CS-UY 4563: Introduction to Machine Learning

(Fall 2019, Updated 7/9/2019)

Overview

This course provides a hands-on approach to machine learning and statistical pattern recognition. The course will describe fundamental algorithms for linear regression, classification, model selection, support vector machines, neural networks, dimensionality reduction and clustering. The course includes demos and labs on real and synthetic data using **Python**. Applications will be demonstrated in audio and image processing, robotic control, gene expression data, neural decodings and text processing. No prior machine learning experience is required.

Students will learn to: Formulate problems using a variety of simple ML models. Use software packages to train and validate models. Analyze the performance of these methods using tools from optimization and probability. Pre-process data and visualize results from different sources (time series, audio, image, text, etc.).

Course instructor: Prof. Gustavo Sandoval (gustavo.sandoval@nyu.edu)

TAs: Will be announced soon

Office Hours:

Gustavo Sandoval (Jay Street Building, Office 846): Tue/Thurs 11:00 – 12:00

Lectures: Tue/Thursday 9:30 – 10:50, Room RGSH 707

Texts:

- Hastie, Tibshirani, Friedman, "Elements of Statistical Learning". http://statweb.stanford.edu/~tibs/ElemStatLearn/printings/ESLII_print10.pdf

- Raschka, "Python Machine Learning", 2015. http://file.allitebooks.com/20151017/Python%20Machine%20Learning.pdf

Supplementary texts and resources

- Bishop, "Pattern Recognition and Machine Learning"
- James, Witten, Hastie and Tibshirani, "An Introduction to Statistical Learning", http://www-bcf.usc.edu/~gareth/ISL/ISLR%20Seventh%20Printing.pdf
- Installing python (need to do this before first recitation): http://jupyter-notebook-beginner-guide.readthedocs.io/en/latest/index.html
- Python tutorial: https://docs.python.org/3/tutorial/

Grading:

- Midterm 1: 25%, Midterm 2: 25%, Labs, HW: 25%, Final project: 25% (two people in one team)
- Labs will involve approximately nine python-based exercises developed by Prof. Sundeep Rangan.
- Midterm and final exams are closed book with cheat sheets. Students will need to be able to write simple python in the exams.
- HW and labs due date as posted on GitHub. No late submission accepted except under **extraordinary circumstances** and must be approved in advance by the instructor. HW are to be handed in at the beginning of the class. **Labs are to be hand-in on NYUClasses**. Solutions will be posted on NYUClasses.

Pre-requisites:

- Undergraduate probability and linear algebra. No ML experience is expected for this class
- Students may NOT enroll in this class if they have taken any one of: CSE-GY 6923 (Grad Intro ML), EE-UY 4563 (UG Intro ML), EL-GY 9133 (Grad Advanced ML)
- Students with prior ML experience are encouraged to take graduate-level Probability (EL-GY 6303) in the Fall and advanced ML in the Spring.
- Programming experience is essential, including some exposure or willingness to learn object-oriented programming. No experience in python is required as python will be taught as part of the class.

Tentative Course Schedule

Lecture notes, demos and labs and HWs organized in course units are available at https://github.com/gussand/introml/blob/master/sequence.md

These materials are developed by Prof. Sundeep Rangan, and I will be updating during the semester.

Week 1 (9/2):

Introduction to machine learning: Examples, types of ML problems. Course logistics. Intro to python and jupyter and GitHub. Single variable linear regression **(Unit 1)** (Brief introduction).

HW: Students should download python, jupyter, github, and do the lab and HW in Unit 1.

Week 2 (9/9):

Linear regression **(Unit2 and 3)**: Linear models, least squares formula; Extensions for non-linear models;

Week 3 (9/16):

Model selection and regularization **(Unit 4)**

Week 4 (9/23):

Understanding underfitting and overfitting with polynomial fitting; Irreducible error due to measurement noise; Bias and variance tradeoff; Cross validation; Lasso and regularization **(Unit 5)**

Week 5 (9/30):

Logistic regression and classification **(Unit 6);**

Week 6 (10/7):

Midterm 1

Week 7 (10/14):

Numerical optimization **(Unit 7):** Unconstrained optimization, gradient descent, global vs. local minima, convexity. Example with logistic regression. Implementation with Python

Week 8 (10/21):

Support vector machines **(Unit 8):** Image classification; SVM formulation, support vectors; Duality, kernel methods

Week 9 (10/28):

Neural networks (Unit 9): Formulation, motivation; Computation graphs, backpropagation; Introduction to tensorflow and keras; Stochastic gradient descent.

Weeks 10-11 (11/4, 11/11):

Convolutional and deep networks **(Unit 10):** Convolutional layers, Pooling, batch normalization; Advanced Tensorflow features; Using GPUs

11/4 Last day to withdraw from a class with a "W" grade

Week 12 (11/18):

Midterm 2

Week 13 (11/25):

Dimensionality reduction **(Unit 11):** Principal component analysis (unsupervised), linear discriminant analysis (supervised), LDA SVDs

Thanksgiving on 11/28 so no class on that day.

Week 14 (12/2):

Unsupervised Clustering **(Unit 12**): K-means, Mixture models, EM methods

Week 15 (12/9):

Tree based methods**(Unit 13**) : Decision tree, Random Forest, Boosting.

Week 16 (12/16): Final exam

Final date is decided schoolwide and assigned to each class, so please don't make travel plans before the end of the week.

Moses Center Statement of Disability

If you are student with a disability who is requesting accommodations, please contact New York University's Moses Center for Students with Disabilities (CSD) at 212-998-4980 or mosescsd@nyu.edu. You must be registered with CSD to receive accommodations. Information about the Moses Center can be found at www.nyu.edu/csd. The Moses Center is located at 726 Broadway on the 3rd floor.

NYU School of Engineering Policies and Procedures on Academic Misconduct – complete Student Code of Conduct here

A. Introduction: The School of Engineering encourages academic excellence in an environment that promotes honesty, integrity, and fairness, and students at the School of Engineering are expected to exhibit those qualities in their academic work. It is through the process of submitting their own work and receiving honest feedback on that work that students may progress academically. Any act of academic dishonesty is seen as an attack upon the School and will not be tolerated. Furthermore, those who breach the School's rules on academic integrity will be sanctioned under this Policy. Students are responsible for familiarizing themselves with the School's Policy on Academic Misconduct.

B. Definition: Academic dishonesty may include misrepresentation, deception, dishonesty, or any act of falsification committed by a student to influence a grade or other academic evaluation. Academic dishonesty also includes intentionally damaging the academic work of others or assisting other students in acts of dishonesty. Common examples of academically dishonest behavior include, but are not limited to, the following:

1. Cheating: intentionally using or attempting to use unauthorized notes, books, electronic media, or electronic communications in an exam; talking with fellow students or looking at another person's work during an exam; submitting work prepared in advance for an in-class examination; having someone take an exam for you or taking an exam for someone else; violating other rules governing the administration of examinations.
2. Fabrication: including but not limited to, falsifying experimental data and/or citations.
3. Plagiarism: intentionally or knowingly representing the words or ideas of another as one's own in any academic exercise; failure to attribute direct quotations, paraphrases, or borrowed facts or information.
4. Unauthorized collaboration: working together on work meant to be done individually.

5. Duplicating work: presenting for grading the same work for more than one project or in more than one class, unless express and prior permission has been received from the course instructor(s) or research adviser involved.
6. Forgery: altering any academic document, including, but not limited to, academic records, admissions materials, or medical excuses.

NYU School of Engineering Policies and Procedures on Excused Absences – complete policy here

A. Introduction: An absence can be excused if you have missed no more than **10 days of school**. If an illness or special circumstance has caused you to miss more than two weeks of school, please refer to the section labeled Medical Leave of Absence.

B. Students may request special accommodations for an absence to be excused in the following cases:

1. Medical reasons
2. Death in immediate family
3. Personal qualified emergencies (documentation must be provided)
4. Religious Expression or Practice

Deanna Rayment, deanna.rayment@nyu.edu, is the *Coordinator of Student Advocacy, Compliance and Student Affairs* and handles excused absences. She is located in 5 MTC, LC240C and can assist you should it become necessary.

NYU School of Engineering Academic Calendar – complete list here.

The last day of the final exam period is _____. Final exam dates for undergraduate courses will not be determined until later in the semester. Final exams for graduate courses will be held on the last day of class during the week of _____. If you have two final exams at the same time, report the conflict to your professors as soon as possible. Do not make any travel plans until the exam schedule is finalized.

Also, please pay attention to notable dates such as Add/Drop, Withdrawal, etc. For confirmation of dates or further information, please contact Susana: sgarcia@nyu.edu