Elements of Machine Learning

https://www.cs.duke.edu/courses/fall23/compsci371/

Introduction and Logistics
ML ≠ ChatGPT

- Large generative models (such as ChatGPT) rely on massive datasets with no annotation
  - ChatGPT has no notion of truth, just plausibility

- More specific discriminative applications with higher stakes require specific, annotated data:
  - Data security, fraud detection, smart cars, medical diagnosis, face recognition, marketing personalization, financial trading,…
  - Use deep learning when there is enough data
  - Need to train simpler predictors when data is scarce

- We’ll cover basics of deep learning but also other predictor types
- Concepts before techniques
Machine Learning in One Slide

• Identify a function \( y = f(x) \):
  \[ x = \text{smartphone face picture}, \quad y = \text{unlock or not} \]

• Give lots of examples (a training set):
  \[ T = \{ (x_1, y_1), \ldots, (x_N, y_N) \} \]

• A learner is another function \( \lambda \):
  It takes \( T \) as input and outputs an approximation to \( f \):
  \[ h = \lambda(T) \]

• Hopefully, \( f \) and \( h \) behave about the same even for previously unseen data:
  \[ h(x) \approx f(x) \]

• That’s the big problem!

• ML is not (just) data fitting
Logistics
Academic Integrity

• *Short version: Cheating will be prosecuted*

• Cheating: Using someone else’s material in your work without giving credit [Lone exception: class materials need not be cited]

• Ditto for making materials available to others

• Giver/receiver are treated the same

• Format for using/making available is immaterial

• Only communication allowed during homework is with your group peers, if any, and with the teaching staff
Notes, Slides, and Videos

• *Notes on the class Syllabus web page are required reading, and are your main source of information*

• *All appendices in the notes are optional reading*

• *Prerecorded videos are for backup, and are optional*

• *In-class lectures are recorded on Panopto, accessible through Sakai*

• *Feel free to integrate with other sources. See Resources web page*
Homework

- Due **most Thursdays by 8:30am** [HW0 due on September 7]
- OK to work in groups of one, two, three [No division of labor!]
- Jupyter notebooks → HTML → PDF [Keep Jupyter cells small]
- Two submissions on Gradescope: PDF, Notebook
- Log in to Gradescope through Sakai
- **One pair of submissions (two files, one PDF, one Notebook) per group. Remember to list all names in the files and through the Gradescope interface!**
- No late homework accepted
- Two worst homework scores (including 0s for no homework) dropped
Exams and Grades

- Exams:
  - Midterm on October 26, in person, during class period
  - Final on December 16 at 9am, 2 hours, in person, not cumulative

- Grades:
  - Homework 60%
  - Midterm 20%
  - Final 20%
Teaching Staff

• **Graduate TAs**: Leslie Dees, Zhehan Qu

• **Undergraduate TAs**: Aakash Kothapally, Angikar Ghosal, Brian Lee, Choonghwan Lee, Dennis Tang, Ishaan Maitra, Luke Triplett, Naomi Gao, Nguyen Nguyen, Soumya Bodavula, Tonya Hu, Jason Shang

• If you like this course, please volunteer to TA next year!

• Each of us holds two office hours per week

• **Check the online calendar before attending office hours**

• We’ll keep listening to Ed Discussion (at reasonable hours)

• **Talk to us!** We are here to help you learn