CS 135
Introduction to Machine Learning

Prof. Michael C. Hughes ("Mike")

Fall 2023, First day of class

As you join, please check out and have open all class:
* **Website:** https://www.cs.tufts.edu/cs/135/2023f/
   Read syllabus, skim schedule, waitlist info, etc.

Many slides attributable to: Emily Fox (UW), Finale Doshi-Velez (Harvard), Erik Sudderth (UCI), and Marty Allen & Liping Liu (Tufts)

Q: How will this semester go?

How will we spend our semester?

If I want more?

Supervised Learning	10 weeks 5 homeworks 2.5 projects	CS 137 – Deep Neural Networks CS 136 – Stat. Pattern Recognition
Unsupervised Learning	2 weeks 0 homeworks 0.5 projects	CS 136 – Stat. Pattern Recognition CS 152 - Bayesian Deep Learning

Week-after-week, students will do the following

- Complete assigned readings to gain a first introduction to key concepts
- Attend in-class live sessions that summarize ideas and make connections
- Attend office hours to get questions answered
- Complete homeworks will build both conceptual and practical skills.
- Complete open-ended practical projects -- often organized like a contest -

Waitlist?

As of 2023-08-29 (a week before class starts), we have 125 students enrolled in the course. This represents the capacity of the assigned lecture hall as well as the max capacity of our assigned TA budget, so we **cannot add any more students**.

Thus, currently, the enrollment list is frozen. No additional students will be automatically enrolled.

That said, some students may drop the course and leave openings for others (usually we see 5-15 openings in the first week of classes as schedules shift).

To be considered for enrollment if a slot opens up, you must do these two things:

- Email the instructor by end of day Thu 9/7 via email with subject "CS 135 Wait List Request"
  - Explain your current state within your degree program (e.g. sophomore undergraduate in CS, Ph.D. student in Math)
  - Explain why taking the course *this* semester would be important to you.
  - Confirm that your hw0 will be been completed by Fri 9/8
- Complete and submit HW0 by end of day Fri 9/8 (**earlier than the deadline for already enrolled students**)
  - This action shows you have the necessary skills and would take the course seriously

Prof. Mike Hughes will make the final decision about all wait list candidates by **noon** on Mon 9/11.

Prereqs

Programming: Students should be comfortable with writing non-trivial programs (e.g., COMP 15 or equivalent). We will use Python, a popular language for ML applications that is also beginner friendly.

Please consult our Python Setup Instructions page to get setup a Python environment for CS 135.

By the first homework (HW0), students will be expected to do the following without much help:

- Load and transform datasets with numpy
- Perform vector mathematical operations in numpy (computing inner products, multiplying matrices, inverting matrices, etc.)
- Create line plots in matplotlib

Prereqs

	Needed	Not needed
Calculus	scalar derivatives vector derivatives	integrals
Linear Algebra	vectors & matrices matrix multiply matrix inverse	eigenvalues determinants
Prob / Stat	Random variable PMF / PDF	Hypo. testing

Tufts CS 135 - Fall 2023 – CC-BY – Prof. Mike Hughes

35

Units of Knowledge

Each one covers ~2 weeks of class

- Unit 1: Regression with linear and neighbor methods
- Unit 2: Classification with linear and neighbor methods
- Unit 3: Neural networks
- Unit 4: Trees and ensembles
- Unit 5: Kernel methods
- Unit 6: PCA and Recommendation Systems

Before each class: *on your own*
- readings from free online textbooks

Deep Learning
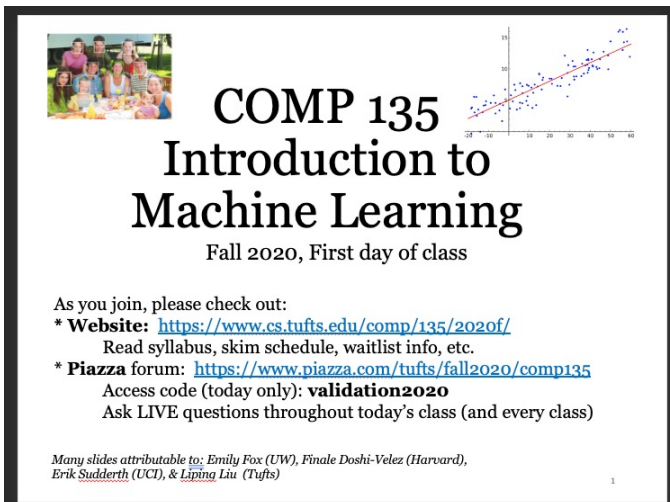
An MIT Press book

Ian Goodfellow and Yoshua Bengio and Aaron Courville

In Class

In class, we will typically have the following structure, all over Zoom:

- First 5 min.: Course Announcements

- Next 40 min.: Lecture for the day

- Next 25 min.: **Interactive** labs, or sometimes more lecture

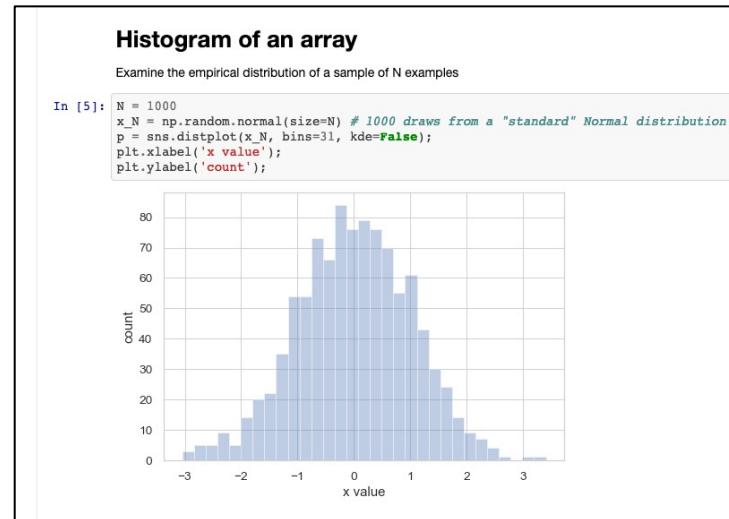- Last 5 min.: Recap of key concepts and lessons learned

Short slide deck: summary of key ideas and sample practice questions

Labs: Jupyter notebook for interactive exploration

What happens each unit?

		M	T	W	Th	F	S
Unit 1		14	class	16	class		
Unit 1		21	class	23	class		
Unit 2		28	class	30	class		
Unit 2		5	class	7	class		

Tufts CS 135 - Fall 2023 – CC-BY – Prof. Mike Hughes

39

Unit-Specific **Homework**

		M	T	W	Th	F	S
Unit 1	HW1 out	14 class		16 class			
Unit 1		21 class		23 class			
Unit 2		28		30	HW1 **due**		
Unit 2		5		7			

HW are individual work!

Due dates are posted on the website's schedule
PDF writeups and Python code will be turned in via Gradescope.
Code will be evaluated by an autograder on Gradescope
Report figures and short answers will be evaluated by TA graders

Homework Late Policy

Each student will have 192 total late hours (= 8 late days) to use throughout the semester across homeworks HW1-HW5. **No late hours are allowed on HW0.**

For each individual assignment, you can submit beyond the posted deadline at most 96 hours (4 days) and still receive full credit. Thus, for one assignment in the course due on Thu 11:59pm ET, you could submit by the following Mon at 11:59pm ET.

This late work deadline is key to our classroom goals. It allows us to always release homework solutions on Tue mornings and discuss the solution in class.

The timestamp recorded on Gradescope will be official. Late time is rounded up to the nearest hour. For example, if the assignment is due at 3pm and you turn it in at 3:05pm, you have used one whole hour.

Projects

Open-ended programming challenges, can do in small groups
3 projects all semester, each one ~4 weeks long
• Due dates are posted on the website's schedule
• Results and relevant code will be turned into Gradescope
• Polished PDF reports will be turned in via Gradescope

Leaderboard   Search 🔍

⬍ RANK	⬍ SUBMISSION NAME	▲ ERROR_RATE	⬍ AUROC
1	Ben	0.14833333333333332	0.9272055555555554
2	Darren	0.15666666666666662	0.9183333333333333
2	=.=	0.15666666666666662	0.9267666666666667
4	PC	0.16000000000000003	0.9214222222222221
5	Ki Ki	0.16166666666666663	0.9147666666666666
5	bellkor's pragmatic chaos	0.16166666666666663	0.9187555555555557

Project Late Policy

Projects turned in by the posted due date will be eligible for up to 100% of the points.

Projects turned in up to four days after the posted due date will be eligible for up to 85% of the points.

Students with unforeseen and exceptional circumstances may contact the instructor to make other arrangements. Without explicit instructor approval for an extension beyond four days, we may score your project zero total points.

Exams

- 1 midterm exam
- 1 final exam

Pencil and paper concept questions

Grading

Final grades will be computed based on a numerical score via the following weighted average:

- 23% average of homework scores (HW0 weighted 3%, HW1-HW5 weighted 5% each)
- 45% average of project scores (ProjA, ProjB, and ProjC, weighted equally)
- 15% midterm exam
- 15% final exam
- 2% participation in class, office hours, and in Piazza discussions

Workload

Here's our recommended break-down of how you'll spend time each week:

- 1.25 hr / wk preparation before Tue class (reading, lecture videos)
- 1.25 hr / wk active participation in Tue class
- 1.25 hr / wk preparation before Thu class (reading, lecture videos)
- 1.25 hr / wk active participation in Thu class
- 6.00 hr / wk on homework or project, whichever is due next

This totals to 11.00 hr / wk

Typically, by assignment

- for each HW you are given 2 weeks from release to due date. We expect about 8 hours are neede
- for each Project you are given 3+ weeks. We expect about 16 hours are needed from each team r

Collaboration Policy

You must write anything that will be turned in -- all code and all written solutions -- **on your own** without help from others. You may not share any code or solutions with others, regardless of if they are enrolled in the class or not.

We do encourage high-level interaction with your classmates. After you have spent at least 10 minutes thinking about the problem on your own, you may verbally discuss assignments with others in the class. You may work out solutions together on whiteboards, laptops, or other media, but you are not allowed to take away any written or electronic information from joint work sessions with others. No notes, no diagrams, and no code. Emails, text messages, and other forms of virtual communication also constitute "notes" and should not be used preparing solutions.

Use of external resources

When preparing your solutions, you may always consult textbooks, materials on the course website, or existing content on the web for general background knowledge. However, you cannot ask for answers through any question answering websites such as (but not limited to) Quora, StackOverflow, etc. If you see any material having the same problem and providing a solution, you cannot check or copy the solution provided. If general-purpose material was helpful to you, please cite it in your solution.

RE: AI assistive technologies such as ChatGPT:

- For homeworks, you cannot use any AI assistance at all.
- For open-ended projects, you may use such technologies to "automate the boring stuff" in terms of code development, but the high-level plan and vision for the project should be yours. You are expected to fully understand any code you use. You should write every word of your report yourself (no AI-assisted writing). Your report should disclose all steps that involved AI assistance.

Questions?

Tufts CS 135 - Fall 2023 – CC-BY – Prof. Mike Hughes

49

Let's Get Started!

- Try today's posted labs (see schedule page on website):
  - Jupyter Notebook: Intro to Numerical Python and NumPy
  - Jupyter Notebook: Intro to Dataframes and Plotting

Course Introduction

Concepts: supervised learning, unsupervised learning, difference between ML and AI

Date	Assignments	Do Before Class	Class Content	Optional Extras
Wed 09/09 day01	- HW0 out	Read course syllabus Videos on Canvas: - day01 - Types of Machine Learning	**Course Overview** - Slides: Why take this course? What will we learn? [PDF] - Slides: How will this semester go? [PDF] - Intro to Numerical Python and NumPy [notebook] - Intro to Dataframes and Plotting [notebook]	

HW0 due NEXT Mon (9/11), 11:59pm AoE
- Assesses if you have relevant programming skills
- Get started early!