UPPSALA UNIVERSITY
Department of Mathematics
Måns Thulin

Multivariate Methods

Course literature:
Johnson & Wichern: *Applied multivariate statistical analysis*, 6th edition, Pearson

Reference literature:
Gut: *An intermediate course in probability*, Springer (Lecture 2)
Hastie, Tibshirani & Friedman: *The elements of statistical learning*, Springer (Lectures 6 and 8-10)
Sharma: *Applied multivariate techniques*, Wiley (Lectures 6-10)
Tan, Steinbach & Kumar: *Introduction to data mining*, Addison Wesley (Lectures 1 and 8-10)

Teacher:
Måns Thulin. E-mail: `thulin@math.uu.se`. Office: 74115 (4th floor, house 7, the Ångström laboratory). Phone: (018) 471 33 89.

Examination:
Take-home exam late May or early June 2011. Oral presentation of a method for clustering. Four mandatory homeworks. In each homework (and in the exam) there will also be non-mandatory problems which, if solved in a satisfactory manner, will give higher grades if you pass the course.

Course goals (from syllabus):
In order to pass the course (grade 3) the student should...

–have a knowledge of methods of visualizing multivariate data sets
–be familiar with the multivariate normal distribution
–know how to perform statistical tests of the mean value vector of a multivariate normal distribution
–know how to perform statistical tests of two or several populations of a multivariate normal distribution
–know methods and techniques for validation of multivariate normal distribution
–be able to use principal component and factor analysis for typical problems
–be able to use canonical correlation analysis
–be able to use classification techniques
–be familiar with methods for multivariate cluster analysis
–be able to present mathematical statistical arguments to others

Turn page over!

Teaching:
15 sessions, divided into 10 lectures, 2 problem solving sessions and 3 computer exercises.
In the preliminary course outline below, L=lectures, C=computer exercises, P=problem solving sessions and E=examination.

Date		Topics	Homeworks
22/3	L1	Introduction. Some multivariate problems. Visualization. (Ch 1)	
24/3	L2	Sample geometry, multivariate moments and multivariate normal distribution (MND). (Ch 2.6 & 3-4.2)	H 1 out
28/3	C1	Visualization and descriptive statistics. Multivariate normal distribution.	
1/4	L3	Estimation of parameters for MND. Wishart distribution. Tests for normality. Outliers. (Ch 4.3-4.8)	HW1 in
6/4	L4	Inference about the mean vector. Hotelling's T^2. Confidence regions. (Ch 5)	
8/4	L5	MANOVA. Testing for equality of covariance matrices. (Ch 6)	HW2 out
11/4	C2	Inference under MND.	
13/4	P1	Problems from Chapters 4-6.	
2/5	L6	Principal Components Analysis (Ch 8)	
4/5	L7	Factor analysis (Ch 9) and Canonical Correlation Analysis (Ch 10)	HW2 in
6/5	P2	Problems from Chapters 8-10.	HW3 out
10/5	C3	PCA, factor analysis and CCA.	
13/5	L8	Classification and discrimination (Ch 11)	HW3 in, HW4 out
17/5	L9	Some problems from Ch 11. More about classification: algorithmic methods and decision trees (not in the book)	
23/5	L10	Cluster analysis (Ch 12). Presented by students.	Oral presentations, HW4 in
??/?	E	Take-home exam	

Computer exercises:
Three scheduled computer exercises using the free software R, which can be downloaded from `http://www.r-project.org/`. Instructions for the computer exercises will be available for download from the student portal.

Course homepage:
All files and documents necessary for the course can be found on the course page at the student portal: `http://studentportalen.uu.se/`
After each lecture, slides (if used) and a brief summary of the lecture along with reading instructions will be posted on the course page.
Note that you will need to be logged in using an Upunet-S account to access some of the functions of the page. See `http://www.uu.se/node327` for more information.