

Math 570 Applied Multivariate Analysis.

Fall 2014

- **Instructor:** [Xingye Qiao](#)
- **Email:** qiao@math.binghamton.edu
- **Phone number:** (607) 777-2593
- **Office:** OW-134
- **Meeting time & location:** TR 10:05 - 11:30 at OW 100E.
- **Office hours:** T 3:00-5:00, F 11:00-12:00

If you need to reach me, please e-mail qiao@math.binghamton.edu.

Please include [Math570] in the subject line of your email, or your email may not be read promptly.

Prerequisite

Math 501 and Math 502, or equivalent. **Graduate students from outside of the mathematical department and senior undergraduate students may take this course with some preparation (please consult the instructor prior to the semester).** One lecture session will be devoted to reviewing linear algebra materials that are useful in this course.

Learning Objectives

1. A review of the theoretical aspect of Multivariate Statistical Analysis, including: multivariate normal distributions, the multivariate Central Limit Theorem, quadratic forms, Wishart distributions, Hotelling's T square, inference about multivariate normal distributions.
2. Modern applied multivariate statistical methods, including: Principal Component Analysis, Canonical Correlation Analysis, Classification (Bayes rule, Linear and Quadratic discriminant analysis, cross-validation, and logistic regression etc.), factor analysis and Independent Component Analysis, clustering and multidimensional scaling.
3. Machine learning approaches, including Classification and Regression Trees, Support Vector Machine and other large margin classifiers, kernel methods, LASSO and sparsity methods, additive models, etc., if time permits.

Recommended Texts

The required texts are **Härdle & Simar 2012** and **Izenman 2013** (see below for details).

▪ Elementary

- Johnson, Richard A & Wichern, Dean W. 2007. Applied multivariate statistical analysis. Upper Saddle River, N.J: Pearson Prentice Hall. [Amazon Link](#)
- Härdle, Wolfgang & Simar, Léopold. 2012. Applied multivariate statistical analysis. Berlin: Springer (also visit here for sample codes). [Amazon Link](#)

▪ Advanced and applied

- Izenman, Alan Julian. 2013. Modern multivariate statistical techniques: Regression, classification, and manifold learning. New York: Springer New York. Amazon Link || Book Home Page (including R, S-plus and MATLAB code and data sets)
 - Hastie, Trevor, Tibshirani, Robert, and Friedman, J. H. 2009. The elements of statistical learning: Data mining, inference, and prediction. New York, NY: Springer New York. Amazon Link
 - James, Witten, Hastie and Tibshirani, 2014. An Introduction to Statistical Learning with Applications in R. Book Home Page. The PDF file of the book can be downloaded for free. There is also a R library for this book.
- **Theoretical**
- Anderson, T. W. 2003. An introduction to multivariate statistical analysis. Hoboken, N.J: Wiley-Interscience. Amazon Link
 - Muirhead, Robb J. 1982. Aspects of multivariate statistical theory. New York: Wiley. Amazon Link
- **Working with R or SAS**
- Everitt, Brian, and Hothorn, Torsten. 2011. An introduction to applied multivariate analysis with R. New York: Springer. Amazon Link
 - Khattree, Ravindra, and Naik, Dayanand N. 1999. Applied multivariate statistics with SAS software. Cary, NC: SAS Institute. Amazon Link
 - Khattree, Ravindra, and Naik, Dayanand N. 2000. Multivariate data reduction and discrimination with SAS software. Cary, NC: SAS Institute. Amazon Link
 - A Little Book of R for Multivariate Analysis

Most of the books listed above have been left on course reserve. You can go to the Newcomb Reading Room to loan the books for up to 1 day a time.

Grading

- Homework (50%): there will be about four to five homework assignments.
- Midterm exam (25% or 15%): a midterm exam focusing on the theoretical part of the course will be administered in the middle of the semester. Students who are not in the Math PhD program will receive a slightly easier set of problems and a smaller weight for the midterm exam than Math PhD students will do.
- Presentation (10%): each student will choose a research topic (either original research or research conducted by other researchers) related to this course and give a 30-minute presentation. The presentation of each student shall be judged by peer students and the instructor.
- Course project (15% or 25%): a final project will be assigned to each student by the end of the semester. Students who are not in the Math PhD program will gain a greater weight for the course project than Math PhD students will do. The guidelines for the final project can be found here.

Software

There is no designated software for this course. You may use the software that makes the most sense for you. Many pharmaceutical companies use SAS for compliance with FDA regulations. Academic intuitions as well as labs often use R and python. Corporations often use MATLAB, Stata, Minitab, S, etc. because of the relatively high reliability despite the cost. However, it is expected that the student immerse herself with use of at least one software.

Used to be expensive, SAS University Edition is now free for download and use.

Course decks

You need to log in this website in order to download these documents. They can also be downloaded from the blackboard (see the content page).

- Multivariate Data Exploration, slides, Reading: HS Ch. 1, Izenman Ch. 4
 1. hidalgostamp.zip
 2. kde_examples.zip
 3. swissbanknotes.zip, swissbank.mat.zip
 4. golub.zip, golub.mat.zip
- Matrix algebra review, notes, Reading: HS Ch. 2, Izenman Sec. 3.2
- Multivariate normal distribution, notes, Reading: HS Ch. 4-5, Izenman Sec. 3.3
- The Wishart distribution, notes, Reading: HS Ch. 4-5, Izenman Sec. 3.4
- Inference for MVN, notes, Reading: HS Ch. 6-7, Izenman Sec. 3.5
- Linear Dimension Reduction: PCA and CCA, slides, Reading: HS Ch. 9-10, 15, Izenman Ch 7
- Classification, slides, Reading: HS Ch. 13, Izenman 8 & ESL Ch. 4.3, 4.4, and 7.10.
- Linear Dimension Reduction: Latent Variable Models, slides, Reading: HS Ch. 11, Izenman 15
- Clustering, slides, Reading: HS Ch. 12, Izenman 12
- Multidimensional Scaling, slides, Reading: HS 16, Izenman 13
- Support Vector Machine, slides, Reading: Izenman 11, ESL 12
- Sparsity, slides, Reading: Izenman 5.6-5.8, ESL 3
- Tree, slides, Reading: Izenman 9, ESL 9.2
- Ensemble methods, slides, Reading: Izenman 14, ESL 7.11, 8.2, 10, 16.

Schedule

- Week 1 (Sept. 2, 4): introduction; matrix
- Week 2 (Sept. 9, 11): MVN; Wishart
- Week 3 (Sept. 16, 18): Wishart; Inference for MVN; PCA
- Week 4 (Sept. 23): PCA; (break)
- Week 5 (Sept. 30, Oct. 2): CCA; Bayes rule, LDA/QDA
- Week 6 (Oct. 7, 9): Logistic Regression; cross validation; other classifiers
- Week 7 (Oct. 14, 16): FA; ICA
- Week 8 (Oct. 21, 23): k-means; Hierarchical clustering
- Week 9 (Oct. 28, 30): Gaussian mixture/EM; manova
- Week 10 (Nov. 4, 6): ; mds; svm;
- Week 11 (Nov. 11, 13): svm; kernel methods; lasso
- Week 12 (Nov. 18, 20): lasso; midterm exam
- Week 13 (Nov. 25, 27): tree; (break)
- Week 14 (Dec. 2, 4): bootstrap, bagging, subsampling; boosting; RF
- Week 15 (Dec. 9, 11): high dimensional learning; multiple testing

▪ Week 16: final week

Presentations

Nov 4	Xiaojie Du (zou06)
Nov 6	Lin Yao (shen07)
Nov 11	Lishun Li (Jung09)
Nov 13	Armin Pillhofer (tibshirani02)
Nov 18	Zach Seymour (Witten10)
Nov 20	Ruiqi Liu (sun12)

From:

<https://www2.math.binghamton.edu/> - **Department of Mathematics and Statistics, Binghamton University**

Permanent link:

<https://www2.math.binghamton.edu/p/people/qiao/teach/570-f2014>Last update: **2016/02/28 22:51**