Applied Machine Learning!!!

W207 Section 9
Rasika Bhalerao
rasikabh@berkeley.edu

Format of this class

- Weekly **synchronous** sessions (Tuesdays 4pm Pacific)

- Weekly **async**hronous material on Canvas
  - Complete before corresponding synchronous class
  - Bring any questions!

- 9 individual **homework** assignments
  - Due Sundays

- **Final group project**

- Weekly **office hours**
  - Tuesdays 6pm - 7pm, open to all W207 sections

Where?

- Canvas "Calendar" section:
  - Synchronous sessions (Tuesdays 4pm Pacific)
  - Office hours

- Slack
  - #datasci-207
  - #w207-9-fall2022

- Github
  - Homework assignments:
    - https://github.com/MIDS-W207/coursework_2022/tree/main/Homework
  - Notes based on async modules:
    - https://github.com/MIDS-W207/coursework_2022/tree/main/Notes
  - Weekly synchronous material:
    - https://github.com/MIDS-W207/coursework_2022/tree/main/Demos
    - https://github.com/MIDS-W207/rasikabh/tree/main/live_sessions

- Contact me on Slack or at rasikabh@berkeley.edu

Introductions!

- Your name (and pronouns)

- What brings you here?

- Name one way machine learning affects your life!

Did everyone fill out the intro form?

https://forms.gle/UpzimpEUfMeFGFuE8

Schedule

Supervised learning methods

	Sync	Topic
2	Aug 30	Linear Regression / Gradient Descent
3	Sep 6	Feature Engineering
4	Sep 13	Logistic Regression
5	Sep 20	Multiclass classification / Eval Metrics
6	Sep 27	Neural Networks
7	Oct 4	KNN, Decision Trees, Ensembles

Unsupervised learning methods

	Sync	Topic
8	Oct 11	KMeans and PCA
9	Oct 18	Text Embeddings
10	Oct 25	CNNs
11	Nov 1	EDA, Real data, Baselines
12	Nov 15	Fairness / Ethics
13	Nov 29	Fancy Neural Networks
14	Dec 6	Final Presentations

Assignment Schedule

Due Date	Assignment
Aug 28	HW1
Sep 4	HW2
Sep 11	HW3
Sep 18	HW4
Sep 25	HW5
Oct 2	HW6
Oct 16	Group project baseline
Oct 23	HW8
Nov 6	HW9
Nov 20	HW10
Dec 4	Final project notebook + presentation

Homework assignments

- Find them here on Github:
  https://github.com/MIDS-W207/coursework_2022/tree/main/Homework
  - You should get access if you fill out the intro form (ask if not)
- You can do them in any IPython environment
  - Recommended: https://colab.research.google.com/
- Submit homework on Canvas / Gradescope
  - Either upload your .ipynb file or submit a link to a Github repo for that assignment
  - If there is an issue of environment, upload the version that runs in default Google Colab
- Lowest 2 assignment grades will be dropped
  - Final homework grade will be the average of best 7 scores out of 9
  - 5 "late days" this semester

Grading

- Instructors want all students to get an A

- We care more about actual learning

- Approximate scale:

A	4.0	94% - 100%
A-	3.7	90% - 93.9%
B+	3.3	86% - 89.9%
B	3.0	83% - 85.9%
B-	2.7	80% - 82.9%

C+	2.3	76% - 79.9%
C	2.0	73% - 75.9%
C-	1.7	70% - 72.9%
D+	1.3	66% - 69.9%
D	1.0	63% - 65.9%

D-	0.7	60% - 62.9%
F	0	< 60%

Final group project

- Each group will pick a **Kaggle competition**

  - One of the 4 default Kaggle projects or choose your own (run it by me just in case)

- In addition to doing the Kaggle **competition task**:

  - Thoroughly **explore the data**

  - Apply the learned **thinking/methods/algorithms**

  - **Analyze and improve** methods

Final group project deliverables

- **Baseline progress update**: .ipynb file to Canvas by Oct 16
  - An end-to-end solution with:
    - Exploratory data analysis
    - Metric of evaluation and why you chose it
    - Split the data into train/dev/test
    - A simple machine learning technique
    - Evaluate results
  - Explain how you will evaluate any challenges
  - Briefly describe what you still plan to do

- **Final project**: .ipynb file to Canvas by last day (Dec 4)

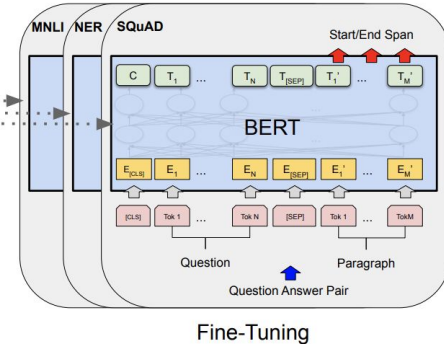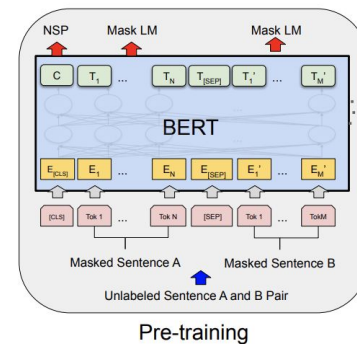- Final **group presentation** on last day

Project groups

- Groups of 2 or 3

- Finding groups
  - Find groups and project topics by week 4 by signing up on this doc:
    https://docs.google.com/document/d/1R3J_X1Rz6WP8eMQ2cyMC0wAr5iQdhMK_httdoNO6L0w/edit?usp=sharing
  - This class has a large variety of backgrounds - use it to your advantage!
  - Please tell me early on if you are having issues

- One person from each group should submit the baseline/final notebooks to Canvas

- Each group member must speak in the final presentation

Behavior expectations

- Healthy disagreement is expected
- Be mindful of one another's schedules
- Be a good listener
- Have fun in a professional manner
- Share related real-world experience
- Ask questions when something is confusing
- Keep it 100 but be respectful
- Be open-minded to new ideas in the real world and when coding
- On time for group meetings

https://www.blog.google/products/google-cloud/cloud-covered-what-was-new-with-google-cloud-in-march/

Pre-training

Fine-Tuning

https://arxiv.org/pdf/1810.04805.pdf

What is AI? What is ML?

https://fortune.com/2016/12/01/facebook-artificial-intelligence-news/

https://blog.dormakaba.com/what-is-facial-recognition-and-how-does-it-work/

https://www.theatlantic.com/technology/archive/2012/02/on-this-day-garry-kasparov-faces-off-with-deep-blue/253230

What are these things?!!?!? 😅

- Machine learning
- Artificial intelligence
- Deep learning
- Data mining
- Statistics
- Natural language processing
- Computer vision
- Big data