CS 760: Machine Learning
Course Overview

Fred Sala

University of Wisconsin-Madison

Sept. 9, 2021

Logistics: Lecture Location

- In-person in Social Sciences 5208

- We will **record** and make these available after class
  - Possibly live-stream; depends on whether it's needed

Logistics: Enrollment

- Currently at capacity, approx. 125 students
  - Most folks on waitlist may not make it in
  - **Sorry** ☹ … will be offered again

Logistics: Teaching Team

Instructor: **Fred Sala**

- Location: CS 5385
- Office Hours: W 1:00-2:30 PM / by appointment

TAs: **Changho Shin, Harit Vishwakarma**

- Changho OH Tues 4:00-6:00pm
- Harit OH Thu 4:00-6:00pm

  - Note: times possibly **subject to change**

Logistics: Teaching Team

Two more assistants:

Felix +

Arthur

Note: if I’m late replying to anything, they’re the **cause** ☺

Logistics: Content

Three locations:

- 1. **Course website**: https://pages.cs.wisc.edu/~fredsala/cs760/fall2021/

- 2. **Piazza**. https://piazza.com/wisc/fall2021/cs760
  - access code: introml
  - **Preferred for questions!**

- 3. **Canvas**

Flickr Photo by Jacob Surland - creative commons licensed (BY-NC)

Logistics: Lecture Format

Typically, 75 minutes

- 1-2 breaks with quizzes (ungraded; for understanding only)
- Can also ask questions

We'll post slides on website **before class**

We'll post quizzes **after class**

Logistics: Assignments & Grades

Homeworks:

- 8 or so, worth 30% total
- Posted after class; due when class starts on due date

Exams:

- Midterm: 20%, Oct. 27
- Final: 20%, Dec. 20

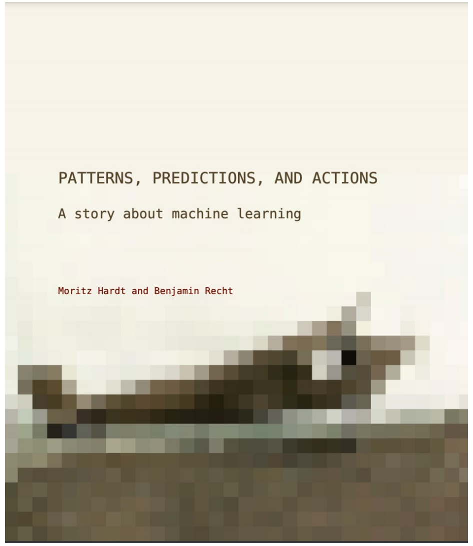
Final Project:

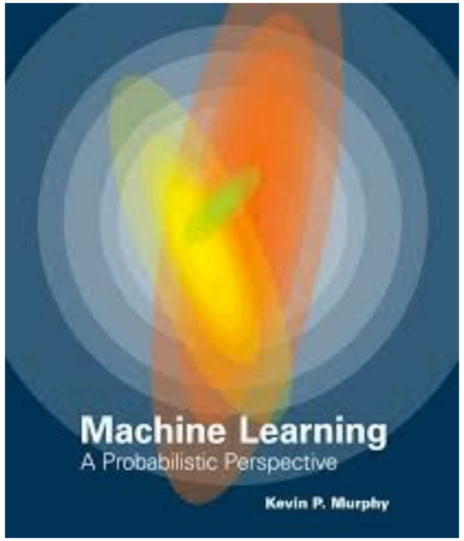
- 30% total, groups of 2-5; proposal midway. **More info soon!**

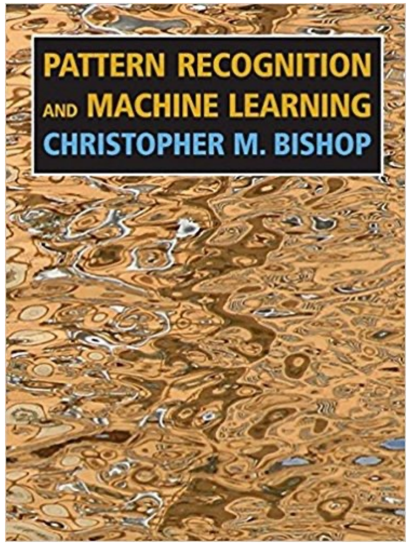
Class Setup: Reading

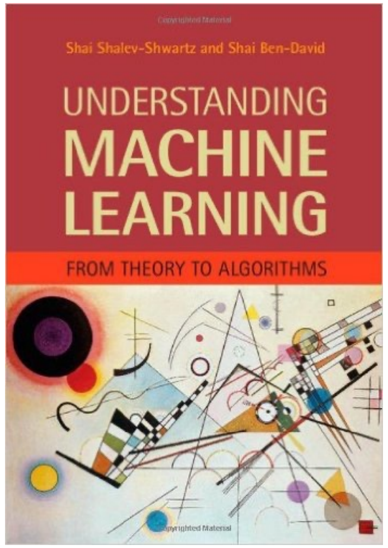
No required textbook, but you should read from the below

- Should all be available online / digital library access
- Will also post articles, papers to read









Class Setup: Background

More on this at the end of class, but

- **Linear algebra** (working with data, linear transformations)
- **Calculus** (for optimization, convergence, etc.)
- **Probability** (dealing with noise, sampling)
- **Programming** (for implementation)

Plenty of resources available

- Just need enough experience/mathematical maturity to pick up missing bits

Class Setup: Goals

Two goals:

- **Understanding** ML
- **Foundation** for future research in ML

If you just want to **use** ML, but do not plan to do research, consider taking:

- CS540
- STAT 451
- ECE/CS/ME 532

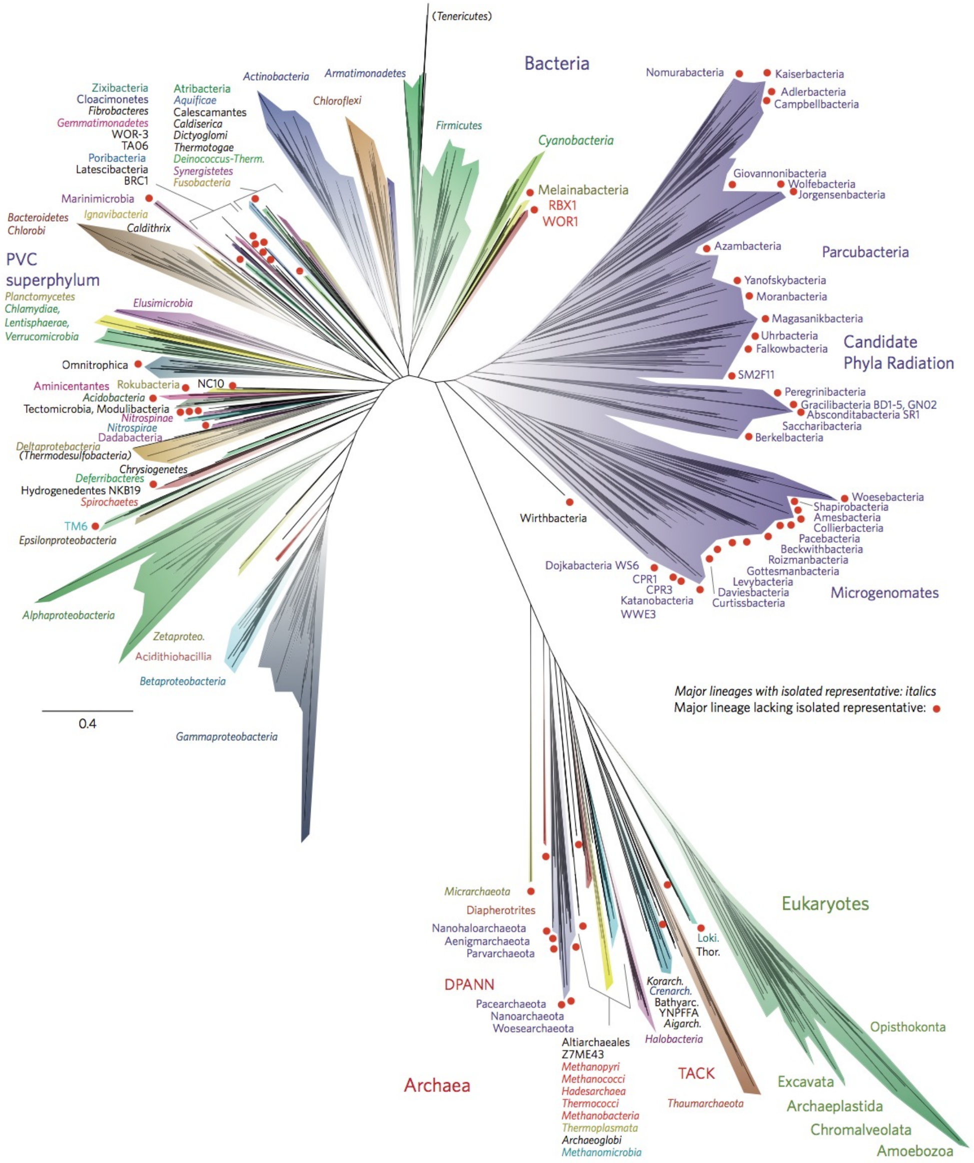
Class Setup: Goals II

Mini-goals:

- **Intuition for each algorithm/model**
- **Big picture/ML ecosystem**

Examples:

- What types & how much data?
- How hard to train?
- What generalizes best?
- Where is the field going?

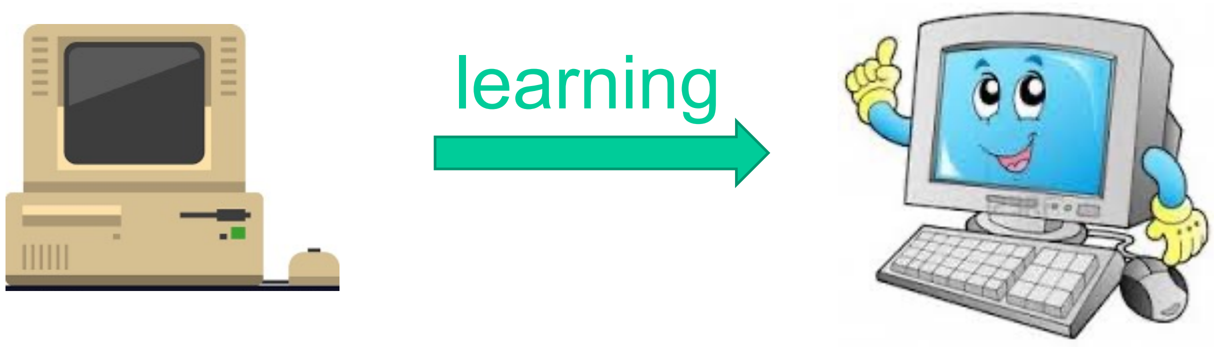


Break & Questions

ML Overview: Definition

What is machine learning?

"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**." *Machine Learning*, Tom Mitchell, 1997



ML Overview: Motivation

Why would we do this?

- We're building a self-driving car. Could just write down rules
  - **Painful**! A lot of cases…

```
/**
 * controls steering of the car
 * @param angle
 * @param trim
 */
void steer(float angle, float trim = 0.0) {
    // seems like 360 right 520 left
    PWMPCA9685Device device = new PWMPCA9685Device()
    device.setPWMFrequency(50) //internet says 50hz for servos is optimal
    Servo servo0 = new PCA9685Servo(device.getChannel( channel: 1))
    LOG.info("steer angle non corrected:${angle} trim:${trim}")
    if (trim != 0) {
        trim = configTrim
        servo0.setTrim(trim)
    }
    servo0.setInput((angle).toFloat())
    System.out.println("configTrim in service=${configTrim}")
    Thread.sleep( millis: 1000) // important to give time for servo to move

}
```

ML Overview: Motivation

Why would we do this?

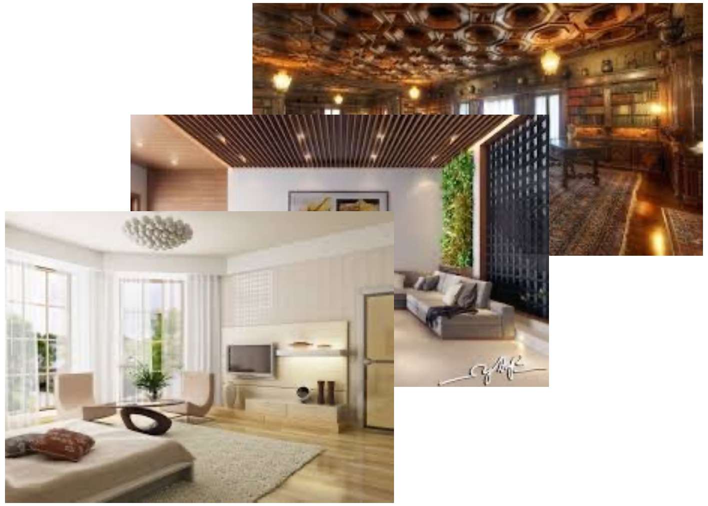
- We're building a self-driving car. Could just write down rules
  - **Painful**! A lot of cases…
  - **Learn from examples** instead

Waymo

ML Overview: Flavors

Supervised Learning

- Learning from examples, as above
- **Workflow**:
  - Collect a set of examples {data, labels}: **training set**
  - “**Train**” a model to match these examples
  - ”**Test**” it on new data
- **Image classification**:



indoor

outdoor

ML Overview: Flavors

Supervised Learning

- **Example: Image classification**
- Recall **T**ask/**P**erformance measure/**E**xperience definition
  - **T**ask: distinguish **indoor** vs **outdoor**
  - **P**erformance measure: probability of misclassifying
  - **E**xperience: labeled examples

Modality: **images**

indoor

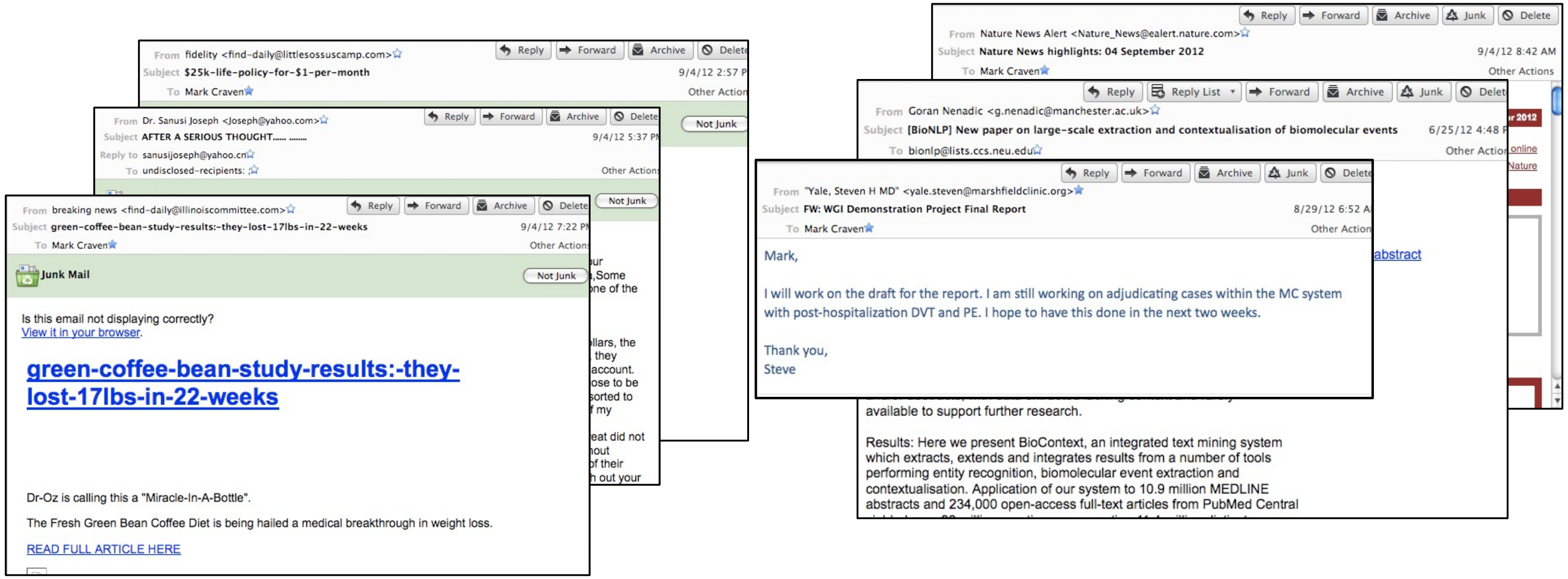
outdoor

ML Overview: Flavors

Supervised Learning

- **Example: Spam Filtering**
  - **T**ask: distinguish **spam** vs **legitimate**
  - **P**erformance measure: probability of misclassifying
  - **E**xperience: labeled examples of messages/emails

Modality: **text**

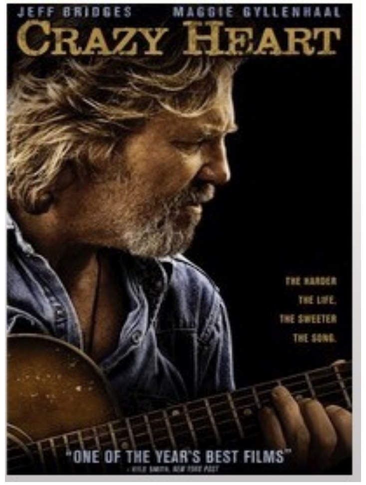


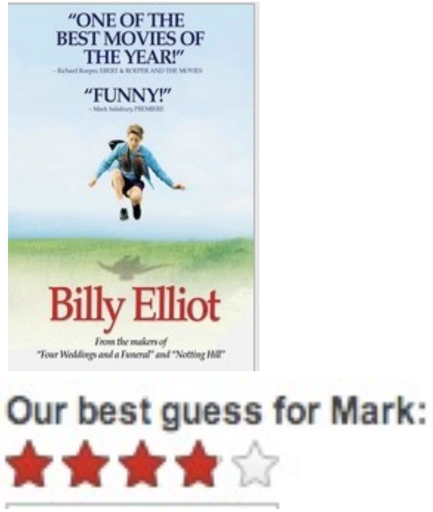
ML Overview: Flavors

Supervised Learning

- **Example: Ratings/Recommendations**
  - **T**ask: predict how much a user will like a film
  - **P**erformance measure: distance from user's rating
  - **E**xperience: previous ratings

Modality: **lots**

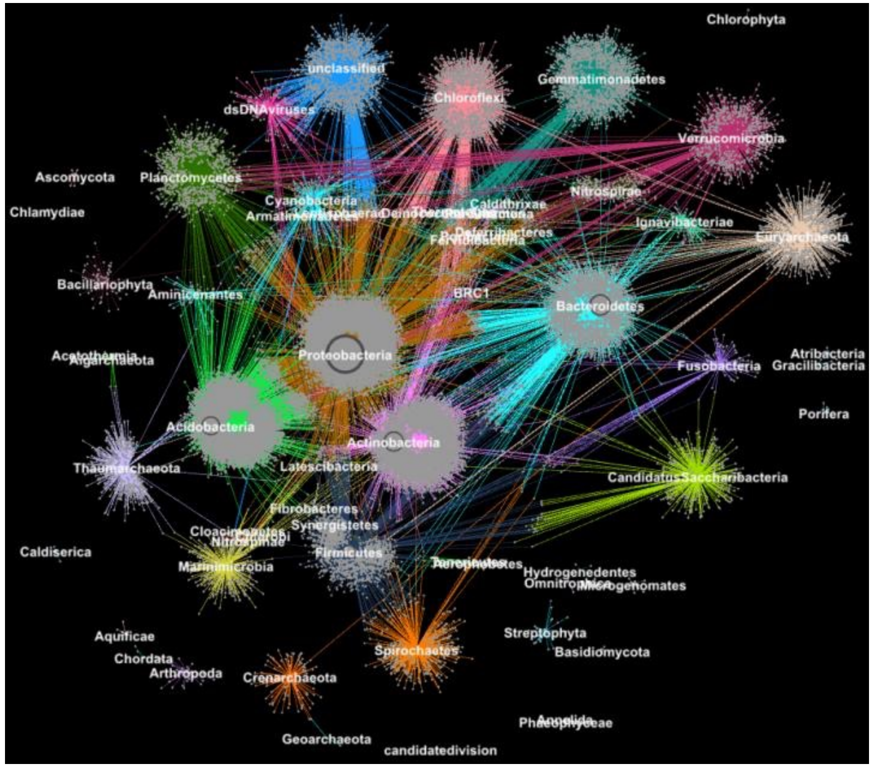




ML Overview: Flavors

Unsupervised Learning

- Data, but no labels. No input/output.
- Goal: get "something": structure, hidden information, more
- **Workflow**:
  - Collect a set {data}
  - Perform some algorithm on it

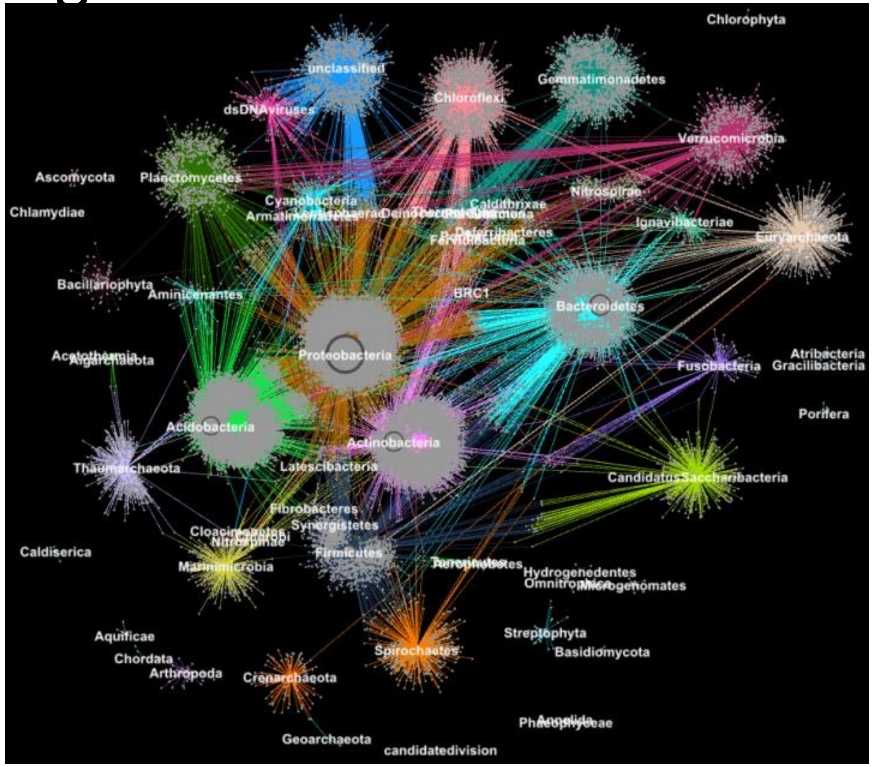


- **Clustering**: reveal hidden structure

ML Overview: Flavors

Unsupervised Learning

- **Example: Clustering**
  - **T**ask: produce distinct clusters for a set of data
  - **P**erformance measure: closeness to underlying structure
  - **E**xperience: available datapoints



Modality: **lots**

ML Overview: Flavors

Unsupervised Learning

- **Example: Generative Models**
  - **T**ask: produce artificial images of faces
  - **P**erformance measure: photorealism
  - **E**xperience: available images

Modality: **images**

StyleGAN2 (Kerras et al '20)

ML Overview: Flavors

Reinforcement Learning

- Agent interacting with the world; gets rewards for actions
- Goal: learn to perform some activity
- **Workflow**:
  - Create an environment, reward, agent
  - **Train**: modify policy to maximize rewards
  - **Deploy** in new environment

- **Controlling aircraft**: learn to fly

ML Overview: Flavors

Reinforcement Learning

- **Example: Controlling aircraft**
  - **T**ask: keep the aircraft in the air
  - **P**erformance measure: reward for following trajectory
  - **E**xperience: state/action/reward from previous flights

Modality: **video/sensor data**

ML Overview: Flavors

Reinforcement Learning

- **Example: Playing video games**
  - **T**ask: play Atari arcade games
  - **P**erformance measure: winning/advancing
  - **E**xperience: state/action/reward from previous gameplay episodes

Modality: **video/sensor data**

Break & Questions

Assignment: Reading

For HW1, article by Jordan and Mitchell on course website

REVIEW

Machine learning: Trends, perspectives, and prospects

M. I. Jordan[1*] and T. M. Mitchell[2*]

Machine learning addresses the question of how to build computers that improve automatically through experience. It is one of today's most rapidly growing technical fields, lying at the intersection of computer science and statistics, and at the core of artificial intelligence and data science. Recent progress in machine learning has been driven both by the development of new learning algorithms and theory and by the ongoing explosion in the availability of online data and low-cost computation. The adoption of data-intensive machine-learning methods can be found throughout science, technology and commerce, leading to more evidence-based decision-making across many walks of life, including health care, manufacturing, education, financial modeling, policing, and marketing.

Machine learning is a discipline focused on two interrelated questions: How can

ance when executing some task, through some type of training experience. For example, in learn-

Assignment: Homework

For HW1, self-diagnostic on background. Topics:

- Linear Algebra
- Calculus
- Probability
- Big-O notation
- Basic programming skills

- If these feel very unfamiliar, talk to us

Assignment: Homework

For HW1, self-diagnostic on background. **Examples**:

Consider the matrix X and the vectors $\mathbf{y}$ and $\mathbf{z}$ below:

$$X = \begin{pmatrix} 9 & 8 \\ 7 & 6 \end{pmatrix} \qquad \mathbf{y} = \begin{pmatrix} 9 \\ 8 \end{pmatrix} \qquad \mathbf{z} = \begin{pmatrix} 7 \\ 6 \end{pmatrix}$$

1. Is X invertible? If so, give the inverse, and if no, explain why not.

2. If $y = \tan(z)x^{6z} - \ln(\frac{7x+z}{x^4})$, what is the partial derivative of y with respect to x?

Assignment: Homework

For HW1, self-diagnostic on background. **Examples**:

Match the distribution name to its probability density / mass function. Below, $|\boldsymbol{x}| = k$.

(a) Laplace

(b) Multinomial

(c) Poisson

(d) Dirichlet

(e) Gamma

(f) $f(\boldsymbol{x}; \boldsymbol{\Sigma}, \boldsymbol{\mu}) = \frac{1}{\sqrt{(2\pi)^k \boldsymbol{\Sigma}}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})\right)$

(g) $f(x; n, \alpha) = \binom{n}{x} \alpha^x (1-\alpha)^{n-x}$ for $x \in \{0, \ldots, n\}$; 0 otherwise

(h) $f(x; b, \mu) = \frac{1}{2b} \exp\left(-\frac{|x-\mu|}{b}\right)$

(i) $f(\boldsymbol{x}; n, \boldsymbol{\alpha}) = \frac{n!}{\Pi_{i=1}^k x_i!} \Pi_{i=1}^k \alpha_i^{x_i}$ for $x_i \in \{0, \ldots, n\}$ and $\sum_{i=1}^k x_i = n$; 0 otherwise

(j) $f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$ for $x \in (0, +\infty)$; 0 otherwise

(k) $f(\boldsymbol{x}; \boldsymbol{\alpha}) = \frac{\Gamma(\sum_{i=1}^k \alpha_i)}{\prod_{i=1}^k \Gamma(\alpha_i)} \prod_{i=1}^k x_i^{\alpha_i - 1}$ for $x_i \in (0, 1)$ and $\sum_{i=1}^k x_i = 1$; 0 otherwise

(l) $f(x; \lambda) = \lambda^x \frac{e^{-\lambda}}{x!}$ for all $x \in Z^+$; 0 otherwise

Assignment: Homework

For HW1, self-diagnostic on background. **Examples**:

Draw the regions corresponding to vectors $\mathbf{x} \in \mathbb{R}^2$ with the following norms:

1. $||\mathbf{x}||_1 \leq 1$ (Recall that $||\mathbf{x}||_1 = \sum_i |x_i|$)
2. $||\mathbf{x}||_2 \leq 1$ (Recall that $||\mathbf{x}||_2 = \sqrt{\sum_i x_i^2}$)
3. $||\mathbf{x}||_\infty \leq 1$ (Recall that $||\mathbf{x}||_\infty = \max_i |x_i|$)

Resources

Probability

- Lecture notes: http://www.cs.cmu.edu/~aarti/Class/10701/recitation/prob_review.pdf

Linear Algebra:

- Short video lectures by Prof. Zico Kolter: http://www.cs.cmu.edu/~zkolter/course/linalg/outline.html
- Handout associated with above video: http://www.cs.cmu.edu/~zkolter/course/linalg/linalg_notes.pdf
- Book: Gilbert Strang. Linear Algebra and its Applications. HBJ Publishers.

Big-O notation:

- http://www.stat.cmu.edu/~cshalizi/uADA/13/lectures/app-b.pdf
- http://www.cs.cmu.edu/~avrim/451f13/recitation/rec0828.pdf

Post others you like!

Thanks Everyone!

Some of the slides in these lectures have been adapted/borrowed from materials developed by Mark Craven, David Page, Jude Shavlik, Tom Mitchell, Nina Balcan, Elad Hazan, Tom Dietterich, Pedro Domingos, Jerry Zhu, Yingyu Liang, Volodymyr Kuleshov