

Istanbul Technical University
Faculty of Computer and Informatics
Computer Engineering Department

BLG 454E
Learning From Data
Homework III

Uğur Uysal - 150140012

March 3rd, 2018

Contents

1	Question I	1
1.1	Part A	1
1.2	Part B	1
1.3	Part C	1
1.4	Part D	1
1.5	Part E	1
2	Question II	2

1 Question I

1.1 Part A

The main motivation of reducing dimension is that finding the simplest model to fit our data and it is easier to find simple models when dimension of data is not too large. Also training cost decrease respect dimension of data.

1.2 Part B

We can calculate that how much variance of data is carried by the PCA in the new form of data. Significantly low dimension and high variance is the point of success. Also PCA will work well when linear components are dominant. To evaluate the performance we can try fit model with different number of components and compare accuracy and simplicity of models.

1.3 Part C

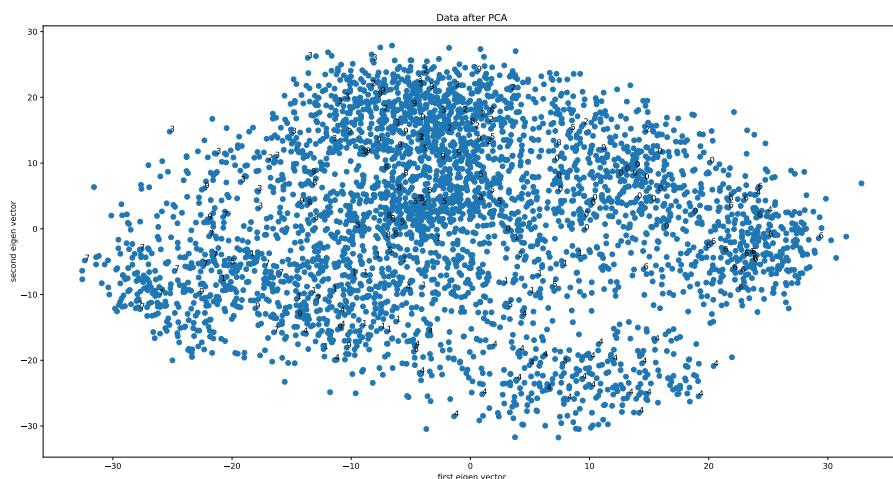
In figure 1, the data is 2-D and they are already clustered as we can observe it. After PCA, we see that is a simple projectile on 1-D and the points are confused and hard to classify. Since PCA does not use class information, for this example it is unsuccessful and unnecessary.

1.4 Part D

PCA finds the linear components but sometimes non-linear components may be useful too.

1.5 Part E

Figure 1: Data Points After PCA



No standardization is used for the PCA. In practice, removing mean from data points did not change the plot.

2 Question II

Figure 2: 1 Rank Too Bad



Figure 3: 5 Rank Too Bad

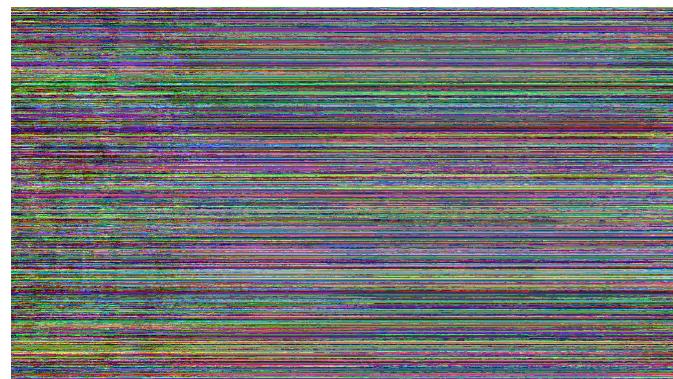


Figure 4: 20 Rank

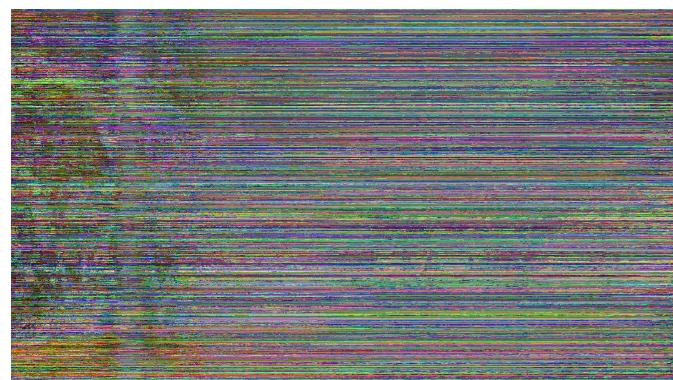


Figure 5: 50 Rank

