Syllabus for CSCI 493.77: Deep Learning

Professor **Susan L. Epstein**
Email: susan.epstein@hunter.cuny.edu
Office hours: Wednesdays and Fridays 4-5 PM
https://us02web.zoom.us/j/86477126359?pwd=V0JSK0lHTHV2R1lyRlo0dlhDUEVHdz09
Class meets: Mondays and Thursdays 1:10-2:25 PM
Department office: 1008 Hunter North

Professor **Tiziana Ligorio**
Email: tligorio@hunter.cuny.edu
Office hours: Tue 12- PM, Thu 3-4 PM
Office: 520H Hunter North
Mode of instruction: P (in person)
Department phone: 212-772-5213

Course website: On Blackboard, accessible through the CUNY Portal with Chrome, Firefox, or Safari.

Course description

Machine learning is the subfield of artificial intelligence that learns to predict and classify from data. Deep learning is a family of machine learning algorithms based on artificial neural networks. This course takes a pragmatic, hands-on approach to material but is rigorously grounded in mathematics. Course work blends theory with practice. You will learn to use a variety of Python-based packages for visualization, machine learning, and deep learning, and address theoretical and computational challenges. As you explore data from many different sources, this course will change forever your ideas about computers and learning.

Prerequisites

Making the future is fun, but it also takes knowledge. All students should have *completed* CSCI 335, CSCI 150, MATH 155, and STAT 213 with grades of C or better. (The Department does not accept CR grades as prerequisites.) In addition to skill in Python programming, you are going to need a solid background in Boolean logic, basic probability theory and combinatorics, big-O complexity analysis, algorithm design, and data structures. A general fondness for mathematics and data is essential.

Required course materials

Text: *Hands-on Machine Learning with Scikit-Learn, Keras & TensorFlow,* A. Géron, **2nd edition**, ISBN 10: 1492032646 / ISBN 13: 9781492032649. No earlier edition matches the accuracy and coverage of this book. Classwork and reading assignments will reference this text. **Accept no substitutes**. Available through the Hunter bookstore @https://hunter.textbookx.com/institutional/index.php?action=browse#/books/3202381/. You can find some corrections here: https://www.oreilly.com/catalog/errata.csp?isbn=0636920142874

Slides will be available immediately after class. You are strongly encouraged to follow the animated version during class for greater clarity.

Google Colab: In this course we will use some of Google's computing power through labs in Jupyter notebooks. If you are unfamiliar with them, you can get an overview here: https://www.tutorialspoint.com/google_colab/google_colab_quick_guide.htm

All programming in this course requires Python and assumes that you know pandas and numpy. If you need help with those, contact Professor Ligorio.

Required code and data may be distributed through your CSCI departmental Linux account.

Students' responsibilities

• **Come to all classes on time and well prepared.** The text is a necessary resource, but the course will go far beyond it. You are responsible for all material in the assigned reading, whether or not it is covered during class. Be ready to ask, and to answer, questions on the reading. Detailed notes are highly recommended.

• **Maintain a Linux account with the Department** and abide by the rules for the Department's labs, found here: http://cs.hunter.cuny.edu/~csdir/. **If you do not have a CSCI Linux account, contact Professor Epstein immediately. To claim your account** log onto eniac and from your home directory issue the command: touch spring 2022 If you have difficulties with an existing account, contact cstechsp@hunter.cuny.edu.

• **Promptly read all email sent by the instructors to your registered Hunter Blackboard address.** Changes to assignments, clarifications, and instructions will often come by email.

• **Keep pace with the course.** The course schedule, required reading, all assignments, quizzes, and the project will be accessed through Blackboard. Be sure to **check it regularly for changes** as the semester progresses.

• **Acknowledge** any help received from other people, and reference in full any material used (e.g., book, paper, journal, web site) to prepare assignments. **Be sure to read "How to avoid plagiarism" on the class website.**

• **Class participation** is strongly encouraged and 10% of your grade. Your presence is necessary but not sufficient. Relevant questions count, as do accurate responses to questions posed in class by the instructors or another student. Questions that help us all understand the material in greater depth are particularly welcome.
• **Time commitment:** The amount of time you devote to this course will depend upon your interest, your mathematical background, your ability to read technical material, and your skill with Python. Plan on at least 8 hours per week outside of class for it.

Anticipated course topics (subject to change)

Introduction to AI and machine learning
Fundamental ideas and methods
Neural and mathematical foundations
Functional units, matrices, and linear algebra
Training and testing artificial neural networks
Deep neural networks
Convolutional neural networks
Practical methodology and techniques
Recursive neural networks
Autoencoders
Generative adversarial networks
Transformers

Assignments

All assignments are designed to increase your understanding of the material, and must be done as the course progresses. Often, one will build upon the next, so **skipping an assignment is *not* an option**. *It is not possible to do well in this course without doing the assignments thoroughly and submitting them on time.*
• **Reading assignments** require hours of careful study**.** Classes are organized by topic; some topics will span several classes. If you do not do the reading *regularly and thoroughly*, you will find the course extremely difficult. Reading is either from the text or posted on the class website. Quizzes will cover reading, labs, and class discussions.
• **Quizzes** occur weekly and must be taken through Blackboard. Your quiz is drawn at random from a very large pool of questions. (For example, the probability that any two students will encounter the identical 10-question quiz from a pool of 50 is 1 in 10,272,278,170.) Study the slides, your notes, and the reading carefully before you take a quiz, because you can only take your quiz once and will only have an hour in which to do so. Each quiz covers the most recent material heavily, but will likely include questions from earlier in the course. *No late quizzes will be accepted.*
[a] **Lab assignments** require *hours* of thought and effort and must be submitted through GradeScope. **Plan on spending time on them *over several days.*** You may discuss your ideas with each other, but you must do your own work. Any written components must be coherent, in full sentences, and use appropriate technical vocabulary. Programs must be fully documented and execute correctly. *If your code does not run you will receive no credit for that question.*
• **Project.** Your project is team-based and you will work in pairs. How much fun you have with it and how ambitious you are will determine how much time it will take. The project will be posted and discussed in class, and require both team members to present orally and write separate documents with full references.
• **Deadlines:** All work must be submitted **on or before its respective due date and time. Quizzes and lab assignments are due at 1PM. Project steps are due at 11PM.**
• **Extensions: Upon request,** you may **delay one lab assignment for one week without penalty**. Otherwise, the instructors may grant a brief extension on an individual basis, *if requested in advance*. Repeated or last-minute requests for extensions will be denied in other than extraordinary circumstances.

Grading

This course includes both theory and practice. You must be able to define important terms and to explain ideas in clear English. Course grades are based on lab assignments, an extensive project, weekly quizzes, and thoughtful, well-prepared class participation. The project will include an original essay of at least 500 words and an in-class presentation. **Grades will not be curved.** You must **complete all lab assignments to pass this course**.

Lab assignments	25%
Quizzes	25%
Project	40%
Class participation	10%

Late lab assignments and project steps lose 10% per 24-hour period after their respective deadlines. For example, any amount of time up to 24 hours loses 10%, 24 to 48 hours loses 20%, and so on. *No late quizzes or extra credit will be accepted.*

Learning Outcomes

This course partially satisfies the following Departmental 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.

By the end of this course, students will be able to:
• Explain what it means for a program to learn and the major theoretical results that address it
• Identify the neuroscience foundations of artificial neural networks and their variants
• Prepare data for deep learning experiments
• Apply and evaluate the performance of multiple deep learning algorithms
• Demonstrate skill with deep learning software
• Design rigorous training and testing environments for deep learning
• Identify algorithms, parameters, and hyperparameters appropriate for learning on a particular dataset
• Analyze and compare the performance of those learning techniques and assess learning results

Intelligent agents communicate

• **Talk to us:** ***Everyone*** is expected to participate in class. Ask questions. Express opinions. In return, we are happy to answer questions, listen to concerns, and talk with any student about topics related to the class (or not). We actually *enjoy* student visits during office hours. You can also make an appointment to see us at other times. We also welcome your feedback throughout the semester about how the course is progressing.
• **Write to us:** You can reach us by email almost every day, but not too late at night or very early in the morning.
• **Course website:** The course website is available on Blackboard and used in a variety of ways. Check it regularly, particularly for *schedule and assignment updates*.
• **Be clear and correct:** Your work must be *legible and unambiguous*.
• **Writing:** In accordance with Departmental requirements for elective courses, the final project includes (but is not limited to) a written analysis of at least 500 words.
• **Project pairs and study groups:** The project is intended for pairs. Plan on finding and staying with a partner who can meet regularly. Although study groups are not required, students who discuss the material and work together typically learn much more than those who work alone. You are encouraged to form study groups to make up and then solve practice problems, or to work on the ones you will find at the ends of the chapters. Such practice is necessary, but not sufficient, preparation for quizzes.

Course policies

• **Attendance:** Students are expected to attend all classes.
• **Lateness:** See the Grading section above for the lateness policy. Persistent lateness to class will eventually be counted as absence.
• **Missed quizzes:** Except in extraordinary circumstances, no makeup quizzes will be given.
• **Extra credit:** Any extra credit opportunities will be specified in the lab assignments and/or project and will be due at the same time as the required material. *No late extra credit will be accepted.*
• **Blackboard: Students are expected to check the class website daily.**
• **Email:** Students are expected to read their Hunter email daily for clarifications and changes in reading and other assignments.
• **For rules, guidance, and instructions for the Linux lab, including remote log in,** everything you need to know can be found here: http://compsci.hunter.cuny.edu/~csdir/
• **Electronics** : All cell phones, pagers, and other devices must be inaudible during class. Laptops and other electronic devices are to be used only for class related activities. Activities *not* related to class include but are not limited to Facebook, Twitter, TikTok, other social networking services, "surfing", email, and video services. Students whose electronic device disrupts the class or is used for anything other than class related activities will lose 5 points from their class participation grade per occurrence.
• **You may not record this class without express permission from the instructors.**

Academic integrity

We take academic integrity seriously. It allows us to grade fairly and to ensure that your grade reflects what you have learned. Fair grading reflects on the quality of our Department, the quality of your degree, and the quality of Hunter College itself. *Giving and receiving output, code, or answers is equally reprehensible.*
• You may discuss your ideas with one another, but you must write your own assignments, run your own experiments, and type and debug your own code, all without directly copying someone else's.
• You may not show your solution to a classmate or ask to see theirs, nor may you ask another student to change your code.
• You may not use any code from the Internet (e.g., StackOverflow). The sole exception is the official code in labs and the text, which you must comment in detail in your own file.
• You may not post your code where it is accessible or seek help from online forums.
• Contract cheating is expressly forbidden. It is a kind of academic dishonesty where students have others do any coursework for them. For clarification see: http://en.wikipedia.org/wiki/Contract_cheating
• **Hunter College Policy on 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.
• For the first incident of cheating or plagiarism your grade will be a 0. For the second incident, you will fail the class. We report all such incidents to the Office of Student Affairs.

ADA policy

ADA policy: 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/or medical conditions. It is recommended that all students with documented disabilities (Emotional, Medical, Physical and/or Learning) consult the Office of AccessABILITY located in Room E1214B to secure necessary academic accommodations. For further information and assistance please call 212-772-4857 or 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.

1. 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).
2. 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. They make all current and future students' experience in CSCI equitable and safe.

Want to learn more?

Students often ask for additional reference material. Here are some suggestions:

• **Do not go web surfing.** Many sites and blogs are inaccurate or shallow or both.

Python

• https://docs.python.org/3/tutorial/

• https://runestone.academy/runestone/books/published/thinkcspy/index.html

Deep learning

• *Artificial Intelligence: A Modern Approach*, Russell and Norvig, 4th edition, Chapter 21

• *Deep Learning,* Goodfellow, Bengio, and Courville. The grad student's go-to, with solid mathematical and theoretical background material.

Other families of machine learning algorithms

• *Bayesian Reasoning and Machine Learning,* David Barber. **Available online** at http://web4.cs.ucl.ac.uk/staff/D.Barber/textbook/240415.pdf with errata at http://web4.cs.ucl.ac.uk/staff/D.Barber/pmwiki/pmwiki.php?n=Brml.Errata

• *The Elements of Statistical Learning*, Trevor Hastie, Robert Tibshirani, and Jerome Friedman. 2008, a solid mathematical approach, **available online** at https://web.stanford.edu/~hastie/Papers/ESLII.pdf

• *Machine Learning,* Tom Mitchell. 1997 but a classic and accessible text

• *Machine Learning,* Peter Flach, 2012, mathematical but reasonably accessible with limited topic coverage

• *Pattern Classification,* Richard Duda, Peter Hart, and David Stark, second edition, 2001. The rigorous statistical origins of much research in machine learning

• *Machine Learning: A Probabilistic Perspective,* Kevin P. Murphy, 2012. The probabilistic compendium, with a heavy mathematical bent.

• *Pattern Recognition and Machine Learning*, Christopher M. Bishop, 2007. An excellent graduate level text with a heavy mathematical bent.

Statistics

• *Introduction to the New Statistics: Estimation, Open Science, and Beyond*, 2017, Cumming and Calin- Jageman, Taylor & Francis,, a basic and clear introduction to statistical testing

• *Understanding the New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis*, 2012, Cumming, Taylor & Francis, a more advanced version that elaborates on the material in *Introduction to the New Statistics*

Mathematics

• Probability: The Analysis of Data, Guy Lebanon. Volume 1 covers most of the math you wish you knew, and is available online http://www.theanalysisofdata.com/probability/0_2.html

- More resources appear on the class website

Want to stay current?

The top conferences and journals in the field are

- *Conference on Neural Information Processing Systems* (NeurIPS)
- *International Conference on Machine Learning* (ICML)
- *Annual Conference on Learning Theory* (CoLT)
- *Journal of Machine Learning Research* (JMLR) available free on line at `www.jmlr.org`
- *Machine Learning* (MLJ) Published by Springer

How to do well in this course

- Allot *substantial* time from your life to this course.
- Do the assigned reading *before* class, so you can ask informed questions.
- Attend class faithfully and take detailed notes.
- Ask questions in class when you don't understand something.
- Study your notes, the reading, and the slides before you take the quizzes.
- Carefully review the labs before you work on the lab assignments each week.
- Submit all assignments on time.
- Abide by the Department's Policy on Academic Dishonesty.

Note: Details of this document, including grading percentages, are subject to change as the need arises.

Acknowledgements

Scholars acknowledge their sources. Thanks to Pierter Abbeel, Jimmy Ba, Yoshua Bengio**,** Carla Brodley, Aaron Courville, Tom Dietterich, Mark Ebden, Tim Finin, Peter Flach, Aurélien Géron, Ian Goodfellow, Geoff Hinton, Marie des Jardins, Thorsten Joachims, Sergey Karayev, Ron Kohavi, Miroslav Kubat, Yann LeCun, Percy Liang, Tom Mitchell, Mehryar Mohri, Andrew Moore, Andrew Ng, Peter Norvig, Smiljana Petrovic, Tomaso Poggio, Afshin Rostamizadeh, Lorenzo Rosasco, Stuart Russell, Russ Salakhutdinov, Padhraic Smyth, Ameet Talwalkar, Josh Tobin, Dan Weld, and Pat Winston, from whose material some of this course was adapted.