New York University Tandon School of Engineering
Computer Science and Engineering

CS-UY 4563 (ECE-UY 4563)
Introduction to Machine Learning, Spring 2020

Professor Christopher Musco
Mondays, Wednesdays 9:00-10:20am, Jacobs Building, JABS 775B.

PROFESSOR CONTACT

Email: cmusco@nyu.edu.

Office: 370 Jay St., #1105 (Floor 11).

Office Hours: Wednesdays, 10:30-12:30pm, or by appointment.

TEACHING ASSISTANTS

Prathamesh Dharangutte: prathamesh.d@nyu.edu.

Zige Wang: zw1923@nyu.edu.

Office Hours: Time and location TBA.

ESSENTIAL LINKS

Course Web page: https://www.chrismusco.com/introml/.

Piazza Forum (for questions): https://piazza.com/class/k5psbvecgix5t2

NYU Classes (for turning in work):
https://newclasses.nyu.edu/portal/site/670a4d41-faa3-423f-b052-79b7f55aaff7

Course Git repo: https://github.com/cpmusco/introml, where you can directly download up-to-date course material (e.g. labs, homeworks, lecture slides). Note that this is a fork of the course repository used in previous years, and material will differ: please make sure you download assignments and labs from this link only!

COURSE DESCRIPTION

This course provides an introduction to machine learning and statistical pattern recognition, through a mixture of hands-on exercises and theoretical foundations. We will cover fundamental algorithms for function fitting (regression), classification methods (logistic regression, support vector machines, decision trees), model selection, generalization, neural networks, basic numerical optimization, dimensionality reduction, unsupervised learning, boosting, and a number of other topics. The course includes programming exercises on real and synthetic data using current software tools. A number of applications are demonstrated in audio and image processing, text classification, and more.

COURSE PREREQUISITES

Modern machine learning uses a lot of math! Probably more than any other subject in computer science that you will study as an undergraduate. Fortunately, however, the breadth of math used is relatively narrow: besides basic calculus, you can get very far with an understanding of just probability and linear algebra. That understanding needs to be *solid* for you to succeed in this course. Formally our prerequisites are as follows, and I list specific topics you should know from probability and linear algebra:

1. **Data Structures and Algorithms**. (CS-UY 1134, ENGR-UH 3510, CS-UH 1050, or CSCI-SHU 210)

2. **Probability** (MA-UY 2222, 2224, 2233, 3012, 3014, or 3514, MATH-UH 2011Q, ENGR-UH 2010Q, or MATH-SHU 235)

- Random variables, discrete and continuous probability distributions, expectation, variance, covariance, correlation, conditional and joint probability, Gaussian random variables, law-of-large-numbers.

3. **Linear Algebra** (MA-UY 2034, 3044, or 3054, MATH-UH 1022 or 1023, or MATH-SHU 140 or 141)

- Matrices and vectors, vector inner and outer products, matrix-vector and matrix-matrix multiplication, vector norms (e.g., Euclidean), matrix norms (e.g., Frobenius, operator), triangle inequality, Cauchy-Schwarz, solving systems of linear equations, linear independence, matrix rank, column/row span, null space, orthogonal matrices, basics of eigenvectors, eigenvalues, and eigendecomposition.

You also need to be a good programmer for this course. All coding exercises and assignments will be in Python. No prior experience specifically with Python is required, but I will not be focused on teaching the language, besides showing you specific tools useful for machine learning. So if you are not familiar with the basics, you will need to spend time familiarizing yourself. See the TAs if you run into any issues!

COURSE STRUCTURE AND GRADING

There are two class meetings per week, involving lecture, demonstrations, and ungraded exercises. You will also work on assignments at home, take two in-class tests, and complete a final project. Details follow:

Class participation (10% of grade). This grade captures how much you contribute to your own learning and that of your peers. Since different students have different styles, there are *many ways* to earn full credit for this part of the course! You can actively and vocally participate in class. You can ask questions, or answer those of your peers, on the class Piazza forum. You can attend and actively contribute to office hours (either mine or those of the TAs). Etc.

Weekly programming labs and written problem sets (30% of grade). These assignments are completed at home and reinforce the material discussed in class. I expect a lot of your learning to occur while working on these exercises, and investing time on them is the best way to prepare for the exams and final project. Assignments and their due dates will be posted on the course webpage. Late assignments will only be accepted if there are extenuating circumstances and you have obtained prior permission from the instructor.

Two in-class midterm exams (20% each, 40% total). For both exams you will be allowed a cheat-sheet (a two-sided piece of paper with whatever information you want on it).

Final project (20% of grade) To be completed in teams of 2. This is a completely open ended machine learning project: the goal is to pick a problem you are interested in that is amenable to machine learning, collect whatever data you need to address that problem, and apply tools learned in class to the problem. More details will be released later in the semester!

COURSE POLICIES

Programming labs: Turned in via NYU Classes.

Written problem sets: Turned in via NYU Classes. While not required, I encourage students to prepare written problem sets in LaTeX or Markdown (with math support.) Students will receive 10% extra credit on the Problem Set 1 for preparing it in LaTeX or Markdown.

Homework collaboration policy:

- *Discussion of high-level ideas for solving problems sets or labs is allowed, but all solutions and any code must be written up independently.* This reflects the reality that machine learning research and implementation are rarely done alone. However, any machine learning practitioner must be able to communicate and work through the details of a solution individually.

- *Students must name any collaborators they discussed problems with at the top of each assignment (list at the top of each problem separately).*

- *Do not write or code in parallel with other students.* If problem ideas are discussed, solutions should be written or implemented at a later time, individually. I take this policy very seriously. Do not paraphrase, change variables, or in any way copy another students solution.

Midterm: The exams will be held during class on **March 9th** and **April 20th**. Make sure you are available on those days. If there is any issue, let me know ASAP.

Grade changes: Fair grading is very important to me. If you feel that you were incorrectly docked points on an assignment or exam, please contact the TAs first, who will escalate to me as necessary. Do not wait until the end of the semester! If you notice something off, better to ask ASAP.

Questions about performance: If you are struggling in the class, contact me as soon as possible so that we can discuss what's not working, and how we can work together to ensure you are mastering the material. It is difficult to address questions about performance at the end of the semester, or after final grades are submitted: by Tandon policy no extra-credit or makeup work can be used to improve a students grade once the semester closes. So if you are worried about your grade, seek help from me and the TAs *early*.

READINGS

There is no textbook to purchase. Primary course material will consist of lecture slides and my handwritten notes from class (I use a tablet as a digital whiteboard and make those notes available.) To supplement that information, optional readings will be posted from each lecture. Many will come from:

- *An Introduction to Statistical Learning*, by James, Witten, Hastie, and Tibshirani. A free version of the book is available here http://faculty.marshall.usc.edu/gareth-james/ISL/.

- *Elements of Statistical Learning*, by Hastie, Tibshirani, and Friedman for some more advanced topics. A free version of the book is available here https://web.stanford.edu/~hastie/ElemStatLearn/.

- *Python Machine Learning* by Raschka is a good reference for issues with Python implementations and the labs. You can access it at http://file.allitebooks.com/20151017/Python%20Machine%20Learning.pdf. Note that the this book may be slightly out of date with the current Python version.

COURSE SCHEDULE

The following schedule is tentative and subject to change. An updated schedule will be maintained at https://www.chrismusco.com/introml/.

UNIT 1: Regression and Function Fitting

1. 1/27 Introduction to Machine Learning
2. 1/29 Simple Linear Regression, Loss Functions
3. 2/3 Multivariate Linear Regression, Data Transformations
4. 2/5 Model Selection, Cross Validation, Regularization

2/9: Last day to drop and not receive a grade of "W".

UNIT 2: Classification

5. 2/10 Classification, Naive Bayes
6. 2/12 The Bayesian Perspective cont.

    2/17 PRESIDENTS DAY, NO CLASS
7. 2/19 Logistic Regression, Support Vector Machines
8. 2/24 Optimization, Gradient Descent, SGD
9. 2/26, The Kernel Trick

10. 3/2, Learning Theory, Generalization
11. 3/4, Overflow/review day

    3/9, Midterm 1

UNIT 3: Beyond Linear Methods: Neural Networks

12. 3/11, Boosting, Decision Trees

    3/16, SPRING BREAK, NO CLASS

    3/18, SPRING BREAK, NO CLASS

13. 3/23, Neural Networks 1
14. 3/25, Neural Networks 2
15. 3/30, Convolution, Feature Extraction, Edge Detection
16. 4/1, Convolutional and Deep Networks
17. 4/6, Time Series and Sequential Prediction

UNIT 3: Unsupervised Learning

18. 4/8, Dimensionality Reduction 1: Principal Component Analysis
19. 4/13, Dimensionality Reduction 2: Semantic Embeddings, Recommendation Systems
20. 4/15, Clustering, K-Means, Gaussian Mixture Models

    4/20, Midterm 2

21. 4/22, Graphs and Spectral Clustering

UNIT 4: Select Topics

22. 4/27, TBA (possible guest speaker)
23. 4/29, Introduction to Online Learning
24. 5/4, Introduction to Reinforcement Learning
25. 5/6, Project Presentations
26. 5/11, Project Presentations

COURSE OBJECTIVES

1. Students will learn how to view and formulate real world problems in the language of machine learning. Categories of problems include those involving prediction, classification, pattern recognition, and decision making.

2. Through in-class demonstrations and at-home programming labs, students will gain experience applying the most popular and most successful machine learning algorithms to example problems. The goal is to prepare students to use these tools in industrial or academic positions.

3. In addition to experimental exploration, students will learn how theoretical analysis can help explain the performance of machine learning algorithms, and ultimately guide how they are used in practice, or lead to the design of entirely new methods.

4. Students will build experience with the most important mathematical tools used in machine learning, including probability, statistics, and linear algebra. This experience will prepare them for more advanced coursework in the subject, or research at the undergraduate or graduate level.

5. A major goal is to prepare students to read and understand contemporary research in machine learning, including papers from NeurIPS, ICML, ICLR, AAAI, JMLR, and other major machine learning venues. Since machine learning is a rapidly evolving field, many of its most powerful tools today may no longer be relevant in 15 years. The goal is to provide students with a theoretical foundation that will allow them to keep up with changes in the field.

At the end of this course, every student should feel confident holding a conversation about machine learning with anyone, no matter their experience in the field. This would include a potential employer, a professor or graduate student working in the field, a stranger on the subway, or a grandparent.

ADDITIONAL INFORMATION

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 here 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.

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 Schools rules on academic integrity will be sanctioned under this Policy. Students are responsible for familiarizing themselves with the Schools 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 persons 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 ones 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.

The complete policy can be found 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

The full calendar can be found here. Please pay attention to notable dates such as Add/Drop, Withdrawal, etc. For confirmation of dates or further information, please contact Susana Garcia: sgarcia@nyu.edu.

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

The full calendar can be found here. Please pay attention to notable dates such as Add/Drop, Withdrawal, etc. For confirmation of dates or further information, please contact Susana Garcia: sgarcia@nyu.edu.

Student Outcomes

The course is a component of the following student outcomes:

☑ 1   an ability to identify, formulate, and solve complex engineering problems by applying principles of engineering, science, and mathematics

☐ 2   an ability to apply engineering design to produce solutions that meet specified needs with consideration of public health, safety, and welfare, as well as global, cultural, social, environmental, and economic factors

☑ 3   an ability to communicate effectively with a range of audiences

☐ 4   an ability to recognize ethical and professional responsibilities in engineering situations and make informed judgments, which must consider the impact of engineering solutions in global, economic, environmental, and societal contexts

☑ 5   an ability to function effectively on a team whose members together provide leadership, create a collaborative and inclusive environment, establish goals, plan tasks, and meet objectives

☑ 6   an ability to develop and conduct appropriate experimentation, analyze and interpret data, and use engineering judgment to draw conclusions

☑ 7   an ability to acquire and apply new knowledge as needed, using appropriate learning strategies.