Statistics 7560: Multivariate Analysis

- Lectures: MWF 1:50 - 2:45 in Scott Lab N0050
- 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 concerned with the *simultaneous* statistical analysis of multiple variables. This is an advanced course intended for doctoral students in statistics and related fields. 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 the following topics:

- Random Vectors and Matrices
- Multivariate Gaussian Theory
- Principal Components Analysis (PCA)
- Canonical Correlation Analysis (CCA)
- Linear Discriminant Analysis (LDA)
- Multivariate methods in Hilbert Space
- Kernel-based methods
- Gaussian Mixture Models
- High-Dimensional Statistics

Prerequisites

Stat 6802, or permission of the instructor. Students are expected to be able to read and write mathematical proofs. Preparation in multivariate calculus, linear algebra, and mathematical statistics is absolutely necessary for this course. Familiarity with the statistical computing environment R or languages such as Matlab is expected. Some of the concepts the instructor assumes that students are familiar with include:

- 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

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, but the following book is recommended for supplemental reading and freely available on the OSU network:

- Izenman, A. J. (2008): *Modern Multivariate Statistical Techniques*.

It provides modern and broad coverage of multivariate methods. It is similar in breadth and style to *The Elements of Statistical Learning* by Hastie, Tisbhirani, and Friedman. The instructor will occasionally recommend reading from Izenman (2008) and other references.

A good (and free) reference for matrix algebra and calculus is:

- Petersen, K.B. and Pedersen, M. S. (2012): *The Matrix Cookbook*. Available at `http://matrixcookbook.com`.

The definitive references for the classical theory multivariate analysis are:

- Anderson, T. W.: *An Introduction to Multivariate Statistical Analysis*.
- Muirhead, R. J.: *Aspects of Multivariate Statistical Theory*.

They are comprehensive in their coverage of Multivariate Normal and Wishart Distribution theories. Many details not covered in lecture can be found in the above two books.

Grading

Evaluation will be based on the following components:

- Homework
- Midterm Exam
- Presentation
- Participation

Homework

There will be occasional homework assignments. They will be posted on the course webpage and collected in lecture on the due date.

Midterm Exam

There will be a midterm exam on **Wednesday, February 12**.

Presentation

Students are required to make a 25 minute presentation in the later part of the course. A list of potential topics will be provided by the instructor. The topic selected by the student must be approved in advance by the instructor.

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.