

Multivariate Statistics: Exercise 11

December 19, 2018

Canonical correlation analysis (CCA):

Use the data set *fitness.csv*, available at our TUWEL course. The data originate from a fitness program of 20 males, and some physiological (weight, waist, pulse) and training (chins, situps, jumps) variables were measured. We are interested in the correlation between these two sub-groups.

- (a) Write a function for robust CCA based on the MCD estimator. It is best to estimate the joint covariance of the complete data set robustly, and then to decompose this covariance matrix into the appropriate parts to perform CCA. Consider an option to do the decomposition based on the robustly estimated correlation matrix. Return also the robustness weights from the MCD estimator.
- (b) Apply the function (a) to the fitness data. How can you interpret the results? Is centering and scaling of the input data needed? Compare with those from classical CCA.
- (c) Plot the first pair of canonical variables for the classical and the robust case. In the latter, use the robustness weights as color for the observations. What do you conclude?
- (d) Use the function `CCAgrid` from the `library(ccaPP)`, with the option `method="M"`, which is resulting in a robust correlation measure. Compare with the previous results from robust CCA (correlations, linear combinations, plots).
- (e) Study deeply this material:
<https://www.r-bloggers.com/merry-christmas-2016-with-r/>

Save your (successful) R code together with short documentations and interpretations of results in a text file (= R script file), named as *Matrikelnummer.11.R* (no word document, no plots). Submit this file to Exercise 11 of our tuwel course (deadline December 18).