CS 7290: Topics in Statistical Inference

Spring 2018

January 9, 2018

Location: TF 3:25 - 5:05pm, Richards Hall 165

Instructor: Olga Vitek, WVH 310, o.vitek@neu.edu
Office hours Thursdays 9:00am-10:00am, or by appointment.

Tecahing assistant: Mrs. Dan Guo, WVH 310 guo.dan@husky.neu.edu
Office hours TBA

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 covers both Bayesian and frequentist approaches to inference.

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: https://ovitek.github.io/CS7290/S18/index.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/spring2018/cs7290 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 4 roughly biweekly 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 any degree of similarity with that of another student (current or past, at Nortehastern or outside) is considered plagiarism, and will not be accepted. The homework will be assigned a grade of 0. Additional consequences are described at
`http://www.northeastern.edu/osccr/pdfs/Resources/Faculty_Guide_to_Academic_Integrity.pdf`

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

Grades: All grades will be distributed via Blackboard.

Re-grades of homeworks and exams: 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.

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

The project grade consists of project proposal (20%), project report (oral 30% and written, 30%), and project review (20%).

Projects having any degree of similarity with work by any other group, or with any other document (e.g., found online) is considered plagiarism, and will not be accepted. The minimal consequence is that all the group members will receive the project score of 0, and the best possible overall course grade will be C. Additional consequences are described at
`http://www.northeastern.edu/osccr/pdfs/Resources/Faculty_Guide_to_Academic_Integrity.pdf`

Breakdown of the final grade: The final grade is based on a total of 500 points broken down into homeworks (100 pts), midterm 1 (100 pts), midterm 2 (100 pts), project (100 pts), final exam (100 pts).

The final letter grades will follow the usual scale:

90-100 = A-range (i..e, A+, A or A-)
80-89 = B-range (i.e., B+, B or B-)
70-79 = C-range (i.e., C+, C or C-)
60-69 = D
0-59 = F
The cutoffs for '+' and '-' grades will be determined at the end of the semester, at the discretion of the intructor. This scale is subject to change at any time, at the discretion of the instructor.

Changes to final course grade: Changes to the final course grade should be requested in writing, within **one week** after receiving the final course grade. The request should contain a technical explanation of why re-grading is necessary. If the request is justified, the instructor will regrade **all the submissions**, including all the homeworks, the exams and the project, to determine the new grade. The new grade can potentially be lower than the original grade.