Homework 1

Daria Laslo

March 15, 2022

Exercise 1. Starting from the objective of PCS as shown the lectures (eq. 7) is equivalent to the formulation found in the homework description (eq. 1 there) by calculating the squared parathesis as follows:

$$\begin{aligned} ||X - UU^T||^2 &= ||X||^2 - 2X^TUU^TX + X^TUU^TUU^TX \\ &= ||X||^2 - X^TUU^TX \end{aligned} \tag{1}$$

Using the property that given two vectors $a, b \in \mathbb{R}$ we have $a^Tb = trace(ab^T)$, we can rewrite the second term (X^TUU^TX). If

$$a = (X^TU)^T = U^TX \tag{2}$$

$$b = U^TX \tag{3}$$

Then:

$$\begin{aligned} X^TUU^TX &= a^Tb \\ &= trace(ab^T) \\ &= trace((U^TX)(U^TX)^T) \\ &= trace(U^TXX^TU) \end{aligned} \tag{4}$$

Going back to our objective, we can rewrite the minimization problem only in terms of the second term as denoted in the equation 1 of this document, as the minimization does not depend on the first term. Furthermore, the trace is nothing else then the sum of the elements of the diagonal and therefore we can express in terms of the elements on the diagonals. By reversing the sign we change the objective to a maximisation problem, this giving us the final objective as defined in equation 2 (in the homework desccription).

$$\begin{aligned} \arg\min -trace(U^TXX^TU) &= \arg\min -trace(\sum_{i=1}^{n} U^Tx_ix_i^TU) \\ &= \arg\max trace(\sum_{i=1}^{n} U^Tx_ix_i^TU) \end{aligned} \tag{5}$$

Exercise 2. **Part 1.** In Figure 1 we can observe a clustering of the samples containing the different labels somewhat separated in the two dimensions defined by the two principal components. Another observation is the higher variance along the first principal component than for the second one which is expected considering how the pricipal components are defined.

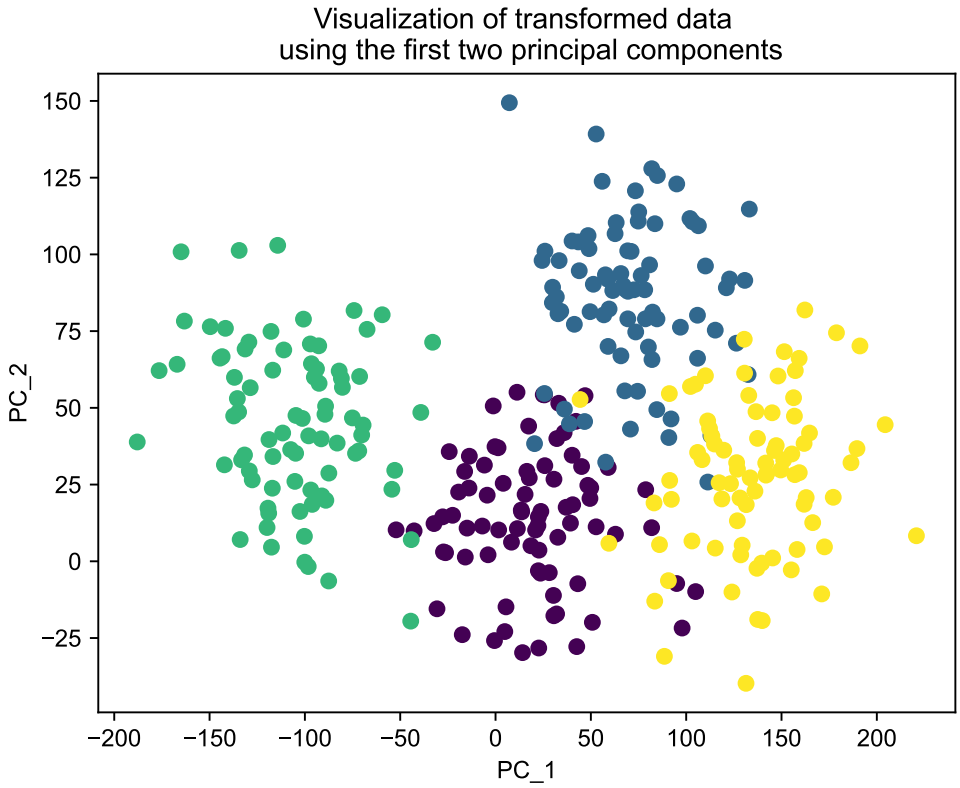

Figure 1: Scatter plot for the first two principal components on the original data.

In Table 1 we can see the variance explained by the first ten principal components. To explain at least 75% of the variance 2 PC are needed. For 85%, 3 are needed, whereas 5 are required for 95%. This can also be observed in Figure 2 which depicts the cumulative sum for increasing number of principal components.

Principal component	Original data	Normalised
1	0.69	0.4
2	0.1	0.29
3	0.08	0.11
4	0.05	0.02
5	0.04	0.01
6	0.02	0.01
7	0.01	0.01
8	0	0.01
9	0	0.01
10	0	0.01

Table 1: Variance explained by the first ten principal components for the original and normalised data.

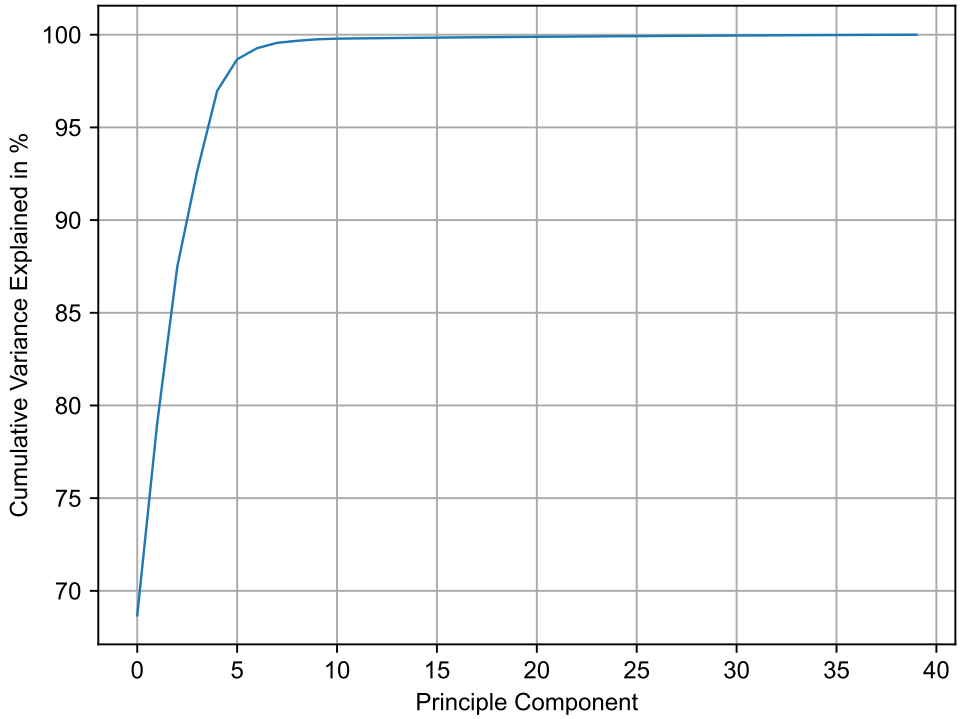

Figure 2: Cumulative plot of the variance explained by all principal components in the original data.

Part 2. When looking at the scatter plot of the data based on the first two components (Figure 3) the separation is much clearer and the clusters can be easily identified.

As we can see in Figure 4 and Table 1, 12 principal components are required to explain 90% of the variance. However, 2 principal components were enough to clearly separate the

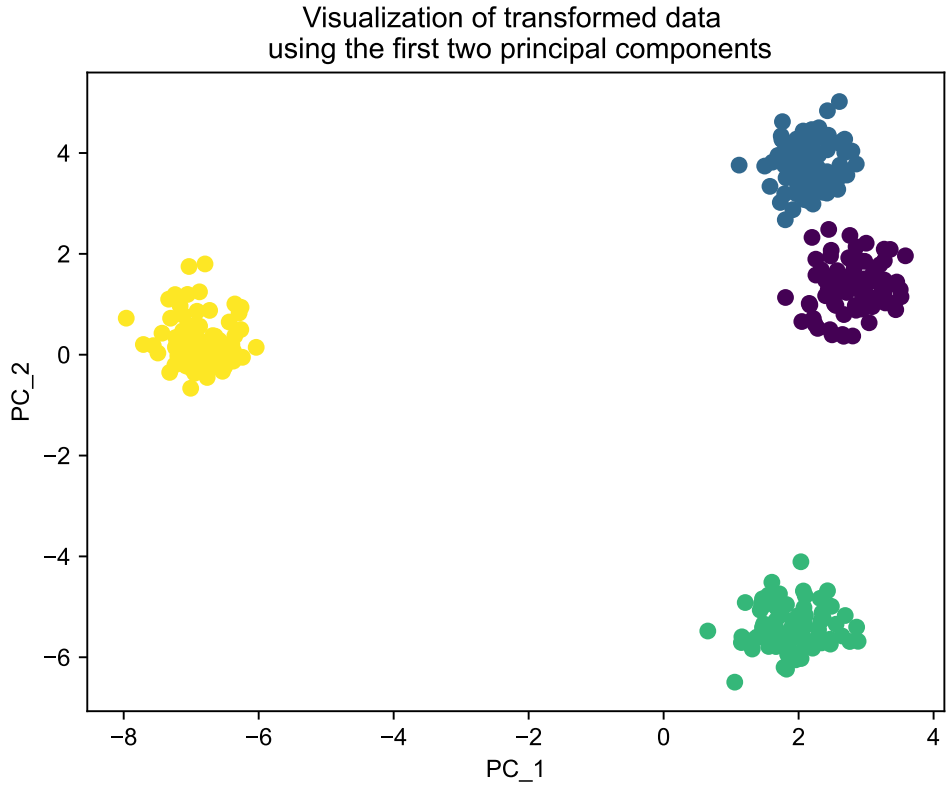

Figure 3: Scatter plot for the first two principal components on the normalised data.

clusters based on the normalised data (as shown in Figure 3). We can conclude from this that performing PCA on the normalised data makes this procedure more efficient and more informative. Therefore, the normalisation can be considered a necessary step in applying PCA.

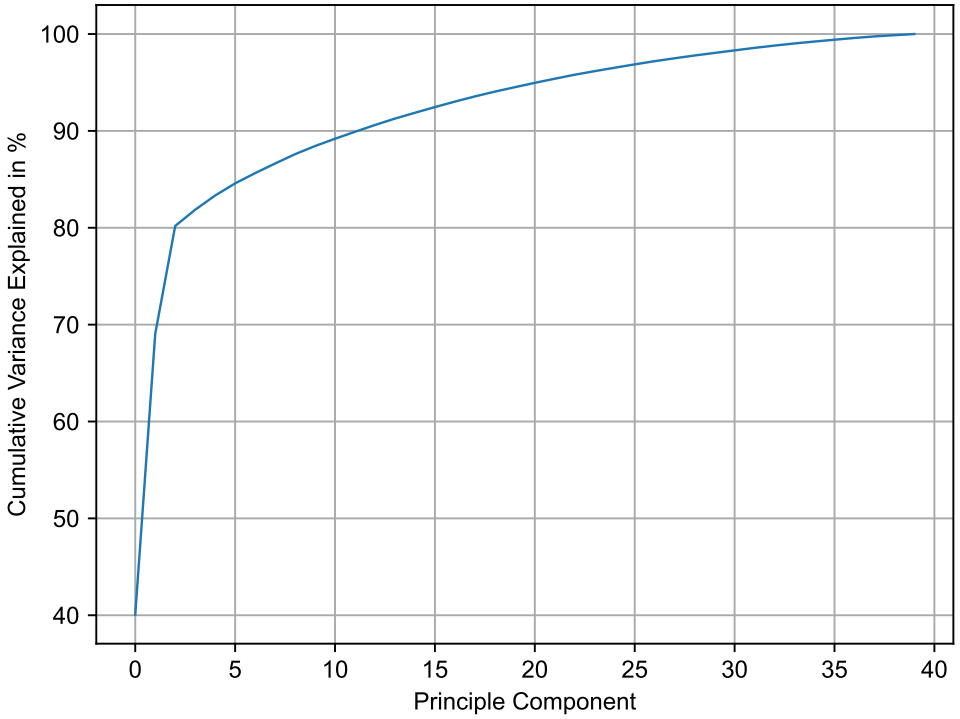

Figure 4: Cumulative plot of the variance explained by all principal components in the normalised data.