Multivariate Statistics: Exercise 8

December 04, 2014

Canonical correlation analysis:

Consider the data set *diabetes.RData* from the web page of our exercises. They can be loaded in R with `load()`, resulting in the data frame `diabetes` with 76 rows (diabetes patients) and 5 columns (relative weight, fasting plasma, glucose intolerance, insulin response to oral glucose, insulin resistance). It is of medical interest to establish a relation between the first 2 variables (X) and the last 3 variables (Y). We will use canonical correlation anaylsis for this task.

(a) Use the function *cancor()* – see help. Center and scale the data (why?). How strong is the linear relationship? How can you interpret the linear combinations for the X and Y data?

(b) Plot the first two canonical variables φ_1 and η_1 against each other. What can you see in the plot? Plot also the second canonical variables φ_2 and η_2 against each other. How can you interpret the plot?

(c) Use the function `CCAgrid` from the `library(ccaPP)` – see help. Use `method="pearson"` and compare the results with those from above. Use `method="spearman"` and compare with the previous results. What is the methodological difference?

(d) Replace cell $[1, 1]$ in the Y data matrix (value 0.81) by a value of 8.1. Perform again the tasks from (c). What changes?

(e) Perform a permutation test for independence as follows: In each of $R = 100$ runs, permute the rows of the Y data randomly, keep the X data fixed, and compute the first canonical correlation coefficient (test statistic). The p-value is the percentage of replicates of the test statistic being larger than the canonical correlation coefficient from then unpermuted data.

Carry out this permutation test for the Pearson and the Spearman correlation measure, for both the original and the modified data. What do you conclude? Compare the Pearson results with the formal test from our course notes.

Save your (successful) R code together with short documentations and interpretations of results in a text file, named as *Familyname8.R*. Send the file as an email attachment to *mehmet.mert@tuwien.ac.at*, at latest Tuesday (02.12).