HUNTER
The City University of New York

Machine Learning

CSCI 353/CSCI 795
Fall 2020

Instructor: Anita Raja, Ph.D., Professor of Computer Science
Office: HN 1090G
Email: anita.raja@hunter.cuny.edu **URL:** https://anraja.commons.gc.cuny.edu/

Course Meeting Time: Tuesdays 5:35pm - 8:15pm

Meeting Format and Location: Fully Online Synchronous Class: Lectures will be held via blackboard collaborate during class time every week. Link will be posted on blackboard discussion board day before class. See Camera Use policy described below.

Office hours: Mondays, Tuesdays 1:00-2:00pm zoom meetings by appointment.
https://huntercollege.zoom.us/j/95863175457?pwd=ZzdYYXlsTWpycExGQk5HTEZCRVBMZz09
Meeting ID: 958 6317 5457
Passcode: 138608

Prerequisites:

- CSCI 235, MATH 150 with a grade of C or better. Background in Linear Algebra, calculus and optimization, probability, algorithm design, and data structures is expected.
- Python programming skills.
- Please be aware that the instructor in this course will require that the camera and audio be on during some class sessions especially during testing.

Course Credits: 3
This 3-credit course requires 3 hours of classroom or direct faculty instruction and on average 2-6 hours of out-of-class student work each week for approximately 15 weeks. Out-of-class work may include but is not limited to: Required Reading, Coding assignments, Written assignments, and studying for quizzes and exams. Class participation will include occasional pop quizzes on the blackboard.

Course webpage: Blackboard; visit it regularly.

Textbook resources: There is no required text book for this class**.** Comprehensive lecture slides will be made available the day before class and videos of the lectures will be available via the blackboard. Students are encouraged to take notes during class. Other readings will be provided on the schedule page.

Though there is no required text for this course, my lectures will draw references substantially from the following books. Several of these books have free authorized versions online that you can find using a quick search.

- Pattern Recognition and Machine Learning (PRML) by Christopher M. Bishop (https://www.microsoft.com/en-us/research/people/cmbishop/prml-book/)
- The Elements of Statistical Learning: Data Mining, Inference, and Prediction by Trevor Hastie, Robert Tibshirani, Jerome Friedman (http://web.stanford.edu/~hastie/ElemStatLearn/)
- Data Mining: Practical Machine Learning Tools and Techniques, 2017, Witten, Frank, and Hall. Morgan Kaufmann, the fourth edition, ISBN 978-0-12-804291-5. https://www.cs.waikato.ac.nz/ml/weka/book.html Available through the Hunter bookstore: http://hunter.textbookx.com/institutional/. Errata for the text are posted at http://www.cs.waikato.ac.nz/ml/weka/errata.html
- Learning From Data, Yaser S. Abu-Mostafa, Malik Magdon-Ismail, Hsuan-Tien Lin: ISBN: 9781600490064. http://amlbook.com/support.html
- Machine Learning: A Probabilistic Perspective by Kevin P. Murphy
- Artificial Intelligence: A Modern Approach by Stuart Russell and Peter Norvig (http://aima.cs.berkeley.edu/)
- Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems by Aurélien Géron

Advanced Topics:

- Reinforcement Learning: An Introduction by Richard S. Sutton and Andrew G. Barto (http://www.incompleteideas.net/book/first/ebook/the-book.html)
- Deep Learning by Ian Goodfellow, Yoshua Bengio, and Aaron Courville https://www.deeplearningbook.org/
- Boosting: Foundations and Algorithms by, Robert E. Schapire, and Yoav Freund

Catalog Description and Topics:

Machine Learning uses interdisciplinary techniques to automatically discover patterns in large volumes of data at high speed that help make predictions or decisions. It is now ubiquitous with applications spanning from medical decision making to financial informatics, weather tracking and homeland security. This course covers the theory, algorithms, and applications of machine learning. Students will study what is involved in learning models from data and how to evaluate these models.

Topics covered will include: (1) supervised learning: generative/discriminative learning, parametric/non- parametric learning, neural networks, and support vector machines; (2) unsupervised learning: clustering, dimensionality reduction, kernel methods; (3) learning theory: bias/variance tradeoffs; VC theory; large margins; (4) reinforcement learning and adaptive control and (5) fairness, bias and machine ethics.

Python will be the programming language for this course. Only scientific libraries such as SciPy, NumPy, Pandas, Matploblib will be allowed to be used in the assignments and final project. Generally students are not allowed to use any machine learning library (scikit-learn, tensorflow, etc.) unless explicitly permitted.

Since this is an advanced elective, we will provide support for all ML topics covered in the courses. Students will be expected to be comfortable with all related coding/compilation/debugging for the programming assignments and projects.

Machine Learning generally involves significant mathematical knowledge and ability to think mathematically. A reference is provided in the blackboard to the mathematical pre-requisites for this course and students are encouraged to review the topics on their own. A subset of mathematical topics will be covered during lectures.

This a cross-listed undergraduate and graduate course. There will be higher expectations with regard to the quality and quantity of work for the 795- level students compared to the 353 level students. As a student in the 795 course, you will be expected to meet or exceed those expectations, which may require more of your time and a more proactive approach to studying and greater diligence in completing assignments. In particular, assignments and exams will contain additional requirements or activities. In the 353 level, these will be opportunities for extra credit/bonus points. However, in the 795 level, these are *required* and will be treated as part of your regular grade. There will higher expectations from 795 students for their final project as well.

Learning Goals:
This class partially satisfies the following learning goals:
(1d) Display knowledge of at least two area disciplines within computer science (for example: artificial intelligence, computer theory, formal methods, etc.)
(2a): Be proficient in writing and reading programs sufficient to implement and study algorithms.
(2b) Be able to apply principles of design and analysis in creating substantive projects involving programs and algorithmic design, and have experience working in teams on projects of moderately realistic scope.
(3a): Be able to communicate technical ideas effectively, both in writing and in oral presentations.
(3b)Demonstrate an understanding of the ethical concerns typically arising in the context of computing.

Learning Outcomes:
By the end of the course, students should be able to

- Understand the process involved in equipping machines to learn models from data.
- Gain knowledge about a wide variety of learning algorithms and how to optimize them.
- Be skilled at evaluating models generated from data.

- Apply the algorithms to a real-world problem and report on the accuracy of the various models.

How to Succeed in this Course:

Plan on doing *all* of the following:

- Do the assigned readings *before* the lecture, *not after* it.
- Make a list of questions before the class.
- Give yourself enough time to complete assignments and the project. Each will need significant amount of effort.
- Submit all assignments on time.
- Solve a set of problems at the end of each chapter.
- Check the blackboard for updates.
- Study for exams.
- Adhere to assignment requirements and the College's Academic Integrity policy.

Syllabus and Readings:

The topics covered and tentative schedule are listed in the schedule document on the blackboard. You are responsible for everything in the listed chapters regardless of how much time we spend on them in class. As noted above, you should read ahead so that you can ask questions in class, email the instructor or use the blackboard discussion forum to clear up anything you find confusing. Responses to posting can be expected within 24 hours of the post during week days.

Course Grading Rubric

Your final course grade will be calculated as follows:

Participation and Pop Quizzes	10%
Assignments (3)	30%
Midterm Exam	20%
Final Exam: Final Project + Short Oral Exam	40% (25% final project; 15% oral exam)

Logistics: Assignments, Projects and Participation

- We will use Blackboard collaborate for lectures. Recordings of the lectures will be posted on the blackboard page.
- We will use the waiting room feature via zoom for office hours and oral exams.
- Participation in class is encouraged. The Blackboard discussion board will be used for discussion for ML related questions. Students are encouraged to participate in these discussions in answering peer questions. The instructor will both respond when needed and supervise the discussion.
- Attendance will not be taken but pop quiz performance as well engagement in class discussion and blackboard discussion board will contribute towards participation grade.
- Students will need access to a camera since cameras must be turned on timed tests and exams.

- There will be one written take-home midterm exam and the final exam will be composed of a final project deliverable and an oral exam. The exams will cover class readings, notes and discussions.
- Assignments will be composed of both written and programming components. Each assignment will require several days of work.
- All submissions must be uploaded to the blackboard on time. Only electronic submissions will be accepted. Any handwritten and scanned components must be neat and legible.
- Assignments are individual submissions. You may discuss the assignments but the solution and code must be your own. Sharing and/or receiving solutions is not allowed. Please read academic integrity notes further below.
- It is imperative that all submitted program listings and executions be thoroughly documented. You are allowed to use the source code that is provided by the book. Acknowledge all code that you use from the textbook.
- All programs must compile and run. Zero credit will be given for programs that do not compile or do not run. Usually the homework assignments will only state the major objectives of the program to be written; it will be often up to you to make design decisions regarding I/O, efficiency, error handling, and so on. Make sure you test the code adequately to indicate the correctness and robustness of your approaches.
- All programming projects must be submitted by **11pm on the due date**.
    - For each late day, 10% is deducted (i.e. ε to 24 hours late results in 10% penalty; 24-48 hours late means 20% penalty and so on). After three late days, the assignment **will not be accepted.**
- If you have questions about grading or believe that points were deducted unfairly, you must **email the instructor within a week after the grade for the assignment is posted**. No further consideration will be given to any assignment a week after the grade is posted.
- The final project will be a group project with 2-3 students with diverse backgrounds on each team. A set of topics and project specifications and criteria will be shared mid-semester. In accordance to Departmental requirements for electives, the final project will include a written component of atleast 500 words.
- Students are encouraged to ask questions in class. The class participation grade will be evaluated using the following criteria:
    - Relevance and intellectual depth of questions.
    - Accurate responses to questions posed in class.
    - Initiating and/or Engaging in discussions in class and via online board that help further understanding of course topics among students.
    - Innovative and neat solution approaches to problems posed in class.

Technical Support

You can use your own machines or the Hunter lab machines (remotely) for this class. The Department linux webpage is here: http://compsci.hunter.cuny.edu/~csdir/ To claim your eniac account, you need to login and issue the command: touch fall.2020 in their home directory NLT September 14, 2020. If you need tech help for any reason, email cstechsp@hunter.cuny.edu about the issue.

You can continue to remotely login to Hunter lab machines as follows:

1) ssh <your_username>@eniac.cs.hunter.cuny.edu
2) type your password
3) Now you are at a gateway machine that is called eniac
4) Do not do any processing on eniac. Just ssh through eniac to one of the machines in the lab (see next step)
5) ssh <your_username>@cslab<X>.cs.hunter.cuny.edu, where <X> is the number 1 through 29. You can pick any machine. If the machine is down you can try another machine. For instance to login to the 2nd machine type: ssh <your_username>@cslab2.cs.hunter.cuny.edu
6) All cslab<X> machines and eniac see the same directories for your account. That means that you see the same files in all machines.
7) If you want to test your programs in one of the cslab machines you can use sftp in order to transfer your code to eniac. Then you can ssh to eniac (see step 1) and after that ssh to any cslab<X> machine (see step 5

Course Material and Communication

All course material (i.e. slides or other sources) will be uploaded on *blackboard*. Homework assignments will be uploaded on *blackboard*. We will also use *blackboard email and discussion board* for communication.

If you have a question about the class, topics covered, homework, etc. please first check the posts in the Q & A section. If you did not find an answer, please post your question there. That will make it easier for all students to see answers. You can also answer questions of fellow students. Do not post any code solutions. For questions involving personal matters you can email me. Talk to me if there are questions or concerns about the course. I also welcome feedback on the course progress through out the semester.

The standards and requirements set forth in this syllabus serve as the course policy. Notice of any modifications/changes to the syllabus will be by announcement in class or by changes to this syllabus posted on the course website along with the change date.

Camera Use and Recordings Policy

Students who participate in this class with their camera on or use a profile image are agreeing to have their video or image recorded solely for the purpose of creating a record for students enrolled in the class to refer to, including those enrolled students who are unable to attend live. If you are unwilling to consent to have your profile or video image recorded, be sure to keep your camera off and do not use a profile image. Likewise, students who un-mute during class and participate orally are agreeing to have their voices recorded. If you are not willing to consent to have your voice recorded during class, you will need to keep your mute button activated and communicate exclusively using the "chat" feature, which allows students to type questions and comments live.

Make-up Policy

All exams must be taken and submitted on time. Failure to take an exam counts as a zero grade on that exam. There will be no make-up exams. In case you must miss an exam or a homework for a unforeseen valid medical/family emergency, you have to inform me as soon as possible (no later than the day of the exam or earlier), provide appropriate

documentation and there will be a conversation on whether and how your exam grade composition will be adjusted.

Academic Integrity

Hunter College regards acts of academic dishonesty (e.g., plagiarism, cheating on examinations, obtaining unfair advantage, and falsification of records and official documents) as serious offenses against the values of intellectual honesty. The college is committed to enforcing the CUNY Policy on Academic Integrity and will pursue cases of academic dishonesty according to the Hunter College Academic Integrity Procedures. More details in the programming rules document on the blackboard.

I take academic integrity very seriously to ensure fairness in grading to all students in the class and to ensure that the grade you get reflects the value of your degree and quality of the institution you will graduate from.

1. While you are encouraged to discuss project assignments with others, **all work submitted must be your own.** You MAY NOT show your solution to a classmate or ask another student to see their solution. You may not ask another student to debug your code.
2. You may not use code from the Internet (e.g. StackOverflow). You can use code from the textbook unless otherwise specified in the assignment. You should properly attribute the code (add a comment citing in detail the source of the code — NOTE: you must always do this whenever you find yourself using others' code).
3. You may not post your code or answers where it is accessible to others, and you may not seek help from online forums. Contract cheating is a form of academic dishonesty in which students get others to complete their coursework for them. Please read more information on Contract cheating from http://en.wikipedia.org/wiki/Contract_cheating
4. As a rule of thumb, **you must type and debug your code without directly copying someone else's code**. For the first incident of cheating or plagiarism your grade will be a 0 and it will not be dropped as the lowest. For the second incident, you will fail the class. We report all incidents to the Office of Student Affairs.

ADA Compliance

In compliance with the American Disability Act of 1990 (ADA) and with Section 504 of the Rehabilitation Act of 1973, Hunter College is committed to ensuring educational parity and accommodations for all students with documented disabilities and / or medical conditions. It is recommended that all students with documented disabilities (Emotional, Medical, Physical and / or Learning) consult the Office of Accessibility located in Room E1124 to secure necessary academic accommodations. For further information and assistance please call (212-772-4857)/TTY (212-650-3230).

Hunter College Policy on Sexual Misconduct

In compliance with the CUNY Policy on Sexual Misconduct, Hunter College reaffirms the prohibition of any sexual misconduct, which includes sexual violence, sexual harassment,

and gender-based harassment retaliation against students, employees, or visitors, as well as certain intimate relationships. Students who have experienced any form of sexual violence on or off campus (including CUNY-sponsored trips and events) are entitled to the rights outlined in the Bill of Rights for Hunter College.
a. Sexual Violence: Students are strongly encouraged to immediately report the incident by calling 911, contacting NYPD Special Victims Division Hotline (646-610-7272) or their local police precinct, or contacting the College's Public Safety Office (212-772-4444).
b. All Other Forms of Sexual Misconduct: Students are also encouraged to contact the College's Title IX Campus Coordinator, Dean John Rose (jtrose@hunter.cuny.edu or 212-650-3262) or Colleen Barry (colleen.barry@hunter.cuny.edu or 212-772-4534) and seek complimentary services through the Counseling and Wellness Services Office, Hunter East 1123. CUNY Policy on Sexual Misconduct Link:
http://www.cuny.edu/about/administration/offices/la/Policy-on-Sexual-Misconduct-12-1-14-with-links.pdf

Thank you for abiding by these policies. Doing so will ensure the learning experience is most productive and fair to everyone taking this class and/or the future offerings of this course.

CUNY Policies: As indicated above**,** online courses are subject to the same CUNY policies as are in-person courses regarding academic integrity, the acceptable use of computer resources, equal opportunity and non-discrimination, sexual misconduct, workplace violence, domestic violence, and reasonable accommodations for persons with disabilities. The associated links are listed below:

- CUNY Academic Integrity Policy: https://www.cuny.edu/about/administration/offices/legal-affairs/policies-procedures/academic-integrity-policy/
- CUNY Policy on Acceptable Use of Computer Resources: https://www.cuny.edu/wp-content/uploads/sites/4/page-assets/about/administration/offices/cis/it-policies/ComputerUsePolicy1.pdf
- CUNY Policy on Acceptable Use of University Data in the Cloud: https://www.cuny.edu/wp-content/uploads/sites/4/page- assets/about/administration/offices/cis/information-security/security-policies- procedures/Acceptable-Use-of-University-Data-in-the-Cloud-2019-8-19a.pdf (and related Data Classification Standard: https://www.cuny.edu/wp-content/uploads/sites/4/page-assets/about/administration/offices/cis/information-security/security-policies- procedures/Data-Classification-Standard-CUNY-2019-8-19a.pdf
- CUNY Intellectual Property Policy: https://www.cuny.edu/wp-content/uploads/sites/4/page-assets/about/administration/offices/legal- affairs/policies-procedures/Intellectual-Property-Policy.pdf
- CUNY information on copyright: https://www.cuny.edu/about/administration/offices/legal-affairs/intellectual-property/copyright-materials/
- CUNY Equal Opportunity and Non-Discrimination Policy: https://www.cuny.edu/about/administration/offices/legal-affairs/policies-procedures/equal-opportunity-and-non-discrimination-policy/
- CUNY Policy on Sexual Misconduct: http://www.cuny.edu/wp-content/uploads/sites/4/page-assets/about/administration/offices/legal- affairs/policies-procedures/Sexual-Misconduct.pdf

- CUNY Campus and Workplace Violence Prevention Policy: https://www.cuny.edu/wp-content/uploads/sites/4/page- assets/about/administration/offices/legal-affairs/CUNY-Campus-and-Workplace- Violence-Prevention-Policy-2.28.11-and-amended-9.26.2011.pdf
- CUNY Domestic Violence and the Workplace Policy: http://policy.cuny.edu/general-policy/article-v/#policy_5.061
- CUNY Procedures for Implementing Reasonable Accommodations and Academic Adjustments: https://www.cuny.edu/about/administration/offices/legal- affairs/policies-procedures/reasonable-accommodations-and-academic- adjustments/

Acknowledgements:
Some of the materials for this course were adapted from the textbooks, online materials associated with the various texts and other online sources. Thanks also to Professors Ansaf Salleb-Aouissi and Mohammad Hasan for the discussions and resources.