



University of Colorado **Boulder**

Department of Computer Science

CSCI 4622: Machine Learning

Chenhao Tan

Lecture 1: Introduction

# Basic Information

- Course location and time: ECCR 265, 3:00-3:50pm (MWF)
- Instructor: Chenhao Tan
- Teaching assistant: Amit Rege
- Graduate student staff: Cory Paik
- Course website: <https://github.com/BoulderDS/CSCI-4622-Machine-Learning-fa20>

# About me

- Research interests: human-centered machine learning, natural language processing, computational social science
- <https://chenhaot.com>

# COVID-19 requirements

- Maintain 6-foot distancing when possible,
- Wear a cloth face covering (over nose and mouth), especially when unable to maintain a distance of at least 12 feet,
- Clean local work area,
- Practice hand hygiene,
- Follow public health orders, and if sick and
  - you live off campus, do not come onto campus (unless instructed by a CU Healthcare professional), or
  - you live on-campus, please alert [CU Boulder Medical Services](#).
- Email me if you have any questions or have to miss classes/homeworks because you are sick or quarantined. No need to share any details.

# Outline

- An overview of machine learning
- Administrivia
- Syllabus

Machine learning is everywhere!

# Reading comprehension

The Normans (Norman: Nourmands; French: Normands; Latin: Normanni) were the people who in the 10th and 11th centuries gave their name to Normandy, a region in France. [...]

Q: In what country is Normandy located?

A: France

**Three lidar systems**

**A forward facing camera**

**Radar sensors**

**Self-driving sensors**





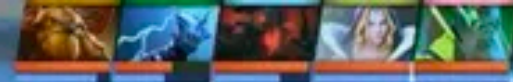


6  
▲ 1k



7:05

1



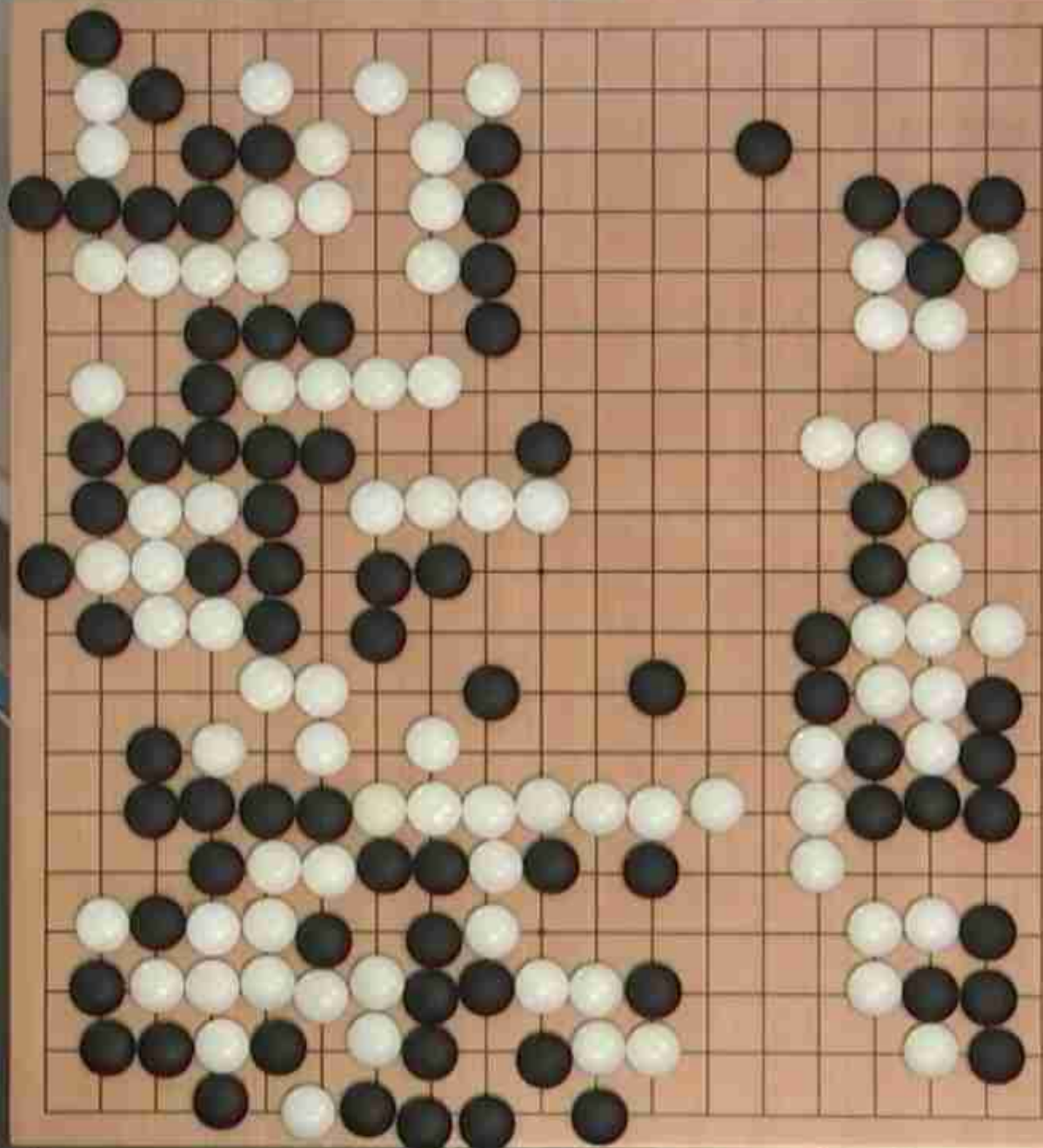
ES ▼

- ⋮
- ⋮
- ⋮





● ALPHAGO  
00:10:29



● LEE SEDOL  
00:01:00



## Popular on Netflix



## Dark Movies



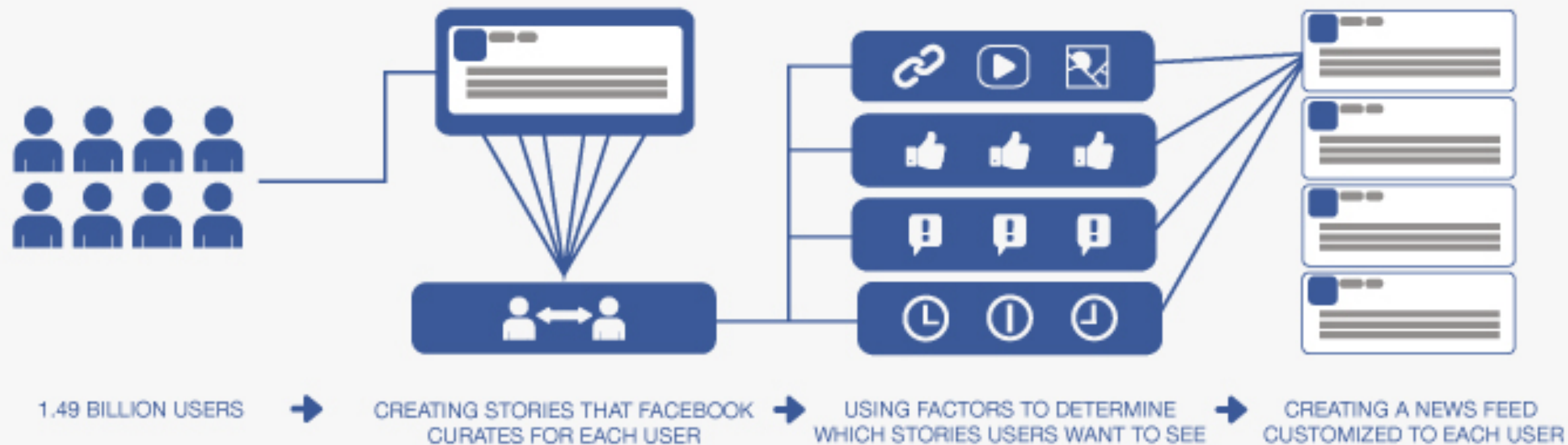
## Romantic Opposites-Attract...



## Emotional Movies



## FACEBOOK'S NEWS FEED & USER EXPERIENCE



# Which tweet will be retweeted more?



Food trucks are the epitome of small independently owned LOCAL businesses! Help keep them going! Sign the petition [bit.ly/P6GYCq](http://bit.ly/P6GYCq)



I know at some point you've have been saved from hunger by our rolling food trucks friends. Let's help support them! [bit.ly/P6GYCq](http://bit.ly/P6GYCq)



# Which tweet will be retweeted more?



cactus\_music  
@cactus\_music



Food trucks are the epitome of small independently owned LOCAL businesses! Help keep them going! Sign the petition [bit.ly/P6GYCq](http://bit.ly/P6GYCq)



cactus\_music  
@cactus\_music

I know at some point you've have been saved from hunger by our rolling food trucks friends. Let's help support them! [bit.ly/P6GYCq](http://bit.ly/P6GYCq)

[GET CHART](#)[COMPARE](#)[EVENTS ▾](#)[TECHNICAL INDICATORS ▾](#)[CHART SETTINGS ▾](#)[RESET](#)

Aug 29, 2012 2:59 PM - 3:04 PM EDT: ■ ^DJA 4416.03



© 2012 Yahoo! Inc.

■ Volume: 571,700

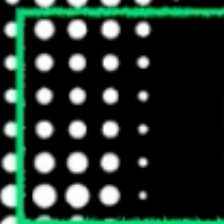
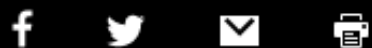


ANNALS OF MEDICINE APRIL 3, 2017 ISSUE

# A.I. VERSUS M.D.

*What happens when diagnosis is automated?*

By Siddhartha Mukherjee





# Machine learning is everywhere!

- Smart city
- Entertainment
- Social
- Finance
- Medical

Email me to introduce yourself, one of your core values, and a machine learning application you care about.

# What is machine learning?

One definition (Mitchell):

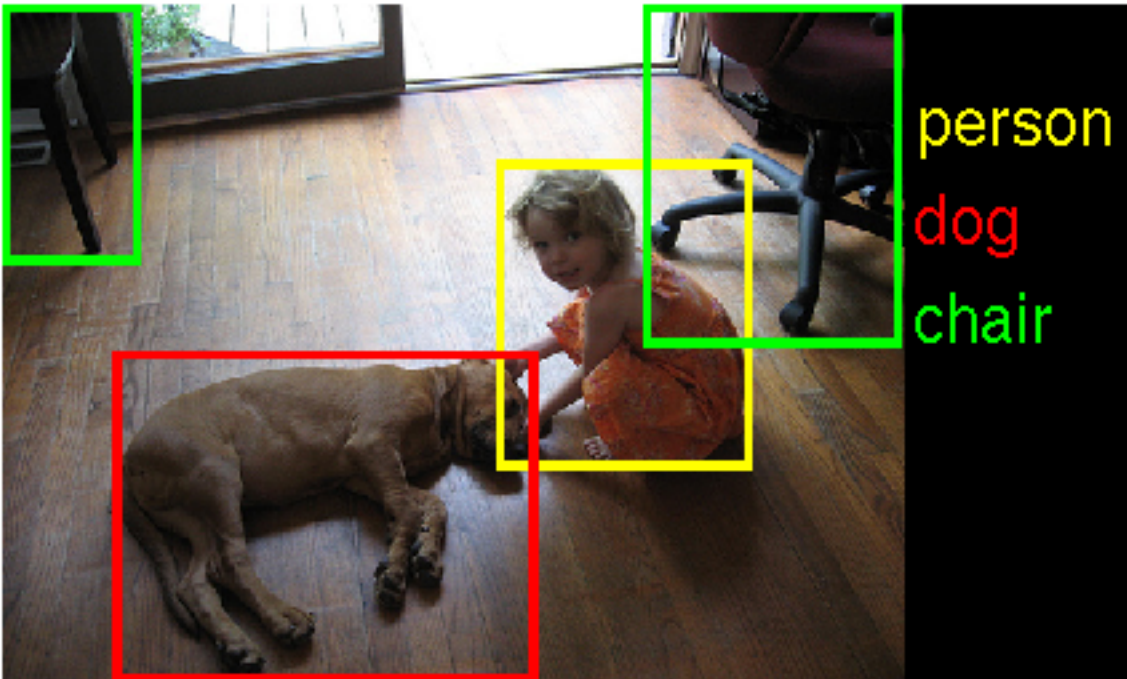
A computer program is said to learn from

- experience  $E$  with respect to some class of
- tasks  $T$  and
- performance measure  $P$ ,

If its performance at tasks in  $T$ , as measured by  $P$ , improves with experience  $E$ .

Let us apply this to classic tasks in machine learning!

# ImageNet/Object recognition



- $T$ : identifying objects in an image
- $E$ : tons of images with annotated objects
- $P$ : how often the objects are identified correctly

# Sentiment analysis



## Hutzler #571 Banana Slicer

The only banana slicer you will ever need.

Gourmac's easy-to-use Banana Slicer provides a quick solution to slice a banana uniformly each and every time. Simply press the slicer on a peeled banana and the work is done. Safe, fun and easy for children to use. Kids just love eating bananas with this as their favorite kitchen tool. The Banana Slicer may also be used as a quick way to add healthy bananas to breakfast cereal or to make uniform slices for a fruit salad or ice cream dessert.

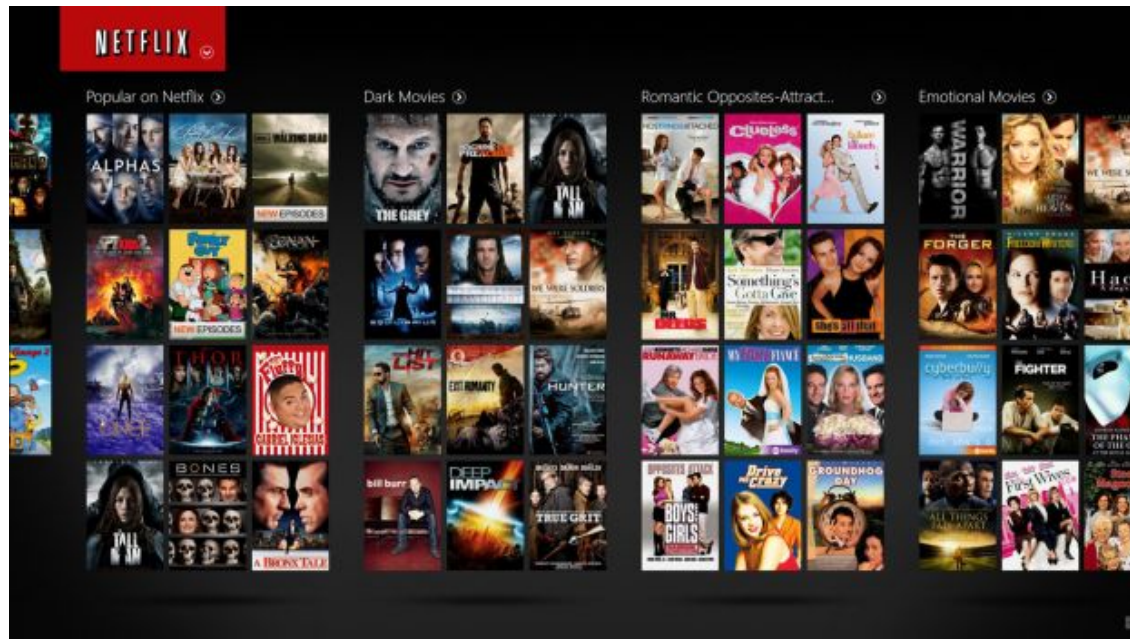
★ ★ ☆ ☆ ☆ **Angle is wrong**

By Jim Anderson on August 1, 2012

I tried the banana slicer and found it unacceptable. As shown in the picture, the slicer is curved from left to right. All of my bananas are bent the other way.

- $T$ : deciding whether a review is positive or negative
- $E$ : reviews with ratings
- $P$ : how often the sentiment is predicted correctly

# Movie recommendation



- $T$ : recommend movies
- $E$ : movie watching history and movie ratings
- $P$ : future ratings of users? User active time on website? User subscription periods?

# Why do we want machines to learn?

- Intellectually satisfying!
- Helping us solve problems

Automate tasks that we know how to perform

- Object recognition
- Driving

Emulate human ability

Explore tasks that we don't know how to perform

- Predict message popularity
- Movie recommendation
- Newsfeed ranking

Extend human ability

# Human-centered machine learning

Help humans solve critical societal challenges

- Medical diagnosis
- Recidivism prediction
- Personalized education



# How does machine learning work as of today?

- Collect or happen upon data ( $X$ , experience in the previous definition)
- Analyze it to find patterns
- Use those patterns to perform some task ( $T$ )

We will study algorithms that find and exploit patterns in data.

# We will study algorithms that find and exploit patterns in data.

- Goal: fluency in thinking about modern machine learning algorithms
- We will learn about
  - When to use them
  - The assumptions they make about data
  - Their capabilities and their limitations
  - Theoretical guarantees
- We will learn a language and process for solving data analysis problems.

# Supervised vs. unsupervised methods



## Hutzler #571 Banana Slicer

The only banana slicer you will ever need.

Gourmac's easy-to-use Banana Slicer provides a quick solution to slice a banana uniformly each and every time. Simply press the slicer on a peeled banana and the work is done. Safe, fun and easy for children to use. Kids just love eating bananas with this as their favorite kitchen tool. The Banana Slicer may also be used as a quick way to add healthy bananas to breakfast cereal or to make uniform slices for a fruit salad or ice cream dessert.

Data:  $X$       Labels:  $Y$

- **Supervised methods** find patterns in **fully observed** data and then try to predict something from **partially observed** data.
- For example, in sentiment analysis, after learning something from annotated reviews, we want to take new reviews and automatically identify sentiments.

# Supervised vs. unsupervised methods



Data:  $X$       Hidden  
Structure:  $Z$

- **Unsupervised methods** find **hidden structure** in data, structure that we can never formally observe.
- For example, modeling topics from a collection of scientific papers; evaluation is usually more difficult.

# Outline

- An overview of machine learning
- **Administrivia**
- Syllabus

# Course information

- Course on github: <https://github.com/BoulderDS/CSCI-4622-Machine-Learning-fa20>
- Piazza (all Q&A should happen here): <https://piazza.com/class/kdz12mi733y1f1>
- Canvas (all grading happens here; all zoom links are here): <https://canvas.colorado.edu/courses/63981>
- Lecture videos: <https://cu-classcapture.colorado.edu/Mediasite/Channel/f5adfe679d034d7fb60d9ffbdda2da5c5f>

# COVID-19 requirements

- Maintain 6-foot distancing when possible,
- Wear a cloth face covering (over nose and mouth), especially when unable to maintain a distance of at least 12 feet,
- Clean local work area,
- Practice hand hygiene,
- Follow public health orders, and if sick and
  - you live off campus, do not come onto campus (unless instructed by a CU Healthcare professional), or
  - you live on-campus, please alert [CU Boulder Medical Services](#).
- Email me if you have any questions or have to miss classes/homeworks because you are sick or quarantined. No need to share any details.

# Grading policy

- Homework (30%)
- Exams (24%)
- Final project (36%)
- Participation (5%)
- Living in 2020 (5%)



# Grading policy

- Homework (30%) (Count the highest four, and NO LATE SUBMISSIONS.)
- Exams (24%)
- Final project (36%)
- Participation (5%)
- Living in 2020 (5%)

# Final project

- Competition or replication or research
  - Kaggle competitions
  - Replicate a research paper
  - Work on something related to your interest

# Final project

- Group formation due: Oct 9
- Final project proposal due: Oct 23
- Final project midpoint report (this can be a short update): Nov 18
- Hackathon: Nov 30 --- Dec 6
- Final project presentation: Dec 7
- Final project report due: Dec 9

# Learning is a two-way process

- Private emails
  - Only for feedback, all Q&A should happen through Piazza
- Periodic survey

# Outline

- An overview of machine learning
- Administrivia
- **Syllabus**

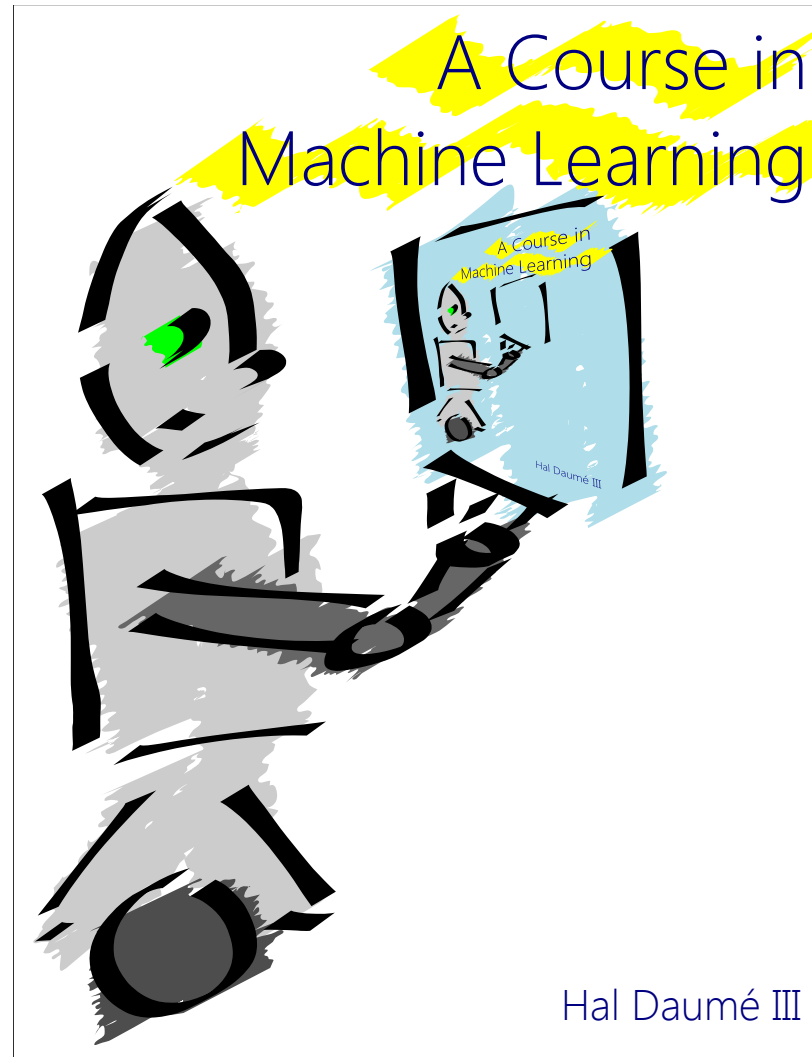
# Preliminary schedule

- Supervised learning
- Unsupervised learning
- Maybe learning theory
- Ethics
- <https://github.com/BoulderDS/CSCI-4622-Machine-Learning-fa20/blob/master/info/schedule.md>

# Prerequisites

- Programming language: Python
- Math background:
  - Probability
  - Linear algebra
  - Calculus
  - *Information theory*
  - *Optimization*

# Course textbook





# Q & A

- Email me to introduce yourself, one of your core values, and a machine learning application you care about.