Machine Learning: Course Overview

CS 760@UW-Madison

Class enrollment

- typically the class was limited to 30
- we've allowed ~100 to register
- the waiting list full

- unfortunately, many on the waiting list will not be able to enroll
- but CS760 will be offered in the next semester!

Teaching team

- Yingyu Liang
  email: yliang@cs.wisc.edu
  office hours: Thu 4-5pm
  office: BBCollaborate Ultra on Canvas

- TAs: Xufeng Cai, Diwanshu Jain
  email: xcui74@wisc.edu, djain23@wisc.edu
  office hours: Thu 8:30-9:30pm (Cai), TBD (Jain)
  office: BBCollaborate Ultra on Canvas

- Graders: Shah Nisarg, Sharma Vedang

- More Information on course website:
  http://pages.cs.wisc.edu/~yliang/cs760_spring21/

Attending lecture synchronously

If you can join CS 760 M/W/F at 2:30pm CT:

- (Optional) download slides
  - Course website → Schedule → [Slides]
- Sign in to Canvas
- Join BBCollaborate Ultra session “Lecture”

- Lecture block 1:
  - Watch video
    - Can stay in BBCollaborate Ultra
    - If that fails, stream in Kaltura Gallery
    - If that fails, download from Kaltura Gallery
  - Q&A: Ask questions by “raise hands”, discuss
  - Quizzes: Short ungraded quizzes to check concepts
- Lecture block 2 (same as 1)
- Lecture block 3 (same as 1)
- After class, check or post questions on Piazza

Attending lecture asynchronously

If you cannot join CS 760 M/W/F at 2:30pm CT:

- (Optional) download slides
- Sign in to Canvas

- Watch 3 lecture videos in Kaltura Gallery
- After class, watch BBCollaborate Ultra recordings to listen to Q&A/Quizzes
  - BBCollaborate Ultra → menu → Recordings
  - Quizzes: download from course website, next to slides
- Check or post questions on Piazza

Where to find content

- Canvas – private materials that should not be shared
  - Videos
  - Assignments
  - Grades
- Course website – public materials
  - Slides
  - Schedule
  - Policies
- Piazza
  - Discussion, questions
  - Announcements

Monday, Wednesday *and* Friday?

- we'll have ~30 lectures in total, just like a standard TR class
- can push the lectures forward (finish early, leave time for review)

- ***see the schedule on the course website***

Course emphases

- a variety of learning settings: supervised learning, unsupervised learning, reinforcement learning, etc.
- a broad toolbox of machine-learning methods: decision trees, nearest neighbor, neural nets, Bayesian networks, SVMs, etc.
- some underlying theory: bias-variance tradeoff, PAC learning, mistake-bound theory, etc.
- experimental methodology for evaluating learning systems: cross validation, ROC and PR curves, hypothesis testing, etc.

Major goals

1. Understand what a learning system should do
2. Understand how (and how well) existing systems work

Emphasize on understanding, laying the foundation for future research in machine learning.

If you just want to **use** machine learning, but do not plan to do **research** in machine learning, better to take:

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

Course requirements

- 7-8 homework assignments: 60%
  - programming
  - computational experiments (e.g., measure the effect of varying parameter x in algorithm y)
  - some written exercises
  - **post on Canvas; submit your solutions on Canvas**
  - will drop the lowest scored one in calculating the final score

- Midterm Exam: 20%
- Final Exam: 20%

Expected background

- CS 540 (Intro to Artificial Intelligence) or equivalent
- good programming skills
- probability/statistics
- linear algebra
- calculus, including partial derivatives

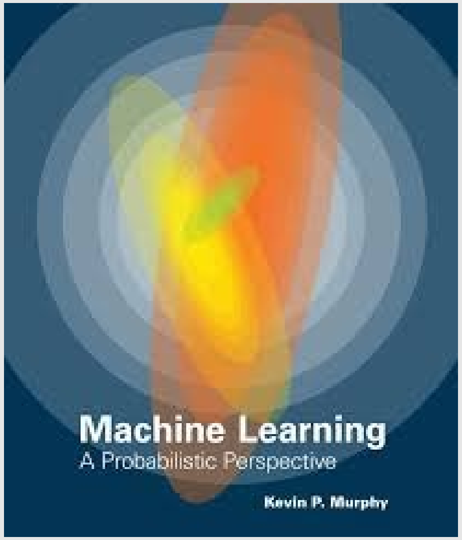
Programming languages

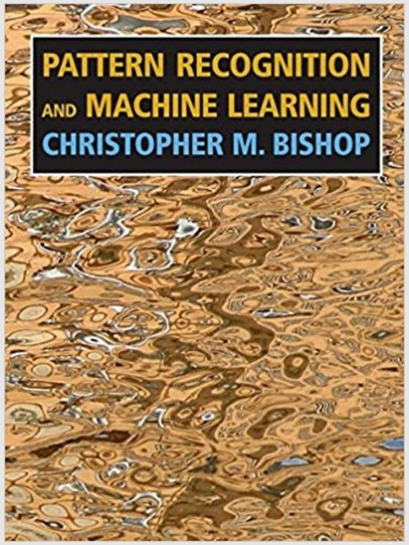
- for the programming assignments, you can use
  - C
  - C++
  - Java
  - Perl
  - Python
  - R
  - Matlab
- **Highly suggest: Python**
- **programs must be callable from the command line and *must run on the CS lab machines (this is where they will be tested during grading!)***

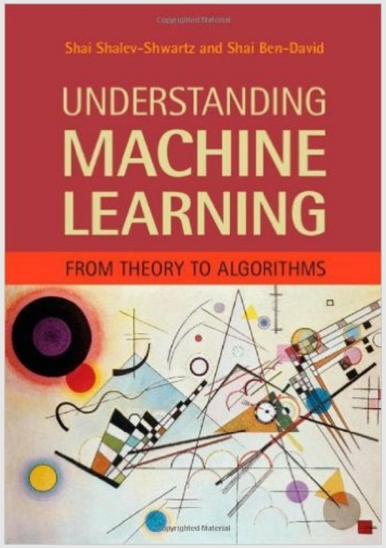
Course readings

Optional but recommend <u>one</u> of the following books

- *Pattern Recognition and Machine Learning.* C. Bishop. Springer, 2011.
- *Machine Learning: A Probabilistic Perspective*. K. Murphy. MIT Press, 2012.
- *Understanding Machine Learning: From Theory to Algorithms.* S. Shalev-Shwartz, S. Ben-David. Cambridge University press, 2014.







Course readings

- the books can be found online or at Wendt Commons Library

- additional readings will come from online articles, surveys, and chapters
- will be posted on course website

Machine Learning Examples

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.”

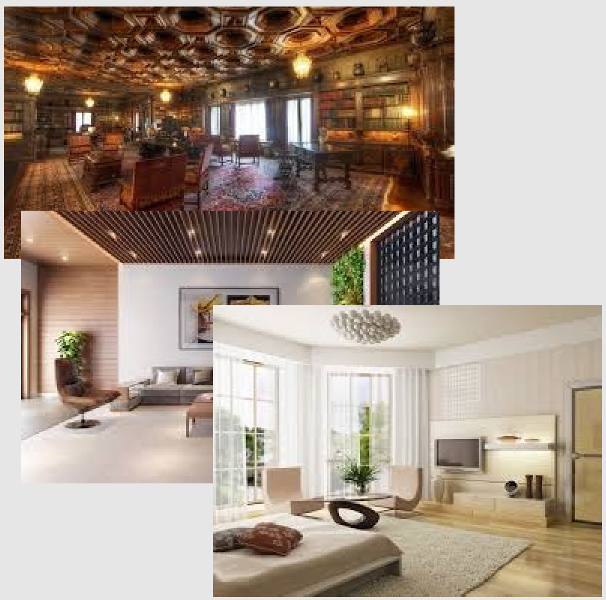
------- *Machine Learning*, Tom Mitchell, 1997

What is machine learning?

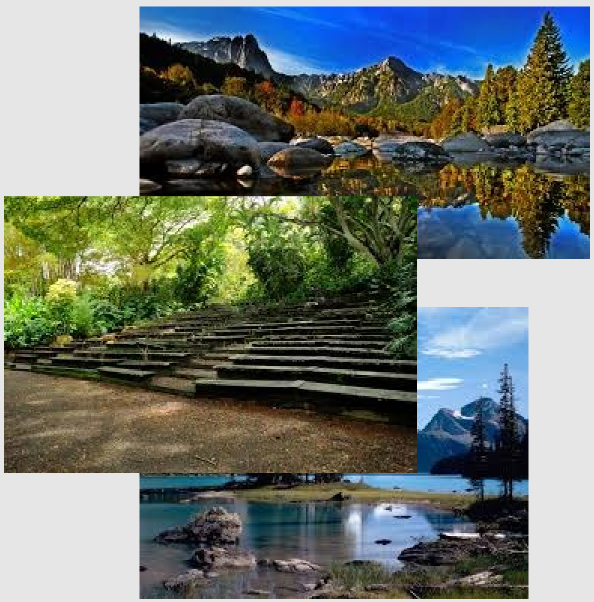
- the study of algorithms that
  improve their performance P
  at some task T
  with experience E

- to have a well-defined learning task, we must specify: $< P, T, E >$

ML example: image classification



indoor

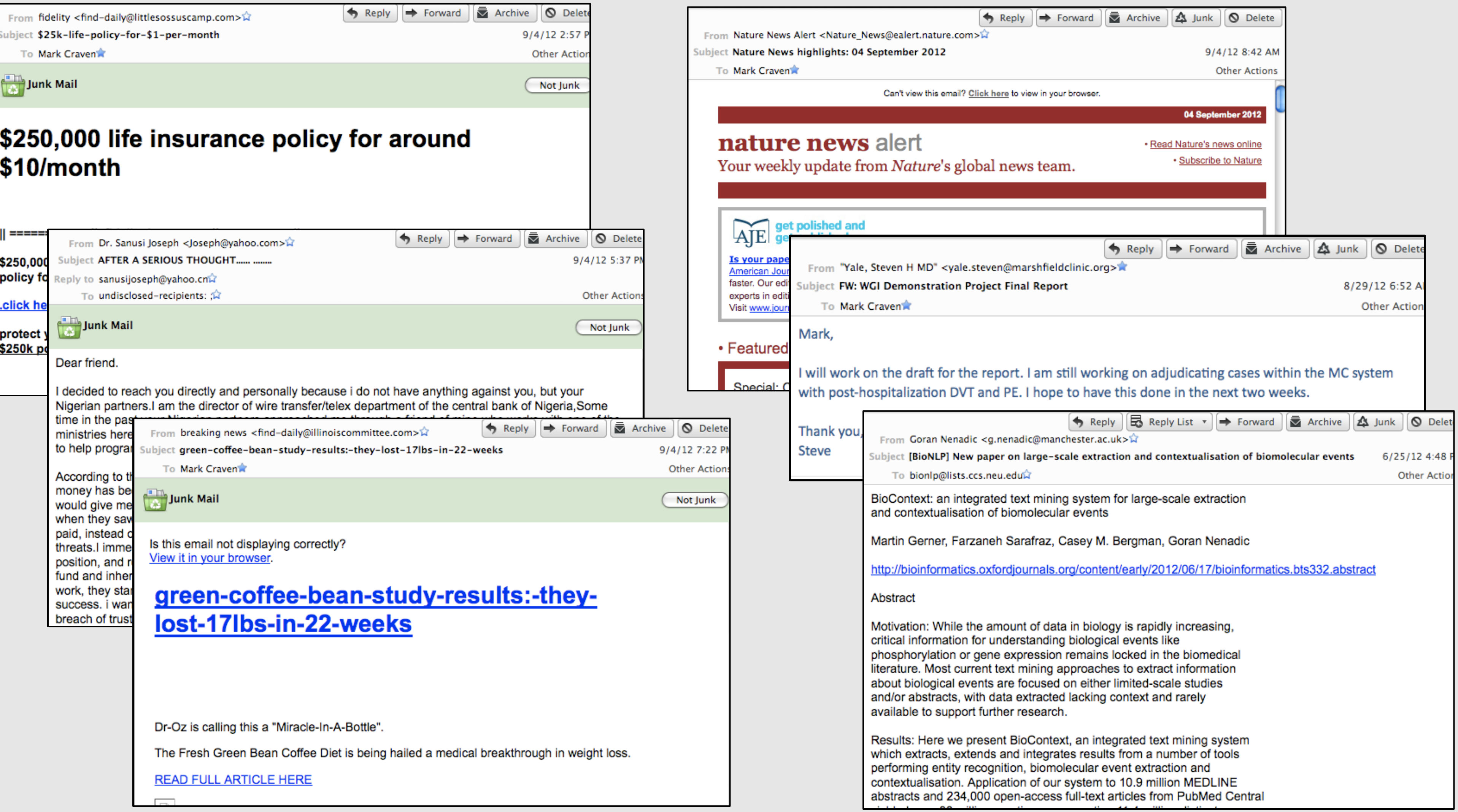


outdoor

ML example: image classification

- T : given new images, classify as indoor vs. outdoor
- P : minimize misclassification costs
- E : given images with indoor/outdoor labels

ML example: spam filtering



ML example: spam filtering

- T : given new mail message, classify as spam vs. other
- P : minimize misclassification costs
- E : previously classified (filed) messages

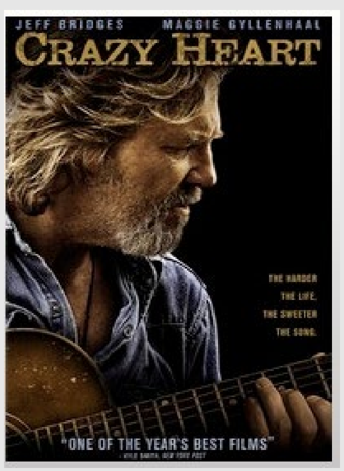
ML example: predictive text input

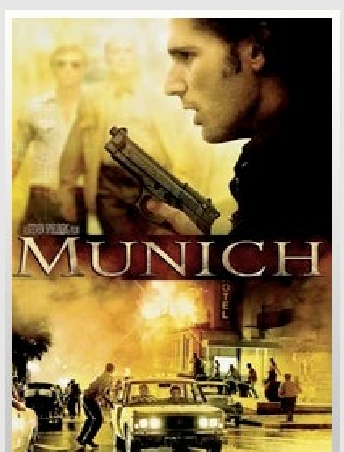
ML example: predictive text input

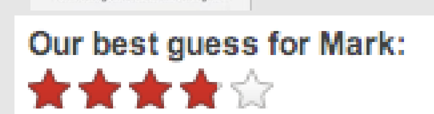
- T : given (partially) typed word, predict the word the user intended to type
- P : minimize misclassifications
- E : words previously typed by the user
  (+ lexicon of common words + knowledge of keyboard layout)

domain knowledge

ML example: Netflix Prize







ML example: Netflix Prize

- T : given a user/movie pair, predict the user's rating (1-5 stars) of the movie
- P : minimize difference between predicted and actual rating
- E : histories of previously rated movies (user/movie/rating triples)

ML example: autonomous helicopter

video of Stanford University autonomous helicopter from http://heli.stanford.edu/

ML example: autonomous helicopter

- T : given a measurement of the helicopter's current state (orientation sensor, GPS, cameras), select an adjustment of the controls
- P : maximize reward (intended trajectory + penalty function)
- E : state, action and reward triples from previous demonstration flights

ML example: Atari Breakout

Google DeepMind's Deep Q-learning playing Atari Breakout
From the paper “Playing Atari with Deep Reinforcement Learning”,
by Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, Martin Riedmiller

ML example: AlphaGo

Assignments

Reading assignment

- read
  - article by Jordan and Mitchell on course website

- course website:
  http://pages.cs.wisc.edu/~yliang/cs760_spring21/

HW1: Background test

- posted on Canvas
- will set up how to submit the solutions on Canvas

- If you find many questions intimidating, we suggest you drop the course and take it again in the future when you are more prepared.

Background test

Topics

- linear algebra
- calculus
- probability
- big-O notations for algorithm analysis
- basic programming skills

Test example

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?

Test example

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

Test example

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 $||\mathbf{x}||_\infty = \max_i |x_i|$)

Useful resources for background

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