Introduction to CIS 419/519 Applied Machine Learning

Dan Roth
danroth@seas.upenn.edu|http://www.cis.upenn.edu/~danroth/|461C, 3401 Walnut

Slides were created by Dan Roth (for CIS519/419 at Penn or CS446 at UIUC),
Some slides were taken with approval from other authors who have made their ML slides available.

Penn Engineering

CIS 419/519 Remote Version

- Weird Times, but the show must go on.

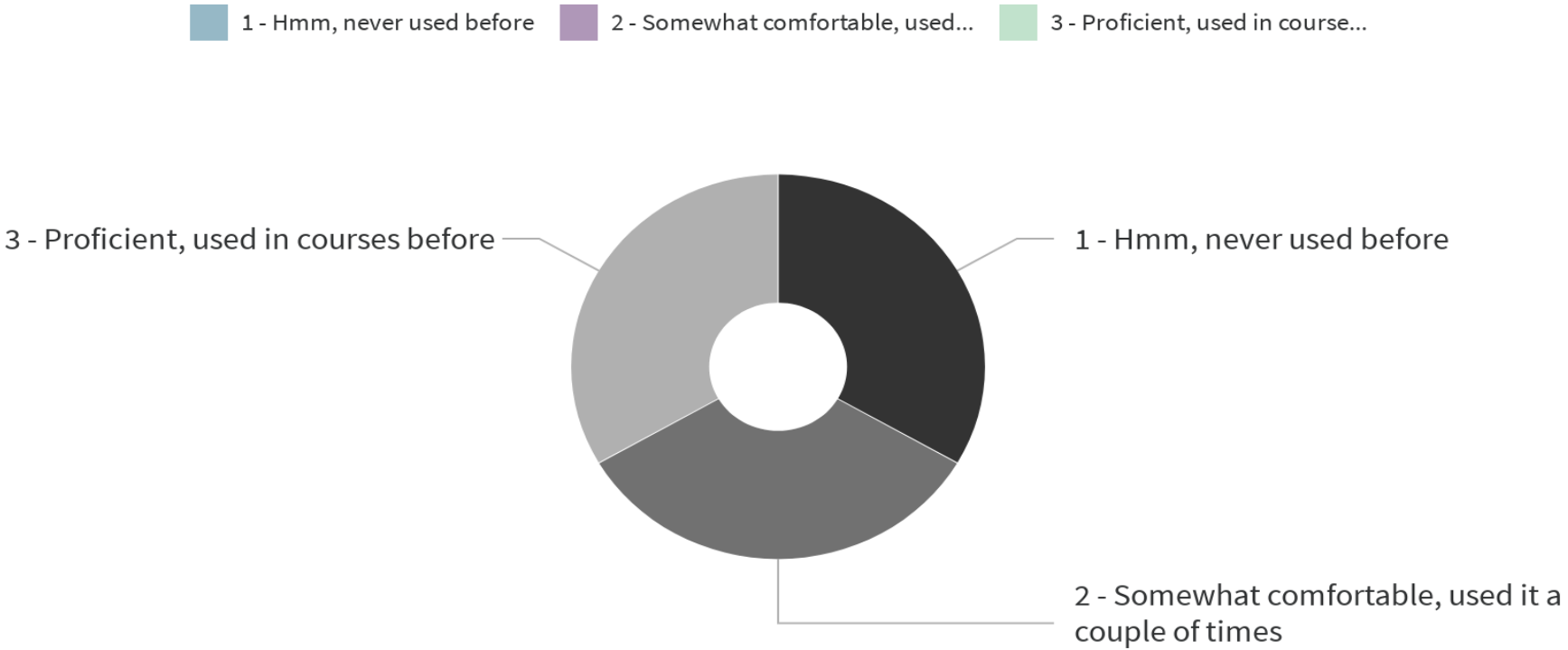
- I will try to run the class as usual, with a few exceptions

  - Exams
  - It could be boring to just sit there and listen to me talking
  - We'll try to get you to participate.
    - Chat
    - Poll Everywhere
      - (we will ask you to login using your upenn account since participation is mandatory)
    - Raise your hand

Where are you currently located?

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

How comfortable are you with using Python?



Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

Describe one machine learning application which interests you...

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

CIS 419/519: Applied Machine Learning

- Monday, Wednesday: 10:30pm-12:00pm On Zoom
- (My) Office hours: Mon 5-6 pm; Tue 12-1pm
- 13 TAs
- Assignments: 5 Problems set (Python Programming)
  - Weekly (light) on-line quizzes
- Weekly Discussion Sessions
- Mid Term Exam (take home)
- [Project] (We'll talk about it later)
- Final (take home)
- No real textbook:
  - Slides/Mitchell/Goldberg/Other Books/Lecture notes /Literatur

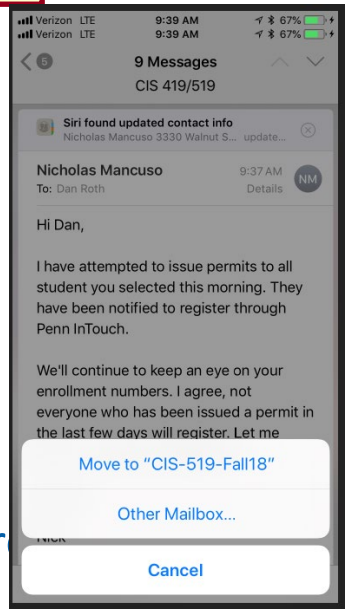
Starting next week TAs Office Hours will also start next week

HW0 !!!

Go to the web site

Be on Piazza

Registration for Class



CIS 419/519: Today

- What is Learning?
- Who are you?
- What is CIS 419/519 about?
- The Badges Game…

Who are you? (one aspect)

First year student

Sophomore

Junior

Senior

MS Student

PhD Student

None of the above

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

Would you mind opening you video so we can see you?

Happy to do it sometimes

No, I want it off

Sure, I can keep it open all the time

Only when I ask a question

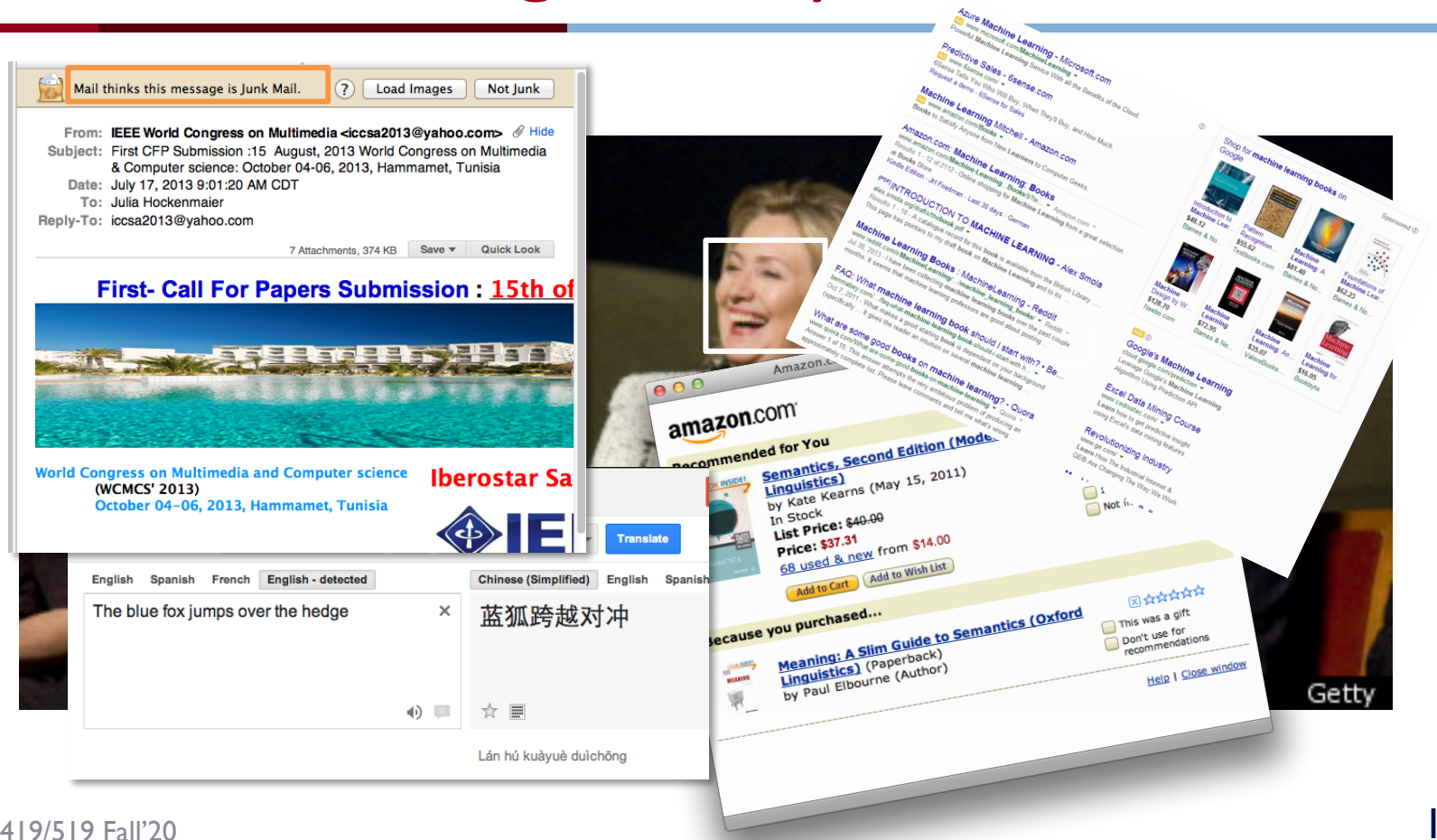
None of the above

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

An Owed to the Spelling Checker

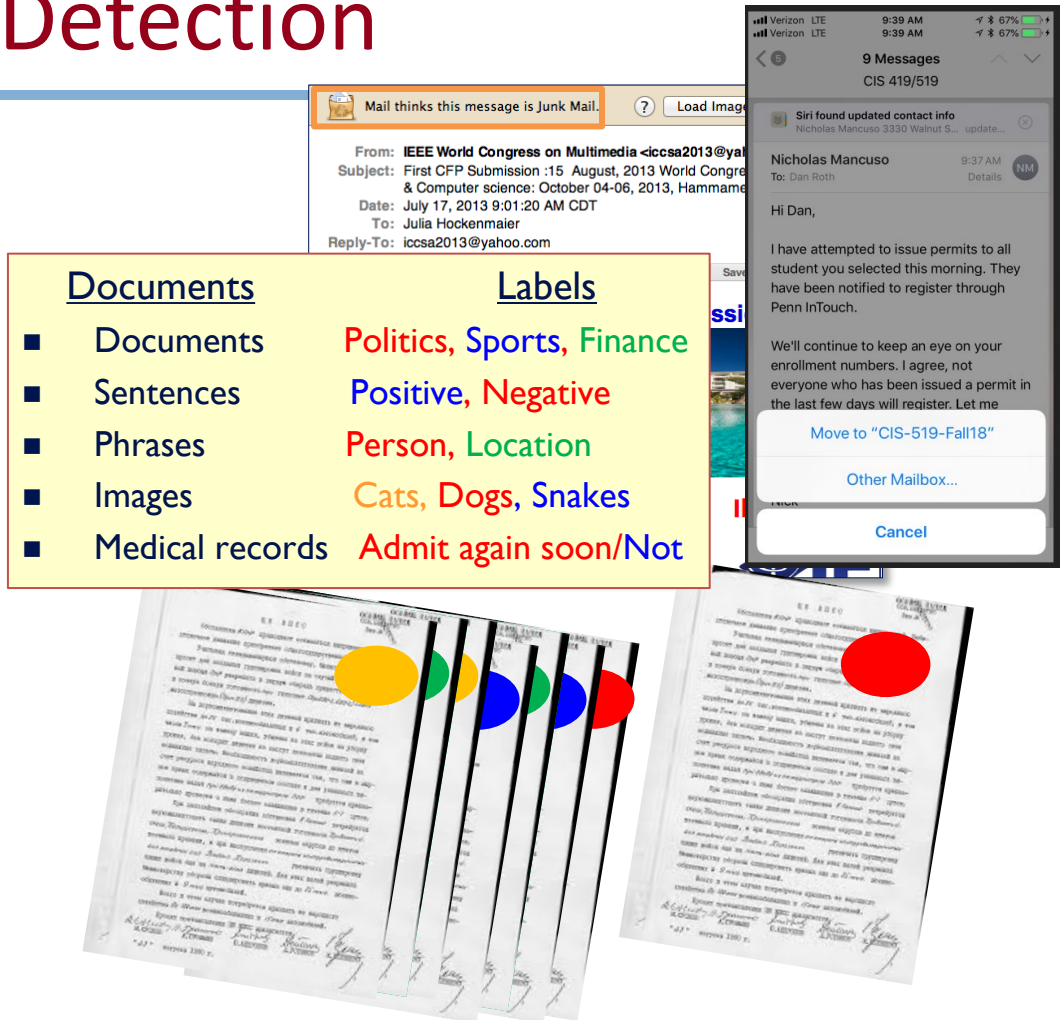
- I have a spelling checker, it came with my PC
- It plane lee marks four my revue
- Miss steaks aye can knot sea.
- Eye ran this poem threw it, your sure reel glad two no.
- Its vary polished in it's weigh
- My checker tolled me sew.
- A checker is a bless sing, it freeze yew lodes of thyme.
- It helps me right awl stiles two reed
- And aides me when aye rime.
- Each frays come posed up on my screen
- Eye trussed to bee a joule...

Machine Learning is Everywhere



Applications: Spam Detection

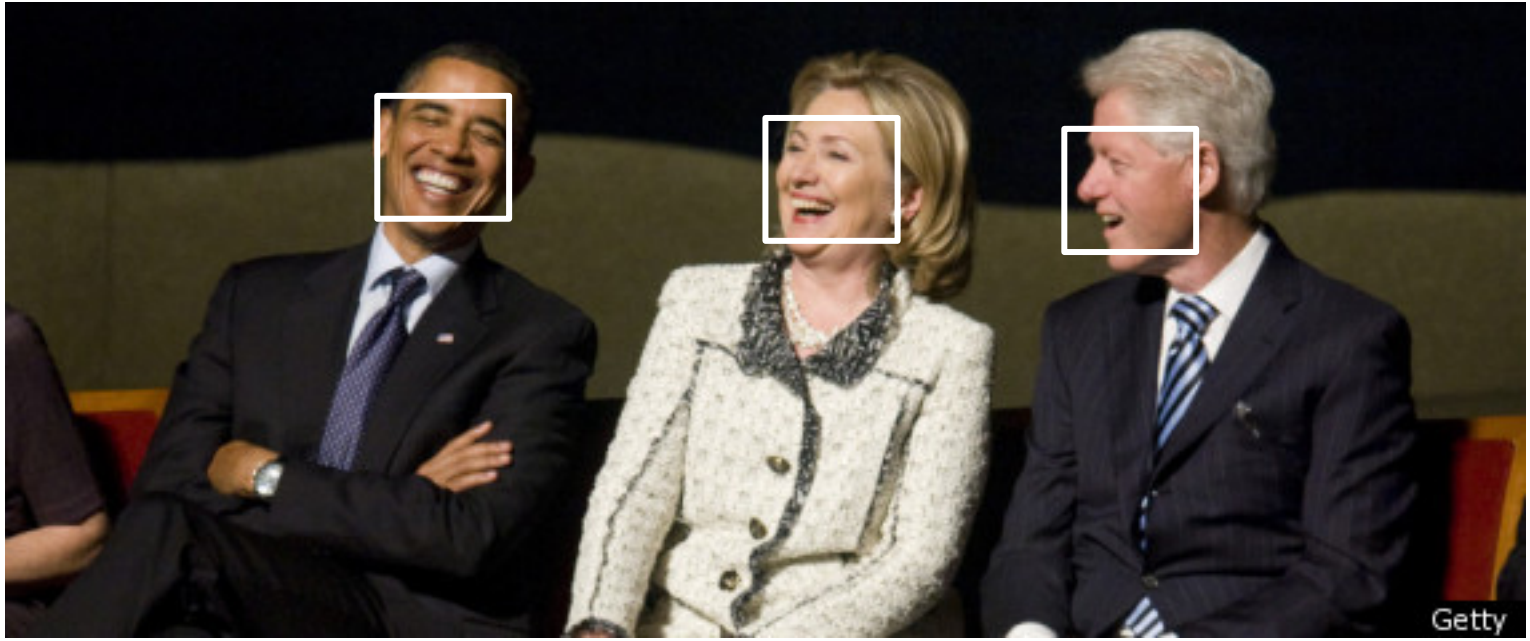
- This is a binary classification task: Assign one of two labels (i.e. yes/no) to the input (here, an email message)
- Classification requires a model (a classifier) to determine which label to assign to items.
- In this class, we study algorithms and techniques to learn such models from data.



Some More Involved Examples

- Driving:
  - https://www.youtube.com/watch?v=_1MHGUC_BzQ
  - E.g., go to 11:48
- Objects:
  - https://www.youtube.com/watch?v=_1MHGUC_BzQ
  - Go to 1:41
- Tesla Accidents:
  - https://www.youtube.com/watch?v=FVgkWii5JdM
  - Go to 1:45

Machine Learning: Any Limitations?



What else do you know given this image? (About the people, or more generally)

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

Some More Involved Examples

- Wikifier:
  - https://www.youtube.com/watch?v=IkryLTdogjw

- Some text generation:
  - https://transformer.huggingface.co/doc/arxiv-nlp

Automated Summarization (And hallucination)

A NYT article on the Beirut explosion

And its summary

Comprehension

(ENGLAND, June, 1989) – Christopher Robin is alive and well. He lives in England. He is the same person that you read about in the book, Winnie the Pooh. As a boy, Chris lived in a pretty home called Cotchfield Farm. When Chris was three years old, his father wrote a poem about him. The poem was printed in a magazine for others to read. Mr. Robin then wrote a book. He made up a fairy tale land where Chris lived. His friends were animals. There was a bear called Winnie the Pooh. There was also an owl and a young pig, called a piglet. All the animals were stuffed toys that Chris owned. Mr. Robin made them come to life with his words. The places in the story were all near Cotchfield Farm. Winnie the Pooh was written in 1925. Children still love to read about Christopher Robin and his animal friends. Most people don't know he is a real person who is grown now. He has written two books of his own. They tell what it is like to be famous.

1. Christopher Robin was born in England.
2. Winnie the Pooh is a title of a book.
3. Christopher Robin's dad was a magician.
4. Christopher Robin must be at least 65 now.

This is an Inference Problem; where is the learning?

Christopher Robin and Chris are the same person

Yes, they are

No, they are not

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

What do you see in this image?

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

Learning

- Learning is at the core of
  - Understanding High Level Cognition
  - Performing knowledge intensive inferences
  - Building adaptive, intelligent systems
  - Dealing with messy, real world data
  - Analytics
- Learning has multiple purposes
  - Knowledge Acquisition
  - Integration of various knowledge sources to ensure robust behavior
  - Adaptation (human, systems)
  - Decision Making (Predictions)

Learning = Generalization

- H. Simon -

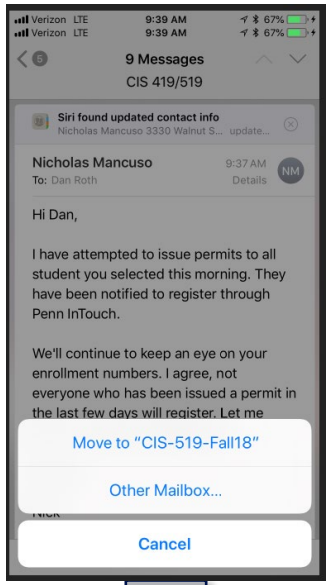
"Learning denotes changes in the system that are adaptive in the sense that they enable the system to do the task or tasks drawn from the same population more efficiently and more effectively the next time."

Remember this!
(and remember the Tesla example)

The ability to perform a task in a situation which has never been encountered before

Learning = Generalization

- The learner has to be able to classify items it has never seen before.



Mail thinks this message is about my Fall 2018 ML Class

How can we expect a program to make predictions on items it has never seen before? What should this program rely on?

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

Learning = Generalization

The ability to perform a task in a situation which has never been encountered before

- Classification
  - Medical diagnosis; credit card applications; hand-written letters; ad selection; sentiment assignment,…
- Planning and acting
  - Game playing (chess, backgammon, go); driving a car

The Badges game

- Skills
  - (A robot) balancing a pole; playing tennis; driving
- Common sense reasoning
  - Natural language interactions

What does the algorithm gets as input? (features)

Generalization depends on the Representation as much as it depends on the Algorithm used.

Same Population?

New Zealand

In ~~New York State~~, the longest period of daylight occurs during the month of _____.

Why Study Machine Learning?

- "A breakthrough in machine learning would be worth ten Microsofts"
  -Bill Gates, Chairman, Microsoft
- "Machine learning is the next Internet"
  -Tony Tether, Former Director, DARPA
- Machine learning is the hot new thing"
  -John Hennessy, President, Stanford
- "Machine learning is going to result in a real revolution"
  -Greg Papadopoulos, CTO, Sun
- "Machine learning is today's discontinuity"
  -Jerry Yang, CEO, Yahoo

Why Study Learning?

- Computer systems with new capabilities.
- AI
- Understand human and biological learning
- Understanding teaching better.
- Time is right.
  - Initial algorithms and theory in place.
  - Growing amounts of on-line data
  - Computational power available.
  - Necessity: many things we want to do cannot be done by “programming”.
  - (Think about all the examples given earlier)

Learning is the Future

- Learning techniques will be a basis for every application that involves a connection to the messy real world
- Basic learning algorithms are ready for use in applications today
- Prospects for broader future applications make for exciting fundamental research and development opportunities
- Many unresolved issues – Theory and Systems
  - While learning is hot, there are many things we don't know how to do
  - And that's why this is NOT a class about deep neural networks

❑ Very active field

❑ What to teach?

- ❑ The fundamental paradigms
- ❑ Some of the most important algorithmic ideas
- ❑ Modeling

And: what we don't know

US Open Highlights

- Over 700 matches over two weeks periods.
- A lot of manpower required to generate video highlights…
- Key: automate using incidental signals
  - Crowd noise, like gasps and cheers.
  - Players' gestures and reactions (e.g. celebratory air punches and fist pumps)
  - Non-trivial: need to deal with biases, etc.

https://venturebeat.com/2019/07/05/how-wimbledon-and-watson-are-using-ai-to-curate-video-highlights/

Course Overview

- Introduction: Basic problems and questions
- A detailed example: Linear classifiers; key algorithmic idea
- Two Basic Paradigms:
  - » Discriminative Learning & Generative/Probabilistic Learning
- Learning Protocols:
  - » Supervised; Unsupervised; Semi-supervised
- Algorithms
  - » Gradient Descent
  - » Decision Trees
  - » Linear Representations: (Perceptron; SVMs; Kernels)
  - » Neural Networks/Deep Learning
  - » Probabilistic Representations (naïve Bayes)
  - » Unsupervised /Semi supervised: EM
  - » Clustering; Dimensionality Reduction
- Modeling; Evaluation; Real world challenges
- Ethics

CIS 419/519: Applied Machine Learning

- Monday, Wednesday: 10:30pm-12:00pm On Zoom
- Office hours: Mon 5-6 pm; Tue 12-1pm
- 13 TAs
- Assignments: 5 Problems set (Python Programming)
  - Weekly (light) on-line quizzes
- Weekly Discussion Sessions
- Mid Term Exam (take home)
- [Project] (look at the schedule)
- Final (take home)
- No real textbook:
  - Slides/Mitchell/Goldberg/Other Books/Lecture notes /Literatur

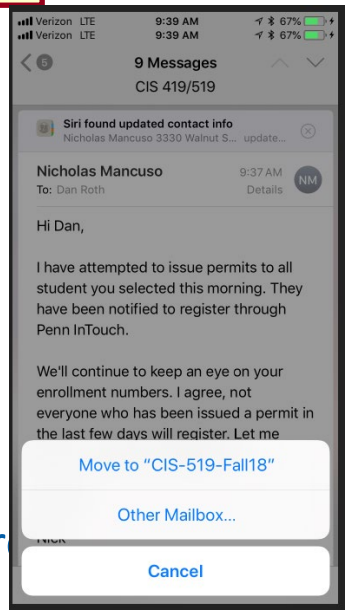
Starting next week TAs Office Hours will also start next week

HW0 !!!

Go to the web site

Be on Piazza

Registration for Class



CIS 519: What have you learned so far?

- What do you need to know:
  - Some exposure to:
    - Theory of Computation
    - Probability Theory
    - Linear Algebra
  - Programming (Python)
- Homework 0
  - If you could not comfortably deal with 2/3 of this within a few hours, please take the prerequisites first; come back next semester/year.

Participate, Ask Questions

Ask **during** class, not **after** class

- **Applied Machine Learning**
- Applied: mostly in HW
- **Machine learning:** mostly in class, quizzes, exams

CIS 519: Policies

- Cheating
  - No.
  - We take it very seriously.
- Homework:
  - Collaboration is encouraged
  - But, you have to write your own solution/code.
- Late Policy:
  - You have a credit of 4 days; That's it.
- Grading:
  - 40% - homework; ; 35%-final; 20%-midterm; 5% Quizzes;
  - Participation in Polls is mandatory (80% of the meetings)
  - [Projects: 20%]
- Questions?

Class' Web Page

Note also the Schedule Page and our Notes

CIS 519 on the web

- Check our class website:
  - Schedule, slides, videos, policies
    - https://www.seas.upenn.edu/~cis519/fall2020/index.html
    - Sign up, participate in our Piazza forum:
    - Announcements and discussions
    - https://piazza.com/class/kec0q01gqceim
  - Check out our team
    - Office hours
    - [Optional] Discussion Sessions

How will you rate today's lecture?

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

Did I ask too many questions?

Yes, way too many?

Just right

Not enough

I'll post my response to this on Piazza later today

None of the above

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

What is Learning?

- The Badges Game…
  - This is an example of the key learning protocol: supervised learning
- First question: Are you sure you got it?
  - Why?
- Issues:
  - Prediction or Modeling?
  - Representation
  - Problem setting
  - Background Knowledge
  - When did learning take place?
  - Algorithm