Introduction

CSCI-P556 Applied Machine Learning
Lecture 1

D.S. Williamson

Agenda & Outline

Today's Topics

- What is this class? What this class is NOT?
- Who am I? What I do?
- Who are you?
- Syllabus and logistics
- Why (Applied) Machine Learning?

Zoom

Virtual and Synchronous Interactions

- We will be conducing P556 via Zoom (lectures and office hours)
- Not easy, nor the preferred means of education — but it is currently the safest way for all of us
- Please read about the zoom comportment in the syllabus

Zoom Etiquette

General

- A short list is below:
  - Keep microphone muted, except during Q&A
  - Keep video on, or add profile picture
  - Be respectful and professional at all times (appearance, comments, environment)
  - Use full names (first and last)
  - Avoid multi tasking
  - One person talks at a time
  - Share the camera

- All lectures are recorded and posted
  - Be mindful of this
  - Do not share these with others.

Zoom Etiquette

Asking Questions: Interaction is key! Please ask and answer questions

- Procedures for asking questions:
  - Type question in chat
  - An instructor will get to your question in order.
  - Designated Q&A periods (Chat or Oral)
    - Post that you have a question in the chat
    - You will then be called on
    - Unmute and ask question(s).
    - May ask you to clarify chat question.
    - Follow-up questions are allowed, time permitting.

- Please do not interrupt the instructor or students
- Chat and “Raise Hand” features are hard to see - Be Patient

What is this class?

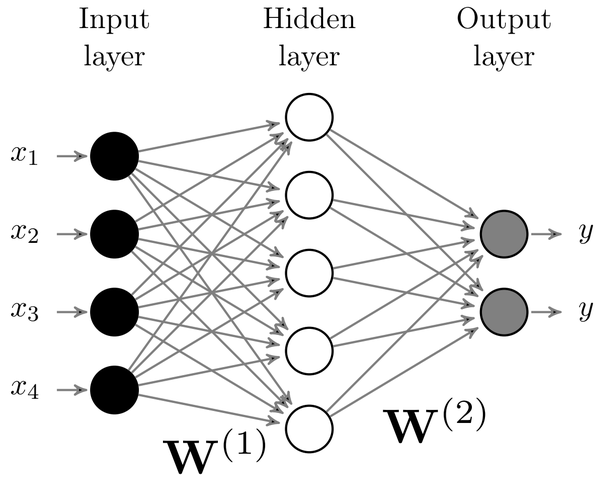
Applied Machine Learning

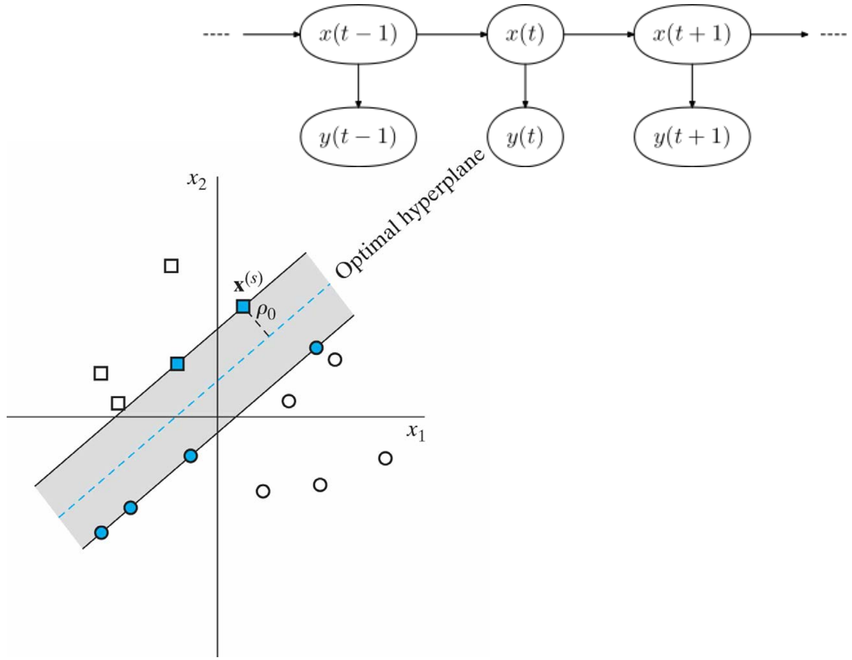
- Graduate course to teach you more about practical aspects of machine learning
- We'll discuss the ***full pipeline*** of machine learning
  - Data collection, processing and labeling
  - ML algorithm theory (some), implementation and evaluation
  - Experimental setup and design
- If you:
  - Are a reasonable (Python) programmer (or willing to learn)
  - Have basic skills in math (Calculus, Linear Algebra, Probability)
  - Want to learn how to tailor machine learning to different real-world problems
- Then P556 Applied Machine Learning is for you!

What is this class?

Learning Objectives

- By the end of the course, you will:
  - Become familiar with the full pipeline of applying machine learning to different real-world problems
  - Gain skills regarding the collection, processing, and labeling of data
  - Understand and implement the most (but not all) popular machine-learning algorithms
  - Improve your theoretical understanding of machine learning
  - Enhance your (Python) programming skills and utilize ML libraries (PyTorch, Scikit, TensorFlow, …)
  - Design experiments that reasonably assess performance
  - Evaluate multiple learning algorithms across several tasks
  - Assess potential societal impacts of the machine-learning algorithms





What is this class?

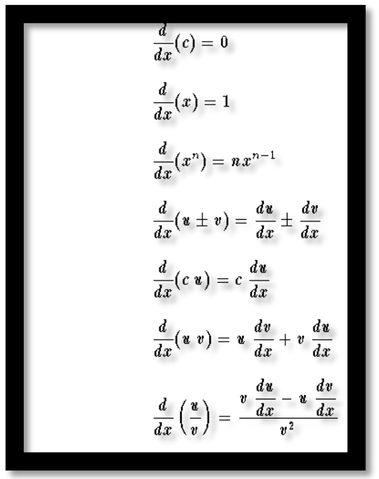
Learning Objectives

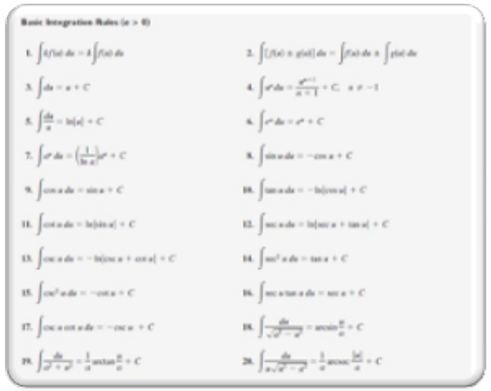
- This is an applied course, but merely knowing how to implement algorithms without some knowledge of the math is useless!
  - We WON'T fully cover Mathematical proofs
  - We WILL cover the theory at a "High Level"

- By the end of the course, you will:
  - Become familiar with the full pipeline of applying machine learning to different real-world problems
  - Gain skills regarding the collection, processing, and labeling of data
  - Understand and implement the most (but not all) popular machine-learning algorithms
  - Improve your theoretical understanding of machine learning
  - Enhance your (Python) programming skills and utilize ML libraries (PyTorch, Scikit, TensorFlow, …)
  - Design experiments that reasonably assess performance
  - Evaluate multiple learning algorithms across several tasks
  - Assess potential societal impacts of the machine-learning algorithms

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \times \begin{bmatrix} 2 & 0 \\ 1 & 2 \end{bmatrix} = \begin{bmatrix} 4 & 4 \\ 10 & 8 \end{bmatrix}$$

$$\begin{bmatrix} 2 & 0 \\ 1 & 2 \end{bmatrix} \times \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 2 & 4 \\ 7 & 10 \end{bmatrix}$$





What is this class?

Learning Objectives

- By the end of the course, you will:
  - Become familiar with the full pipeline of applying machine learning to different real-world problems
  - Gain skills regarding the collection, processing, and labeling of data
  - Understand and implement the most (but not all) popular machine-learning algorithms
  - Improve your theoretical understanding of machine learning
  - Enhance your (Python) programming skills and utilize ML libraries (PyTorch, Scikit, TensorFlow, …)
  - Design experiments that reasonably assess performance
  - Evaluate multiple learning algorithms across several tasks
  - Assess potential societal impacts of the machine-learning algorithms

What is this class?

Learning Objectives

- By the end of the course, you will:
  - Become familiar with the full pipeline of applying machine learning to different real-world problems
  - Gain skills regarding the collection, processing, and labeling of data
  - Understand and implement the most (but not all) popular machine-learning algorithms
  - Improve your theoretical understanding of machine learning
  - Enhance your (Python) programming skills and utilize ML libraries (PyTorch, Scikit, TensorFlow, …)
  - Design experiments that reasonably assess performance
  - Evaluate multiple learning algorithms across several tasks
  - Assess potential societal impacts of the machine-learning algorithms

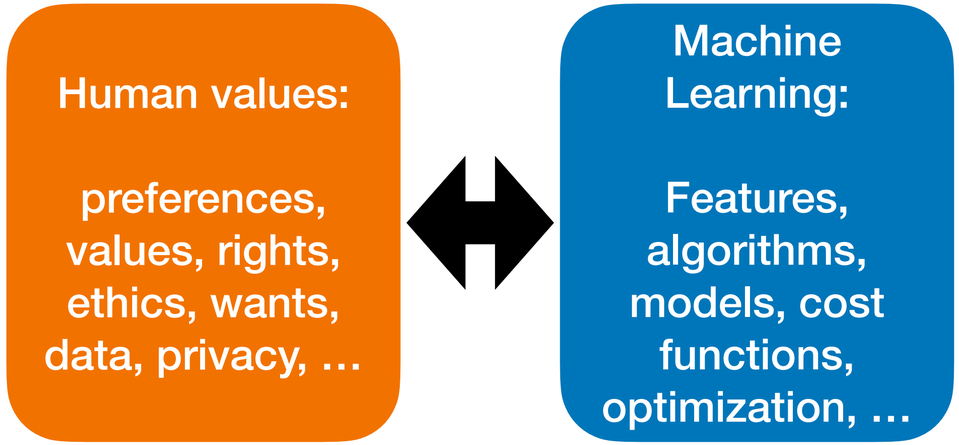
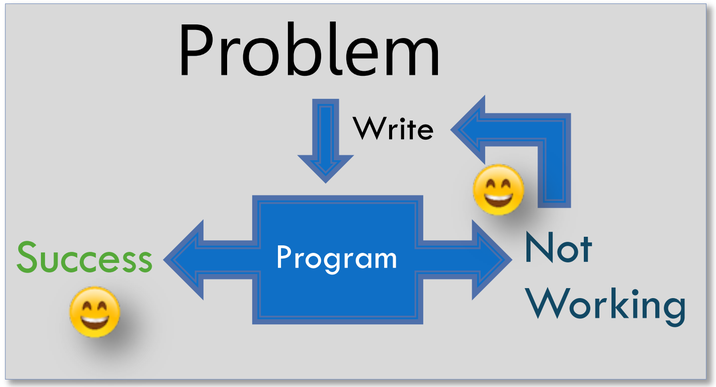


Image adopted from Rahwan (2018)

What this class is NOT?

Applied Machine Learning

- An introduction to programming (or Python), calculus (derivatives/integrals), linear algebra and probability
  - You are expected to know these already (at least at the undergraduate level)
  - External resources will be provided as needed (or potentially reviewed in class)
- A debugging class.
  - You are expected to resolve your own bugs.
  - You are graduate students and conceivably could be working at a company.
- A class to merely learn how to call functions, and use ML packages without knowing what is going on or what you are doing!
  - We will not merely run examples in Jupyter Notebook.
  - Theory/math (though not heavy) will be included in the lectures (at least)
  - Intuition will be tested in the assignments



Agenda & Outline

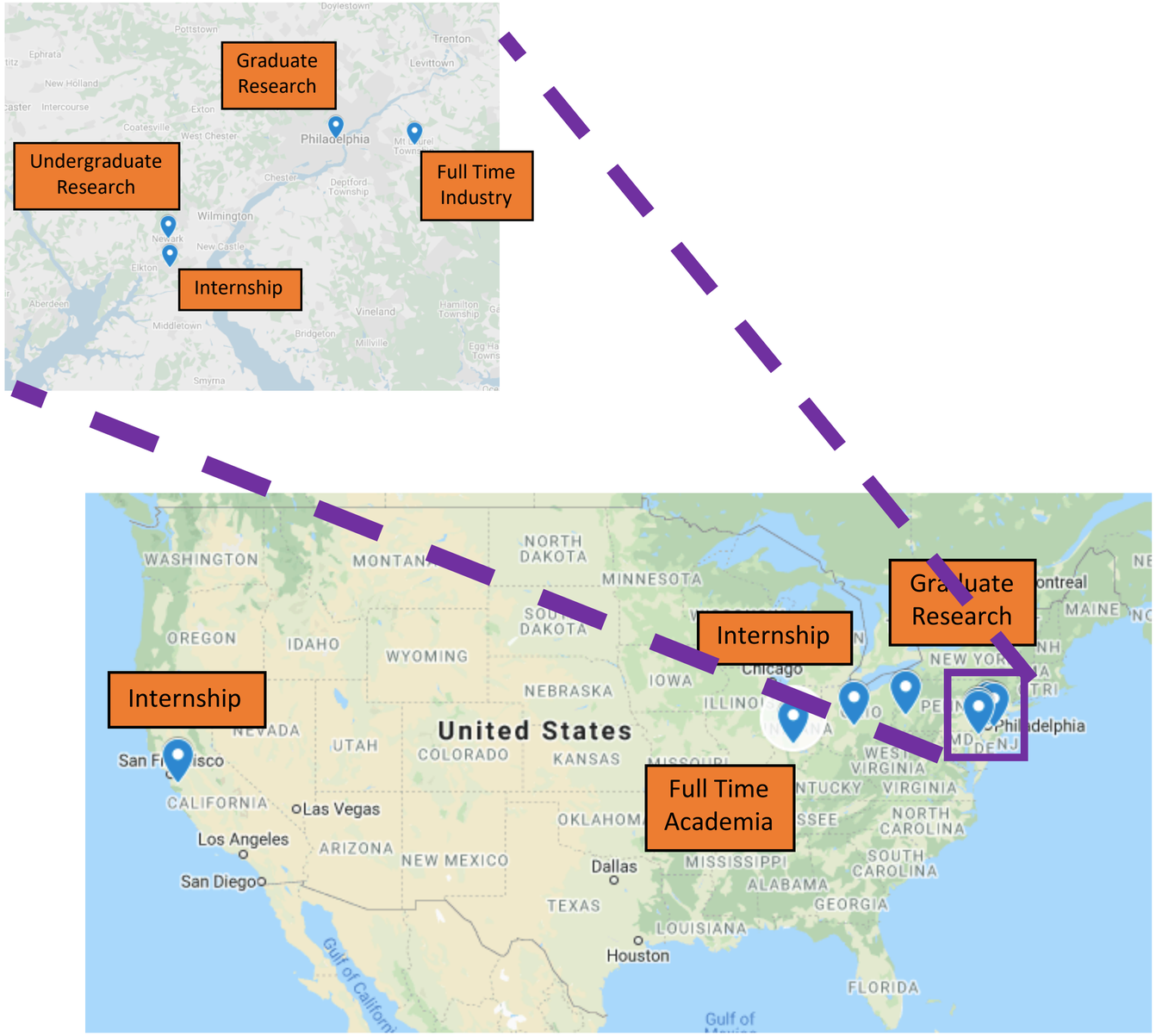
Today's Topics

- What is this class? What this class is NOT?
- Who am I? What I do?
- Who are you?
- Syllabus and logistics
- Why (Applied) Machine Learning?

Who am I?

A little about me

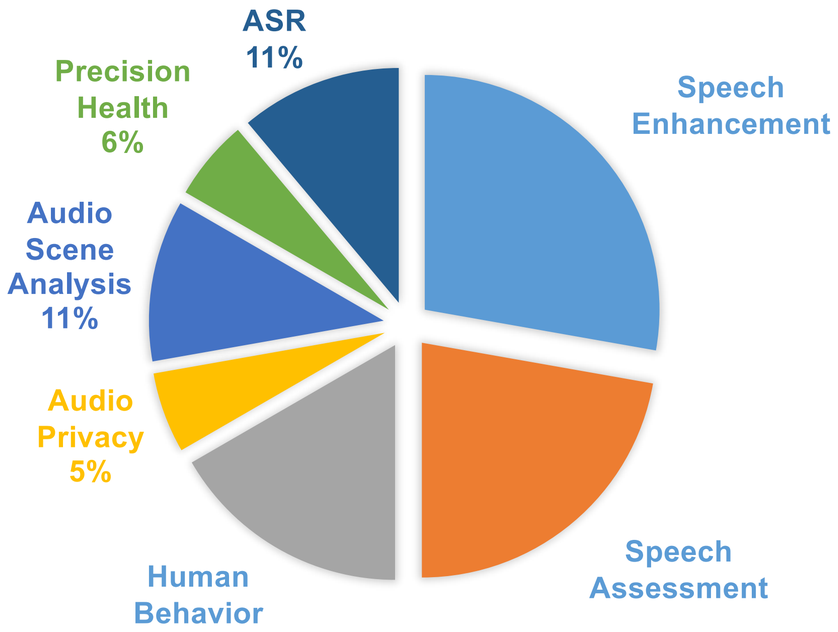
- Grew up in New Castle, Delaware
  - I work in data science, AI/ML, engineering
  - I work in signal processing, speech processing, hearing science, room acoustics, music information retrieval, human-computer interaction, human-in-the-loop machine learning, privacy and security
- Ph.D., Computer Science and Engineering, The Ohio State University
- Professional Industry Experience
  - Dade Behring Inc., Newark, DE
  - Bettis Atomic Power Laboratory, Pittsburgh, PA
  - Lockheed Martin, Moorestown, NJ
  - Audience, Inc. (Knowles), Mountain View, CA



What I do?

Here at IU

- Director of the ***Audio, Speech, and Information Retrieval (ASPIRE)*** Research group
- **Purpose**: Extract meaningful information from signals to improve HCI, make predictions, or to help with decision making



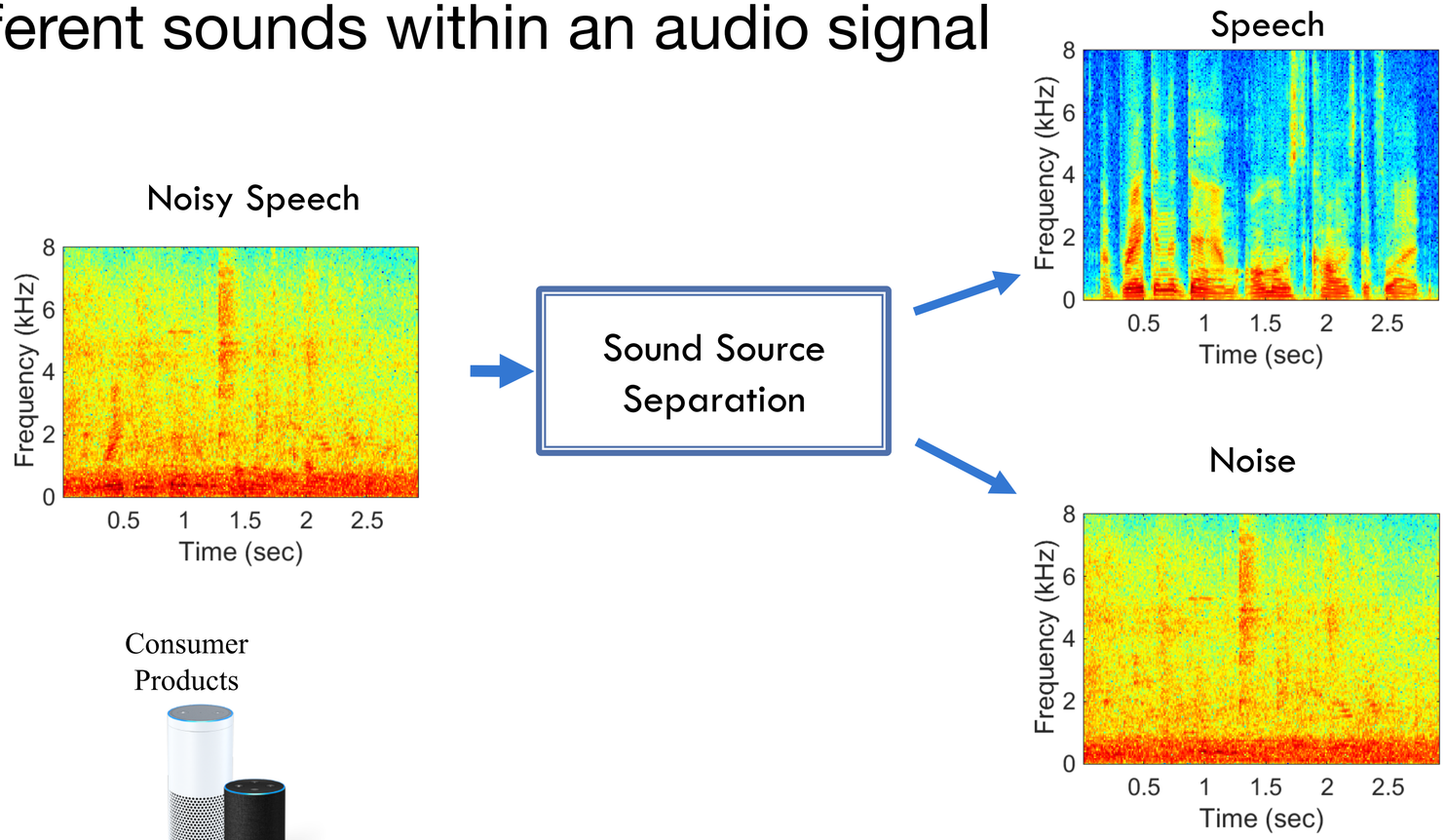
Approaches incorporate:

- Machine Learning
- Signal Processing
- Reinforcement Learning
- Human Studies
- Computing (e.g. GPUs)
- Programming (Python or MATLAB)
- Acoustics
- …

What I do?

Sound Source Separation/Enhancement

- Remove/separate different sounds within an audio signal



- Important for:

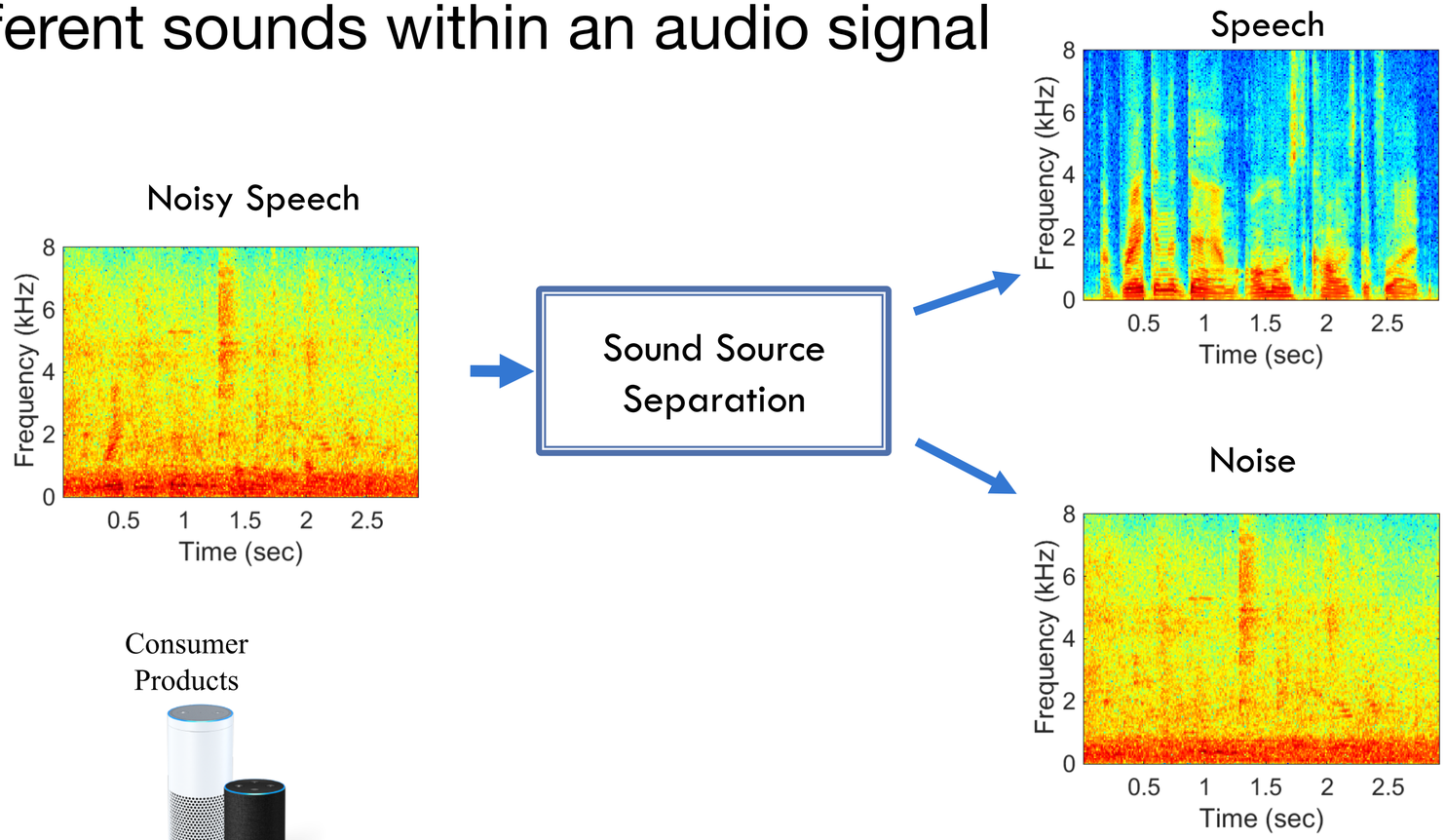
Hearing Impaired

Consumer Products

What I do?

Sound Source Separation/Enhancement

- Remove/separate different sounds within an audio signal



- Important for:

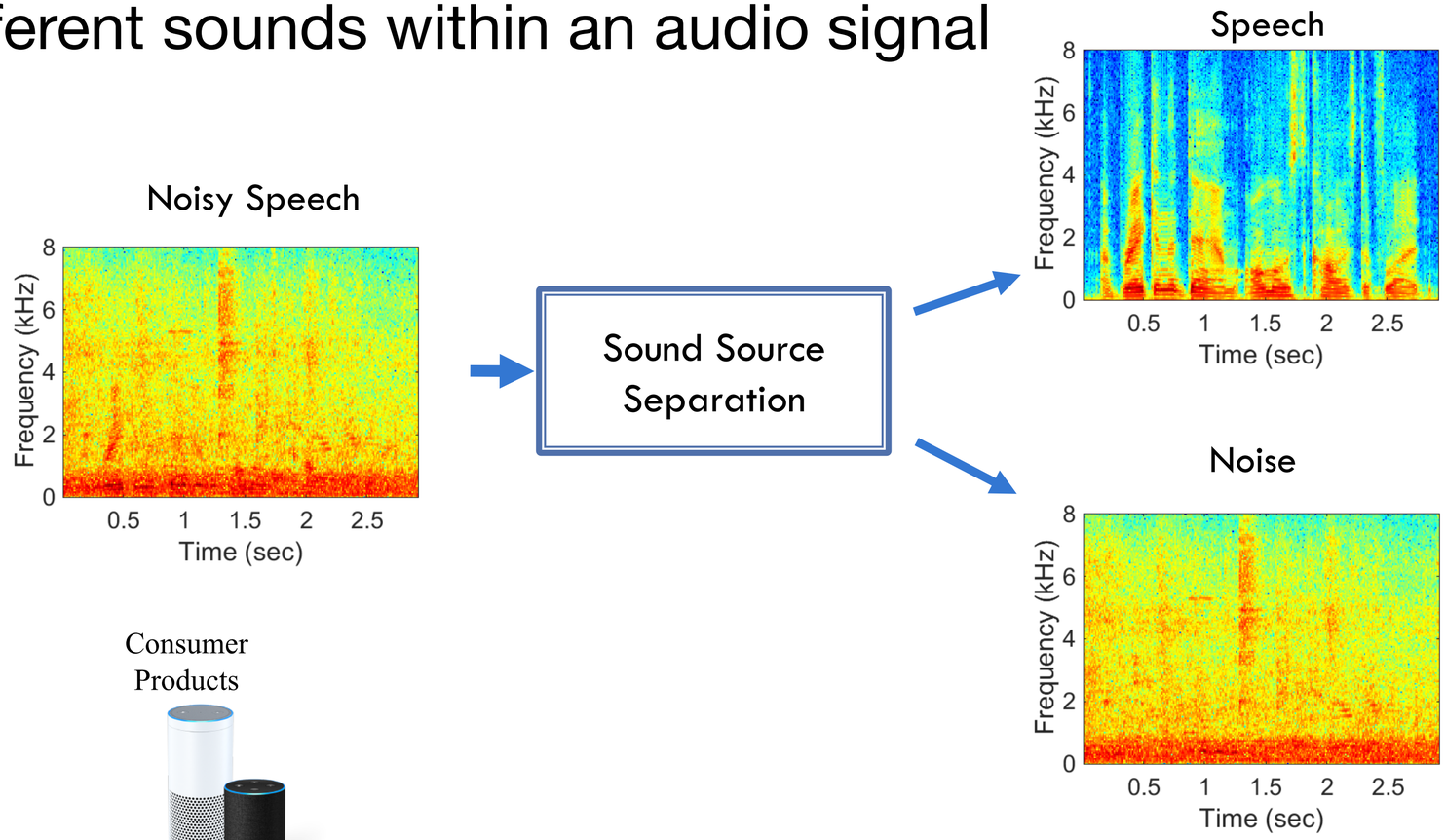
Hearing Impaired

Consumer Products

What I do?

Sound Source Separation/Enhancement

- Remove/separate different sounds within an audio signal



- Important for:

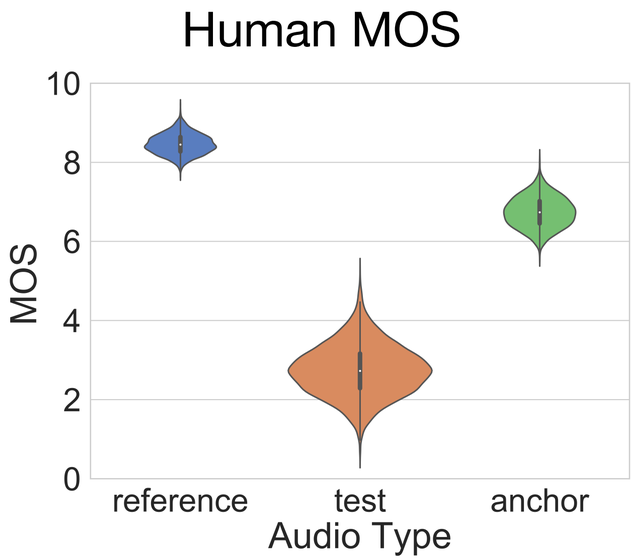
Hearing Impaired

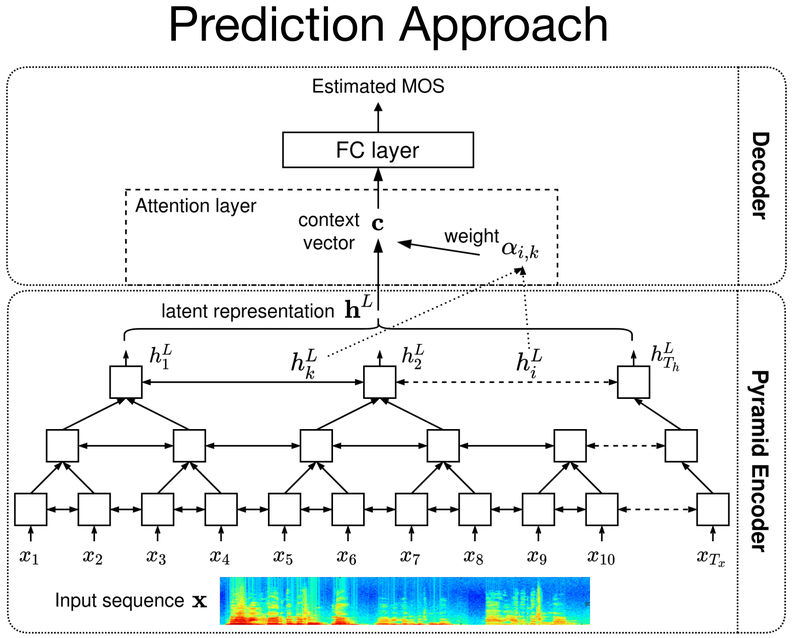
Consumer Products

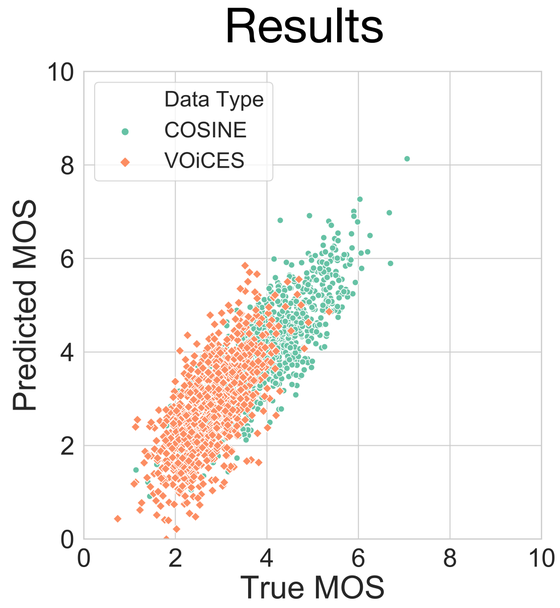
What I do?

Human speech perception models using deep learning

- Predict how people may assess a given audio signal
- Mean opinion score (MOS): Average user rating







What I do?

Audio and Privacy

- Devices with “always-on” microphones are becoming ubiquitous

- Not much of a problem for video/images

- How do we ensure audio privacy is preserved?
  - Turning off the device is not the answer
  - What does that light really mean?
  - Can we trust the manufacturers?
- Currently pursuing solutions.

Agenda & Outline

Today's Topics

- What is this class? What this class is NOT?
- Who am I? What I do?
- Who are you?
- Syllabus and logistics
- Why (Applied) Machine Learning?

Agenda & Outline

Today's Topics

- What is this class? What this class is NOT?
- Who am I? What I do?
- Who are you?
- Syllabus and logistics
- Why (Applied) Machine Learning?

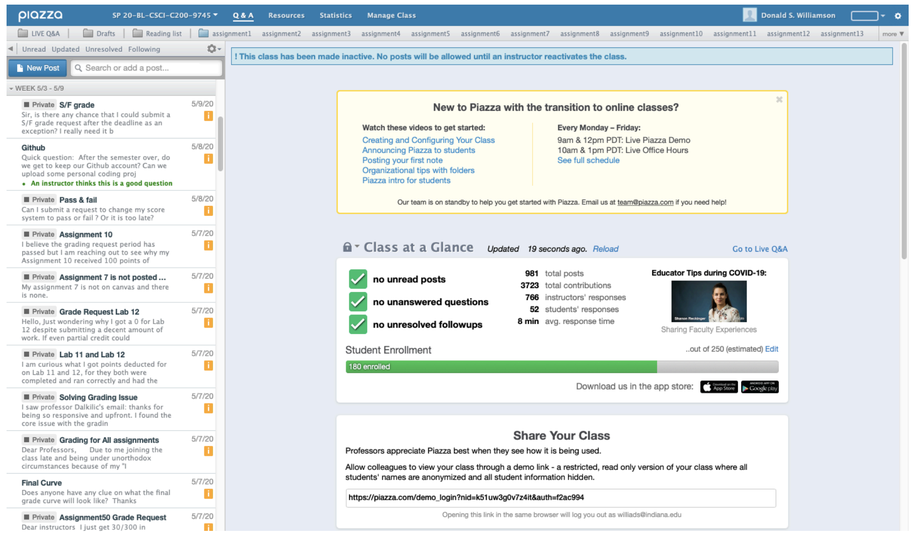
Online Learning Tools

Where all the goods are

- We will use Canvas to only hold materials
- Github (IU account) to house your programs
- Piazza for ***ALL*** messaging about the class

Piazza is no longer free!
A ***$5 charge*** will show on your account, ***but*** the CS department will ***reimburse*** this!

NO EMAIL - unless urgent ***and*** extremely private



Office Hours and Associate Instructors

- Office hours will be conducted through Zoom. See Canvas for the links

Instructor: Dr. Donald S. Williamson	**Time:** TR 4:55pm – 6:10pm
Email: Piazza (required), williads@indiana.edu	**Place:** Zoom (see Canvas for link)
Office Hours: TBD	**Course Page**: Canvas
AI: Junyi Fan	**Email:** Piazza (required), fanjun@iu.edu
Office Hours: T 3:00 - 4:00PM	**Place:** Zoom (see Canvas for link)

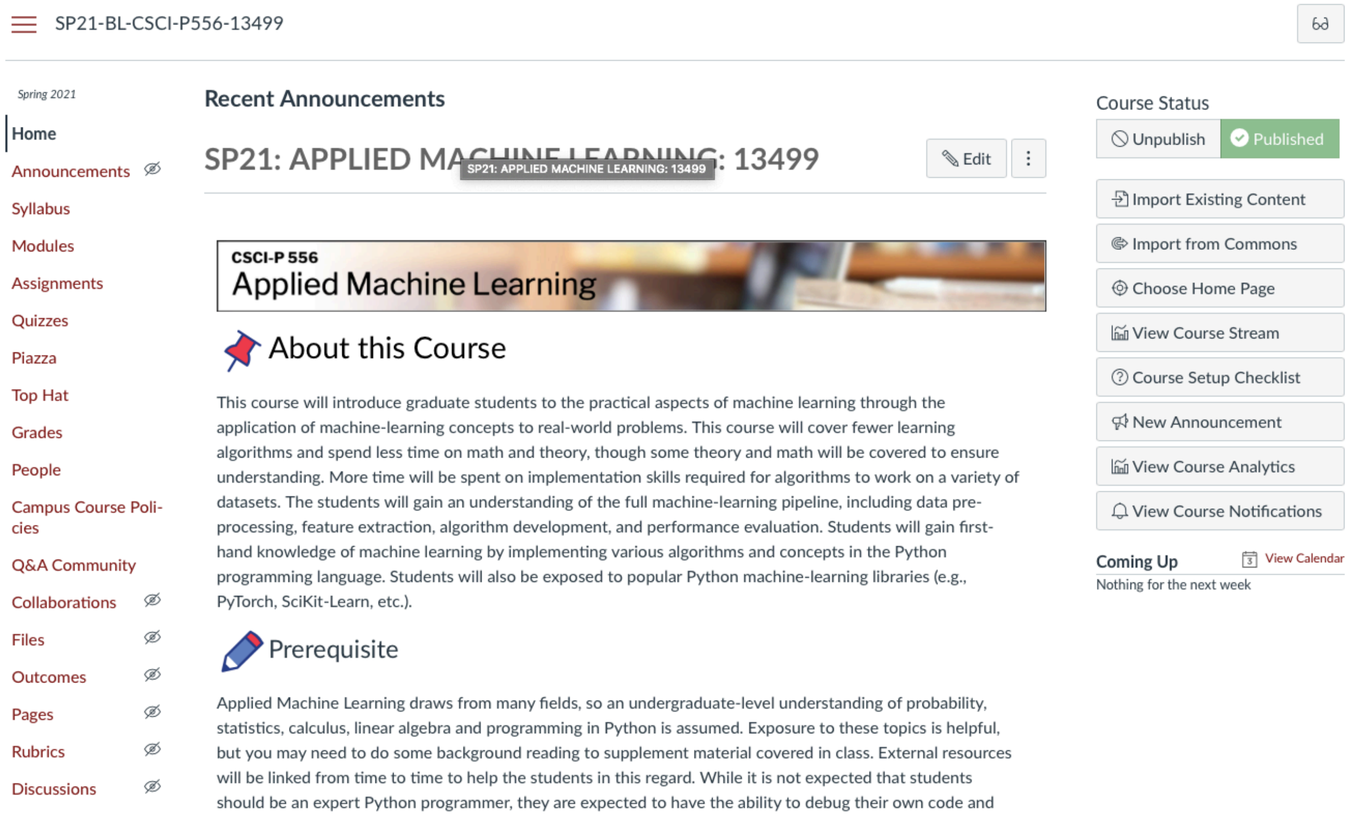
Note Takers

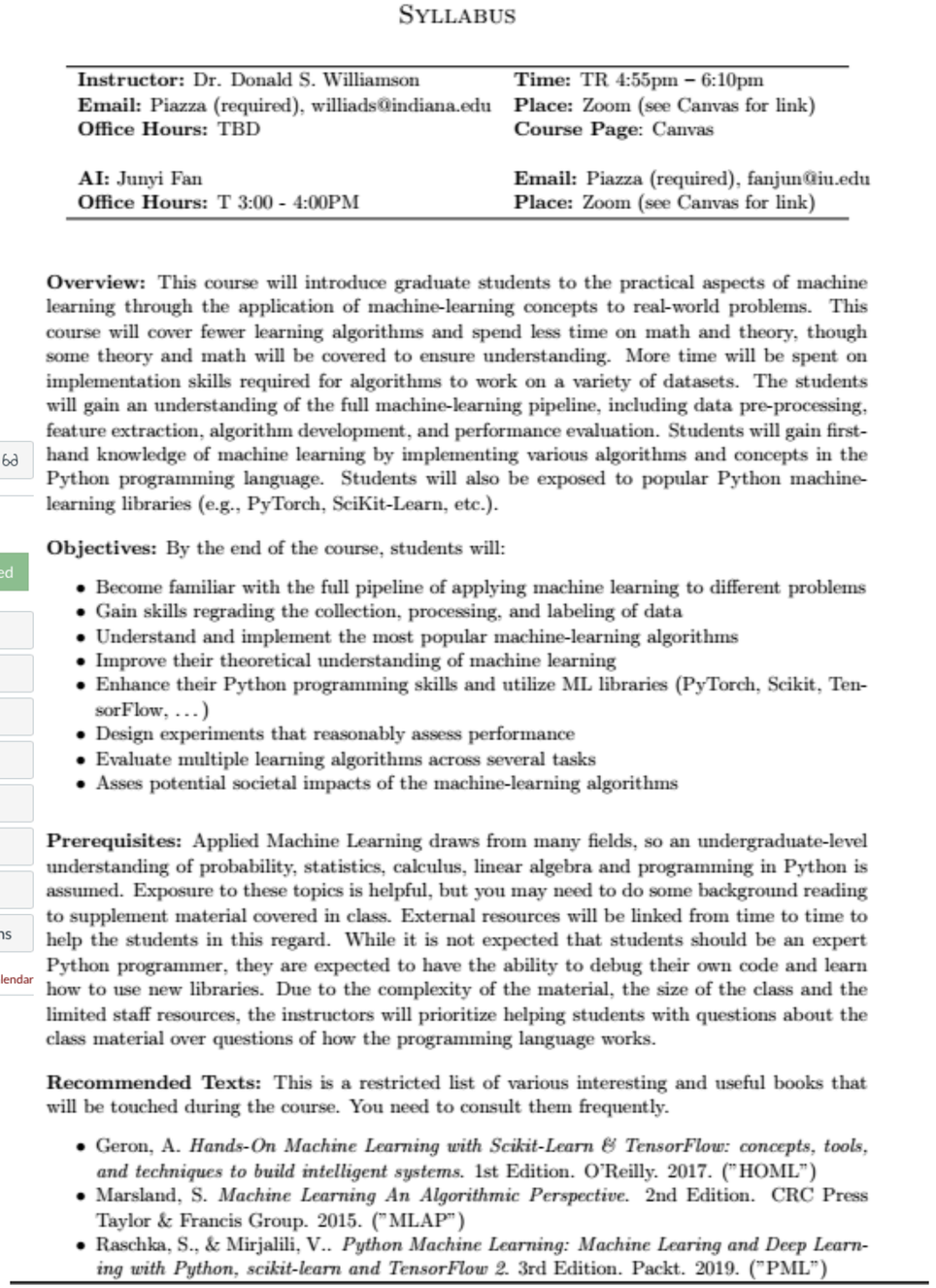
- For some modicum of extra credit, up to three students can volunteer to take notes and then publish them for the class, as an additional help for everyone
  - Notes that will be posted must be typed (no scanning)
  - Contact the Lead AI about submitting the notes (Likely through Piazza)
- The students will take notes for the week (e.g. two lectures)
- Helps with understanding the material

Syllabus

Found on Canvas

- You must read the entire syllabus!





SYLLABUS

Instructor: Dr. Donald S. Williamson	**Time:** TR 4:55pm – 6:10pm
Email: Piazza (required), williads@indiana.edu	**Place:** Zoom (see Canvas for link)
Office Hours: TBD	**Course Page:** Canvas
AI: Junyi Fan	**Email:** Piazza (required), fanjun@iu.edu
Office Hours: T 3:00 - 4:00PM	**Place:** Zoom (see Canvas for link)

Overview: This course will introduce graduate students to the practical aspects of machine learning through the application of machine-learning concepts to real-world problems. This course will cover fewer learning algorithms and spend less time on math and theory, though some theory and math will be covered to ensure understanding. More time will be spent on implementation skills required for algorithms to work on a variety of datasets. The students will gain an understanding of the full machine-learning pipeline, including data pre-processing, feature extraction, algorithm development, and performance evaluation. Students will gain first-hand knowledge of machine learning by implementing various algorithms and concepts in the Python programming language. Students will also be exposed to popular Python machine-learning libraries (e.g., PyTorch, SciKit-Learn, etc.).

Objectives: By the end of the course, students will:

- Become familiar with the full pipeline of applying machine learning to different problems
- Gain skills regrading the collection, processing, and labeling of data
- Understand and implement the most popular machine-learning algorithms
- Improve their theoretical understanding of machine learning
- Enhance their Python programming skills and utilize ML libraries (PyTorch, Scikit, TensorFlow, . . .)
- Design experiments that reasonably assess performance
- Evaluate multiple learning algorithms across several tasks
- Asses potential societal impacts of the machine-learning algorithms

Prerequisites: Applied Machine Learning draws from many fields, so an undergraduate-level understanding of probability, statistics, calculus, linear algebra and programming in Python is assumed. Exposure to these topics is helpful, but you may need to do some background reading to supplement material covered in class. External resources will be linked from time to time to help the students in this regard. While it is not expected that students should be an expert Python programmer, they are expected to have the ability to debug their own code and learn how to use new libraries. Due to the complexity of the material, the size of the class and the limited staff resources, the instructors will prioritize helping students with questions about the class material over questions of how the programming language works.

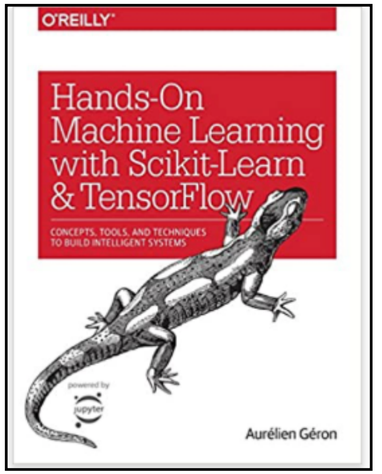
Recommended Texts: This is a restricted list of various interesting and useful books that will be touched during the course. You need to consult them frequently.

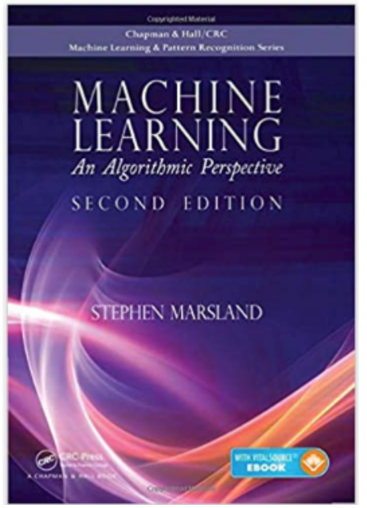
- Geron, A. *Hands-On Machine Learning with Scikit-Learn & TensorFlow: concepts, tools, and techniques to build intelligent systems.* 1st Edition. O'Reilly. 2017. ("HOML")
- Marsland, S. *Machine Learning An Algorithmic Perspective.* 2nd Edition. CRC Press Taylor & Francis Group. 2015. ("MLAP")
- Raschka, S., & Mirjalili, V.. *Python Machine Learning: Machine Learing and Deep Learning with Python, scikit-learn and TensorFlow 2.* 3rd Edition. Packt. 2019. ("PML")

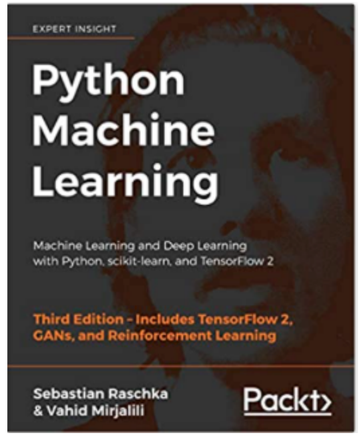
Course Materials

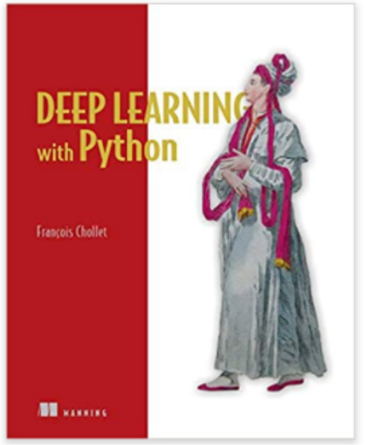
Recommended Textbooks

- Geron, A. *Hands-On Machine Learning with Scikit-Learn & TensorFlow: concepts, tools, and techniques to build intelligent systems*. 1st Edition. O'Reilly. 2017
    - Free Download Available from IU Libraries
- Marsland, S. *Machine Learning: An Algorithmic Perspective*. 2nd Edition. CRC Press Taylor & Francis Group. 2015
- Raschka, S., & Mirjalili,. V. *Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn and TensorFlow 2*. 3rd Edition. Packt. 2019.
- Chollet, F. *Deep Learning with Python*. 1st Edition. Manning. 2018. [Optional and least used]
    - Freely Available through IU Libraries









Grading Policy

Four main components

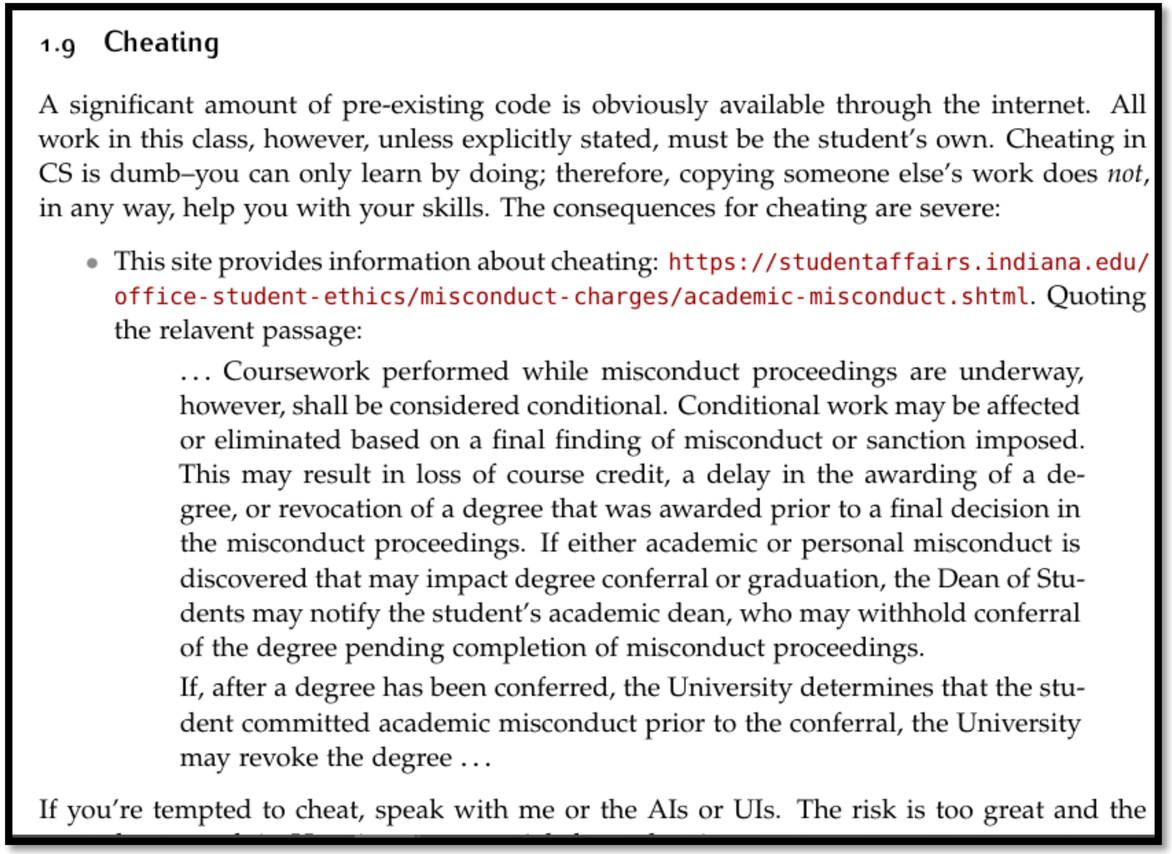
- Grading Plan
  - Participation (10%)
  - Homework (30%)
  - Quizzes (30%)
  - Project (30%)
- Grading Rubric
  - 99-100 A+
  - 93-98+ A
  - 90-92+ A-
  - 88-89+ B+
  - ….
- Grades may be curved at the end, if needed

Grading Components

- Participation: expected to attend class, ask/answer questions
- Homework:
  - Likely 5 (may change) assignments
  - Mostly programming (Python), but may have intuitive related questions
  - Will be completed in pairs, where pairs we be randomly assigned per assignment.
    - To foster community and learn from each other
    - Pairs will each submit assignments, though they may be the same
- Quizzes:
    - Announced/unannounced short quizzes throughout the semester
    - Based on recent lectures and/or homework
- Project:
  - Team (up to 3) assignment
  - More details later on (See syllabus also)

Cheating

- Cheating may be tempting, but DON'T do IT!
  - You don't make progress
  - It doesn't help you learn
  - We'll catch you if you do.
  - It catches up to you



1.9 Cheating

A significant amount of pre-existing code is obviously available through the internet. All work in this class, however, unless explicitly stated, must be the student's own. Cheating in CS is dumb–you can only learn by doing; therefore, copying someone else's work does *not*, in any way, help you with your skills. The consequences for cheating are severe:

- This site provides information about cheating: `https://studentaffairs.indiana.edu/office-student-ethics/misconduct-charges/academic-misconduct.shtml`. Quoting the relavent passage:

  > ... Coursework performed while misconduct proceedings are underway, however, shall be considered conditional. Conditional work may be affected or eliminated based on a final finding of misconduct or sanction imposed. This may result in loss of course credit, a delay in the awarding of a degree, or revocation of a degree that was awarded prior to a final decision in the misconduct proceedings. If either academic or personal misconduct is discovered that may impact degree conferral or graduation, the Dean of Students may notify the student's academic dean, who may withhold conferral of the degree pending completion of misconduct proceedings.
  >
  > If, after a degree has been conferred, the University determines that the student committed academic misconduct prior to the conferral, the University may revoke the degree ...

If you're tempted to cheat, speak with me or the AIs or UIs. The risk is too great and the

Tentative Class Schedule

This will likely change!

INTRODUCTION	
Week 1	Course introduction, logistics and Machine Learning basics
Week 2	Data collection, pre-processing and labeling
Week 3	Evaluation and methodology
SUPERVISED LEARNING	
Week 4	Probability review and Naive Bayes
Week 5	Linear and Nonlinear Regression
Week 6	Neural Networks and Deep Learning I
Week 7	Neural Networks and Deep Learning II
Week 8	Neural Networks and Deep Learning III
Week 9	Support Vector Machines
Week 10	Decision Trees, Ensemble Learning and Random Forests

UNSUPERVISED LEARNING	
Week 11	Clustering, KMeans and Nearest Neighbor
Week 12	Gaussian Mixture Models and Nonnegative Matrix Factorization
Week 13	Dimensionality Reduction and Societal impacts of ML
Week 14	Reinforcement Learning (time permitting)
Project	
	Project Presentations I (if needed)
Week 15	Project Presentations II
Week 16 (Finals Week)	Project Presentations III (if needed)

Agenda & Outline

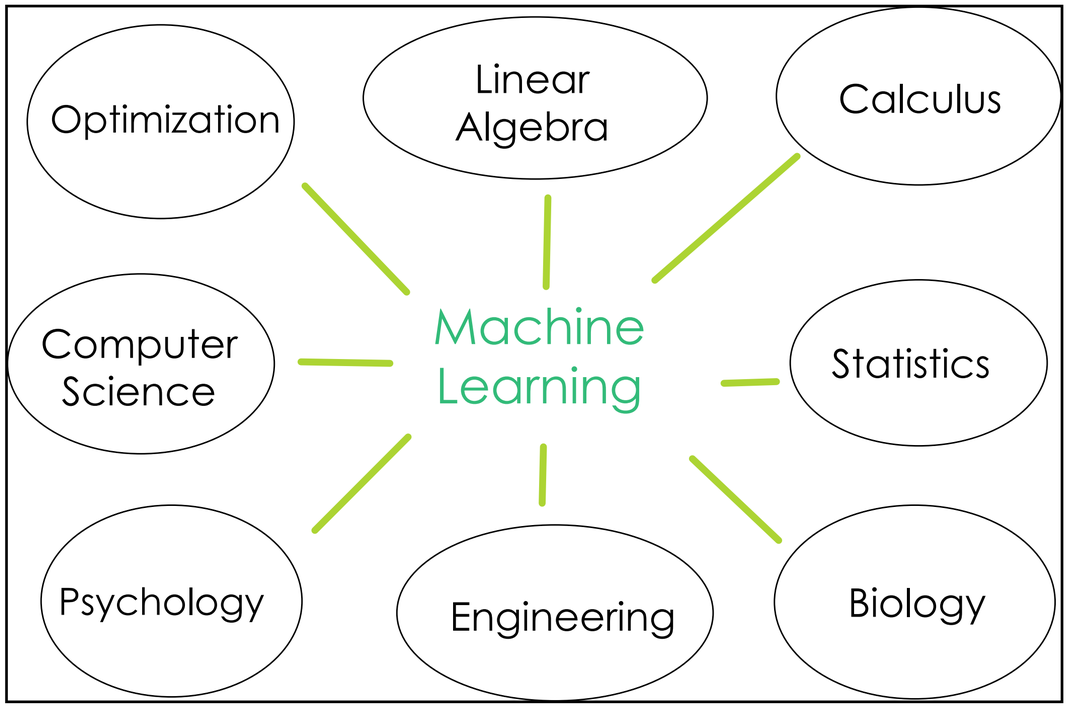
Today's Topics

- What is this class? What this class is NOT?
- Who am I? What I do?
- Who are you?
- Syllabus and logistics
- Why (Applied) Machine Learning?

What is Machine Learning?

It's Multifaceted

- "A set of methods that can automatically detect patterns in data, and then use the uncovered patterns to predict future data, or to perform other kinds of decision making under uncertainty" ~ Kevin P. Murphy
- Concerned with developing, analyzing, and applying algorithms that make useful inferences when provided with data
- Provides a framework for solving hard problems!



Why is Machine Learning Important?

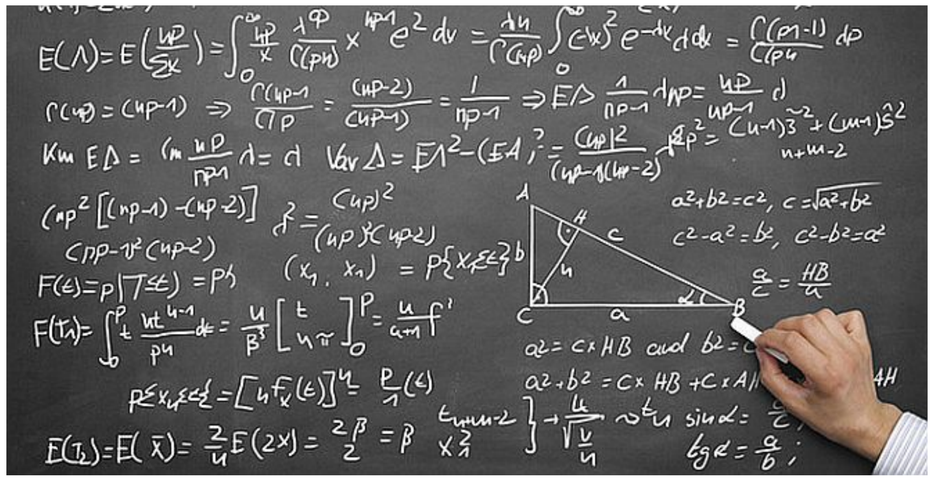
Data is everywhere

- Developing methods to analyze data and perform predictions is useful in many ways
  - Weather prediction
  - Cancer (and other disease) research
  - Consumer products (human computer interaction)
  - E-commerce
  - Marketing
  - Etc.!

What Makes it Challenging?

- Identifying patterns in data is hard!

- Knowing how to apply machine learning to real world problems is …. Hard!



- Mathematics, Probability, Engineering, etc. are hard!

Why Learn to Apply ML?

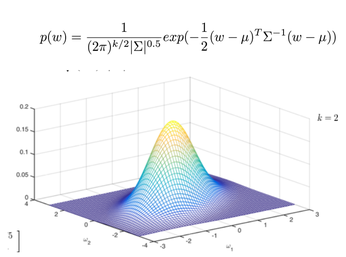
You won’t be students forever

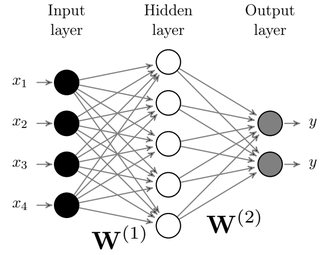
- There are several problems that need to be solved, and ML can help!
- Applying ML helps you better understand ML
  - There is only so much you can learn from theory
- Companies expect this of you!
- There are implementation details that you only learn when you apply solutions to a problem!

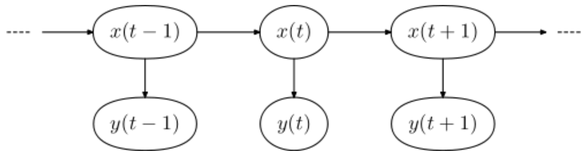
Applications of Machine Learning

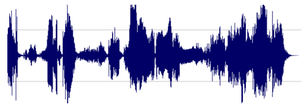
Automatic Speech Recognition

- Used in home assistants and cell phones (e.g. Hey Siri, Hey Google, Hey Alexa)
- Consist of several ML algorithms
  - Probability
  - Deep Neural Networks
  - Hidden Markov Models
  - Natural Language Processing
  - …









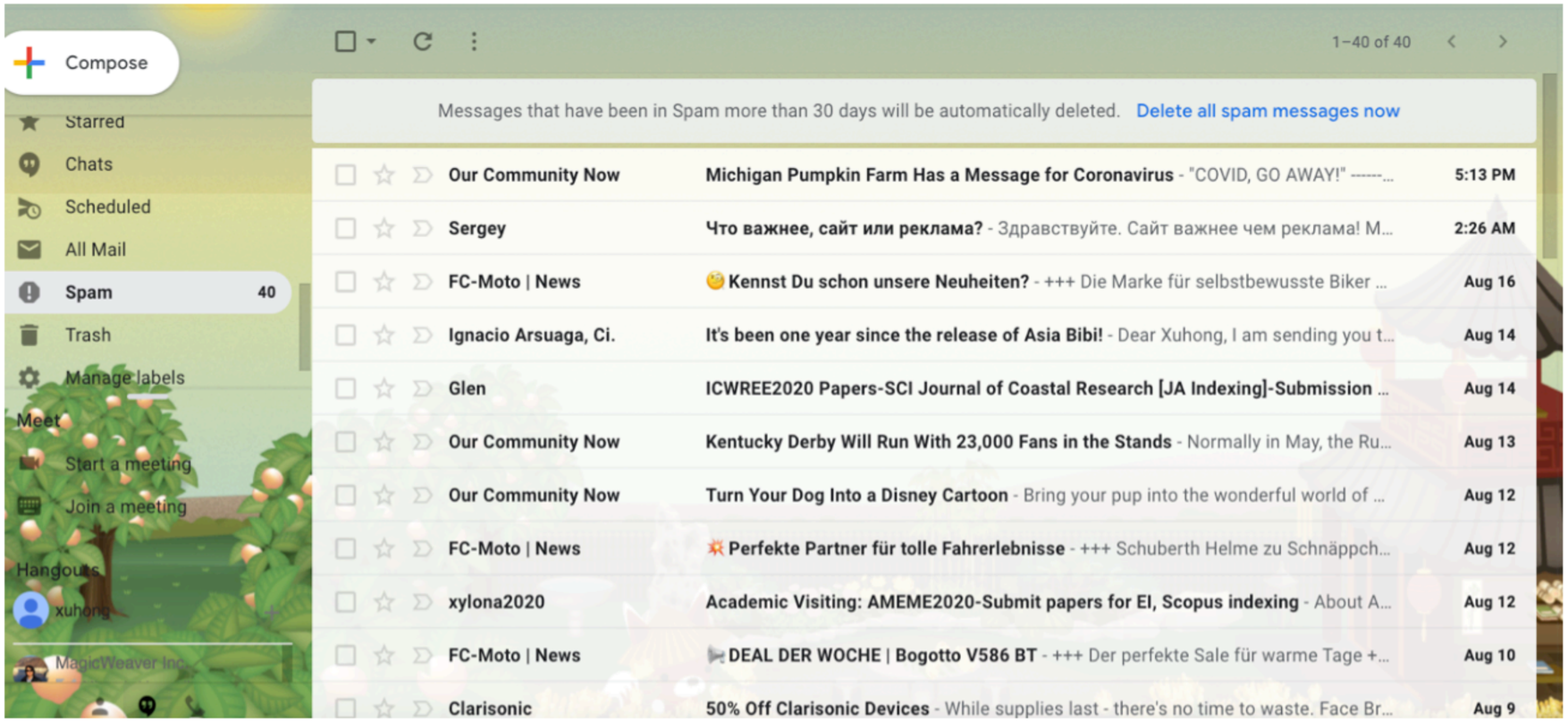
Speech Recognition

"Will you be able to…"

Applications of Machine Learning

Spam Filtering

- Block unwanted emails through classification!
- Utilize:
  - Neural Networks
  - Naive Bayesian classification
  - Support Vector Machines

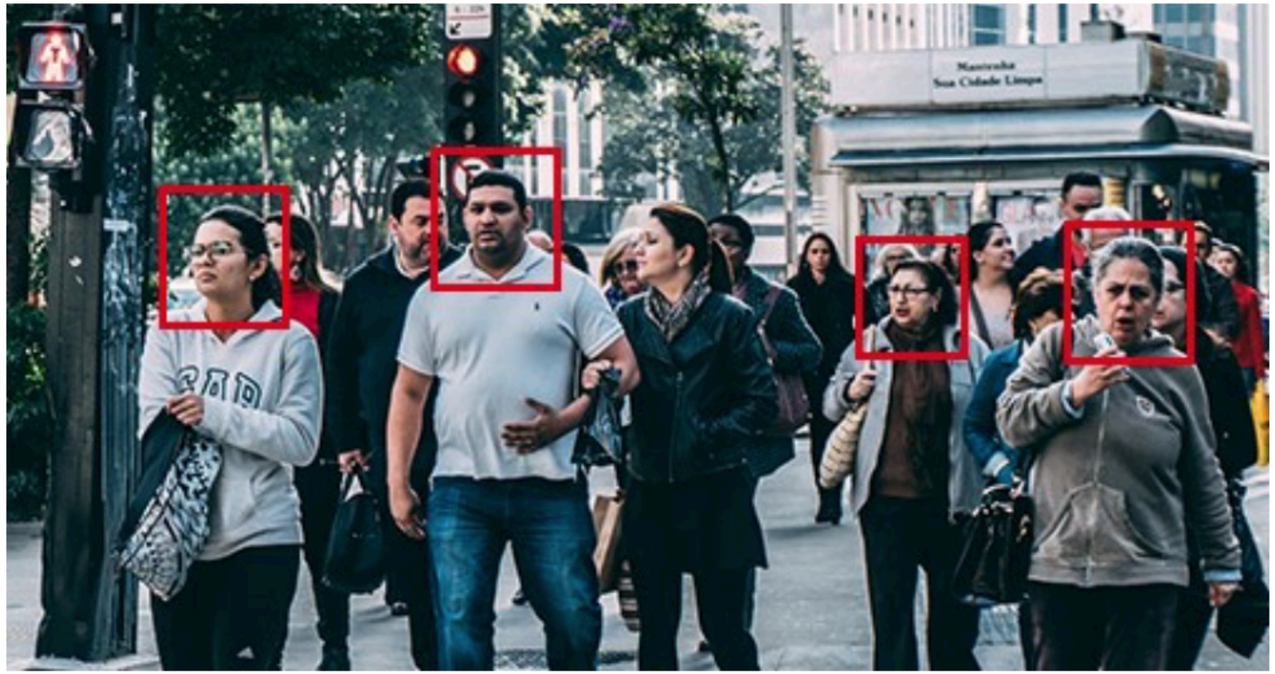


Applications of Machine Learning

Face Detection

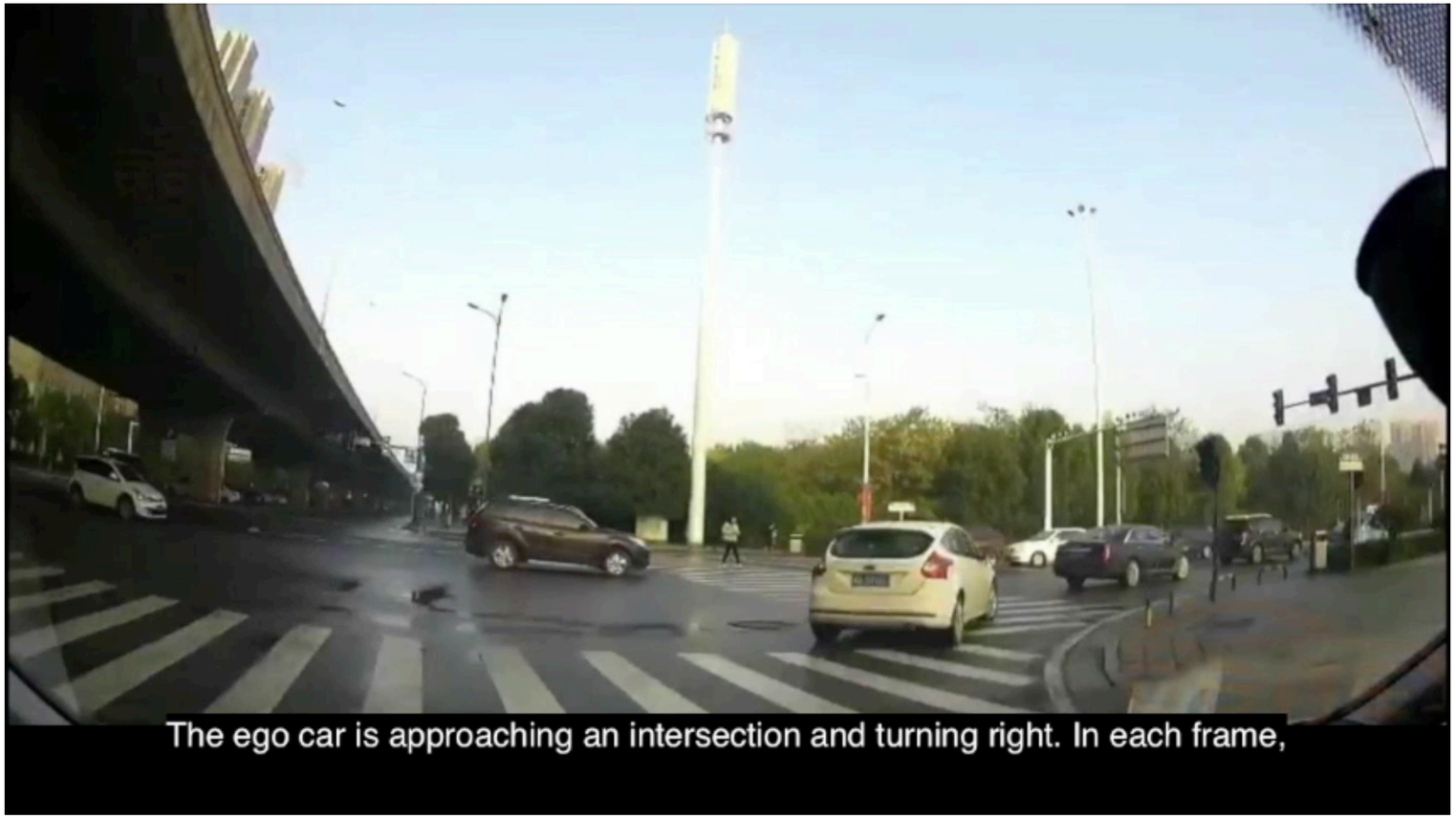
- Locate and localize one or more faces in a photograph

- Many different strategies
  - Feature based
  - Appearance based
  - Knowledge based
  - Template matching



Applications of Machine Learning

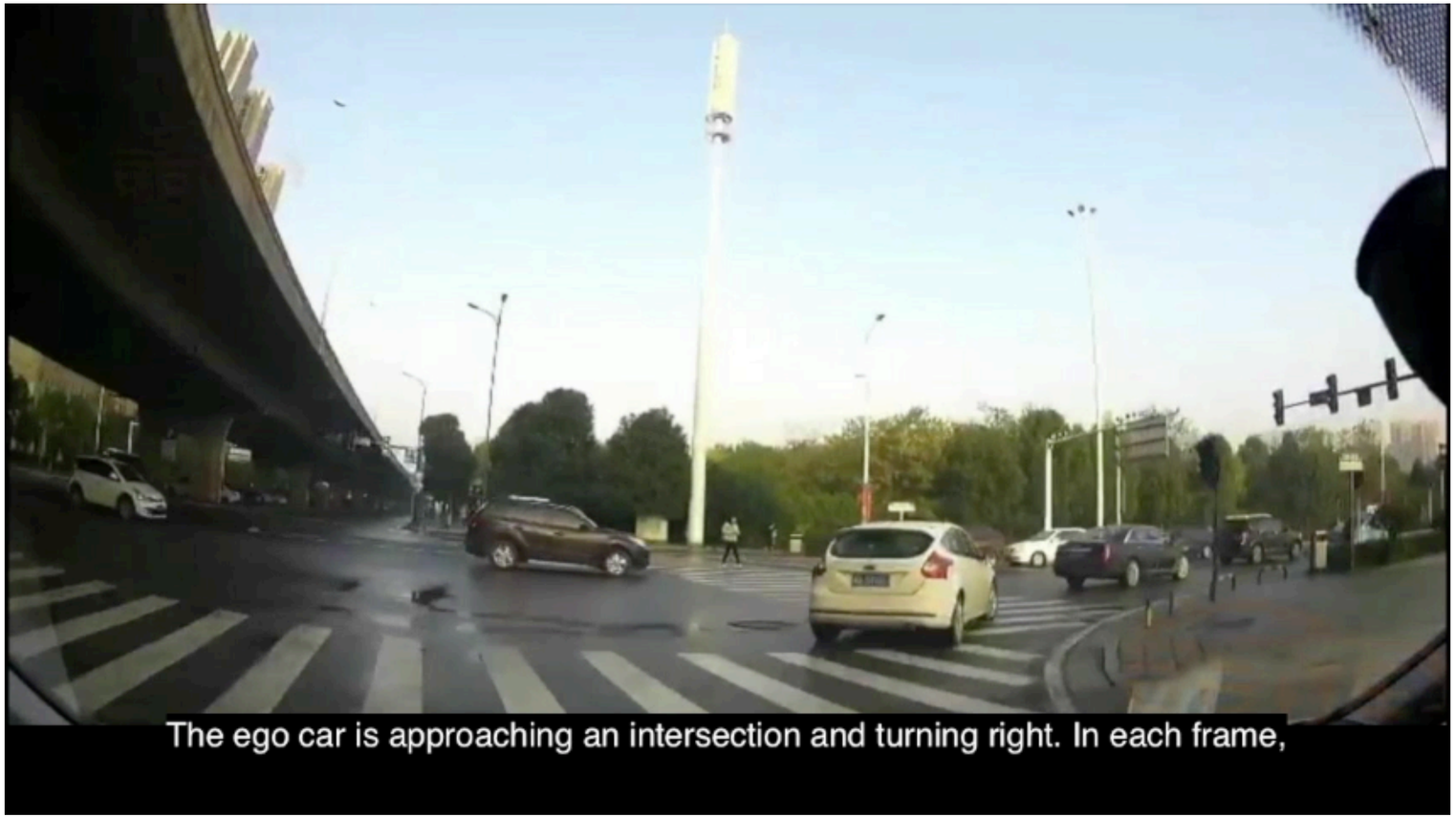
Future Vehicle Localization and Event Prediction



Yao et al, Egocentric Vision-based Future Vehicle Localization for Intelligent Driving Assistance Systems, ICRA 2019.
Yao et al. Unsupervised Traffic Accident Detection in First-Person Videos, IROS 2019.

Applications of Machine Learning

Future Vehicle Localization and Event Prediction



Yao et al, Egocentric Vision-based Future Vehicle Localization for Intelligent Driving Assistance Systems, ICRA 2019.
Yao et al. Unsupervised Traffic Accident Detection in First-Person Videos, IROS 2019.

Applications of Machine Learning

Not so Fast!

- We WON’T cover these specific applications
- We will cover some of the component algorithms
- Assignments and Project will enable you to implement solutions for less-challenging/complicated application

Conclusion

- Participate in Class!
  - Be prepared. Read!
  - Ask questions
  - Learning is about effort
- Check Piazza for announcements and updates
- Next Class
  - Machine Learning basics
  - Check syllabus/Canvas for required readings and homework 0 (setup working environment)