Multivariate Statistics: Exercise 6

November 14, 2018

Principal component analysis:

1. Use the data set *data(yarn)* from the `library(pls)`, and from this list only the element `yarn$NIR`.

   (a) Visualize the data with `matplot()` such that each observation forms a line in the plot.

   (b) Compute the first 5 principal components of the centered (not scaled) data. For this task, use the functions

   - `princomp()`
   - `prcomp()`
   - `PCAgrid()` from the `library(pcaPP)`, with the scale measure (`method="sd"`)

   .

   (c) Visualize the resulting directions (`$loadings`) from (b) in terms of lines, in analogy to (a). Interpret these results.

2. We consider the data set *xray.pnm*, which is available from the TUWEL course of our exercises.

   Install the R package `pixmap`. Read the data into R with

   ```
   x <- read.pnm("xray.pnm")
   ```

   and visualize the object with `plot(x)`. You can see an X-ray false color image with a foreign body in the second finger.

   `str(x)` shows the content of the object. It contains the matrices of the pixels for the color ranges red/green/blue.

   Our aim is to compress the information with principal component analysis. Construct a new object with the same structure as `x`, but where the slots "red", "green", and "blue" are not the original matrices, but reconstructed data using the first k principal components. Thus, apply on each of the 3 matrices the function `prcomp()`, compute loadings and scores, and reconstruct the data with the first k PCs. The resulting new object can be visualized with `plot()`.

   - How do you have to reconstruct correctly in order to obtain the same colors as the original image?
   - How do you solve the problems that occur when you plot the resulting image?

- Which number of components k do you need to select in order to see the necessary details in the image?
- Compute a measure of information loss when reducing the information to k PCs.
- Compute a "compression factor" informing about the achieved image compression with k PCs.

Save your (successful) R code together with short documentations and interpretations of results in a text file (= R script file), named as *Matrikelnummer_6.R* (no word document, no plots). Submit this file to Exercise 6 of our tuwel course (deadline November 13).