Statistics 7560: Multivariate Analysis

- Lectures: MWF 1:50 - 2:45 in Denney Hall 0238
- Instructor: Dr. Vincent Q. Vu (vqv@stat.osu.edu)
- Office Hours: M 2:45 - 3:45 in CH 325 or by appointment
- Course Webpage: http://vince.vu/courses/7560

Overview

Multivariate analysis is a graduate level course in statistical methods for analyzing multivariate data. It is concerned with the simultaneous statistical analysis of multiple variables. The course will introduce students to methodology, theoretical foundations, and computational aspects of multivariate data analysis from a modern perspective. Theoretical derivations will be presented together with practical aspects and intuition.

The course will include topics from the following:

- Random Vectors and Matrices
- Multivariate Normal Distribution Theory
- Multivariate Regression
- Linear Dimension Reduction: Principal Components Analysis, Canonical Correlation Analysis
- Linear Discriminant Analysis
- Latent Variable Models: Factor Analysis, Independent Component Analysis
- Covariance Estimation
- Nonlinear Dimension Reduction: Polynomial PCA, Principal Curves and Surfaces, Kernel PCA

Prerequisites

Stat 6802, or permission of the instructor. Preparation in multivariate calculus and linear algebra is absolutely necessary for this course: students are expected to do background reading to ensure that they understand those concepts. The instructor assumes that students are familiar with:

- Convergence in probability and convergence in distribution
- Maximum likelihood, Fisher information
- Loss function, risk of an estimator
- Bias and variance
- Trace, determinants, eigenvalues, and eigenvectors
- Gradients and Hessians

Students are expected to be able to read and write mathematical proofs and R computer programs. The first homework will have a few review problems that indicate the level of preparation this course requires. If you find them too difficult, then you will probably have difficulty with the rest of the course.

Recommended Books

There are no required books for this course, but the follow book is recommended for supplemental reading:

It provides modern and broad coverage of multivariate methods. It is similar in breadth and style to *The Elements of Statistical Learning* by Hastie, Tishbirani, and Friedman.

The appendix of


is an excellent reference for the prerequisite background on matrix calculus and linear algebra. Another good (and free) reference for matrix algebra and calculus is:


The definitive references for the classical theory multivariate analysis are:

- Muirhead, R. J.: *Aspects of Multivariate Statistical Theory*.

They are comprehensive in their coverage of Multivariate Normal and Wishart Distribution theories.

**Grading**

Your final grade will be based on the following components:

- Homework: 25%
- Midterm Exam: 25%
- Project: 40%
- Participation: 10%

**Homework**

There will be six homework assignments. They will be posted on the course webpage, due Fridays at 1:50 PM, and collected in lecture. Your assignments will sometimes involve data analysis or computer simulations. Do not hand in your commands or raw computer output. You should explain briefly but clearly (in well written prose) what you did, what conclusions you reached and how you reached them. You may cite and include specific results (e.g., plots or tables) from your output as needed. You will almost always make and examine more plots and summaries than you hand in; your write-up should focus on the important points and key pieces of evidence underlying your analysis.

**Midterm Exam**

The date is **Friday, March 8**.

**Project**

Students will work alone or in teams of two. The project will consist of:

1. A one page proposal due **Friday, February 8**.
2. A three page progress report due **Friday, April 5**.
3. A two minute spotlight presentation due **Friday, April 19**, and delivered **Monday, April 22**.
4. A final report (maximum 8 pages) due **Friday, April 26**.

A complete description of the project is on the course webpage.
Computing

You may use any programming language you want. However, the instructor cannot help unless you use R or Matlab.

Policy on Collaboration

Collaboration on homework assignments with fellow students is encouraged. However, such collaboration should be clearly acknowledged, by listing the names of the students with whom you had discussions. You may not, however, share written work or code with others. Your homework submission should be written by you alone.