AS-74.4191 Multivariate Regression Methods
Examination January 4, 2013 / OH, VS

You can answer the following questions in Finnish, English or Swedish.

1. Explain briefly the basic idea of the following methods

   a) Mean centering (2 p)

   b) Subspace identification (SSI) (2 p)

   c) NIPALS (2 p)

2. The covariance matrix of a predictor data set X and the cross-covariance matrix of the predictors X and responses Y are

$$R_{xx} = \begin{pmatrix} 3 & -0.02 & 0.01 & -2 \\ -0.02 & 5 & 1.5 & 0.02 \\ 0.01 & 1.5 & 2 & 0.01 \\ -2 & 0.02 & 0.01 & 2 \end{pmatrix} \text{ and } R_{xy} = \begin{pmatrix} 5.02 & 0.02 & 0.10 \\ 0.01 & 0.01 & -3.25 \\ -0.02 & -0.04 & -2.50 \\ -4.51 & -0.05 & -0.02 \end{pmatrix},$$

   respectively. Explain what happens and what kind of prediction results are to be expected when

   a) Multivariate linear regression (MLR) or

   b) Principal component regression (PCR)

   is applied to the data set. Numerical calculations are not required. (6 p)

3. Figures i-iii represent the results of singular spectrum analysis of a time series.

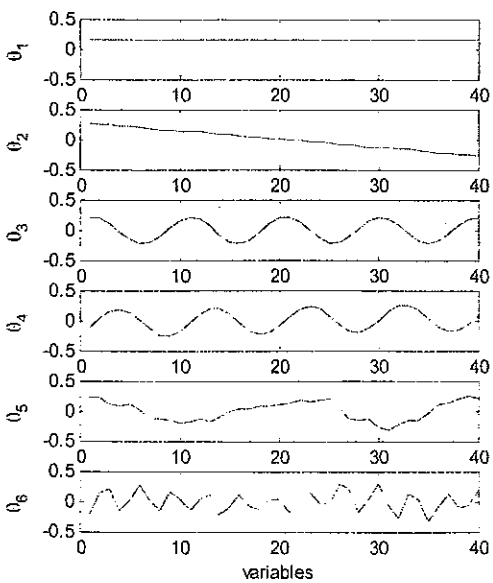

Fig. i: SSA loadings

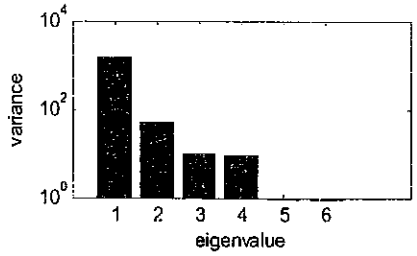

Fig. ii: SSA component variances

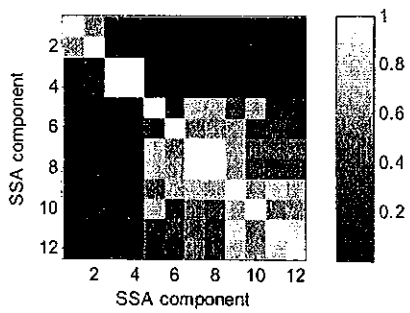

Fig. iii: w-correlations

   a) Explain the meaning of each figure (3 p)

   b) Describe the appearance of the time series based on the figures. (3 p)

TURN

4. a) Why is data preprocessing important and what kind of methods are typically used? (3 p)
b) What kind of preprocessing methods can be used for data collected in spectroscopy? (3 p)

5. What is the difference between
a) principal component analysis (PCA) and factor analysis (3 p)
b) principal component analysis and independent component analysis (ICA) (3 p)