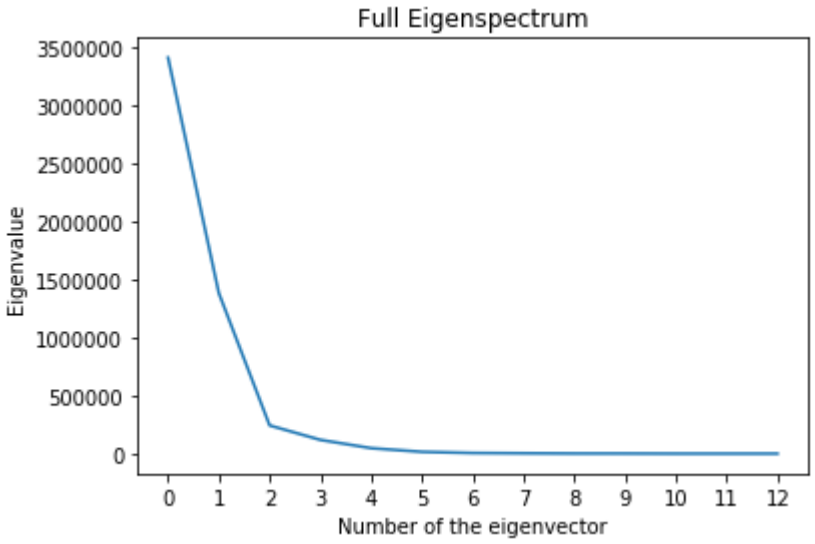
See code.

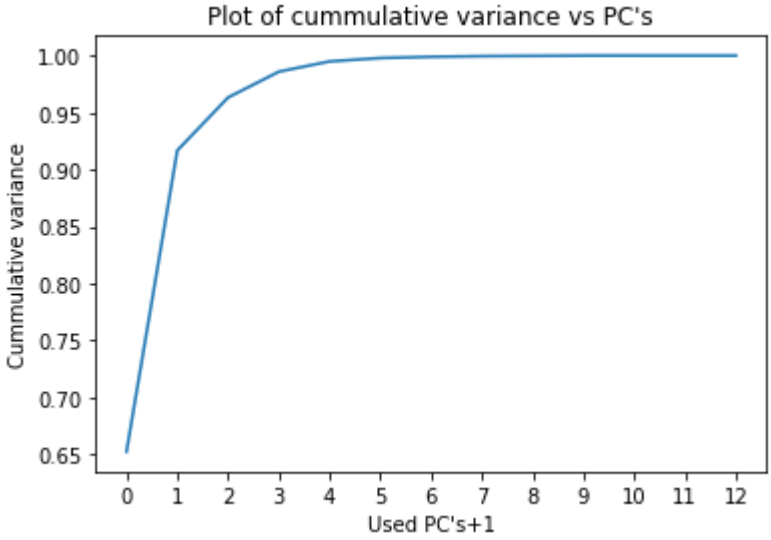
1.2 b



We at least see from the data, that there seem to be a relation between murder rate and unemployment rate.

1.3 c

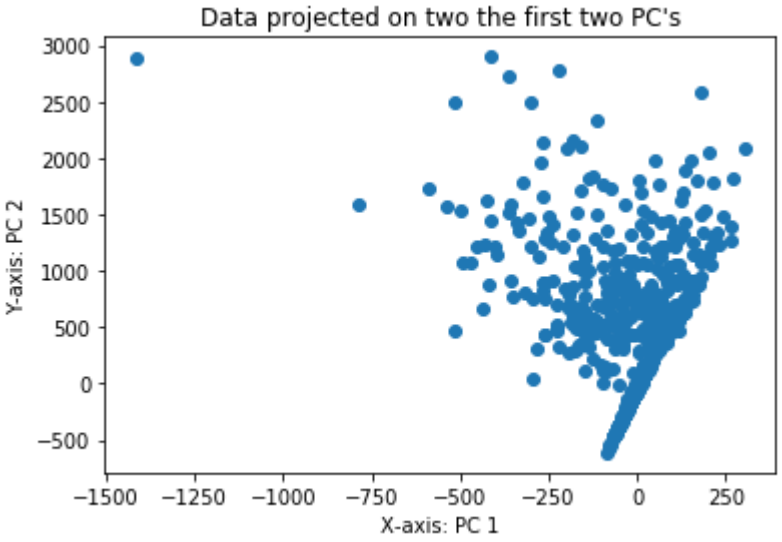




- We need 2 principal components in order to capture 90% of the variance
- We need 3 principal components in order to capture 95% of the variance

See code for a formalized measure.

2 Exercise 2: Visualization in 2D



3 Exercise 3: Clustering

For the clustering i used the ***KMeans***-function from *sklearn.cluster*. I simply fitted the model to the data with parsed parameters to be 2 clusters, giving initial centroids, and thus obtained the clusters-centers to be:

```
My final clusters are:
[[  5.70726496e+00   4.93012821e+01   7.92408120e+02   3.85595940e+03
    3.38821368e+03   1.35652778e+03   2.91737179e+02   1.29989316e+02
    6.86111111e+01   3.81880342e+01   1.87692308e+01   4.13461538e+00
    4.42307692e-01]
 [  2.19924812e+00   1.40018797e+01   1.73727444e+02   1.40094549e+03
    3.18759962e+03   2.62043985e+03   1.00147368e+03   6.31413534e+02
    4.95295113e+02   2.95238722e+02   1.45689850e+02   2.91466165e+01
    2.82330827e+00]]
```