Applied Machine Learning

2019 Fall CS5785 Cornell Tech

Fall 2019 is here! Here's what you need to know.

Course Description

Learn and apply key concepts of modeling, analysis and validation from Machine Learning, Data Mining and Signal Processing to analyze and extract meaning from data. Implement algorithms and perform experiments on images, text, audio and mobile sensor measurements. Gain working knowledge of supervised and unsupervised techniques including classification, regression, clustering, feature selection, association rule mining and dimensionality reduction.

Prerequisites

CS 2800 or equivalent, Linear Algebra, and experience programming with Python or Matlab, or permission of the instructor.

Room & Time

Tuesdays and Thursdays, 12:30PM-1:45PM, Bloomberg Center 131, Cornell Tech

Class number: 12453

Links: CMS (https://cms.csuglab.cornell.edu/web/auth/?action=loginview) for homework submission, Slack (https://cs5785fall2019.slack.com/) for discussions. To join slack, please use this link (https://join.slack.com/t/cs5785fall2019/shared_invite/enQtNzQwNTA4MTc5NDkyLWUyYzQ5NjRiNDM0YzJ

Textbooks (Available for free)

Required:
T. Hastie, R. Tibshirani and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction (2nd edition) (http://web.stanford.edu/~hastie/ElemStatLearn/), Springer-Verlag, 2008.
Recommended:
Yaser S. Abu-Mostafa, Malik Magdon-Ismail, Hsuan-Tien Lin: Learning from Data (http://work.caltech.edu/textbook.html), AMLBook, 2012.
P. Harrington, Machine Learning in Action (http://www.manning.com/pharrington/), Manning, 2012.

A. Rajaraman, J. Leskovec and J. Ullman, Mining of Massive Datasets
(http://infolab.stanford.edu/~ullman/mmds.html), v1.1.
H. Daumé III, A Course in Machine Learning (http://ciml.info/), v0.8.

Course Requirements and Grading

- **Grade Breakdown:** Your grade will be determined by the assignments (30%), one prelim (30%), a final exam (30%), and in-class quizzes (10%).

- **Homework:** There will be four assignments and an "assignment 0" for environment setup. Each assignment will have a due date for completion. Half of the points of the lowest-scoring assignment will count as extra credit, meaning the points received for homeworks 1, 2, 3, and 4 is calculated as (sum of scores) / 3.5.

- **Late Policy:** Each student has a total of **one** slip day that may be used without penalty.

- **External Code:** Unless otherwise specified, you are allowed to use well known libraries such as *scikit-learn, scikit-image, numpy, scipy,* etc. in the assignments. Any reference or copy of public code repositories should be properly cited in your submission (examples include *Github, Wikipedia, Blogs*). In some assignment cases, you are NOT allowed to use any of the libraries above, please refer to individual HW instructions for more details.

- **Collaboration:** You are encouraged (but not required) to work in groups of no more than 2 students on each assignment. Please indicate the name of your collaborator at the top of each assignment and cite any references you used (including articles, books, code, websites, and personal communications). If you're not sure whether to cite a source, err on the side of caution and cite it. You may submit just one writeup for the group. Remember not to plagiarize: all solutions must be written by members of the group.

- **Quizzes:** There will be surprise in-class quizzes to make sure you attend and pay attention to the class.

- **Prelim: October 22** in class. The exam is closed book but you are allowed to bring one sheet of written notes (Letter size, two-sided). You are allowed to use a calculator.

- Past Prelim 0 (prelims/prelim-0.pdf)
- Past Prelim 1 (prelims/prelim-1.pdf)
- Past Prelim 2 (prelims/prelim-2.pdf)

- **Final Exam: December 3-10.** The final exam will be hosted on Kaggle. You will develop an algorithm, prepare a professional paper, submit an anonymized version to the EasyChair conference system, and peer-review the work from other groups. You are strongly encouraged to work in a group of three students.