CS 7280: Topics in Statistics and Data Analysis

Fall 2016

September 6, 2016

Location: WF 11:45 - 1:25am, Richards Hall 165

Instructor: Olga Vitek, WVH 310, `o.vitek@neu.edu`
   Office hours Wednesdays 1:30-2:30pm, or by appointment.

Tecahing assistant: Sara Taheri, WVH 310 `mohammadtaheri.s@husky.neu.edu`
   Office hours Tuesdays 3:45-4:45pm, or by appointment

Goals of the course: This is a basic course in applied statistics. The course introduces concepts of statistical modeling, inference, and experimental design that are particularly relevant to computer scientists, and which complement other courses such as machine learning. The topics include the concepts of random sampling, point and interval estimation, hypothesis testing and prediction, and optimal allocation of resources for data collection. The course focuses on frequentist approaches to inference. Bayesian inference will be briefly discussed if time allows.

   The course will use the programming language R. In many cases the course will rely on the existing implementations of statistical methods, but some programming effort will also be required. When appropriate, the course connects statistical methods to current events, and emphases challenges and opportunities for inference with large and/or complex datasets.

   At the end of the course the students will be able to (1) recognize the problems of inferential nature and understand the underlying principles, (2) use statistical inference in data analysis, and (3) draw valid conclusions and clearly present the results.

Pre-requisite: The course is designed for graduate students in computer science, but is also open to students from other majors. The course attempts to be as self-contained as possible and can be taken by students without an extensive prior training in statistics. However, the mathematical and computational literacy at the beginner graduate student level is expected. Prior exposure to R is desirable but not required.

Software: The data examples, the case studies, the homeworks and the projects will use the programming language R. Access to R is required. Please install R from `http://lib.stat.cmu.edu/R/CRAN/` prior to the course. Instructions for using statistical methods in R will be provided during the course.

Course web page: `http://www.ccs.neu.edu/course/cs7280f16/CS7280-Fall16.html`
   Daily updates on the schedule, handouts and homework assignments will be posted on the course page.

Attendance: Attendance is optional, but you are responsible for all the material covered in class.

Communication: The course will be using the discussion board Piazza
   `piazza.com/northeastern/spring2016/cs7280/home` You are encouraged to ask and answer questions on the discussion board. All important announcements will be made through Piazza. Once the course begins, course-related email inquiries will be left unanswered.

Textbook: The key textbooks are:
   Agresti and Finley (2008). *Statistical Methods for the Social Sciences*, 4th Ed, Prentice Hall.
   Kutner, Nachtsheim, Neter & Li (2005). *Applied Linear Statistical Models*, 5th Ed, McGraw-Hill.

   Pages from additional texts will be distributed on the course website.

Homework: Expect weekly homeworks during the semester. Extensions to homework deadlines can be obtained if requested at least 48 hours before the deadline, and duly justified. Homeworks turned in after the deadline will not receive credit.

Although some aspects of the homeworks can be discussed with your colleagues and on Piazza, each homework should be done independently. A homework having high level of similarity with that of another (current or past) student is considered plagiarism, and is not accepted.

Exams: One in-class midterm exam, and one final exam.

Project: At the end of the semester groups the students will perform a group project analyzing a real-world problem.

Grades: All grades will be distributed via Blackboard.

Re-grades: All re-grading requests should be made in writing, within one week after receiving the grade. The request should state the specific question that needs to be re-graded, as well as a short (1-2 sentences) explanation of why re-grading is necessary. The new grade can potentially be lower than the original grade.

Breakdown of the final grade: The final grade is based on a total of 400 points broken down into homeworks (100 pts), midterm (100 pts), project (100 pts), final exam (100 pts).

The final letter grades will follow the usual scale:
90-100 = A, 80-89 = B, 70-79 = C, 60-69 = D, 0-59 = F.
Half-points (i.e., '+' and '-'), may also be allocated.

The scale is subject to change at any time, at the discretion of the instructor.