

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!



Instructor



- Yingyu Liang
email: yliang@cs.wisc.edu
office hours: Mon 4-5pm
office: 6393 Computer Sciences



- Ying Fan

email: yingfan@cs.wisc.edu

office hours: Tues 1-2pm, Thu 1-2pm

office: CS 1351



Monday, Wednesday *and* Friday?



- we'll have ~30 lectures in all, just like a standard TR class
- will push the lectures forward (finish early, leave time for projects and review)
- ***see the schedule on the course website:***
http://pages.cs.wisc.edu/~yliang/cs760_spring20

Course emphases



- **a variety of learning settings**: supervised learning, unsupervised learning, reinforcement learning, active 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.

Two major goals



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

Course requirements



- 7-8 homework assignments: 30%
 - programming
 - computational experiments (e.g. measure the effect of varying parameter x in algorithm y)
 - some written exercises
- Midterm Exam #1: 20%
- Midterm Exam #2: 20%
- final project: 30%
 - project group: 3-5 people

Expected background



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

Programming languages



- for the programming assignments, you can use

C

C++

Java

Perl

Python

R

Matlab

- suggest: Python

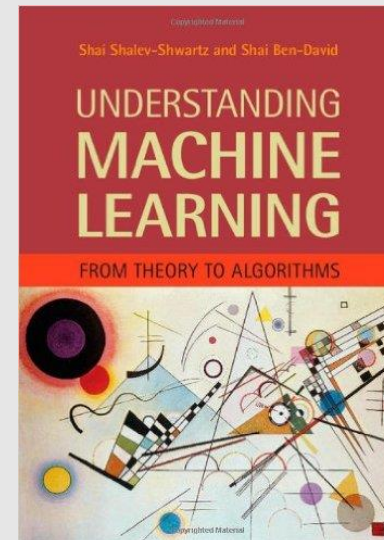
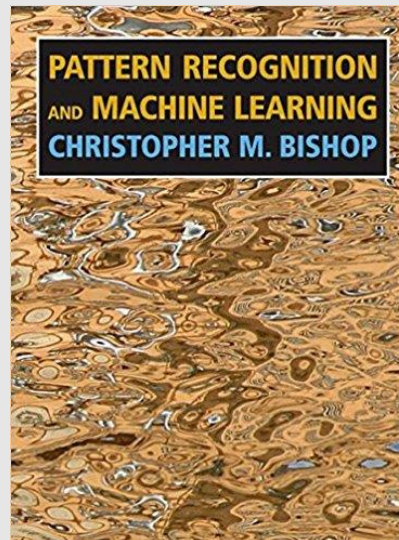
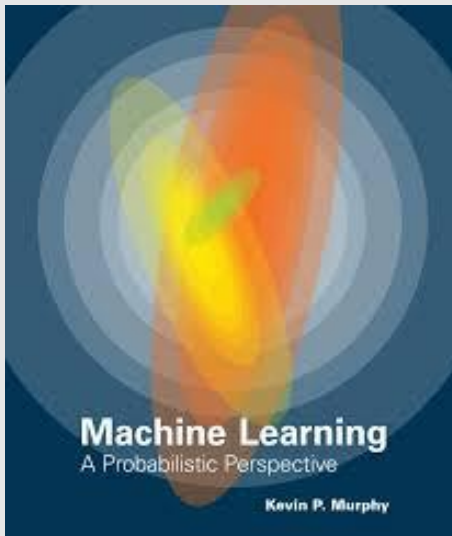
- programs must be callable from the command line and *must run on the CS lab machines (this is where they will be tested during grading!)*

Course readings



Recommend to get one 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

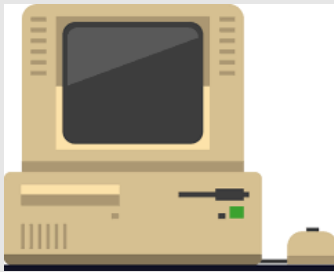


What is machine learning?



- “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



learning
→

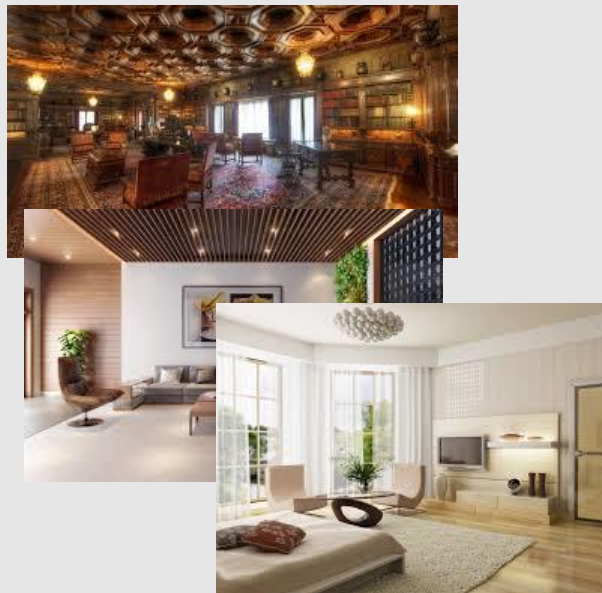


What is machine learning?

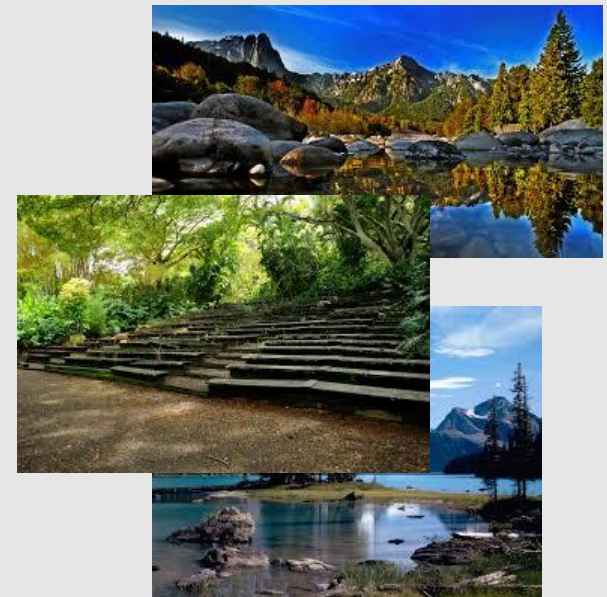


- the study of algorithms that improve their performance P at some task T with experience E
- to have a well-defined learning task, we must specify: $\langle P, T, E \rangle$

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



From fidelity <find-daily@littlesossuscamp.com>☆
Subject \$25k-life-policy-for-\$1-per-month
To Mark Craven★
9/4/12 2:57 PM
Other Actions

Junk Mail Not Junk

\$250,000 life insurance policy for around \$10/month

From Dr. Sanusi Joseph <Joseph@yahoo.com>☆
Subject AFTER A SERIOUS THOUGHT.....
Reply to sanusijoseph@yahoo.cn★
To undisclosed-recipients: ☆
9/4/12 5:37 PM
Other Actions

Junk Mail Not Junk

Dear friend.

I decided to reach you directly and personally because i do not have anything against you, but your Nigerian partners.I am the director of wire transfer/telex department of the central bank of Nigeria,Some time in the past my partners here in the ministries here to help program

From breaking news <find-daily@illinoiscommittee.com>☆
Subject green-coffee-bean-study-results:-they-lost-17lbs-in-22-weeks
To Mark Craven★
9/4/12 7:22 PM
Other Actions

Junk Mail Not Junk

Is this email not displaying correctly?
[View it in your browser.](#)

[green-coffee-bean-study-results:-they-lost-17lbs-in-22-weeks](#)

Dr-Oz is calling this a "Miracle-In-A-Bottle".

The Fresh Green Bean Coffee Diet is being hailed a medical breakthrough in weight loss.

[READ FULL ARTICLE HERE](#)

From Nature News Alert <Nature_News@ealert.nature.com>☆
Subject Nature News highlights: 04 September 2012
To Mark Craven★
9/4/12 8:42 AM
Other Actions

Can't view this email? [Click here](#) to view in your browser.

nature news alert
Your weekly update from *Nature's* global news team.

04 September 2012

[Read Nature's news online](#)
[Subscribe to Nature](#)

From "Yale, Steven H MD" <yale.steven@marshfieldclinic.org>★
Subject FW: WGI Demonstration Project Final Report
To Mark Craven★
8/29/12 6:52 AM
Other Actions

Mark,

I will work on the draft for the report. I am still working on adjudicating cases within the MC system with post-hospitalization DVT and PE. I hope to have this done in the next two weeks.

Thank you
Steve

From Goran Nenadic <g.nenadic@manchester.ac.uk>☆
Subject [BioNLP] New paper on large-scale extraction and contextualisation of biomolecular events
To bionlp@lists.ccs.neu.edu★
6/25/12 4:48 PM
Other Actions

BioContext: an integrated text mining system for large-scale extraction and contextualisation of biomolecular events

Martin Gerner, Farzaneh Sarafraz, Casey M. Bergman, Goran Nenadic

<http://bioinformatics.oxfordjournals.org/content/early/2012/06/17/bioinformatics.bts332.abstract>

Abstract

Motivation: While the amount of data in biology is rapidly increasing, critical information for understanding biological events like phosphorylation or gene expression remains locked in the biomedical literature. Most current text mining approaches to extract information about biological events are focused on either limited-scale studies and/or abstracts, with data extracted lacking context and rarely available to support further research.

Results: Here we present BioContext, an integrated text mining system which extracts, extends and integrates results from a number of tools performing entity recognition, biomolecular event extraction and contextualisation. Application of our system to 10.9 million MEDLINE abstracts and 234,000 open-access full-text articles from PubMed Central

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

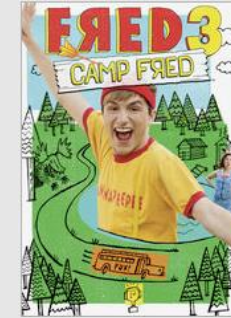
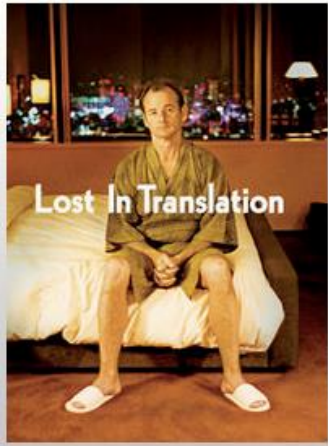


- T : given (partially) typed word, predict the word the user intended to type
- P : minimize misclassifications
- E : words previously typed by the user
(+ lexicon of common words + knowledge of keyboard layout)

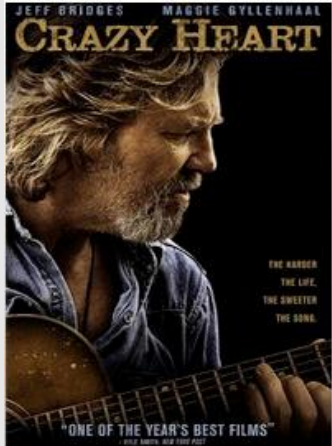
domain knowledge



ML example: Netflix Prize



Our best guess for Mark:
★★★★☆



Our best guess for Mark:
★★★★☆

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
 - Chapter 1 of Murphy
 - article by Jordan and Mitchell on course website
- course website:
http://pages.cs.wisc.edu/~yliang/cs760_spring20/

HW1: Background test



- posted on course website
- will set up how to submit the solutions on Canvas
- Two parts: minimum and medium tests
- if pass both: in good shape
- if pass minimum but not medium: can still take but expect to fill in background
- if fail both: suggest to fill in background before taking the course

Minimum background test



- 80 pts in total; pass: 48pts
- linear algebra: 20 pts
- probability: 20 pts
- calculus: 20 pts
- big-O notations: 20 pts

Minimum test example



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

1. What is the inner product of the vectors \mathbf{y} and \mathbf{z} ? (this is also sometimes called the *dot product*, and is sometimes written as $\mathbf{y}^T \mathbf{z}$)
2. What is the product $X\mathbf{y}$?
3. Is X invertible? If so, give the inverse, and if no, explain why not.
4. What is the rank of X ?

Minimum test example



1. If $y = 4x^3 - x^2 + 7$ then what is the derivative of y with respect to x ?
2. If $y = \tan(z)x^{6z} - \ln\left(\frac{7x+z}{x^4}\right)$, what is the partial derivative of y with respect to x ?

Medium background test



- 20 pts in total; pass: 12 pts
- algorithm: 5 pts
- probability: 5 pts
- linear algebra: 5 pts
- programming: 5 pts

Medium test example



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

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

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

(a) Laplace

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

(b) Multinomial

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

(c) Poisson

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

(d) Dirichlet

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

(e) Gamma

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

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



THANK YOU

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, and Pedro Domingos.

