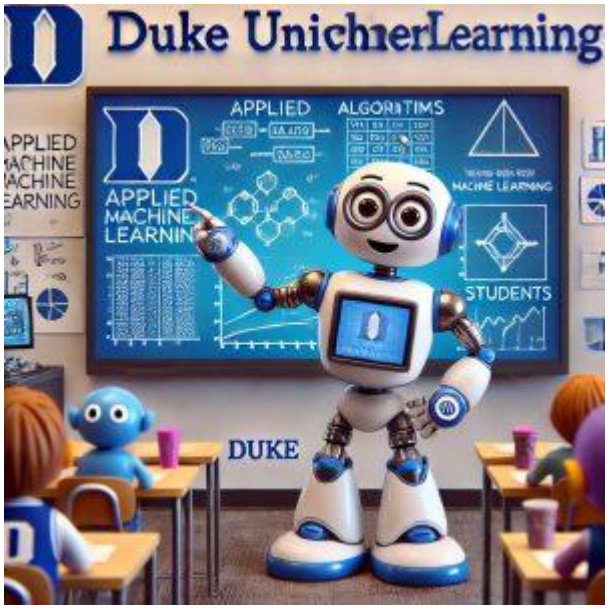
Overview

- Class Meeting. Mondays and Wednesdays 1:25 – 2:40 pm in Biological Sciences 111 (recordings on Canvas).



- Instructor. Brandon Fain.
- Graduate Teaching Assistants: Henry Bell, Yuzhe (Jason) Fu, and Liyang Zhu
- Undergraduate Teaching Assistants. Eric Lee, Hala Mohammed, Nicholas Maroulis, and Weston Slayton
- Course Platforms:
    - Canvas learning management system
    - Ed Discussion forum for questions (accessible from Canvas)
    - Gradescope submission and grading (accessible from Canvas)

Course Description

Machine Learning (ML) studies techniques to automatically learn patterns from data rather than explicitly programing a behavior. This course explores applications of machine learning in tabular data, computer vision, human language, and reinforcement learning, including recent generative artificial intelligence such as transformer language models and diffusion image generation models. Linear models and deep artificial neural networks of different architectures including multilayer perceptrons, convolutional neural networks, and transformers, will be utilized. Students will apply all techniques on real data using modern software.

Primary Course References

- BB: *Deep Learning Foundations and Concepts* by Christopher M. Bishop with Hugh B [BB online access link].
- JM: Speech and Language Processing (3rd edition) by Dan Jurafsky and James H. M [JM online access link]
- SB: Reinforcement Learning: An Introduction by Richard S. Sutton and Andrew G. Barto [SB online access link].

Topics

The course covers the following 5 major topics with a heavy emphasis on deep learning with neural networks. In addition, integrated treatment of societal and ethical issues in machine learning, open-source models, and practical programming.

1. *Introduction to Machine Learning.* Introduction to Python and NumPy programming, linear models including logistic regression, training, testing, validation, regularization.
2. *Artificial Neural Networks, Convolutions, and Image Recognition.* Introduction to deep learning with neural networks, PyTorch programming, multilayer perceptrons and convolutional neural networks, training with minibatch stochastic gradient descent, backpropagation, and automatic differentiation, image recognition and object detection, and transfer learning.
3. *Transformers and Language Models*. Conceptual intro to recurrent neural networks followed by detailed study of attention mechanisms and transformer architectures, encoder and decoder language models, generative large language models, HuggingFace, model alignment and fine-tuning, prompt engineering and in-context learning.
4. *Diffusion and Image Generation.* Continued study of transformers with contrastive learning and vision transformers. Diffusion process for image generation and multimodal vision-language models.
5. *Reinforcement Learning*. Introduction to value-based reinforcement learning in the tabular setting with monte carlo and dynamic programming solutions, deep value-based reinforcement learning (deep q-learning), gymnasium environment API, policy-based reinforcement learning with reinforce and actor-critic.

Background and Prerequisites

COMPSCI 201 Data Structures and Algorithms is a required and enforced prerequisite. Students are expected to have experience programming small to medium sized software projects, to be familiar with standard data structures such as arrays, lists, Strings, and maps, to be able to read code and documentation, and to be able to debug a program.

Students should also have mathematics background at least at the level of first semester Calculus. Students should be comfortable with algebra, summations, logarithms, exponential functions, derivatives, etc. No mathematical proof writing required. No prior machine learning experience required.

Experience with basic matrix/vector math, Python programming, and probability will all be very helpful but are not strictly required; expect to spend more time on assignments and reviewing material if these topics are new, beginning with the recommended background material below.

The following is recommended for all students to complete by the end of the first week of classes.

- Probability: Deep Learning Book Applied Math Basics 3.1-3.11
- Vectors and Matrices: Deep Learning Book Applied Math Basics 2.1-2.5
- Numerical Python: Stanford Python NumPy Tutorial (with Jupyter and Colab)

Assignments

References and Reading

You should expect to read to deepen your understanding of a given topic after lecture. You will get the most out of the course depending on what you put in, and the course references (below) are rich sources of information beyond the minimum that we can introduce during lecture. Specific suggestions for readings are marked on the schedule.

Homework Assignments

There will be regular weekly homework assignments, each as a set of Jupyter Notebooks including programming in Python and working with data to train, test, apply, and evaluate machine learning models. Homework assignments may also include test-style questions and integrated societal and ethical considerations with readings and short answer writing exercises. See the schedule for due dates.

Collaboration

You can work on homework assignments alone or in a group of two (that is, with a single partner). If working with a partner, you should not split up the assignment, but should actively collaborate (for example by pair-programming, by discussing shared approaches, by reviewing one another's code, etc.) Working with a partner is optional and we do not assign or match partners. You can switch partners between assignments, but once you begin working with a partner for a particular assignment you may not change for that same assignment.

Generative AI/LLM Usage

You may use generative AI (such as DukeGPT) to help you complete homework assignments but should conform to the following guidelines of responsible usage:

- Approved: Asking for conceptual clarification, help debugging and editing, suggesting strategies for implementation, giving examples of API usage.
- Not approved: Uploading entire assignment parts/tasks for AI completion, submitting code that was written by AI, submitting any AI generated text to short answer questions.

Submission

Homework assignments should be submitted on the course Gradescope, which will be accessible through a link on Canvas. If you work with a partner, your group should make a single submission on gradescope and ensure both members have been added using the group feature.

Exams

There will be at two written in-class exams during the semester. Exams will be 75 minutes long, closed book and closed notes. Exams will focus on conceptual understanding, critical evaluation, and practical implementations and will be based primarily on lecture and assignment materials. Questions may be any combination of multiple choice, fill in the blank, and short answer.

Final Project

At the end of the course, students will complete a final project of their choice using any of the techniques from the course. You might choose to build a web or mobile application powered by AI such, attempt a data science / machine learning competition or hackathon style task, such as the featured challenges regularly posted on Kaggle, or reproduce the results of a research study published in venues like NeurIPS or ICML.

You are strongly encouraged to use state-of-the-art machine learning models for building your final project. While some projects may necessitate building and training a model completely from scratch, it is much more likely that most projects will be best served by fine-tuning, adapting, integrating, or otherwise working with existing models as a starting point. Similarly, you are strongly encouraged to select a project that can be completed using only publicly available data – that is, that will not require you to collect your own data given the time constraints of the semester.

Final Project AI Usage

You may use generative AI (such as DukeGPT) to help develop your final project, even including the use of AI-generated code (unlike for homework assignments). However, you must attribute any AI-generated code (for example, by providing details about the authors and

generations in the comments) and you are fully responsible for the work you submit (that is, “but that’s what the AI said” would not be a valid defense if one were to lose credit due to errors or inconsistencies in the code).

Final Project Collaboration

You can work on the final project alone or in a group of two (that is, with a single partner). Partner projects will be expected to demonstrate more substantial results. For this reason, you should work with a partner if you want to work on a more ambitious project, not just to reduce the amount of work.

Project Workshops

Several class periods near the end of the semester are dedicated as “project workshop.” This is additional dedicated time set aside for you to work on your projects in-class with easy access to instructor and TA discussion, help, and support. Attendance is required unless you have already completed and submitted your final project.

Submission Requirements

The final project should be submitted as a project repository including the following components. You will submit a simple link to this repository on Gradescope.

- Software: This should be a publicly accessible Github or Gitlab repository with all project code, data, and documentation. The repository must follow the organizational structure outlined below to ensure clear presentation of your work and efficient grading.
- Project Demo: This is a 3–5-minute video that shows what your project does and why that matters. Think of this as the video you would show a non-specialist to pitch what you built. There is no reason to show any code in this video – you can use slides with visualizations or diagrams to provide motivation, show the running application, show experimental results, etc. This video should be included in your repository and linked prominently in your README file.

- Technical Walkthrough: This is a 5–10-minute video that shows and discusses how your project works. Think of this as the video you would show a fellow ML engineer to explain how you accomplished what you did. This video should help orient a grader to understand how your code works and where the machine learning concepts are being applied. It should also help a grader understand what was challenging about the project and where the significant technical contributions can be found. Like the project demo, this video should be stored in your repository and clearly linked in your README file.

Repository Organization Requirements

Your project repository must include the following structure and files to facilitate clear evaluation of your work:

- Required Files. The following markdown files should appear in the top level repository.
    - Your repository must contain a README.md file that serves as the main project overview and follows the template requirements specified below.
    - You must also include a SETUP.md file that provides clear installation and setup instructions for running your project. If your project uses external APIs or services, you should provide clear instructions for graders on how they can test your system.
    - Additionally, you must include an ATTRIBUTION.md file that contains detailed attribution of AI-generated code, external libraries, datasets, and any other resources used in your project.
- Directory Structure. Your repository should follow this organizational structure:
    - a src directory containing any of your source code that is not in Jupyter notebooks,
    - a data directory for any data files or data access scripts,
    - a models directory for any trained models, model configurations, or model loading scripts
    - a notebooks directory for Jupyter notebooks used for exploration or analysis,
    - a videos directory containing your demo and technical walkthrough videos,
    - a docs directory for any additional documentation,
    - and a txt file or environment.yml file for any dependency management.
- md File Requirements: Your README.md file must include these sections in the specified order. Where appropriate, you should include sample outputs or screenshots in your README file to demonstrate your project's functionality. Your README.md should be a concise reference and usage guide – your videos should present the primary explanation and discussion.
    - Project Title and short (1-3 sentence) description of what your project does,
    - a What it Does section that describes in one paragraph what your project does,
    - a Quick Start section that concisely explains how to run your project,

- a Video Links section with direct links to your demo and technical walkthrough videos,
- an Evaluation section that presents any quantitative results, accuracy metrics, or qualitative outcomes from testing,
- and an Individual Contributions section for group projects that describes who did what.

Suggestions, Timeline, and Grading

You are not expected to be working on the project during the first half of the semester, though you are encouraged to begin imagining what you might be interested in building — if there are particular data domains of interest, methodologies, competitions, applications, etc. that you might like to work on for personal interest or as part of curating your own professional portfolio.

More detailed suggestions on possible projects and the complete grading rubric will be released as a separate handout by midterm (after the first exam). At this point you should begin concretely planning for your project and discussing ideas with prospective partners, if you are considering working with a partner. You should begin looking for data sources, exploring relevant models and APIs, experimenting, etc. It is very important that you have some concrete ideas for what will do in your project by the time of the second exam.

After the second exam, rest of the semester is dedicated entirely to working on your project, and you should expect to spend a substantial amount of time in and out of class working on your project during the last weeks of the course.

Help

There are many resources for help, and we very much want to help! Please take advantage to make your course experience and personal learning as fruitful as possible.

Office Hours

Office hours are dedicated times set aside by the instructor and teaching assistants where you can ask questions about the course content. You can come to ask for help with assignments, reviewing material from class or before exams, etc. You are also welcome to consult with the instructor more broadly about machine learning, performance in the course, personal goals and challenges, etc.

Regular office hours will begin on Tuesday, September 2 (second week of the semester) and will end on Friday, December 5 (the last day of classes). Regular office hours will not be held during reading or exams periods, university holidays/breaks or on the day immediately before such breaks begin (e.g., the Friday before Fall break or the Tuesday before Thanksgiving break begins). Office hours also will not be held immediately following exams (e.g., if there is an exam on Wednesday and no homework due that week then there will not be office hours on Thursday and Friday).

You can find the updated calendar of office hours times and locations on (in-person and virtual options available) on the course Canvas site.

Edstem Asynchronous Question and Discussion

Ed Discussion is an online question/answer/discussion forum we will use in this class. You can access it through the link on the course Canvas page or directly at edstem.org. You should feel free to not only ask questions on Ed but answer questions (subject to the same guidelines as the general collaboration policy).

If you choose to use Ed, we ask that you follow the guidelines established for this course.

- You should never post your own code publicly for other students
- Be respectful of all other members of the course at all times
- Try things on your own and search for other similar questions before posting your own questions
- Be specific in your question and provide any relevant context

DukeGPT

Duke offers state of the art AI access through DukeGPT at no cost to students. You are welcome to use DukeGPT and other AI tools to help you learn, however you should do so responsibly and in compliance with the course policies regarding AI use as outlined in the Assignments and Policy sections.

Grades

Overall numerical course grades will be determined as the weighted average of the following three categories. Note that there is no final exam, only the final project.

Category	Grade Weight	Note
Homework Assignments	25	8 total, weighted equally
Exam 1	25	
Exam 2	25	
Final Project	25	No final exam, just project

Course letter grades will be assigned based on numerical course grades according to the following threshold function. There is no curve, meaning your grade does not depend on the grades of your peers. There is no rounding; a numerical course grade must be strictly greater than or equal to the given threshold to earn the given letter.

Letter Grade	Threshold	Letter Grade	Threshold
A+	97	C+	77
A	93	C	73
A-	90	C-	70
B+	87	D+	67
B	83	D	63
B-	80	D-	60

Homework will be graded for completeness, correctness, and clarity at the task level. The cover sheets of individual homework assignments will specify the number of points for that homework assignment broken down by part and task. Tests will be graded for correctness.

The final project will be evaluated for (i) completing the submission requirements and specifications, (ii) incorporating relevant methodologies, and (iii) demonstrating meaningful applications in machine learning. A separate and detailed item-level rubric will be provided for the final project as a handout at midterm after the first exam.

Policies

Student Disabilities Accommodations

Duke University is committed to providing equal access to students with documented disabilities. Students with disabilities may contact the Student Disability Access Office (SDAO) to ensure your access to this course and to the program. There you can engage in a confidential conversation about the process for requesting reasonable accommodations both in the classroom and in clinical settings. Students are encouraged to register with the SDAO as soon as they begin the program. Please note that accommodations are not provided retroactively.

More information can be found online at access.duke.edu or by contacting SDAO at 919-668-1267, SDAO@duke.edu.

Religious Accommodations

University policy permits students to be absent from class to observe a religious holiday. Accordingly, Trinity College of Arts & Sciences and the Pratt School of Engineering have established procedures for students to notify their instructors of an absence necessitated by the observance of a religious holiday. Please submit requests for religious accommodations at the

beginning of the semester so we can work to make suitable arrangements well ahead of time. You can find the policy and relevant notification form here:

- Trinity College: Religious Observance & Holidays
- Pratt School: Religious Observance & Holidays

Attendance, Absence, Illness, and Emergencies

For regular classes (not exam or project workshop days) in-person attendance is expected but absence is not penalized. We trust you to make your own best determination about your ability to attend class. If you miss a class, it is your responsibility to view the class recording at your earliest opportunity and to keep up with all other course materials (such as readings, assignments, etc.) There is no hybrid or zoom option.

On the project workshop and exam days, in-person attendance is required. Absence will be excused at the discretion of the instructor and in coordination with the student's academic dean, generally for one of the following reasons.

1. Academic and professional necessity (such as presenting work at a conference)
2. Duke varsity athletic competition (in which case you must complete the procedures documented here proactively, well in advance)
3. Illness or medical necessity (such as a contracting a communicable disease or a panic attack)
4. Bereavement or personal emergency (such as a death in the family)

Please see more details on university policy regarding short- and long-term medical issues, instances of personal distress or emergency, university sick leave policies, and academic support here:

- Trinity College
- Pratt School

The following (non-exhaustive) are not reasons for an excused absence and will generally not be approved by the instructor: (a) interviews that can be scheduled at different times, (b)

informational or networking sessions, (c) personal travel, (d) need to study or complete work for other classes, and (d) elective participation in Duke club activities other than representing Duke in varsity athletic competitions.

Duke Testing Center

This class will use the Testing Center (testingcenter.duke.edu/) to provide testing accommodations to undergraduates registered with and approved by the Student Disability Access Office (SDAO) and to administer make-up tests for students with excused absences. The center operates by appointment only, and appointments must be made at least four (4) consecutive days in advance. However, please schedule your appointments as far in advance as possible.

For those taking make-up tests for excused absences

First make sure to communicate with the instructor to request an excused absence. Depending on the circumstances, you may be asked to secure a dean's excuse (from your academic dean).

You can book a makeup test with the Duke Testing center using the center's website or by calling at (919)-684-1601 (calling is required if you need to complete a make-up test in less than four (4) days).

Makeup tests should be completed within one week of the original test date.

For those with SDAO testing accommodations

You will not be able to make an appointment until you have submitted a Semester Request with the SDAO and your accommodations have been approved. If you have not already done so, promptly submit a Semester Request to the SDAO to ensure you can make your appointment

on time. For instructions on how to register with SDAO, visit their website at https://access.duke.edu/students/ .

You can book accommodations testing sessions with the Duke Testing Center using the center's website. Barring excused absence, Accommodations testing sessions must be scheduled at the normal test date and time (possibly extended depending on the accommodations).

Recordings and Captions

This class will be recorded using the Panopto capture software. Recordings will be made available to current students through the course Canvas. Automatically generated captions are available in the recording.

The recording of lectures is permitted for private study only. You cannot distribute lecture recordings to anyone else and unauthorized distribution of course recordings is a cause for disciplinary action.

Late Work and Extensions

Late submissions will be accepted for up to 72 hours (3 days) from the original due date at a cumulative penalty of 3% per day. Submissions will not be accepted more than 72 hours after the deadline.

An extension may be granted in cases of incapacitation or a dean's excuse for emergencies or extenuating circumstances. Extensions will not, in general, be given for other personal circumstances including travel or extracurricular activities. Extensions are not guaranteed, and you should never wait to receive a response to an extension request before continuing work or submitting what you have so far.

If you need to request an extension, please do so with a private post on Ed discussion — this will be visible only to the teaching team. Please do not disclose sensitive private details — for example, if you are ill, simply note that you are ill and have submitted a STINF, you don't need to go into details.

Generative AI/LLM Usage

Detailed guidelines of responsible usage on homework assignments and the final project are discussed in those sections.

Collaboration

Collaboration is an encouraged part of the course, but only within boundaries that will ensure your learning and maintain academic integrity. The homework assignments and the final project can be completed in groups of size 2 (that is, you can work with a partner).

This is optional: you can by yourself if you like, and we do not assign mandatory partners. You can switch partners between assignments, but once you begin working with a partner for a particular assignment you may not change for that same assignment.

Electronic Device Use in Class

Class time has been reserved for our coming together and provides an opportunity to focus. More than attend, you are expected to be attentive and engaged during class. You are welcome to use electronic devices to take notes, consider code, or otherwise engage in materials related to the course. You should not use electronic devices for other purposes during class that are distracting to yourself and your peers, including but not limited to social media, instant messaging, or video streaming.

Regrade Requests

If you believe any of your work has been graded incorrectly or unfairly, you may submit a regrade request. Such requests will only be accepted through the Gradescope regrade request feature and usually only within a 72 hour window after grades are returned and an announcement is made. Such requests should politely and concisely explain why you believe your answer is substantially correct or was mis-evaluated according to the given rubric. Requests may only reference what is included in the original submission and may not contribute new or additional material.

If you submit a regrade request, a different grader than the original, either a graduate teaching assistant or the instructor, will review the request and regrade the submitted work. The new grade could be higher, the same, or lower than the original (while we are not looking for reasons to lower grades, it does happen at times that a regrade request identifies submissions that were incorrectly graded *too highly* originally). You should only submit a request if you sincerely believe a mistake was made in the original grading.

Academic integrity

The Duke Community Standard always applies as a guiding principle of integrity for the class.

Duke University is a community dedicated to scholarship, leadership, and service and to the principles of honesty, fairness, respect, and accountability. Citizens of this community commit to reflect upon and uphold these principles in all academic and non-academic endeavors, and to protect and promote a culture of integrity.

To uphold the Duke Community Standard:

- I will not lie, cheat, or steal in my academic endeavors
- I will conduct myself honorably in all my endeavors
- I will act if the Standard is compromised