

COMS 4721 (Machine Learning)

Prof. Daniel Hsu

About COMS 4721 (=COMS 4771)

- ▶ Basic principles and methods of supervised machine learning
 1. Appetizer: decision trees (a “non-parametric” method)
 2. Statistical models for prediction
 3. Linear models and inductive bias
 4. Optimization approaches to machine learning
 5. Nonlinear models
- ▶ This is not a course about how to use sklearn, tensorflow, etc.
- ▶ Also not about latest nonsense on arXiv
- ▶ ML beyond COMS 4721:
 - ▶ COMS 4252, 4773: Mathematical theory of learning
 - ▶ COMS 4774: Unsupervised learning
 - ▶ COMS 4775: Causal inference
 - ▶ COMS 4995, 6998: Advanced topics

Key ideas in ML

- ▶ Plug-in principle
- ▶ Inductive bias
- ▶ Linearity
- ▶ Mathematical optimization

About me

- ▶ Professor Daniel Hsu
 - ▶ Informal: “Daniel”
 - ▶ Formal: “Professor Hsu”
 - ▶ Pronouns: he/him
 - ▶ Columbia faculty since 2013
 - ▶ Research interests: algorithms, statistics, & combining the two
 - ▶ Good at: \LaTeX hacking
 - ▶ Bad at: slides

About you

- ▶ You have fluency in
 - ▶ multivariable calculus,
 - ▶ linear algebra, and
 - ▶ elementary probability (no measure theory needed)
- ▶ 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 “HW0” available on course website
 - ▶ This is **required**, due in < 1 week

Administrative stuff

- ▶ Website: <https://www.cs.columbia.edu/~djhsu/ML>
 - ▶ Schedule for office hours/lectures/homework/quizzes/exam
 - ▶ Syllabus
- ▶ Course assistants (CAs):
 - ▶ Katie, Han, ...
- ▶ Office hours: TBD, see calendar on website
- ▶ 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 (**except HW0**) in groups of ≤ 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 \Rightarrow automatic fail

- ▶ Required
 - ▶ Get it from Courseworks
 - ▶ Review of prerequisite material
 - ▶ Due January 24, 11:59 PM