AMS 205: Introduction to Classical Statistical Learning (Winter 2019)
Course Description

January 16, 2019

Lectures: Tuesday, Thursday 11:40 am-1:15pm (Jack Baskin Engineering, Room 169)
Office hours: Tuesday 10:00-11:20am
Course description and background: This course will offer, at the graduate level, an introduction to statistical learning and data mining, focusing on the frequentist (classical) approach to inference. The course will introduce modeling data with probability distributions and estimating their parameters, using popular techniques such as the maximum likelihood estimation (MLE). For computing MLEs, Newton-Raphson, Fisher Scoring and EM algorithms will be introduced with multiple examples. Stochastic gradient descent algorithms for computing MLEs based on big data will also be covered. The course will further introduce notions of statistical hypothesis testing, including p-values, Wald test, Score test, Likelihood Ratio test and Pearson chi-square test. Multiple testing methods with false discovery rate will also be taught. Additionally, the course will introduce the notion of confidence intervals for parameters and develop different techniques to construct confidence intervals. Finally, the course material will include linear regression techniques with a large number of covariates, including lasso and ridge regression. Emphasis will be placed upon the development of a background for statistical modeling and executing such models in real data examples with standard software, such as R or Matlab. Students will obtain experience by working on problems that illustrate the theory, methods and applications discussed during the lectures. Homework problems will involve theoretical and methodological questions, as well as computational problems for implementing the techniques taught in class to perform simulation studies or analysis of real datasets.
Course grade: The course grade will be based on three homeworks, one midterm and a final project. Homeworks are given bi-weekly and selecetd homework problems will be graded. Some of the homework problems will be on theoretical topics covered in class, while

other homework problems include data modeling and computation of methods discussed in class. The final project will consist of expository review of a specific part of the relevant literature in statistics. The project topic can be on the theoretical side, expanding on the material covered in class. It may alternatively involve statistical modeling and inference methods, including illustration with appropriate data sets/case studies. The project topics will be chosen in collaboration with the instructor. A written project report will be required; moreover, there will be in-class project presentations.

Midterm: February 21

Grade distribution: homework 20%, midterm 30%, final project 50%.

Reading/References: The lectures will be largely based on material taken from two following textbooks, though the instructor may add some other relevant topics.

- Wasserman, L. (2004). *All of Statistics: A Concise Course in Statistical Inference*, Springer, NY.
- Casella, G. and Berger, R. L. (2002). *Statistical Inference (2nd Edition)*, Thompson Learning.