Dina

Dina Research School:
Workshop December 6-7, 2001

Tune Landboskole, Greve.

Edited by Anders Ringgaard Kristensen

Dina Notat No. 95 ● December 2001

This note is also available on www at URL:
http://www.dina.dk/phd/w/w9/notat95.pdf

Dina Research School
Department of Animal Science and Animal Health
Royal Veterinary and Agricultural University
Grønnegårdsvej 3, DK-1870 Frederiksberg C

Contents

Program

Thursday, December 6

11.00:
Arrival and registration

12.00:
Lunch

13.00:
Welcome to the workshop, presentation of the participants
Anders Ringgaard Kristensen, The Dina Research School

13.15:
What is visualization, general introduction:
Per Grove Thomsen, The Dina Research School

14.00:
Short break

14.15:
Visualization tools in Matlab
Per Grove Thomsen, The Dina Research School

15.00:
Coffee break

15.30:
Theory session I: Multivariate data analysis using Principal Component Analysis
Frans van den Berg, Department of dairy and Food Science, KVL

16.10:
Short break

16.20:
Theory session II: Multivariate regression modeling using Partial Least Squares
Frans van den Berg, Department of dairy and Food Science, KVL

17.00:
Computer exercises I

18.00:
Dinner

19.00:
Theory session III: Chemometrics from a statistical perspective
Per Bruun Brockhoff, Department of Mathematics and Physics, KVL.

20.00:
Computer exercises II

22.00:
Coffee and sandwich

Friday, December 7

8.00:
Breakfast

8.45:
How to deal with all sorts of data formats
Stefan Mayer, Informatics and Mathematical Modeling, DTU

9.30:
Short break

9.40:
IBM Data Explorer - and similar advanced workbench tools for visualization
Stefan Mayer, Informatics and Mathematical Modeling, DTU

10.25:
Coffee break

10.45:
Computer exercises III

11.45:
Closing and evaluation
Anders Ringgaard Kristensen, The Dina Research School

12.00:
Lunch and departure

Abstracts

What is visualization, general introduction

Per Grove Thomsen

The priciples of modern visualization, examples from applications in science context.

Visualization tools in Matlab

Per Grove Thomsen

Matlab is a modern environment for the scientist, it makes many processes very simple and provides a wealth of toolboxes for applications in many areas of science. One of the strong toolboxes is for visualization of data in a very general context. These tools will be presented and a large variety of examples are given.

Multivariate data analysis using Principal Component Analysis

Frans van den Berg

In the field of chemometrics a popular technique for studying multivariate data tables is Principal Component Analysis. The main objective of PCA is to get a low dimensional - usually graphical - overview on large amounts of information. The bilinear PCA is a model free or soft method, meaning that no (hard) physical or theoretical relations are imposed during model building. We will study both the (essential) mathematics and the practical issues of PCA modeling, using a leading example to illustrate decisions and outcomes. The data is consumer preferences for foods collected in different countries.

If time allows we will briefly look at more 'advanced' methods e.g. multi-block PCA, where several data table with one mode in common are analyzed simultaneously, and multi-way PARAFAC, where data problems of three or more dimensions are analyzed.

Multivariate regression modeling using Partial Least Squares

Frans van den Berg

Another important tool in chemometrics is Partial Least Squares regression for model building between a multivariate descriptor and a uni- or multivariate response block. PLS, like PCA, is a bilinear and model free method. Next to the predictive performance, the (graphical) diagnostics for model interpretation are explained. We will examine both the (essential) mathematics and the practical issues of PLS modeling, using the example to illustrate choices and results.

Computer exercises I

In the computer exercises the participant can get some hands-on experience in PC Analysis using 'The Unscrambler' software. There is also opportunity to further 'discuss by example' the theory behind multivariate modeling.

Chemometrics from a statistical perspective

Per Bruun Brockhoff

Traditionally, chemometrics and classical statistics are seen as rather disjoint scientific fields and historical as well as current contrasts between the fields do exist. However, from a methodological point of view the chemometrical and statistical toolboxes are deeply connected. Some of the relations between PCA and PLS to what may be termed classical statistics will be clarified. If time allows also the multiway (3-way) methods are put in a statistical perspective.

Computer exercises II

In the second computer session the hands-on experience is continued with emphasis on regression model building using PLS.

How to deal with all sorts of data formats

Stefan Mayer

One of the most common hurdles in practical visualisation work is the problem of how to bring the data of interest into the software one wants to use. In particular this problem can get nasty, if the data are binary and perhaps generated on another platform than the one intended for visualisation purposes. In this session an attempt is made, from experience with such problem to give some guidelines, and bring to attention some very simple, but useful tools.

IBM Data Explorer - and similar advanced workbench tools for visualization

Stefan Mayer

In visualization problems including (very) large sets of data often the available simple tools are insufficient. This is partly due to their limited functionality, but may also be due to the often suboptimal computational efficiency. In such cases either visualization software has to be written in the form of ordinary programs using various visualization libraries, or special software packages can be used, allowing graphical programming using visualisation tools. Both approaches will be briefly addressed, and examples of the second approach will be given in more detail from application of the Data Explorer.

Computer exercises III

The exercises on day 2 will use the Matlab problem solving environment. The exercise will illustrate how to use the Matlab graphics tools for working with data fields in 2 and 3 dimensions as well as images. Some experience with the use of Matlab is a clear advantage but not a requirement. A short introduction to Matlab will be provided.

List of participants

Name	Email	Phone
Charlotte Møller Andersen	cma@kvl.dk	35283542, 45252542
Idi Assoumane	idi@kvl.dk	+45 35 28 27 93
Awuah Baffour	baffour.awuah@helsinki.fi	
Karsten Bjerre *Room without shower*	karsten.bjerre@risoe.dk	+45 4677 4115
Kirsten Bjørkestøl	Kirsten.Bjorkestol@hia.no	+47 38 14 15 97
Per Bruun Brockhoff *Only Thursday*	per@dina.kvl.dk	35282361
Ingeborg Callesen	ica@fsl.dk	
Lene Krøl Christensen	lkc@kvl.dk	+45 3528 3568
Jarka Cloupkova	jarka@kvl.dk	+45-35-28 36 43
Søren Døygaard	sd@kvl.dk	35 28 28 01
Tobias Emilsson	tobias.emilsson@lt.slu.se	+46 (0)40 415154
Poul Møller Hansen	poul.moeller.hansen@risoe.dk	+45 46774113
Preben Klarskov Hansen	PrebenK.Hansen@agrsci.dk	58 11 33 97
Martin Hvidberg *Room without shower*	mhv@fsl.dk	45 17 83 27
Michael Karsten Höhle	hoehle@dina.kvl.dk	35 28 30 84
Alaa Nour Eddin Juma	juma@kvl.dk	+45 32 88 67 62
Rasmus Nyholm Jørgensen	Rasmus.Joergensen@risoe.dk	+45 46 77 41 57
Katja Kiviaho	katja.kiviaho@helsinki.fi	+358 400 456 165
Anders Ringgaard Kristensen	ark@dina.kvl.dk	35 28 30 91
Janne Laine	janlaine@operoni.helsinki.fi	+358 40 5179365
Stefan Mayer	stm@imm.dtu.dk	45253072
Søren Saxmose Nielsen	ssn@kvl.dk	
Juhana Nieminen	juhana.nieminen@helsinki.fi	358-9-19150050
Hee-Bok Park	Heebok.Park@bmc.uu.se	+46-18 471 4564

Peter Stoltze Petersen	pepe@kvl.dk	+45 35 28 36 33
Carl-Göran Pettersson	Carl-Goran.Pettersson@evp.slu.se	+46 (0)18 67 12 24
Jussi Rasinmäki	jussi.rasinmaki@helsinki.fi	
Thomas Algot Søllested	tas@kvl.dk	+45 35 28 30 84
Bruce Edward Talbot	bet@fsl.dk	
Per Grove Thomsen	pgt@imm.dtu.dk	45 25 30 73
Helge Tippmann	helge.tippmann@risoe.dk	
Nils Toft	nt@dina.kvl.dk	35 28 30 24
Frans van den Berg	fb@kvl.dk	
Hanne Østergård	hanne.oestergaard@risoe.dk	46 77 41 11