DS 5220: Supervised Machine Learning

Spring 2019

January 7, 2019

Location: Tuesdays and Fridays 9:50am - 11:30am, WVH 108

Instructor: Olga Vitek, WVH 310F, o.vitek@neu.edu
Office hours after the class, or by appointment.

Teaching assistants:
Mr. Tyler Nguyen, nguyen.tua@husky.neu.edu. Office hours: Mondays 4-5pm, Location TBA.
Mr. Nalin Gupta, gupta.nal@husky.neu.edu. Office hours: Thursdays 3-4pm, 304 Kariotis Hall.

Goals of the course: Supervised machine learning is the study and design of algorithms, which enable computers/machines to learn from data, given examples of data with a known outcome. This course is an introduction to supervised machine learning. It provides a broad view of models and algorithms for supervised decision making. The course discusses the methodological foundations behind the models and the algorithms, as well as issues of practical implementation and use, and techniques for assessing the performance.

At the end of the course the students will (1) understand and implement the common supervised learning methods, (2) recognize the problems that are amenable to supervised learning, and perform appropriate data analysis, and (3) recognize failure points and threats to validity of the results.

Pre-requisite: The course is designed for graduate students in data science, but is also open to students from other majors. The mathematical literacy (multivariable calculus, probability, linear algebra) and computational literacy (programing languages such as R, Python or MATLAB) at the beginner graduate student level is required.

Software: The data examples will use either R or Python. Homework assignments and term projects can also use R or Python..

Course web page: https://ovitek.github.io/DS5220/S19/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/spring2019/ds5220 You are encouraged to ask and answer questions on the discussion board. All important announcements will be made through Piazza. Please use public messages for all content-related posts, and private messages for personal matters. **Please do not email the instructor.** Course-related inquiries through personal email will be left unanswered.

Textbook: Required textbooks:
An Introduction to Statistical Learning. G. James, D. Witten, T. Hastie, R. Tibshirani, Springer 2013.
Elements of Statistical Learning. T. Hastie, R. Tibshirani and J. Friedman, Springer, 2009.
Pattern Recognition and Machine Learning. C. M. Bishop, Springer 2006.

Optional textbooks:
Machine Learning: A Probabilistic Perspective. Kevin P. Murphy, MIT Press 2012.
Pattern Classification, 2nd Edition. R. O. Duda, P. E. Hart, D. Stork, Wiley and Sons, 2001.
Machine Learning. T. Mitchell, McGraw-Hill, 1997.

Homework: Expect 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 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 Northeastern 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: One in-class midterm exam, 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 of approximately 3 students will work on a term project. The project can address methodological issues related to those discussed in the class, can implement supervised learning methods, or analyse a real-world dataset.

The project grade consists of project proposal (20%), project report (oral and written, 60%), 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 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-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 an 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.