Quiz 6 (slides 6 and 7):

Assume that you have categorical variables and that you apply multiple correspondence analysis to conduct dimension reduction and to analyze associations between the categories of different variables.

6 a) In PCA you maximize Euclidean distances. What are the distances maximized in MCA?

6 b) How do you interpret the plots produced in MCA?

6 c) Is MCA a robust method?

6 d) It is often advised that in MCA one should use categorical variables that have at least roughly about the same number of categories. Why?

6 e) If you have continuous variables, can you apply MCA?

Quiz 7 (slides 8):

We here consider canonical correlation analysis (CCA).

7 a) Describe the theoretical maximization problem in canonical correlation analysis?

7 b) How are the theoretical canonical vectors and canonical correlations calculated?

7 c) How is sample CCA conducted?

7 d) Assume that you have two groups of variables. There are 2 variables in the first group and 4 variables in the second group. You perform CCA. How many canonical correlations you obtain?

Answers:

6 a) MCA is based on maximizing chi-square distances. (Chi-square distances can be seen as weighted Euclidean distances.)

6 b) You look at the angles. If the angle is smaller than 90 degrees, it indicates that the corresponding attraction repulsion index is larger than 1 (positive association). If the angle is larger than 90 degrees, it indicates that the corresponding attraction repulsion index is smaller than 1 (negative association). Note that one has to be cautious in interpreting the results as dimension reduction may play a role here. If an arrow is long, it might indicate that the corresponding category is rare (or that the corresponding individual has atypical row).

6 c) No, MCA is not a robust method. Small categories have a big effect on the analysis.

6 d) If one variable has very many categories compared to the other variables, then that creates categories that have low probability masses and that has a large effect on the analysis as MCA is not a robust method.

6 e) Yes, one can apply MCA even if the variables are continuous. However, one then has to categorize the observations first.

7 a) See page 5 of the Lecture slides 8.

7 b) See pages 11-12 of the Lecture slides 8.

7 c) The population covariance matrix (of "z") is replaced by the sample covariance matrix and the correlation coefficients are replaced by sample estimates.

7 d) You obtain 2 canonical correlations.