Welcome to

COMS 4774 Spring 2021
Today

► About COMS 4774
► Lecture 1: probability review
Lectures are being recorded and will be available on Courseworks.

Please, by default, keep your microphone muted.

If you have a question:
  - Type the question into the chat; or
  - Type “I have a question about ...” (fill-in the blank) into the chat, and I will call on you at a suitable time to un-mute and ask verbally.

Camera on if possible, but not required!
About COMS 4774: nuts and bolts

▶ COMS 4774 “Unsupervised learning”
  ▶ Perhaps: “Beyond Supervised Learning (COMS 4771)”
  ▶ But with a focus on topics that some people have called “unsupervised”
▶ Course website + syllabus: https://www.cs.columbia.edu/~djhsu/UL
  ▶ Read it today
▶ Gradescope
  ▶ Will sync Gradescope with Courseworks roster
  ▶ Account linked to email address listed on Courseworks; use this account
  ▶ If you have another account already, merge it
▶ Slack workspace for the class
  ▶ Will invite registered participants shortly
▶ Piazza (???)
  ▶ Are they showing you ads? Selling your data?
  ▶ I am soliciting suggestions...
▶ Office hours:
  ▶ Daniel Hsu (me): Tuesdays, 2:35pm–4:35pm
  ▶ Chris Alberti (TA): Fridays, 10am–noon
  ▶ Zoom links will be posted on Courseworks
About COMS 4774: cast of characters

About me

- Prof. Daniel Hsu
- At Columbia since 2013
- Before: Microsoft Research, Rutgers Univ, Univ of Penn, UC San Diego, UC Berkeley
- Been thinking about “machine learning” for a while...

About you

- You have fluency in
  - Multivariable calculus, linear algebra, elementary probability
  - Enough discrete math to know about graphs (vertices and edges)
  - Enough algorithms/complexity to know about Big-O notation and poly vs exp
- You mathematical maturity to
  - write mathematics in complete sentences and paragraph form
  - state and prove theorems
  - (see pointers on course website)
- If any questions about prereqs, please email me
- Tell me more: fill out student survey (link on course website)
About COMS 4774: a play in three acts

1. High-dimensional data
   ▶ probability in high dimensions
   ▶ random linear maps
   ▶ high dimensional Gaussian populations
   ▶ effects of random projections
   ▶ subspace embeddings

2. Low-rank approximations
   ▶ singular value decomposition
   ▶ applications to mixture models
   ▶ sums of random matrices
   ▶ planted partition models
   ▶ spectral graph theory
   ▶ semi-supervised learning

3. Higher-order interactions
   ▶ model identifiability from higher-order moments
   ▶ multivariate moment tensors
   ▶ tensor decompositions
Flavor

Example: Why PCA?

▶ What do the singular values/singular vectors of data matrix tell us?

\[ A := \begin{bmatrix}
    \langle x_1^T \rightarrow \\
    \langle x_2^T \rightarrow \\
    \vdots \\
    \langle x_n^T \rightarrow 
\end{bmatrix} \in \mathbb{R}^{n \times d} \]

▶ COMS 4771 answer: something about regularization, inductive bias in regression, etc.

▶ Or, something about capturing variance in data, but without reference to a concrete purpose for doing so

▶ Suppose data are iid draws from a mixture of \( k \) Gaussian subpopulations

▶ Rank \( k \) PCA projection of the data increases the separation between subpopulations

▶ Suppose \( A \) is adjacency matrix of social network with \( k \) “close knit” communities

▶ Top \( k \) singular vectors “reveal” the community structure

Caveats: (1) gap between theory & practice, (2) data models are unrealistic
Pay-off: clarity & precision
Flavor

Example: Why PCA?

- What do the singular values/singular vectors of data matrix tell us?

\[
A := \begin{bmatrix}
  x_1^T & \rightarrow \\
  x_2^T & \rightarrow \\
  \vdots & \\
  x_n^T & \rightarrow 
\end{bmatrix} \in \mathbb{R}^{n \times d}
\]

- COMS 4771 answer: something about regularization, inductive bias in regression, etc.
- Or, something about capturing variance in data, but without reference to a concrete purpose for doing so
- Suppose data are iid draws from a *mixture of k Gaussian subpopulations*
  - Rank k PCA projection of the data increases the separation between subpopulations
- Suppose \( A \) is adjacency matrix of *social network with k “close knit” communities*
  - Top k singular vectors “reveal” the community structure

Caveats: (1) gap between theory & practice, (2) data models are unrealistic
Flavor

Example: Why PCA?

▶ What do the singular values/singular vectors of data matrix tell us?

\[
A := \begin{bmatrix}
\leftarrow x_1^T & \rightarrow \\
\leftarrow x_2^T & \rightarrow \\
\vdots & \\
\leftarrow x_n^T & \rightarrow 
\end{bmatrix} \in \mathbb{R}^{n \times d}
\]

▶ COMS 4771 answer: something about regularization, inductive bias in regression, etc.
▶ Or, something about capturing variance in data, but without reference to a concrete purpose for doing so
▶ Suppose data are iid draws from a *mixture of k Gaussian subpopulations*
  ▶ Rank k PCA projection of the data increases the separation between subpopulations
▶ Suppose \( A \) is adjacency matrix of *social network with k “close knit” communities*
  ▶ Top k singular vectors “reveal” the community structure

Caveats: (1) gap between theory & practice, (2) data models are unrealistic
Pay-off: clarity & precision
Focus

We will focus on:

- Theoretical analysis of methods for unsupervised learning
  - Consider statistical models of data
  - State and prove mathematical theorems
- Also mathematical tools that are useful for the above
  - Probability and (multi)linear algebra
  - Example:
    - Let $X_1, \ldots, X_n$ be iid random $d \times d$ matrices
    - What can be said about the singular values of $S := \sum_{i=1}^{n} X_i$?
Focus

We will focus on:

▶ Theoretical analysis of methods for unsupervised learning
  ▶ Consider statistical models of data
  ▶ State and prove mathematical theorems
▶ Also mathematical tools that are useful for the above
  ▶ Probability and (multi)linear algebra
  ▶ Example:
    ▶ Let $X_1, \ldots, X_n$ be iid random $d \times d$ matrices
    ▶ What can be said about singular values of $S := \sum_{i=1}^{n} X_i$?

We will not focus on:

▶ numpy, pandas, pytorch, scikit-learn, . . .
▶ Julia, MATLAB, R, . . .
Focus

We will focus on:

- Theoretical analysis of methods for unsupervised learning
  - Consider statistical models of data
  - State and prove mathematical theorems
- Also mathematical tools that are useful for the above
  - Probability and (multi)linear algebra
  - Example:
    - Let $X_1, \ldots, X_n$ be iid random $d \times d$ matrices
    - What can be said about the singular values of $S := \sum_{i=1}^n X_i$?

We will not focus on:

- numpy, pandas, pytorch, scikit-learn, ...
- Julia, MATLAB, R, ...

Nevertheless...

Very useful to learn how to “do numerical linear algebra” (e.g., vector arithmetic, matrix-vector multiply) in your favorite computing environment.
Getting a grade

- Problem sets (≈3 of them, not including “HW0”): 35%
  - Can be done individually or in pairs
- Final project: 35%
  - Read and understand a substantial research paper on machine learning
    - Write a review
    - Add something new (e.g., examples, corollaries, empirical studies)
  - Can be done individually or in pairs
  - Instructions on website
- Class participation: 30%
  - Write scribe notes
  - Edit scribe notes
  - We’ll start on Thursday
  - Instructions on course website
- Academic rules of conduct
  - Don’t cheat. Don’t plagiarize.
  - Do ask questions, and let us know if difficulties arise!
Lecture logistics

- For lectures (after this part), I’m planning to use tablet software called “Write”  
  - [http://www.styluslabs.com](http://www.styluslabs.com)
  - I think it is free for Android, iOS (beta version), Linux, MacOS, Windows
  - $5 for non-beta iOS version (???)

- If you have “Write”, you can connect to shared whiteboard  
  - 2. Remind me to setup the shared whiteboard and share the whiteboard ID
  - 3. Connect to the shared whiteboard
  - 4. Now you can scroll up and down the whiteboard

- I’ll eventually post the whiteboard pdf after each lecture to Courseworks  
  - May be some delay…
  - Not a substitute for taking your own notes
Homework 0

- Required
  - Problem 1: Read the syllabus
  - Problem 2: Fill-out the student survey (link on webpage)
  - Problem 3: Introduce yourself on Piazza (see survey)
Questions?