389I: Healthcare & AI

ter.ps/389iweek1

Topics

- Syllabus Week + Intro to ML Algorithms
- Machine Learning for Healthcare
- Early Disease Detection
- Electronic Health Records (EHRs)
- Drug Discovery
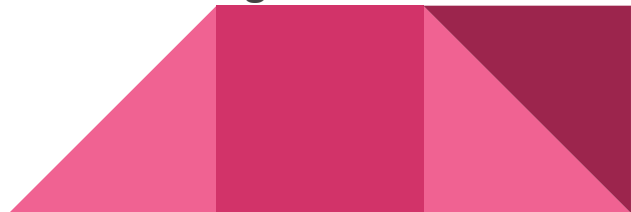- Telemedicine
- Mental Health

Learning Outcomes

- Understand the central role of data in improving healthcare

- Understand healthcare data-specific ethics and challenges

- Be able to apply basic ML techniques to analyze healthcare datasets

Assignments

- Lectures
  - Overview of relevant AI algorithms & domain specific case studies
- Quizzes
  - In class on lecture content or assigned readings
- Codelabs
  - Guided projects on analyzing specific datasets to achieve set results
- Final project
  - Chose a dataset in a relevant domain to analyze & present findings to class

Codelabs

- We will provide datasets in a specific domain to analyze

- Given starter code with directions in Jupyter notebook

- Apply ML techniques & write analysis

- 2 this semester + final project

Instructors

- **Max Leiserson**
  - mdml@cs.umd.edu
  - website
- **Sanna Madan**
  - smadan12@umd.edu
- **Kyle Liu**
  - kliu1234@umd.edu
- **Office Hours**
  - TBD / by appointment
- **Piazza**

Machine Learning

What is machine learning?

- "Labeling machine." Teaching a computer how to **classify**.
  - complex, hard-to-see patterns

- More specifically:

  training of a model from data

  that generalizes a decision

  against a performance measure.

What is "artificial intelligence"?

- Much broader than machine learning

- Computers mimicking the cognitive functions of humans
  - intelligence
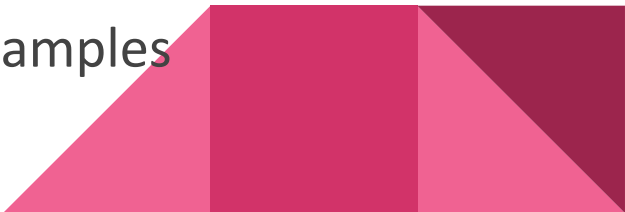
- We're not quite there yet

Machine learning

- Focuses on learning from datasets provided to the computer

- Identify patterns and make decisions from these datasets

- Self driving cars, Netflix recommendations, fraud detection, etc.
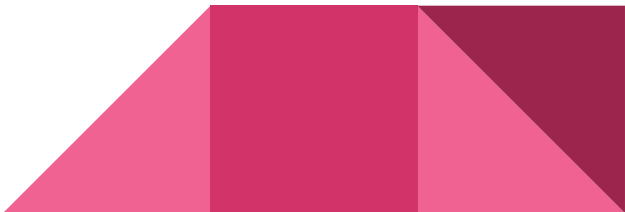
Deep learning

- Even *deeper* - a subset of machine learning

- Multi-layer machine learning model with hidden layers

- Blackbox
  - is this ideal for healthcare?

- Needs much larger datasets, often millions of samples

In this course, we are focusing on **machine learning**.
Not deep learning, not AI.

Why?

Supervised learning

- Most practical machine learning is **supervised**

- Input variables and output variables are given: $Y = F(x)$

- Learns mapping function from input $\rightarrow$ output

- Labels are **known** beforehand

- Classification, regression

Unsupervised learning

- Only input data, with no corresponding output labels: ? = F(x)

- No correct answers to learn from

- Algorithms find their own "structures" in the data instead

- Clustering, association

Data

- Machine learning doesn't matter if the data is of poor quality!!!!!!!!

- If the data is flawed, then the machine learning will be flawed

- Cleansing data may even take up most of the "machine learning" process

- To trust the insights of the data, you must first trust the data

Quality datasets are therefore the first step to ML.

Data cont'd

- Amount of data is also crucial

- More is more

- Large datasets in healthcare/bio once rare, now up and coming

Data cont'd

- Bias in data must also be considered

- e.g., people who get their DNA sequenced tend to be wealthier

- Machine may be getting a skewed perception of reality

Summary

- Machine learning is only a subset of artificial intelligence

- There are many different approaches to machine learning

- Quality, quantity, and bias in data are crucial considerations

Links

- Github
- Survey
- Slides: ter.ps/389iweek1
- Tips for ML