About COMS 4771

- Basic principles and methods of supervised machine learning
  1. Appetizer: decision trees (a “non-parametric” method)
  2. Statistical model for prediction
  3. Linear models and inductive bias
  4. Optimization methods
     - Convex optimization, neural networks
  5. Maybe one other topic if time permits . . .

- This is not a course about how to use sklearn, tensorflow, etc.
- Also not about latest nonsense on arXiv
- Stuff beyond COMS 4771:
  - COMS 4252, 4773: Mathematical theory of learning
  - COMS 4774: Unsupervised learning
  - COMS 4775: Causal inference
  - . . .
Key statistical/algorithmic ideas in ML

- Plug-in principle
- Inductive bias
- Linearity
- Mathematical optimization
About me

- Professor Daniel Hsu
  - Okay to call me “Daniel”!
  - “Professor Hsu” also okay
    - But “Professor Daniel” is a little weird
  - Pronouns: he/him
  - At Columbia since 2013
  - Previously at Microsoft Research, Rutgers, UPenn, UC San Diego, UC Berkeley, ...
  - Research interests: algorithms, statistics, & combining the two
  - Good at: \LaTeX\ hacking
  - Bad at: slides
About you

- I assume you have fluency in
  - multivariable calculus,
  - linear algebra, and
  - elementary probability (no measure theory needed)

- I also assume you can read and write programs in Python
  - (and read online documentation to learn, e.g., how to do I/O with CSV files)
  - See Courseworks for a “Python basics” Jupyter notebook to brush up on Python, Numpy, etc.

- See “Calibration Homework” available on course website
- Let me know why you are interested in ML!
  - Part of HW 1.
Administrative stuff

- Website: https://www.cs.columbia.edu/~djhsu/ML
  - Schedule for office hours/lectures/homework/quizzes/exam
  - Syllabus

- Office hours: Thursdays 2:30-4:30pm, in CS Courtyard
  - Except if raining, then online

- Course assistants (CAs):
  - Abhinava, Arshiya, Maxime
  - Links for online office hours will be posted on Courseworks

- Technology:
  - Ed: communicate with course staff (replacement for Piazza)
  - Courseworks: retrieve assignments, data files, etc.
  - Gradescope: submit homework write-ups, code, quizzes

- Disability services:
  - Please make arrangements with disability services ASAP
Academic rules of conduct

▶ See syllabus
▶ **Cheating**: don’t do it
  ▶ If unsure about something, ask me ASAP
  ▶ Consequence is automatic fail
▶ **Cheating out of desperation** is also cheating
  ▶ Instead: get help early
  ▶ We are here to help
▶ Okay to work on homework in groups of \( \leq 2 \)
  ▶ No collaboration across groups
  ▶ No diffusion of responsibility
  ▶ No diffusion of learning
  ▶ *I personally think you will learn more if you solve homework problems by yourself*
▶ No collaboration at all on quizzes or exams
  ▶ Any collaboration or unauthorized assistance ⇒ automatic fail
Reading assignments

- There are some required reading assignments (mostly from handouts posted on website)
- Unfortunately, most textbooks on ML are not appropriate for this course
  - Closest is “A Course in Machine Learning” by Daumé
  - I have selected some optional reading assignments from a few books that may be used to supplement the lectures
  - All books available online