Syllabus: Introduction to Machine Learning

Class Meetings for Spring 2023:

- In-class Sessions: Tue and Thurs 10:30-11:45 am EST

Course Websites:

- Piazza Discussion Forum: **https://piazza.com/class/lcwv1h9p2a11ai/ (https://piazza.com/class/lcwv1h9p2a11ai/) (https://canvas.tufts.edu/courses/20476/)**
- **(https://github.com/tufts-ml-courses/comp135-20f-assignments)** Schedule, Assignments, and Syllabus: **https://canvas.tufts.edu/courses/44718/pages/review-syllabus-and-schedule (https://canvas.tufts.edu/courses/44718/pages/review-syllabus-and-schedule) (https://www.cs.tufts.edu/comp/135/2020f/)**

Instructor: **Joseph Robinson, Ph.D. (https://www.jrobs-vision.com)** , Assistant Professor of Computer Science

- Contact: Please use Piazza. For extreme personal issues only: robinson(AT)cs.tufts.edu

Teaching Assistants (TAs):

- Ethan Harvey, Yinkai Wang, Si Liu
- For help, come to our **[Office Hours] (https://piazza.com/class/lcwv1h9p2a11ai/post/6)** or post on Piazza

---

Description and Objective: An investigation of programs that can dynamically adapt their behavior. The course focuses on two main general ideas: supervised learning and unsupervised learning. In supervised learning, a set of already-known correct responses to already-seen inputs is provided and is used to train the program to correct responses to new inputs. For instance, a facial recognition program could be supplied with several photographs labeled with the person's name in the image and learn to recognize new images of those persons. In unsupervised learning, a program seeks to find hitherto unknown patterns in data without pre-judgment about what those patterns might be. For example, a music recommendation program might try to find similarities between groups of songs so that when a user likes one such song, others in a similar group can be recommended. The course looks at various computational and mathematical models and techniques that can be applied to such problems.

Objectives for the Course

By the end of the semester, a successful student will be able to do all of the following things:

- Identify the differences between supervised, semi-supervised, and unsupervised learning techniques and how and when each can be employed.
- Identify machine learning problems (e.g., clustering, regression, etc.) that correspond to a real-world problem of interest.

- Translate data sets into properly formed input for a machine learning algorithm.
- Design and implement basic machine learning algorithms, using them to solve tasks of interest. This will involve coding algorithms from scratch and implementing standard algorithms found in modern professional machine-learning libraries.
- Compare and contrast various algorithms and variations on a single algorithm and the results they generate on a given data set.
- Compare and contrast methods for evaluating the success of a program engaged in some machine learning task.

Course Materials

1. **Textbook:** No textbook purchase is required. We will be using many online resources, accessible freely by browser and in PDF form:
   - *A Course in Machine Learning*. Hal Daumé III. http://ciml.info. [**link (http://ciml.info/)** ; **PDF (http://ciml.info/dl/v0_99/ciml-v0_99-all.pdf)**]
   - *Introduction to Statistical Learning*. Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani. Springer, 2013. Corrected 8th printing, 2017. [**PDF (https://static1.squarespace.com/static/5ff2adbe3fe4fe33db902812/t/6009dd9fa7bc363aa822d2c7/1611**]
   - *Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Trevor Hastie, Robert Tibshirani, and Jerome Friedman. 2nd Edition, Springer, 2009. Corrected 12th printing, 2017. [**link (https://web.stanford.edu/~hastie/ElemStatLearn/)** ; **PDF (https://web.stanford.edu/~hastie/ElemStatLearn/printings/ESLII_print12.pdf)**]
   - *Deep Learning*. Ian Goodfellow, Yoshua Bengio, and Aaron Courville. MIT Press, 2016. [**link (https://www.deeplearningbook.org)**]
   - Mastering Python for Data Science [**Dropbox (https://www.dropbox.com/s/s58e0vy6y7qcb0k/Mastering%20Python%20for%20Data%20Science.pdf?dl=0)**]
2. **Lecture notes:** When appropriate, these will be made available in the **notes section (https://canvas.tufts.edu/courses/40350/pages/lecture-notes)** of the class website.

Prerequisites and Expected Competencies

Programming: Students should be able to comfortably write substantial programs (i.e., at the level expected by the end of COMP 15 or an equivalent course). For some assignments, the choice of programming language is left open, as we will be most interested in the analysis of the results of those programs. For other assignments, we will use Python, a popular language for ML applications that is also beginner friendly. While no background in Python is expected, students should be prepared to learn the necessary aspects of the language as they go through the exercises for which it is required.

Mathematics: Comfort with mathematical formalisms is absolutely necessary to the understanding of some core ML algorithms and models. Basic familiarity with multivariate calculus (integrals, derivatives, vector derivatives) is expected. Prior experience with linear algebra and probability theory will also be helpful.

Requirements & Grading

Grades will be based on the following:

1. 35% Homework (4 assignments)
2. 30% Projects (2 assignments + 1 optional/extra-credit)
3. 25% Midterm examination
4. 10% Participation (in-class responses to PollEverywhere)

Letter Grades: COMP 135 uses the following breakdown of letter grades and percentages:

98–100% A+	87–89% B+	77–79% C+	67–69% D+
93–97% A	83–86% B	73–76% C	63–66% D
90–92% A–	80–82% B–	70–72% C–	60–62% D–

Homework

Homework will be assigned regularly in the course. The homework release and due dates are listed on the **schedule page**. In general, students will have about one week for an assignment; assignments are relatively small and lightweight and serve to practice techniques for use in larger projects. This short turn-around time will allow you to practice techniques in a relatively restricted setting before using them in a more open-ended fashion in projects. Homework will be submitted using the Gradescope system, information about which will be provided as necessary.

Assignments will consist of coding and written work, including the analysis of results. Work is expected to be presented professionally, with all work in typed-up form, with graphics and charts correctly rendered.

Regrade requests for all homework must be submitted within a week of the grades being released.

Projects

Projects are meant to be open-ended, simulate case studies in "the real world," and encourage creativity.

Each project will usually be due about 3 weeks after being handed out. Projects will generally center around a particular methodology and task and involve significant programming (with some combination of developing core methods from scratch and using existing libraries). You will be expected to consider some conceptual issues, write a program to solve the task and evaluate your program experimentally.

The main deliverable will be a short report, assessed based on effort, technical sophistication, clear explanation, supporting evidence, and overall performance. Note that it is the report that is key: an implementation that is highly effective on some ML task (in terms of classifier accuracy and the like) but which is presented with little attention to explanatory detail will receive little credit, while a detailed set of experiments that clearly explain why a particular approach did not solve the task well could receive more credit. When working on projects, students should remember that completing the coding part of the task is only really the first step; once the code is complete, time must be budgeted to prepare the presentation of the results. Code will be of less interest than analysis here.

Regrade requests for all projects must be submitted within a week of the grades being released.

Midterm Exam

The exam will be in written format, mid-semester. The exam will be timed and asynchronous; there will be some flexibility regarding the exact time you write the exam. This is a closed-book exam, but **students are allowed 1 page of 8.5x11" paper (front and back)**. Example exams will be distributed before each exam to show the format and type of question. No make-up exams will be given.

Final Essay-Exam

There will be no final exam.

Class Participation (Check-Points) & Attendance

Attending classes is essential. PollEveryWhere will be used to track attendance. This should be an easy 10% for everyone. Please let the instructor know if you plan to be absent and why.

Policy on Late and Missing Work

For late assignments, handed in **within 24 hours after** the time at which it was due, a reduction of 10% will occur; if handed in **within 48 hours** of the expected time, a reduction of 20% will occur; **within 72 hours,** the reduction will be 40%. Without a documented reason, no credit is given for assignments submitted after that point.

Students with extremely particular circumstances and only with prior approval by their academic dean must meet with the professor to make other arrangements for the scheduled homework and exams. The academic dean must initiate emails regarding the situation.

Policy on Collaboration

I encourage you to work together on the material. This is a great way to learn and share ideas. However, to learn something, you must complete the real work of programming and analysis unless specifically directed otherwise. It is perfectly fine for you to discuss the general approach to a problem, work out how to understand an algorithm or model, and help one another with things like getting the software we will

use to work correctly on your computer. However, it is **not okay** to copy code and other materials from anyone inside or outside of the class. While you can use online references to explain key concepts and learn programming techniques, you must not simply copy answers or code you find online, and you should cite any such materials you consulted. This is the only way actually to learn the material.

Note: Obviously, there are different standards for any course components involving group work. Specifically, your collaboration in such cases will necessitate that some of you do certain parts of the work and others do others. The general guidelines above still apply to members of different groups, however.

Piazza & Collaboration

When using the Piazza forum, the same considerations about collaboration are in play when posting questions and providing answers.

Questions may be posted as either **private** (viewable only by yourself and course staff) or **public** (additionally viewable by all students for the course registered on Piazza). Some issues warrant public questions and responses, such as misconceptions or clarifications about the instructions, conceptual questions, errors in documentation, etc. Some issues are better with private posts, including debugging questions that include extensive amounts of code, questions that reveal a portion of your solution, etc.

Please use your best judgment when selecting private vs. public. If in doubt, make it private.

Academic Misconduct

Students should read the Tufts handbook on academic integrity located on the **judicial affairs website (https://students.tufts.edu/student-affairs/student-life-policies/academic-integrity-policy)** . If a student does not understand these terms or any of the material listed on this page, it is his/her responsibility to talk to the professor. To be brief: do your own work, cite any sources from which you take ideas, and give credit where credit is due.

Inclusivity

Respect is demanded at all times throughout the course. Participation is required in the classroom, and everyone is expected to be treated with dignity and respect. We realize everyone comes from a different background with different experiences and abilities. Our knowledge will always be used to better everyone in the class. As an instructor, I have my specific background and perspective on life, along with my history of mental and physical challenges. I don't presume that my experience is the same as anyone else's, but I will always do my best to meet all my students where they live to the best that I can. I'll probably make some mistakes, but I will be trying.

Policy on Sharing

This course is designed for everyone to feel comfortable participating in discussions, asking questions, learning, and facilitating the learning of others. For that atmosphere to be maintained, the recordings of our conversations will only be shared with the enrolled students in the class (not posted publicly), and it is prohibited for any of us who have access to the video to share it outside the course. Additionally, some readings are provided on a fair-use basis while taking this course and are not to be distributed otherwise. This includes posting or sharing readings, videos, or other recordings on publicly accessible websites or forums. Any such sharing or posting could violate copyright law or law that protects the privacy of student educational records.

Writing Support

The StAAR Center for accessibility and academic resources offers friendly, experienced, non-judgmental **writing support (https://students.tufts.edu/staar-center/writing-support)** to writers at all levels of expertise through any stage in the writing process (for free!), and I highly recommend that you take advantage of this excellent opportunity.

Student Resources

Accommodations for Students with Disabilities: Tufts University values the diversity of our students, staff, and faculty and recognizes each student's important contribution to our unique community. Tufts is committed to providing equal access and support to all qualified students by providing reasonable accommodations so that each student may fully participate in the Tufts experience. If you have a disability requiring reasonable accommodations, please get in touch with the StAAR Center (formerly Student Accessibility Services) at **StaarCenter@tufts.edu (mailto:StaarCenter@tufts.edu)** or 617-627-4539 to make an appointment with an accessibility representative to determine appropriate accommodations. Please be aware that accommodations cannot be enacted retroactively, making timeliness a critical aspect of their provision.

Academic Support at the StAAR Center: The StAAR Center (formerly the Academic Resource Center and Student Accessibility Services) offers a variety of resources to all students (both undergraduate and graduate) in the Schools of Arts and Science, Engineering, the SMFA, and Fletcher; services are accessible to all enrolled students. Students may schedule to work on any writing-related project or assignment, attend subject tutoring in various disciplines, or meet with an academic coach to hone fundamental academic skills like time management or overcoming procrastination. Students can make an appointment for any of these services by visiting the **StAAR Center (https://students.tufts.edu/staar-center)** website, which can be accessed by the URL of <go.tufts.edu/StAARCenter>.

Mental Health Support: As a student, there may be times when personal stressors or emotional difficulties interfere with your academic performance or well-being. The Counseling and Mental Health Service (CMHS) provides confidential consultation, brief counseling, and urgent care at no cost for all Tufts undergraduates and graduate students who have paid the student health fee. To make an appointment, call 617-627-3360. Please visit the CMHS website: **http://go.tufts.edu/Counseling (http://go.tufts.edu/Counseling)** to learn more about their services and resources.

Course Summary:

Date	Details	Due
Thu Jan 19, 2023	Reading: Introduction (https://canvas.tufts.edu/courses/44718/assignments/313375)	due by 11:59pm
Tue Jan 31, 2023	Reading: Python's Numpy and Pandas (https://canvas.tufts.edu/courses/44718/assignments/313377)	due by 11:59pm