Helsinki University of Technology
Laboratory of Computer and Information Sciences

30.11.2000

Tik-61.184 Special Course in Information Technology IV
Autumn 2000: Unsupervised Learning

INSTRUCTIONS FOR THE PROJECT WORK

1 General Information

The purpose of this individual project work is to give the student hands-on experience in implementing and using a couple of the algorithms that were discussed in the course. Although we give here a very specific problem, students who have their own idea for the project work are encouraged to contact the lecturer or the assistant.

The project work is to be returned by **January 31st, 2001**. The deadline for handing in the solutions to the weekly exercises is the same.

In the case of problems or questions, please contact the assistant: `patrik.hoyer@hut.fi`

2 Tools

You are free to use any numerical software you want. We strongly recommend using Matlab, which is well suited to this kind of work.

3 Task

On the course homepage (`http://www.cis.hut.fi/Opinnot/Tik-61.184/`) we have given four artificially generated datasets. Each dataset is given as a matrix, where observations (data vectors) are stored in the columns of the matrix. Each dataset was generated using either the Independent Component Analysis (ICA) model, or the Topographic ICA (TICA) model.

Your task is to, for each dataset, estimate the ICA and TICA models. You may use the FastICA Matlab package (if you want) for performing the ICA estimation. The TICA algorithm (as described in Ref. 2) you should implement yourself, including setting appropriate parameter values. For each data set and both algorithms, give the estimated mixing matrix. Then, for each data set, identify the model that was used in generating the data (ICA or TICA?). The second step is not described in the papers, you must yourself figure out how to do it.

4 Reporting the Results

A written report is to be made, describing the exact algorithms used (include all parameter settings) as well as how you deduced which data was generated with which model. Write this report in plaintext (ASCII) format. The complete (sufficiently commented) source code should be provided as well. Finally, include your estimates of the mixing matrices (in a Matlab '.mat' file). Make a tar-archive of everything, and send it by email to the assistant: `patrik.hoyer@hut.fi`

5 Recommended Reading

1. Hyvärinen, A and Oja, E. Independent Component Analysis: Algorithms and Applications. *Neural Networks* 13, 411-430 (2000).

2. Hyvärinen, A, Hoyer, P.O., and Inki, M. The Independence Assumption: Analyzing the Independence of the Components by Topography. In Advances in Independent Component Analysis, Girolami, M., editor. Springer-Verlag (2000).

Note: These were part of the course material.