Syllabus for CSCI 493.77: Deep Learning

Professor **Susan L. Epstein**
Email: susan.epstein@hunter.cuny.edu
Office hours: T 11:30AM -12:30PM F 2:15-3:15PM
Class meets: Tuesdays and Fridays 1:00 - 2:15 PM
Room: 1000 J Hunter North

Professor **Tiziana Ligorio**
Email: tligorio@hunter.cuny.edu
Office hours: Tuesdays and Fridays 11:30 – 12:30AM
Mode of instruction: P (in person)
Attendance is required

Course website: On Blackboard, accessible through the CUNY Portal with Chrome, Firefox, or Safari
Department office: 1008 Hunter North
Department phone: 212-772-5213

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, computer vision, and neurobiology. Coursework blends theory with practice. You will learn to address theoretical and computational challenges and to use a variety of Python-based packages for visualization, machine learning, and deep learning. Exploration of data from many different sources will change forever your ideas about computers and learning.

Prerequisites

Making the future is fun, but it also takes knowledge. 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 or P 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. While CSCI 353 Machine Learning and MATH 260 Linear Algebra would be helpful, they are not required. We do, however, strongly encourage you to review in advance NumPy and Pandas, with the following tutorials

- NumPy
- Pandas

Required course materials

Required text: *Hands-on Machine Learning with Scikit-Learn, Keras & TensorFlow,* A. Géron, **3rd edition**, ISBN 13: 9781098125974. 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. You can find some corrections to the text here: https://www.oreilly.com/catalog/errata.csp?isbn=0636920652915

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

Google Colab: This course uses some of Google's computing power through labs in Jupyter notebooks. If you are unfamiliar with them, there is an overview at https://research.google.com/colaboratory/faq.html At some point in the course you will need more space and time to run your experiments. Once that happens, you must sign up for **Colab Pro** which charges a small monthly fee https://colab.research.google.com/signup

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

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

How this course works

This course consists of theory and practice. Each new topic builds on all the preceding ones. To keep you on pace with this highly technical material and to avoid high-stakes exams, there are a variety of assessments.

• A **reading assignment** for each class is posted on our website. Much of it is from the text.

• The 13 **quizzes** cover reading, labs, and class discussions throughout the course. Each one must be taken within 1 hour through **Blackboard** and can be taken only once. A quiz is usually available immediately after a lecture. Your personal quiz is drawn at random from a large pool of questions. (For example, the probability that any two students will encounter an identical 10-question quiz from a pool of 50 is about 1 in 10 billion.) About half the questions on each quiz will be on material from earlier in the semester; the remainder will be on the most recent lecture and lab and the reading for them.

• **Regrade requests** must be submitted **within 48 hours** after the quiz deadline with this form: https://forms.gle/NFnFRCFWfHrroJyu5.
• There are **4 lab assignments**. Each one requires *hours* of thought and effort and must be submitted through **Gradescope**. If you are unfamiliar with Gradescope, clear instructions are available at https://gradescope-static-assets.s3.amazonaws.com/help/submitting_hw_guide.pdf. 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. Code segments must be preceded by explanatory paragraphs in your notebooks.
• Once the lab assignments end, the **project** begins. It is for pairs of students and staged in 6 parts to keep you on track. How much fun you have with it and how ambitious you are will determine how much time it will take and how successful you will be. The project will be discussed in class and requires both team members to code, to present orally and to write separate documents with full references.

How to manage this course

• **Complete and submit the entry form** https://forms.gle/gXvXtmmjBPLR8G7AA
• **Read** your @myhunter email daily.
• Watch the **deadlines** on the website carefully. All work must be submitted on or before its respective due date and time. Quizzes and lab assignments are due at 12:30 PM. Project steps are due at 11PM.
• Get the text and do the **reading** regularly and thoroughly *before class.*
• *Before* you take a quiz, **review, and study** the slides, your notes, and the reading carefully.
• *After* you take a quiz, go back to it and learn from your errors. Similar questions often appear in later quizzes.
• Do **lab assignments** *over several days***.** Start early so you can ask questions in class or during office hours.
• The project is intended for two students. Plan on finding and staying with a partner who can meet regularly. Complete the "**find a partner**" spreadsheet:
https://docs.google.com/spreadsheets/d/17Ui6q_W_8eRuRXqGjCHx5c7NOQKY4Y50hIg4PCCNKAU/edit?usp=sharing

Grading

You must be able to define important terms and explain ideas in clear, grammatical English. Course grades are based on lab assignments, an extensive project, quizzes, and thoughtful, well-prepared class participation. The project will include an original essay of at least 1000 words and an in-class presentation. **Grades will not be curved.** You must **complete all lab assignments to pass this course**.

Lab assignments	10%
Quizzes	35%
Project	50%
Class participation	5%

Late policy

Course work is designed to increase your understanding of the material and must be done as the course progresses. Often, a topic will build upon several previous ones, 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.
• **Quizzes** *No late quizzes will be accepted.* No exceptions, no discussion. We will drop your two lowest quiz scores.
• **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.
• **Lab extensions:** Life happens. If you contact Professor Ligorio by **email** with your request for an extension ***before*** the deadline, you may delay **one** lab assignment for one week without penalty. Save that "one" for emergencies.
• **Lateness:** Persistent lateness to class will eventually be counted as absence.

Course policies and students' responsibilities

• **Check that your CUNY First email address is up to date.** All communication about this course must be done through your @myhunter address, and it is also how Gradescope and Blackboard will recognize you. **If you are a transfer student**, Blackboard doesn't automatically update your email when you change campuses. To change your Blackboard mailing address, see https://writingcraft.commons.gc.cuny.edu/files/2020/07/ChangeEmailAddressSP13.pdf

• **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.

• **Attendance is required.** 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://compsci.hunter.cuny.edu/~csdir/. **To claim your existing account** log onto eniac and from your home directory issue the command: touch spring 2024. If you have difficulties with an existing account, contact cstechsp@hunter.cuny.edu. **If you do not have a CSCI Linux account,** contact Professor Epstein immediately.

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

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

• **Class participation** is strongly encouraged and is 5% 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.

• **Extra credit:** Extra credit will be offered only for projects at the top of the project leaderboard.

• **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 web surfing, email, games, videos, and social networking web sites (such as Facebook, Twitter, TikTok). 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.**

Anticipated course topics (subject to change)

- Ethical AI and machine learning
- Data and learning
- Neural and mathematical foundations
- Building blocks: functional units, matrices, linear algebra
- ANNs (artificial neural networks)
- DNNs (deep neural networks)
- CNNs (convolutional neural networks)
- Practical methodology
- ANNs and NLP (natural language processing)
- Attention and transformers
- Autoencoders
- GANs (generative adversarial networks) and diffusion

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 (or unrelated) to the class. We actually *enjoy* student visits during office hours. You can also make an appointment to see us at other times. We 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 1000 words.
• **Study groups:** 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 (see the partner meetup form) to make up and then solve practice problems, or to work on the ones at the ends of the chapters. Such practice is necessary, but not sufficient, preparation for quizzes.

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 write 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.
• If you consult the Internet (e.g., StackOverflow) for code, you must cite your sources, be able to explain it thoroughly, and be able to modify it. 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.
• **Plagiarism is unacceptable.** Read the plagiarism handout on our website to be certain you understand it.
• 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
• We strongly discourage the use of automated writing programs, including generative ANNs, to comment code or to write your report. This course is your opportunity to learn how to write and speak about deep learning. If you choose to ignore this, you must indicate clearly the generator, where you have used it, and what your prompt was. (And yes, we can tell.)
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.

• **We report all incidents of cheating or plagiarism** to the Office of Student Affairs. For the first incident your grade will be a 0. For the second incident, you will fail the class.

Policy on bullying

Bullying, cyber bullying, online hate, initialization intimidation, threats, harassment, and pressure to share schoolwork are all forms of violence, CUNY holds a zero-tolerance stance toward all such acts. The University is committed to prevention of any form of bullying, will respond promptly to threats and/or acts, and will protect victims of bullying from retaliation. As a criminal matter the New York Attorney General defines cyberbullying as the use of email, websites, instant messaging, chat rooms. text messaging, and digital cameras to antagonize and intimidate others. Disrupting a teleconferencing platform (such as Zoom, Skype, Blackboard Collaborate Ultra) is a federal crime.

ADA compliance 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 212-772-5478.

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 Laura Nelson (Lne@hunter.cuny.edu or 212-396-6662) and seek complimentary services through the Counseling and Wellness Services Office, Hunter East 1121. CUNY Policy on Sexual Misconduct Link: http://www.cuny.edu/wp-content/uploads/sites/4/page-assets/about/administration/offices/legal-affairs/policies-resources/Sexual-Misconduct.pdf

Thank you for abiding by these policies. They make current and future students experiences equitable and safe.

Personal protective equipment and other COVID matters

Any CUNY or Hunter rules will be enforced. Remember that these rules can change with little warning.

How to learn more

Students often ask for additional reference material. **Beware of web surfing.** Many sites and blogs are inaccurate or shallow or both. Here are some valuable and reliable alternatives:

Python

• https://colab.research.google.com/github/ageron/handson-ml3/blob/main/index.ipynb
• https://docs.python.org/3/tutorial/
• https://runestone.academy/runestone/books/published/thinkcspy/index.htm

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
• *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
• Additional resources appear on the class website.

How 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 online at https://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 Pieter 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.