

Logistics

Mladen Kolar (mkolar@chicagobooth.edu)

► Introduction

► Logistics

Teaching assistants



Xingyue Fang



Boxiang (Shawn) Lyu



JungHo Lee



Simiao Jiao

Logistics

Slides, homework, extra resources

- ▶ available from Canvas

Ed Discussion Forum

- ▶ post all your questions here
- ▶ be active and answer your classmates' questions
(you will even get credit for this)
- ▶ often a good answer is “Google it”
- ▶ link from Canvas

Gradescope

- ▶ used to submit homeworks, projects, and midterm
- ▶ link from Canvas

Prerequisites

Basic probability and statistics.

Linear regression at the level of BUS 41100.

Some experience with R or another programming language are a plus.

Grades

Homework (20%)

Participation on Ed Discussion (10%)

Midterm (30%)

Final project (40%)

Note on grading: Chicago Booth classes have requirements on the average grade — it cannot be above B+ across all sections.

Homeworks

5 assignments — best 4 will count towards your grade

Can be done in groups (suggested size 3)

- ▶ Homeworks are made to reinforce and *complement* material introduced in lectures.
- ▶ The problems are often open-ended, designed to explore and struggle with issues that often manifest in everyday practice.

Midterm (30%)

Individual

Take-home

Out in week 6, due in week 7

Will require data analysis. You can use any statistical tool you are comfortable with.

Final project (40%)

Choose your own problem, bring your own data

Proposal due in **week 5**

Write up and a presentation due in **week 10**

- ▶ **must** be done in group

You will need to fill the peer evaluation form at the end of the class.

Textbook

There are no required textbooks.

We suggest checking out **Introduction to Statistical Learning** by James, Witten, Hastie, and Tibshirani. Get it online at
<https://statlearning.com/>.

There are a number of other suggested books on Canvas.

Why are you here?

Because

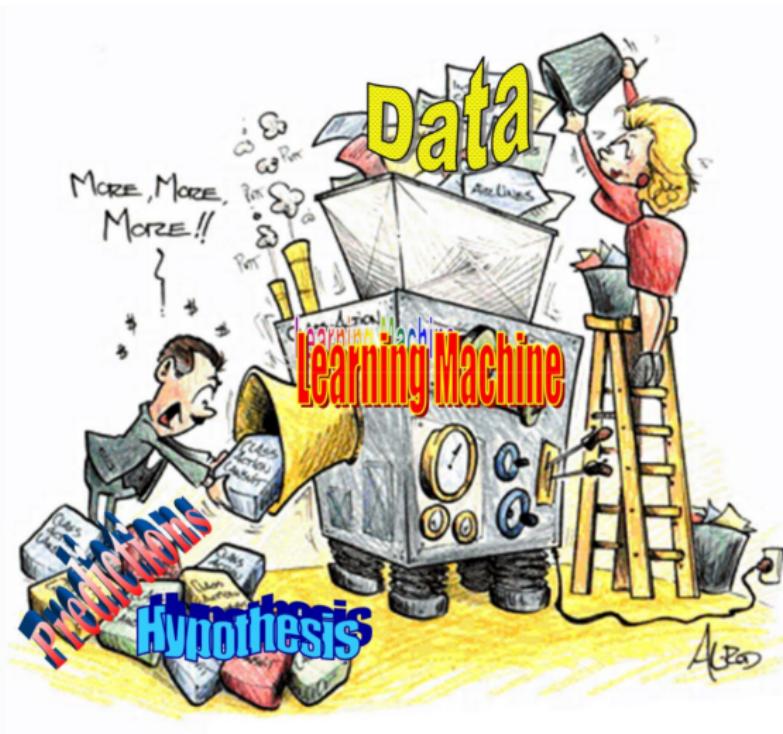
- ▶ you are passionate about the subject
- ▶ you want to learn about one of the transformative technologies of the 21st century
- ▶ ...

No matter the reason, we hope you will get something useful out of the class.

Remember that the more effort you put into the class, the more you will get out.

What is machine learning?

What is machine learning?



What is machine learning?

Machine learning is a technology that allows computers to

- ▶ improve their performance
- ▶ at some task
- ▶ with experience

Machine learning is the science of discovering structure and making predictions in (large) data sets.

Search and Ads

The screenshot shows a Google search results page for the query "machine learning". The results are filtered by "Web" and show approximately 61,000,000 results. The first result is a snippet from Wikipedia, followed by several links from educational platforms like Coursera and Stanford University. On the right side of the results, there are two ads: one for "Machine Learning for All" and another for "Qualcomm Machine Learning". Below the ads, there's a link to "What is Machine Learning?" and another for "Learn Machine Learning". Further down, there's a job listing for "Google Ads Team Is Hiring". The interface includes standard Google search controls like back, forward, and search bar.

machine learning - Google Search - Google Chrome

https://www.google.com/webhp?sourceid=chrome-instant&ion=1&espv=2&ie=UTF-8#safe=off&q=machine+learning

Google machine learning

Web News Videos Books Images More Search tools

About 61,000,000 results (0.47 seconds)

Machine learning is a subfield of computer science that evolved from the study of pattern recognition and computational learning theory in artificial intelligence. Machine learning explores the study and construction of algorithms that can learn from and make predictions on data.

Machine learning - Wikipedia, the free encyclopedia
https://en.wikipedia.org/wiki/Machine_learning Wikipedia

More about Machine learning

Feedback

Machine Learning - Stanford University | Coursera
https://www.coursera.org/learn/machine-learning Coursera

Machine learning is the science of getting computers to act without being explicitly programmed. In the past decade, machine learning has given us self-driving cars, practical speech recognition, effective web search, and a vastly improved understanding of the human genome.

Work Smarter, Not Harder - ML course - Linear Regression with ... - Multiple Features

Machine Learning - University of Washington | Coursera
https://www.coursera.org/course/machinelearning Coursera

Machine learning (also known as data mining, pattern recognition and predictive analytics) is used widely in business, industry, science and government, and there is a great shortage of experts in it.

Machine learning - Wikipedia, the free encyclopedia
https://en.wikipedia.org/wiki/Machine_learning Wikipedia

Machine learning is a subfield of computer science that evolved from the study of pattern recognition and computational learning theory in artificial intelligence.

List of machine learning - Supervised learning - Computational learning theory

Machine Learning for All
www.mathworks.com/Machine_Learning

MATLAB Speaks Machine Learning
Intuitive & Powerful. See Examples.

Qualcomm Machine Learning
www.qualcomm.com/WhyWait

4.3 ★★★★☆ rating for qualcomm.com
Qualcomm is Teaching Machine to Act More Human. See How We Do It.

What is Machine Learning?
www.sas.com/

Learn About Machine Learning & How It Can Help Predict Outcomes.

Learn Machine Learning
www.generationchicago.com/Chicago

Want Results? Learn From The Best! Graduate Job Ready In 2.5 Months.
9 20 N. Upper Wacker Drive, 12th Floor (877) 348-5665

Google Ads Team Is Hiring
www.google.com/jobs/12

Have math skills?
Submit your resume

Machine Learning
www.wise.io/

Find Predictive Signals In Your Customer Data. Request a Free Demo.

Spam Filtering

Protest and Free Expression

Robert J. Zimmer and Eric D. Isaacs president@uchicago.edu via chicago Booth.edu
to president-prov. Jun 7

To: Campus Community
From: Robert J. Zimmer and Eric D. Isaacs
Subject: Protest and Free Expression
Date: June 7, 2015

Within the past week, two protests on campus have violated the University's long-standing commitment to free expression, as expressed over the years by multiple faculty committees and reports, most recently in the Report of the Committee on Freedom of Expression. We write to reaffirm these principles in the context of these recent events.

The Report, reflecting 125 years of University tradition and commitment, forcefully articulates the importance of an environment of free expression of ideas, whether or not others may find this speech disturbing. It is a shared obligation of our community to support such freedom. Antithetical to such freedom are actions that prevent speech on the part of others, obstruct the ability of members of our community to listen, or prevent people in the University from carrying out their work. The two events this week were directly antithetical to the University's values for these reasons.

BUSINESS PROPOSAL

DESCO ENGINEERING <descozambia@mail.zamtel.zm>
to Jun 7

12:23 AM (15 hours ago)

⚠ Be careful with this message. Many people marked similar messages as phishing scams, so this might contain unsafe content. [Learn more](#)

BUSINESS PROPOSAL

I got your contact from a business directory. I decided to contact you for a business with my company. The company I work with is into manufacturing of pharmaceutical materials. There is a raw material which the company used to send me to India to buy. Right now I have been promoted to the post of manager. The company can not send me to India again; they will send a more junior staff. The director has asked for the contact of the supplier in India. I need a person I will present to the company as the supplier in India. You will now buy the product from the local dealer and supply to my company. The profit would be shared between you and I. Why I don't want the company to have direct contact of the local dealer is that, I don't want the company to know the actual price I was buying the product. If you are interested kindly contact me for more details. Through this email id : Julietmark78@hotmail.com
Thanks! Mrs. Juliet Mark

Spam or Ham

Personal recommendation

Who to follow

Follow more people from the suggestions below, tailored just for you.

Search using a person's full name or @username

Search Twitter



Burr Settles @bursettles

Machine learning, music, and mischief. Data dude at
@Duolingo and founder of @FAWM.

Followed by Rammath B., Bert Huang and Tom M Mitchell.



+ Follow



Misha Denil @notmisha

Research Scientist at Google DeepMind. I tweet
about things that interest me, mostly machine learning
things.



+ Follow



Jeffrey P. Bigham @jeffbigham

associate professor of tweeting at @cmuhci. human
computer, #a11y hacker, & social computing
participant. multitasking sockpuppet ninja run by a
crowd within.

Followed by Tomasz Malisiewicz, Jean Yang and Matt
Marge.



+ Follow



John Platt @johnplattml

Using ML, AI, and data systems to benefit humanity.
Now @Google.



+ Follow



Duncan Watts @duncanwatts

Principal Researcher @ Microsoft

Followed by brendan o'connor, petricek and Lada Adamic.



+ Follow



ML @ REDDIT @mxdearn

/r/MachineLearning: Research, News, Discussions,
Software on: Machine Learning, Data Mining, Text
Processing, Information Retrieval, Search Computing
and alike

Followed by Babis Tsourakakis, petricek and Aapo Kyrola.



+ Follow

People to follow on Twitter
(also People You May Know on Facebook)

Personal recommendation

Netflix Prize: View Leaderboard - Google Chrome

Netflix Prize: View L. x Mladen

www.netflixprize.com/leaderboard

NETFLIX

Netflix Prize COMPLETED

Home Rules Leaderboard Update

Leaderboard

Showing Test Score. [Click here to show quiz score](#)

Display top 20 leaders.

Rank	Team Name	Best Test Score	Improvement	Best Submit Time
1	BellKor's Pragmatic Chaos	0.8567	10.06	2009-07-26 18:18:28
2	The Ensemble	0.8567	10.06	2009-07-26 18:38:22
3	Grand Prize Team	0.8582	9.90	2009-07-10 21:24:40
4	Opera Solutions and Vandelay United	0.8588	9.84	2009-07-10 01:12:31
5	Vandelay Industries I	0.8591	9.81	2009-07-10 00:32:20
6	PragmaticTheory	0.8594	9.77	2009-06-24 12:06:56
7	BellKor in BigChaos	0.8601	9.70	2009-05-13 08:14:09
8	Dace	0.8612	9.59	2009-07-24 17:18:43
9	Feedz2	0.8622	9.48	2009-07-12 13:11:51
10	BigChaos	0.8623	9.47	2009-04-07 12:33:59
11	Opera Solutions	0.8623	9.47	2009-07-24 00:34:07
12	BellKor	0.8624	9.46	2009-07-26 17:19:11

\$1M prize!

Personal recommendation

The screenshot shows a web browser window with the title 'Mapping Love with Hadoop | eHarmony Engineering | David Gevorkyan - Google Chrome'. The page content includes:

- Header:** eHarmony: Get Matches, Blog, Technology, Team, Careers, Dating Advice.
- Title:** Mapping Love with Hadoop
- Author:** David Gevorkyan (September 24, 2014)
- Tags:** ARTICLES, MEETUP, TECH TALKS, AFFINITY, COMPATIBILITY, DATA SCIENCE, HADOOP, MACHINE LEARNING, MATCHING, MONGODB, NOSQL, SEAMICRO, SPRING BATCH, SPRING BATCH ADMIN, VOLDEMORT
- Social sharing:** Facebook (246), Twitter (40), LinkedIn (Shares 36), Google+ (5), Share (78).
- Text:** In this talk, I discuss how Hadoop helps us to process over a billion possible matches into several highly compatible matches for each of our users per day.
- Image:** A portrait of David Gevorkyan with the caption: David Gevorkyan is a Principal Software Engineer at eHarmony's Matching Team.
- Text:** eHarmony was founded to give people a better chance at finding happy, passionate, and fulfilling relationships. Did you know that we are already responsible for 5% of all new US marriages, and that more than 600,000 people met their spouses on eHarmony?
- Text:** During this talk I describe how we go about creating highly compatible matches, and how we leverage Big Data technologies to accomplish that goal.
- Text:** eharmony.com/engineering/author/dgevorkyan/
- Right sidebar:** Search bar, Recent Posts (One Year Anniversary Swift Meetup, Taking MongoDB to Production, In Pursuit of Messaging Brokers, Redis at eHarmony as a Store and Cache, eH Automation, an Overview of Front End Regression Testing), and Tags (eharmony, affinity, akka, apache, kafka, apollo, apple, backbone, benchmark, broker).

Fall in love with machine learning

Jeopardy!



IBM's Watson

Autonomous Cars



Carnegie Mellon Boss, the self-driving SUV
1st place in the DARPA Urban Challenge (2007)

Handwritten Character Recognition



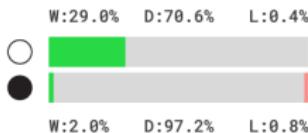
Reading postal address
Processing tax returns

AlphaZero

Chess



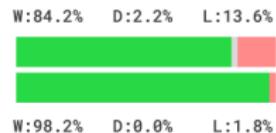
AlphaZero vs. Stockfish



Shogi



AlphaZero vs. Elmo



Go



AlphaZero vs. AGO

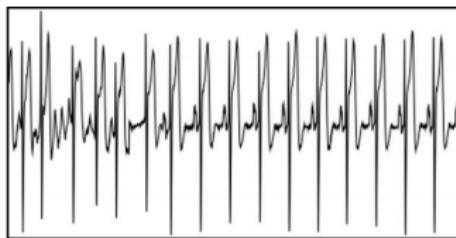


AZ wins AZ draws AZ loses AZ white AZ black

Medical applications

Cardiologist-level arrhythmia detection

www.nature.com/articles/s41591-018-0268-3

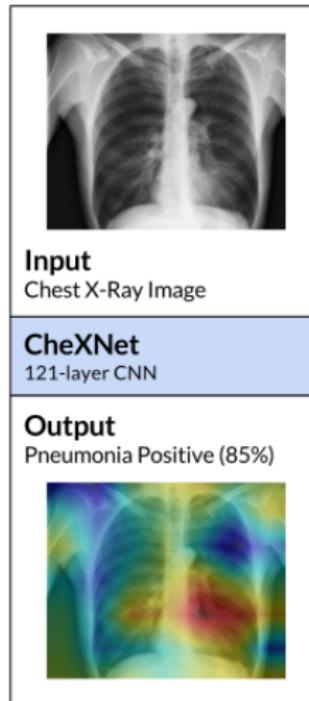


- ▶ If confirmed in clinical settings, this approach could reduce the rate of misdiagnosed computerized ECG interpretations and improve the efficiency of expert human ECG interpretation by accurately triaging or prioritizing the most urgent conditions.

Medical applications

Radiologist-Level Pneumonia Detection

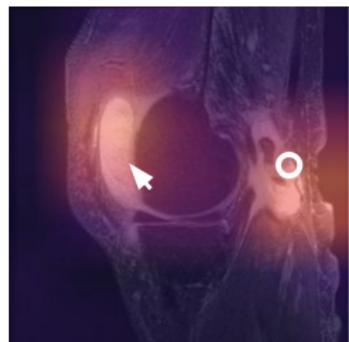
<https://stanfordmlgroup.github.io/projects/chexnet/>



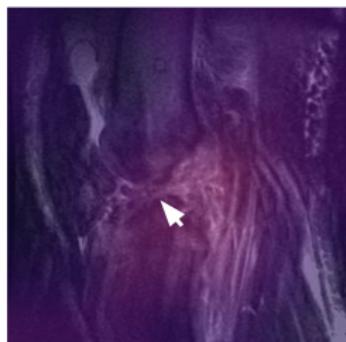
Medical applications

Assisted diagnosis for knee magnetic resonance imaging

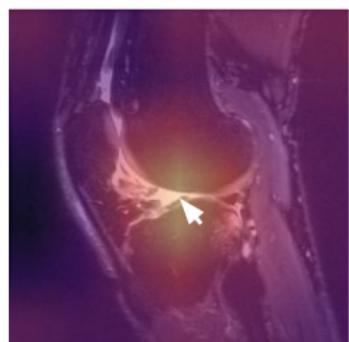
<https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1002699>



(a)



(b)



(c)



(d)

Machine Learning in Action

Many, many more application areas

- ▶ speech recognition
- ▶ natural language processing
- ▶ medical outcomes analysis
- ▶ wearable technology
- ▶ quality of life technology
- ▶ financial forecasting
- ▶ online marketing
- ▶ social media analysis
- ▶ anomaly detection
- ▶ ...

What is this class about?

You will learn about:

- ▶ applying machine learning tools
- ▶ which tools to apply
- ▶ think about what will work and what will not work
- ▶ it is important to understand the ideas behind the various techniques, in order to know how and when to use them
- ▶ one has to understand the simpler methods first, in order to grasp the more sophisticated ones

This class will not be about:

- ▶ data wrangling
- ▶ different implementations and frameworks of machine learning algorithms

Machine learning is a fundamental ingredient in the training of a modern *data scientist*.

Two parts of the class

Supervised learning algorithms

- ▶ regression
- ▶ classification

Unsupervised learning

- ▶ looking for structure without predictive goal

Tentative Schedule at Glance

- Week 1:** Nearest Neighbors; Bias-Variance Trade-Off; Cross-Validation
- Week 2:** Classification and regression trees
- Week 3:** Bagging and boosting
- Week 5:** Evaluating classifiers
- Week 5:** Variable selection
- Week 6:** Principal Component Analysis
- Week 7:** Support Vector Machines; Class Imbalance
- Week 8:** Deep learning
- Week 9:** Deep learning
- Week 10:** No class; Project write-up due