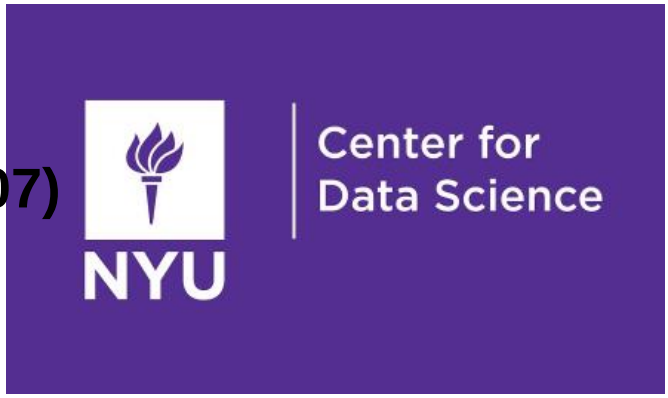
Introduction to Machine Learning (DS-GA 3001.007)



Term:	Fall 2019
Instructor:	Christopher Policastro
Section Leader:	Serkan Karakulak
Grader:	Ravi Choudhary
Contact:	NYU Classes via Messages or Forums
Website:	https://newclasses.nyu.edu/portal https://iml-f19.jupyter.hpc.nyu.edu https://wp.nyu.edu/imlf19/ https://cims.nyu.edu/~policast/textbooks/iml-f19
Lectures:	Wednesdays from 1:00pm-2:40pm at 12 Waverly Place, Room L120
Labs:	Thursdays 2:25-3:15pm at 60th 5th Avenue, Room 115.
Office Hours (Instructor):	Fridays 10am-11am (60 Fifth Ave., 7th floor, room 763) Tuesdays 10am-11am (NYU Classes via Zoom Conferencing)
Office Hours (Section Leader):	Mondays 2:30pm-4:30pm (60 Fifth Ave., 7th floor, room 763)

Course Description

The course introduces statistical and computational methods in machine learning with a focus on supervised learning (regression and classification) and unsupervised learning (dimension reduction and clustering). Students will formulate problems and implement algorithms to make predictions and inferences about data.

The principles and techniques covered in the course will help prepare students for courses on deep learning, causal inference, natural language processing, or optimal control. The practical experience with the tools needed for implementing algorithms will enable students to solve problems with data.

Course Objectives

Students will study principles and techniques for exploring patterns in data, predicting unknown information from known information, and estimating likelihoods. The techniques include both concepts and methods. Concepts teach practices about modelling. Methods teach implementation of models through programming. Learning outcomes include

- Determining Models
    - Bias and Variance
    - Error and Noise
    - Regularization and Validation
- Exploring Patterns
    - Clustering
    - Hidden Variables
    - Kernels
- Making Predictions
    - Linear/Logistic Regression
    - Support Vector Machine
    - Neural Network
- Quantifying Uncertainty
    - Boosting
    - Bagging
    - Expectation Maximization
- Gaining experience
    - Programming in Python
    - Querying in SQL
    - Typesetting in Latex

Throughout the semester, the course will be conducted in Python. However, the course will not focus on the specifics of Python. Instead, students will learn programming skills that should apply to different languages.

Course Requirements

The instructor will hold lecture once a week for 100 minutes. Lecture will combine instructional lessons and interactive activities. The section leader will conduct section once a week for 50 minutes. Section will include both discussion and lab. Discussions will give the class the opportunity to review the lecture guided by questions about the material. Labs will assess students' progress. Working in groups, students will solve problems related to the homework assignments alongside the section leader. Students will be expected to spend time studying outside of class. Grading will be determined by assignments, exams, and a project.

Resources

- Hastie, Tibshirani, Friedman, *Elements of Statistical Learning*, Second Edition, Springer-Verlag, 2009.
- Shalev-Shwartz and Ben-David, *Understanding Machine Learning: From Theory To Algorithms*, 2014.
- David Barber, *Bayesian Reasoning and Machine Learning*, Cambridge University Press, 2012.
- Kevin Murphy, *Machine Learning: A Probabilitic Perspective*, MIT Press, 2010.
- Christopher Bishop, *Pattern Recognition and Machine Learning,* Springer, 2007.
- Yaser Abu-Mostafa et al., *Learning from Data*, Packt Publishing, 2012.

Course Prerequisites

Students should have experience with programming. Some familiarity with algorithms and data structures could be helpful. Students should have experience with linear algebra and calculus. Some familiarity with probability could be helpful.

Course Policies

The grade will be based on assignments, exams, and a project:

- Assignments
    - *Homework*: 6 assignments combining code and short answer questions.
        - Submissions must be uploaded to Gradescope before 11:59pm on the due date.
        - Upload a single PDF document containing all the code, images, and text required for each problem.
        - Where homework assignments are divided into sections, please begin each section on a new page.
        - You will then select the appropriate page ranges for each homework problem, as described in the instructional video: "For students: submitting homework"
    - *Labs*: Weekly problems for groups in section. The problems will outline components of the homework. The lowest score will be dropped from the final grade. Must be completed on JupyterHub before the end of class.
- Exams
    - *Midterm*: Held the 8th week of the semester. If a student misses the midterm, then the final will be worth 25% (see grading distribution below).
    - *Final*: Held the 15th week of the semester. Time and place to be determined by the registrar.
- Project

- Students will work together in groups to put the concepts and methods from class into practice. Groups will gather data, determine a model, and implement an algorithm to make deductions about a problem. The project will not be graded on technicality or novelty. The project is an opportunity for groups to explore applications.
- *Project Proposal*: due the 6th week of classes. 1-2 page write-up addressing the motivation, approaches, and intended experiments of the project.
- *Project Milestone*: due the 10th week of classes. 2-3 page write-up addressing relevant datasets, preliminary experiments and possible next steps.
- *Project Poster*: due the 14th week of classes. Summary of datasets, features, models, and algorithms along with discussion of results. Please see instructions on NYU Classes for guidelines and template along with sample projects. Posters will be uploaded to https://wp.nyu.edu/imlf19/ for questions and comments.

Surveys

The instructors will ask students to complete three anonymous surveys. Surveys are accessible through links on NYU Classes. Survey 1 will be posted in Week 1 to learn the background and interests of the class. Survey 2 will be posted in Week 4 to gather suggestions about the class. Survey 3 will be posted in Week 9 to follow up on the suggestions and to gauge the midterm exam.

Lab Access

Any student adding the course between September 3rd and September 16th must contact the High Performance Computing Group at NYU for access to https://iml-f19.jupyter.hpc.nyu.edu. Please indicate the instructor as faculty sponsor. For more information see https://wikis.nyu.edu/display/NYUHPC/Getting+or+renewing+an+HPC+account.

Collaboration

Students can collaborate on homework and labs. However, students are responsible for mentioning their collaborators' contributions in their submission. Homework or labs without acknowledgements violates course policies.

With the exception of packages, students should avoid including duplicating code in their homework, labs, and project. If students duplicate code in their programs, then they must provide comments about the source with attributions.

Late Assignments

Labs must be completed in class. The project must be submitted the 15th week of the semester.

For homework, each student gets 5 extension days. Extension are rounded up to the nearest day. For example, 1 minute late means 1 extension day. After 5 extension days are used, any homework handed in late will be marked off 20% per day late, rounded up to the nearest number of days. No homework will be accepted more than 5 days late.

Final Grades

The following weights will be used in the assignment of final grades:

Homework	35%
Labs	15%
Midterm	10%
Final	15%
Project	25%

Final grades will be assigned based on the following scale:

A	**93-100**
A-	**90-92**
B+	**87-89**
B	**83-86**
B-	**80-82**
C+	**77-79**
C	**73-76**
C-	**70-72**
D+	**67-69**
D	**60-66**
F	**<60**

Schedule of Classes

Date	Lecture	Lab	Homework	Project
Sep 04	Overview Data Science	Lab 1	HW1 Posted	Project Posted
Sep 11	Overview Machine Learning	Lab 2	Survey 1 Due	
Sep 18	Gradient Descent and Risk	Lab 3	**HW1 Due**, HW2 Posted	
Sep 25	Regularization	Lab 4	Survey 2 Due	

Oct 02	Loss Functions	Lab 5	**HW2 Due**, HW3 Posted	
Oct 09	Support Vector Machine	Lab 6		**Project Proposal Due**
Oct 16	Feature and Model Selection	Review	**HW3 Due**	
Oct 23	**Midterm**			
Oct 30	Kernels and Multiclass	Lab 7	HW4 Posted, Survey 3 Due	
Nov 06	Trees and Bagging	Lab 8		**Project Milestone Due**
Nov 13	Boosting	Lab 9	**HW4 Due**, HW5 Posted	
Nov 20	Neural Networks	Lab 10		
Nov 27			**HW5 Due**, HW6 Posted	
Dec 04	Clustering	Lab 11		**Project Poster Due**
Dec 11	Hidden Variables	Review	**HW6 Due**	
TBD	**Final**			

University Policies

Academic Integrity

The course conforms to NYU's policy on academic integrity for students: (http://www.nyu.edu/about/policies-guidelines-compliance/policies-and-guidelines/academic-integrity-for-students-at-nyu.html

This policy prohibits plagiarism and cheating.

- Plagiarism: presenting others' work without adequate acknowledgement of its source, as though it were one's own. Plagiarism is a form of fraud. We all stand on the shoulders of others, and we must give credit to the creators of the works that we incorporate into products that we call our own. Some examples of plagiarism:
    - a sequence of words incorporated without quotation marks
    - an unacknowledged passage paraphrased from another's work
    - the use of ideas, sound recordings, computer data or images created by others as though it were one's own
- Cheating: deceiving a faculty member or other individual who assess student performance into believing that one's mastery of a subject or discipline is greater than it is by a range of dishonest methods, including but not limited to:
    - bringing or accessing unauthorized materials during an examination (e.g., notes, books, or other information accessed via cell phones, computers, other technology or any other means)

- providing assistance to acts of academic misconduct/dishonesty (e.g., sharing copies of exams via cell phones, computers, other technology or any other means, allowing others to copy answers on an exam)
- submitting the same or substantially similar work in multiple courses, either in the same semester or in a different semester, without the express approval of all instructors
- submitting work (papers, homework assignments, computer programs, experimental results, artwork, etc.) that was created by another, substantially or in whole, as one's own
- submitting answers on an exam that were obtained from the work of another person or providing answers or assistance to others during an exam when not explicitly permitted by the instructor
- submitting evaluations of group members' work for an assigned group project which misrepresent the work that was performed by another group member
- altering or forging academic documents, including but not limited to admissions materials, academic records, grade reports, add/drop forms, course registration forms, etc.

Learning Accommodations

NYU provides appropriate accommodations to students with documented disabilities. The Henry and Lucy Moses Center for Students with Disabilities works with undergraduate and graduate students with learning disabilities, attention deficit-hyperactivity disorders, physical disabilities, sensory impairments, psychiatric disorders, and other such disabilities in order to help students achieve their academic and personal potential. They facilitate equal access to the educational programs and opportunities offered at NYU and coordinate reasonable accommodations for eligible students. These services are designed to encourage independence and self-advocacy with support from the Moses Center staff. The Moses Center staff will facilitate the provision of accommodations on a case-by-case basis.