36-464/36-664: Applied Multivariate Methods
Spring 2014

Instructor: Rebecca Steorts, 132K Baker Hall, beka@cmu.edu
Course TA: Rafael Stern, 8115 Wean Hall, rafaelst@stat.cmu.edu
Course TA: Nicolas Kim, Porter 117, nicolask@andrew.cmu.edu
Course Time: T/Th: 1:30–2:50
Course Location: PH 125C

Steorts Office Hours: Tues: 3–4 pm, Wed: 11:30-12:30
Rafael Stern Office Hours: Monday: 12–1 pm; Thursday: 10–11 am
Nicolas Kim Office Hours: Tuesday: 11 am–12 pm; Wed: 3-4 pm
Course webpage: `http://www.stat.cmu.edu/~rsteorts/`

Applied multivariate methods are an increasingly important tool in applied machine learning and statistics. We will start with reviewing important facts from matrix algebra and learning about the multivariate normal distribution. Then we will delve into unsupervised and supervised learning approached for multivariate data, where the data may or may not be normally distributed. The data is often highly dimensional in the covariates or parameter space, calling for dimension reduction. We will explore a range of approaches starting with factor analysis, principal component analysis, and then moving along to data mining techniques such as classification and clustering methods. Finally, we will explore Bayesian methods for multivariate data and the strengths and weaknesses of both approaches.

Prerequisites: Statistics 401. Students are expected to be very familiar with R and will be expected to have learned LaTex by the end of the course. All reports, scribe notes, exams, etc. should be submitted in Latex pdf format.

Required Texts:

An Introduction to Statistical Learning with Applications in `R` , Gareth James, Daniela Whitten Trevor Hastie, and Robert Tibshirani, (2013), Springer.

Highly Recommend Texts:

Analysis of Multivariate and High-Dimensional Data, Inge Koch (2013), Cambridge.
The Bayesian Essentials with R Second Edition, Jean-Michel Marin and Christan Robert, (2013), Springer.

Note: The Springer texts are free online via Springer Link via the CMU connection. The James, Whitten, et. al (2013) book will be available via the authors webpage in January 2013. The multivariate book is not free online. I highly recommend buying this.

Grading Policy:

Homework	60%
Exams	20%
Final Exam	20%

Topics covered (which are subject to change)

- Review of matrix algebra

- Multivariate data and distributions

- Factor analysis

- Principal Components Analysis

- Classification methods

    - LDA and QDA
    - How SVM replaced LDA and QDA
    - Clustering and classification (regression trees and classification trees)
    - Bagging and Random Forests
    - Boosting

- Introduction to Bayesian methods

- Gibbs sampling

- Bootstrapping and the Bayesian bootstrap

    - Stein estimation
    - Bayesian GLMs
    - Mixture models
    - Image segmentation

Course Policies: Homework assignments will be announced in class (along with the due date). It must be turned in at the beginning of the lecture on the due date. Late homework will not be accepted.

All homework's and take home exams *must* be submitted through the blackboard website and must be neatly typed LaTex and well-explained or points will be deducted. It must be uploaded in .pdf format. All other formats will not be graded. Submissions via email to the TA's or instructor will not be accepted for credit. See below for more information about LaTex.

Scribing is a form of taking notes. Most of you will scribe once during the semester and this will count as a homework grade. Each class will have two scribes. Please prepare one set of notes for scribing that will be uploaded to the course webpage for the course to view. Please use LaTex to

prepare scribe notes, and please use the template file on the course webpage.[1] Two of you will be randomly chosen to scribe on the day of lecture and you will have one week to prepare the notes with your classmate. The combined scribed notes should be emailed to the instructor and TAs by 10 am one week after the course. You are not allowed to switch with other students on the day you are scheduled to scribe. All students that scribe and do an adequate job will have their lowest homework grade dropped. The best scribe will receive an extra award at the end of the semester.

Makeup exams must be approved before the time of the exam and will be given only in case of medical or family emergencies (which must be appropriately documented). All work turned in for a grade must be entirely your own. This particularly relates to homework. You are encouraged to talk to each other regarding homework problems or to the instructor/TA, however the write up and solution *must* be entirely your own solution and work. Furthermore, you are responsible for everything from lecture. Do not depend on the course web page for announcements regarding due dates for homework, changes in schedules, etc. Such announcements will be made in class. Homework assignments will be uploaded to the course webpage along with course readings (please check here frequently for updates).

Please use the Google group for questions and discussions online (the instructor and TAs will answer in a timely fashion). Do not post inappropriate comments online or you will be blocked from the group. Also, please be considerate regarding the amount of emails you send to the instructor and TAs since this is a large class.

Cell phones should be turned off (or set on silent). Laptops are allowed when we are doing applied examples or labs in class, but otherwise should not be out or being used.

Academic Honesty: Carnegie Mellon University requires all members of its community to be honest in all endeavors. Cheating, plagiarism, and other acts diminish the process of learning. When students enroll at CMU they commit themselves to honesty and integrity. Your instructor fully expects you to adhere to the academic honesty guidelines you signed when you were admitted to CMU. For more information on the CMU Honor Code, please go to http://www.cmu.edu/academic-integrity/defining/index.html.

Students with Disabilities: Students who require special accommodations in class or during exams should follow the procedures outlined by the Disability Resources Program http://www.cmu.edu/hr/eos/disability/index.html. Please see the instructor during office hours early in the semester to discuss your accommodation letter confidentially.

Privacy Policies: Student records are confidential. For more information please go to http://www.cmu.edu/policies/.

---

[1]If you are not familiar with Latex (please see http://www.latex-project.org/ for more information and downloading for your OS). This is a great way to write up reports and display mathematical equations and graphical plots.