STAT 205
Introduction to Classical Statistical Learning
Winter, 2021

Instructor: Professor Zehang "Richard" Li
Office: Remote
Office Hours: TBD
E-mail: lizehang@ucsc.edu

Course Meeting Times: TuTh 9:50AM-11:25AM
Location: Zoom
Website: Through `Canvas`

Course Description: This course offers a modern introduction to statistical learning, focusing on the frequentist (classical) approach to statistical inference at graduate level:

- Develop an understanding of modern statistical inference through a wide variety of topics based on fundamental principles of learning and making informed decisions from data.
- Cover various practical methods of parameter estimation, hypothesis testing, and interval estimation. Both important concepts and computational methods will be examined in this class.
- Introduce modern developments of statistical learning with practical examples.

Emphasis will be placed upon the development of a conceptual background for statistical inference and modeling, as well as executing and interpreting such models in real data examples. Students will obtain experience by working on problems that illustrate the concepts, methods, and applications discussed during the lectures. Homework problems will involve conceptual, methodological, and computational problems.

References: There is no required textbook. Lecture notes will be distributed roughly bi-weekly. The lectures will mostly follow Larry Wasserman, *All of Statistics: A Concise Course in Statistical Inference*. We will cover Chapter 6–10 and 12–13 of the book. Additional topics will be introduced based on selected chapters of the following two textbooks:

- George Casella and Roger Berger, *Statistical Inference*, 2nd Edition.
- Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani, *An Introduction to Statistical Learning with Applications in R*. You can find a downloadable PDF copy at `http://faculty.marshall.usc.edu/gareth-james/ISL/`

Prerequisites: Single-variable calculus and algebra. Previous coursework in statistics and probability (STAT 203 or equivalent). Previous programming experience in any programming language.

Evaluation and Grading:

- *Participation (5%)*
- *Homework (35%):* There will be 4–5 homework sets. Homework includes both derivation and computation questions. NO LATE HOMEWORKS ACCEPTED.
- *Midterm (25%):* Feb 9, 10AM–11:30AM.
- *Final (35%):* Mar 17, 8AM–10AM.

Both midterm and final are "take-home" exams. You can request a different starting time of the exam within a 24 hour window. You will still have 1.5 hours to finish the midterm and 2 hours to finish the final. You are allowed to look up the textbook and class note, but are not allowed to work together on the exam or share the exam questions with others.

Computing: We will be using the R programming language (www.r-project.org) throughout this course.

Communication: The course webpage (through `Canvas`) will serve as an archive of homework, lecture notes, and other materials. Announcements concerning course logistics will also be placed on the webpage.

Discussion Board: We will be using `Canvas` discussion board through the course website. Please use this discussion board to ask questions about homework or other course topics.

Schedule The most up-to-date schedule by day is maintained on `Canvas`. The following tentative schedule is subject to change.

- *Week 1, 01/04:* **Introduction and Review** Properties of random samples. Types of convergence. Law of large numbers. Central limit theorem.
- *Week 2, 01/11:* **Principles of data reduction.** The sufficiency principle. Exponential family.
- *Week 3, 01/18:* **Point estimation.** Method of moments, maximum likelihood estimation.
- *Week 4, 01/25:* **Properties of MLE.** Score function. Fisher information. Consistency. Asymptotic normality.

- *Week 5, 02/01:* **Computing MLE.** Newton-Raphson. The EM algorithm. Stochastic gradient descent.
- *Week 6, 02/08:* **Hypothesis testing and confidence interval.** Wald test. Score test. Likelihood ratio test. Chi-square test. Permutation test. Multiple testing.
- **Midterm Exam: no class on Feb 9**
- *Week 7, 02/15:* **Statistical decision theory** Risk function. Bayes estimator. Minimax rule.
- *Week 8, 02/22:* **Linear regression.** OLS. Multiple regression. Resampling methods. Bootstrap.
- *Week 9, 03/01:* **Model selection and regularization.** AIC and BIC. Lasso. Ridge regression.
- *Week 10, 03/08:* **Classification**

Exam accommodation: UC Santa Cruz is committed to creating an academic environment that supports its diverse student body. If you are a student with a disability who requires accommodations to achieve equal access in this course, please submit your Accommodation Authorization Letter from the Disability Resource Center (DRC) to me privately during my office hours or by appointment, preferably within the first two weeks of the quarter. At this time, I would also like us to discuss ways we can ensure your full participation in the course. I encourage all students who may benefit from learning more about DRC services to contact DRC by phone at 831-459-2089 or by email at drc@ucsc.edu.

Campus advocacy resources & education: Title IX prohibits gender discrimination, including sexual harassment, domestic and dating violence, sexual assault, and stalking. If you have experienced sexual harassment or sexual violence, you can receive confidential support and advocacy at the Campus Advocacy Resources & Education (CARE) Office by calling (831) 502-2273. In addition, Counseling & Psychological Services (CAPS) can provide confidential, counseling support, (831) 459-2628. You can also report gender discrimination directly to the University's Title IX Office, (831) 459-2462. Reports to law enforcement can be made to UCPD, (831) 459-2231 ext. 1. For emergencies call 911.

Academic misconduct: Cheating, plagiarism and other forms of academic dishonesty will not be tolerated. Any violation will be prosecuted to the fullest extent as set out in the UCSC Code of Judicial Conduct, section 105.15.