NYU Tandon School of Engineering
CS 6923 Machine Learning
Course Syllabus
Spring 2020

Prof: Linda Sellie
Office Hours: Wednesday 3:10 - 5:00 and by appointment.
Office: 370 Jay Street, room 848
Contact Information: Please send a message to me on NYU Classes. Do not send email.
Class communication: We will use the course site on PIAZZA. NYU Classes is for the syllabus, posting homework assignments, lecture notes, and grade posting.

Description

This course is an introduction to the field of machine learning, covering fundamental techniques for classification, regression, dimensionality reduction, clustering, and model selection. A broad range of algorithms will be covered, such as linear and logistic regression, neural networks, deep learning, support vector machines, tree-based methods, expectation maximization, and principal components analysis. The course will include hands-on exercises with real data from different application areas. Students will learn to train and validate machine learning models and analyze their performance.

Materials

Required:
Many of the lectures will follow the approach used in the lecture notes from Stanford:
http://cs229.stanford.edu/syllabus.html

The main textbook for this course is Introduction to Machine Learning, by Ethem Alpaydin, Third Edition. Published by MIT Press. (available online through the NYU library). It covers many of the topics in the course clearly, but I often follow the approach found in the Stanford notes.

Additional resources specific to each lecture will be found in the syllabus below. This information will be updated during the semester; please check it regularly.
Background:
Stanford’s machine learning course has a nice review of:

  Last semester we didn’t use sections 3.4, 2.10, 3.12, 3.13, 4.5, 4.6


For mathematical notation see: https://sebastianraschka.com/pdf/books/dlb/appendix_a_math_notation.pdf (unfinished draft)

The coding aspect of the course will be in Python/numpy.

Optional:
For a more intuitive and accessible approach to some of the topics in the course, see:

- Daume: A Course in Machine Learning by Hal Daume (unfinished book draft), http://ciml.info/

- WHT: An Introduction to Statistical Learning with Applications in R by James, Witten, Hastie, and Tibshirani, http://www-bcf.usc.edu/~gareth/ISL/


The following books are more comprehensive than the Alpaydin text and assume that the reader has a stronger math/statistics background:

- HTF: Hastie, Tibshirani, Friedman, Elements of Statistical Learning, Second Edition Published by Springer. (available online through the NYU library)

- Bishop: Pattern Recognition and Machine Learning, by Christopher Bishop. Published by Springer.

Prerequisites

Graduate status with undergraduate level probability theory.

Please note that Machine Learning is more mathematical than most other graduate CS courses. Students often have difficulty with this course (and risk getting a grade of C or lower) if they know how to program but have not taken much math, or have not done well in their math courses. You should have taken an undergraduate course that covered probability and statistics. You should know about the probability density function (pdf), cumulative density function (cdf), continuous probability distributions, conditional probability, and expected values. You should know about partial derivatives and gradients (or be prepared to learn about them on your own). You should also know the basics of linear algebra.
Grading Policy

The midterm will be 40% of your grade, the final 45%, and the homework/project 15%. No late homework assignments will be accepted.

The tentative schedule for the midterm is March 13, 2020. The final is May 15, 2020.

Attendance at exams is mandatory. Make-up exams will only be given in the case of a emergency, such as illness, which must be documented, e.g. with a doctor’s note. In such cases, you must notify me as early as possible, preferably before the exam is given. If you miss an exam without a valid excuse, you will receive a grade of zero for that exam. The exams will be closed book and no notes.

Course Work: Homework assignments (approximately weekly) will be posted on NYU Classes. Announcements, and the occasional helpful hint will be posted on Piazza. You are responsible for being aware of any information posted there, so you should check it regularly.

Although the homework makes up a relatively small percentage of the final grade and is a lot of work, it is a key component to mastering the course material. Experience has shown that you will not do well on the exams if you have not done the homework.

Tentative Schedule

The schedule is tentative and subject to change. We may not get through all the material in this schedule.

- Topic 01: Introduction, Bayesian Decision Theory
  - Alpaydin: Chapter 3 sections 1-4, chapter 4 sections 1-5 (except 4.3), chapter 5 sections 2, 4 & 5
  - Ng: Section 1 http://cs229.stanford.edu/notes/cs229-notes2.pdf
  - Background
    * mean, variance, covariance, basic probability (e.g. joint probability, conditional probability, sum rule, conditional independence)
    * Multivariate Gaussian (including isocontours) http://cs229.stanford.edu/section/gaussians.pdf
    * Section 4 in https://ocw.mit.edu/courses/mathematics/18-05-introduction-to-probability-and-statistics/readings/MIT18_05S14_Reading10b.pdf
  - Additional resources
    * Probability vs likelihood https://www.youtube.com/watch?v=pYxNSUDSFH4

* Gentle introduction to maximum likelihood:
  - [https://www.youtube.com/watch?v=XepXtl9YKwc](https://www.youtube.com/watch?v=XepXtl9YKwc)
  - [https://www.youtube.com/watch?v=Dn6b9fCIUpM](https://www.youtube.com/watch?v=Dn6b9fCIUpM)

* Discriminant function:
  - [https://www.projectrhea.org/rhea/index.php/Discriminant_Functions_For_The_Normal(Gaussian)_Density](https://www.projectrhea.org/rhea/index.php/Discriminant_Functions_For_The_Normal(Gaussian)_Density)
  - [https://www.projectrhea.org/rhea/index.php/Discriminant_Functions_For_The_Normal(Gaussian)_Density___Part_2](https://www.projectrhea.org/rhea/index.php/Discriminant_Functions_For_The_Normal(Gaussian)_Density___Part_2)

* Sections 2.1 & 2.2 in Mitchell's on-line draft chapter [https://www.cs.cmu.edu/~tom/mlbook/Joint_MLE_MAP.pdf](https://www.cs.cmu.edu/~tom/mlbook/Joint_MLE_MAP.pdf)

- Topic 02: Naive Bayes, k-NN, ML Experiments
  - Ng: Section 2 [http://cs229.stanford.edu/notes/cs229-notes2.pdf](http://cs229.stanford.edu/notes/cs229-notes2.pdf)
  - Sections 1 & 2 in Mitchell's on-line draft chapter on Naive Bayes: [https://www.cs.cmu.edu/~tom/mlbook/NBayesLogReg.pdf](https://www.cs.cmu.edu/~tom/mlbook/NBayesLogReg.pdf)
  - Background:
    * basic probability (e.g. product rule, conditional independence, conditional probability)

- Topic 03: Linear Regression
  - Alpaydin: Chap. 4.6
  - Background:
    * Basic linear algebra (e.g. basic notation, vector vector products, matrix vector products, matrix matrix products, inverse of a matrix, transpose of a matrix, etc). See [http://cs229.stanford.edu/summer2019/cs229-linalg.pdf](http://cs229.stanford.edu/summer2019/cs229-linalg.pdf)
    * In the lecture, we will use the fact that if $f(x) = x^TAx$ where $A$ is a matrix and $x$ is a vector then $\nabla f(x) = x^T(A + A^T)$. You will not be expected to prove this fact in a homework assignment or on an exam. A proof can be found of this fact in proposition 8 in [https://atmos.washington.edu/~dennis/MatrixCalculus.pdf](https://atmos.washington.edu/~dennis/MatrixCalculus.pdf)
  - Additional resources
    * Section 4.1 from [https://www.cs.princeton.edu/courses/archive/spr09/cos513/scribe/lecture09.pdf](https://www.cs.princeton.edu/courses/archive/spr09/cos513/scribe/lecture09.pdf)

- Topic 04: MLE, Bias/Variance, Regularization
  - Sections 1 and 2 in [https://people.cs.umass.edu/~domke/courses/sml2010/10theory.pdf](https://people.cs.umass.edu/~domke/courses/sml2010/10theory.pdf)
  - [https://www.cs.cmu.edu/~avrim/ML14/inequalities.pdf](https://www.cs.cmu.edu/~avrim/ML14/inequalities.pdf)
Alpaydin: Chap. 4 sections 7 \\& 8

Background:
* Union bound: https://en.wikipedia.org/wiki/Boole%27s_inequality

For fun:
* Overfitting blog post: http://gregpark.io/blog/Kaggle-Psychopathy-Postmortem/

• Topic 05: Logistic Regression
  - Ng: Part 2, section 5 http://cs229.stanford.edu/notes/cs229-notes1.pdf
    * https://eight2late.wordpress.com/2017/07/11/a-gentle-introduction-to-logistic-regression-
    * https://www.stat.cmu.edu/~cshalizi/uADA/12/lectures/ch12.pdf
    * Sigmoid function https://computing.dcu.ie/~humphrys/Notes/Neural/sigmoid.html

• Topic 06: Decision Trees and Random Forests
  - Alpaydin: Chap. 9
  - Additional resources
    * Definition of non-parameteteric https://sebastianraschka.com/faq/docs/parametric_vs_nonparametric.html

• Topic 07: Data Preprocessing, Dimensionality Reduction
  - Intuitive introduction to PCA by Tim Roughgarden from Stanford:
  - https://ro-che.info/articles/2017-12-11-pca-explained-variance
  - Alpaydin: Chap. 6
  - Additional resources
  - Background
    * How to project a point onto line (https://mathinsight.org/dot_product)
• Topic 08: Perceptrons and Neural Nets
  - Alpaydin: Chap. 11
  - Background
    * Derivatives of composite functions: https://www.khanacademy.org/math/ap-calculus-ab/ab-differentiation-ab-3-1a/a/chain-rule-review
  - Additional resources

• Topic 09: Deep Learning
  - https://cs231n.github.io/convolutional-networks/

• Topic 10: Kernel Machines, SVMs
  - Alpaydin: Chap. 13
  - background
    * The concept of strong duality and Slater’s condition: https://www.shivani-agarwal.net/Teaching/CIS-520/Spring-2018/Lectures/Reading/optimization.pdf. We will use both strong duality and Slater’s condition in the derivation presented in class.
    * For constrained optimization problems there exists polynomial time solvers. See https://courses.csail.mit.edu/6.867/wiki/images/a/a7/Qp-cvxopt.pdf or https://scaron.info/blog/quadratic-programming.html
  - On your own

• Topic 11: Clustering
  - Alpaydin: Chap. 7
  - Additional resources
  - Background:
    * How to find the parameters for GDA (A refresher can be found on page 6 of http://cs229.stanford.edu/notes/cs229-notes2.pdf)

• Topic 12: Ensemble Methods
- Alpaydin: Chap. 17 (not 17.5)
- Additional resources:
  * http://web.engr.oregonstate.edu/~tgd/publications/mcs-ensembles.pdf
- On your own:
- Topic 13: Reinforcement Learning
  - https://inst.eecs.berkeley.edu/~cs188/fa18/assets/notes/n4.pdf
  - https://inst.eecs.berkeley.edu/~cs188/fa18/assets/notes/n5.pdf
- Alpaydin: Chap. 16

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 mosecsd@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

The complete Student Code of Conduct can be found here: https://www.nyu.edu/registrar/calendars/university\protect\discretionary{\char\hyphenchar\font}{}academic-calendar.html#1198

1. **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.

2. **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:
   
   (a) 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.

(b) Fabrication: including but not limited to, falsifying experimental data and/or citations.

(c) 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.

(d) Unauthorized collaboration: working together on work meant to be done individually.

(e) 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.

(f) 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


1. 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.

2. Students may request special accommodations for an absence to be excused in the following cases:
   • Medical reasons
   • Death in immediate family
   • Personal qualified emergencies (documentation must be provided)
   • 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 https://www.nyu.edu/Registrar/calendars/university-academic-calendar.html#1198 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