

# Course Overview

Mladen Kolar and Rob McCulloch

1. Logistics
2. What is machine learning?
3. What will this class be about?

# Logistics

Course website (slides, homework, extra resources))

[http://chicagoboothml.github.io/MachineLearning\\_Fall2015/](http://chicagoboothml.github.io/MachineLearning_Fall2015/)

Discussion forum on Piazza

<https://piazza.com/chicagobooth/fall2015/bus41204/home>

- ▶ post all your questions here
- ▶ be active and answer your classmates' questions
- ▶ often a good answer is "Google it"

Chalk

- ▶ used only to submit homework

# Grades

- ▶ Homework (20%)
  - ▶ 8 weekly assignments
  - ▶ can be done in groups (max size 4)
  - ▶ top 7 count
- ▶ Midterm (40%)
  - ▶ individual, take-home, out in week 5, due in week 6
- ▶ Final project (40%)
  - ▶ proposal due in week 7
  - ▶ write up due in the finals week
  - ▶ choose your own problem, bring your own data
  - ▶ more on this later

# People

## Instructors:

- ▶ Mladen Kolar `mkolar@chicagobooth.edu`
- ▶ Rob McCulloch `Robert.McCulloch@chicagobooth.edu`

## TAs:

- ▶ Daniel Hedblom `Hedblom@uchicago.edu`
- ▶ Juan Yrigoyen `JYrigoyen@chicagobooth.edu`
- ▶ Vinh Luong `MBAlearnsToCode@uchicago.edu`

## Textbook

There are no required textbooks.

We suggest checking out *Introduction to Statistical Learning* by James, Witten, Hastie, and Tibshirani. Get it online at <http://www-bcf.usc.edu/~gareth/ISL>.

There are a number of other suggested books on the course web-page.

# Prerequisites

Basic probability and statistics.

Linear regression at the level of BUS 41000 or BUS 41100.

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

# 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
- ▶ you eventually want to earn big \$\$\$
- ▶ ...

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.



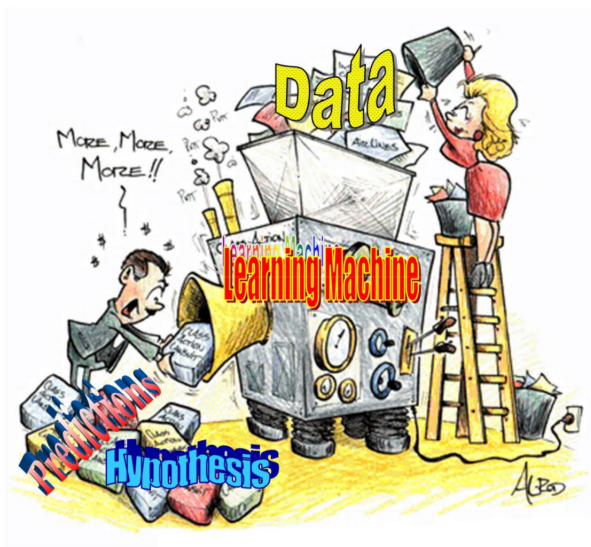
## Feedback please

This is our FIRST class at Booth!

Let us or the TAs know if you have comments, concerns, suggestions!

You can leave (anonymous) feedback at  
<http://www.admonymous.com/boothml>

# 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.

# Spam Filtering

## Protest and Free Expression



**Robert J. Zimmer and Eric D. Isaacs** [president@uchicago.edu](mailto:president@uchicago.edu) [via chicagobooth.edu](http://www.chicagobooth.edu)

Jun 7



to [president-prov.](#)

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

Spam



**DESCO ENGINEERING** <[descozambia@mail.zamtel.zm](mailto:descozambia@mail.zamtel.zm)>

12:23 AM (15 hours ago)



to

**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 being 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 : ([julemark78@hotmail.com](mailto:julemark78@hotmail.com) )  
Thanks/Mrs. Juliet Mark

Spam or Ham

# Personal recommendation

NETFLIX

## Netfix Prize

COMPLETED

Home Rules Leaderboard Update

### Leaderboard

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

Display top  leaders.

| Rank   | Team Name   | Best Test Score | % Improvement | Best Submit Time    |
|--|---|-----------------|---------------|---------------------|
| <b>Grand Prize - RMSE = 0.8567 - Winning Team: BellKor's Pragmatic Chaos</b> |   |                 |               |                     |
| 1  | <a href="#">BellKor's Pragmatic Chaos</a>           | 0.8567          | 10.06         | 2009-07-26 18:18:28 |
| 2  | <a href="#">The Ensemble</a>                        | 0.8567          | 10.06         | 2009-07-26 18:38:22 |
| 3  | <a href="#">Grand Prize Team</a>                    | 0.8582          | 9.90          | 2009-07-10 21:24:40 |
| 4  | <a href="#">Opera Solutions and Vandelay United</a> | 0.8588          | 9.84          | 2009-07-10 01:12:31 |
| 5  | <a href="#">Vandelay Industries I</a>               | 0.8591          | 9.81          | 2009-07-10 00:32:20 |
| 6  | <a href="#">PragmaticTheory</a>                     | 0.8594          | 9.77          | 2009-06-24 12:06:56 |
| 7  | <a href="#">BellKor in BigChaos</a>                 | 0.8601          | 9.70          | 2009-05-13 08:14:09 |
| 8  | <a href="#">Dace</a>                                | 0.8612          | 9.59          | 2009-07-24 17:18:43 |
| 9  | <a href="#">Feeds2</a>                              | 0.8622          | 9.48          | 2009-07-12 13:11:51 |
| 10   | <a href="#">BigChaos</a>                            | 0.8623          | 9.47          | 2009-04-07 12:33:59 |
| 11   | <a href="#">Opera Solutions</a>                     | 0.8623          | 9.47          | 2009-07-24 00:34:07 |
| 12   | <a href="#">BellKor</a>                             | 0.8624          | 9.46          | 2009-07-26 17:19:11 |

\$1M prize!

# Personal recommendation

Mapping Love with Hadoop | eHarmony Engineering | David Gevorkyan - Google Chrome

www.eharmony.com/engineering/mapping-love-with-hadoop/#.Vfs3qHUVhBd

eHarmony: Get Matches Blog Technology Team Careers Dating Advice

eHarmony Engineering

## Mapping Love with Hadoop

David Gevorkyan September 24, 2014

ARTICLES, MEETUP, TECH TALKS AFFINITY, COMPATIBILITY, DATA SCIENCE, HADOOP, MACHINE LEARNING, MATCHING, MONGODB, NOSQL, SEAMICRO, SPRING BATCH, SPRING BATCH ADMIN, VOLDEMORT

Like 246 Tweet 40 Share 36 G+ 5 Share 78

*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.*

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?

During this talk I describe how we go about creating highly compatible matches, and how we leverage Big Data technologies to accomplish that goal.

eharmony.com/engineering/author/dgevorkyan/

### Search

SEARCH

### 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

### Tags

advertising affinity akka apache kafka apollo apple backbone benchmark broker

Fall in love with machine learning

# Handwritten Character Recognition



Reading postal address

Processing tax returns

# 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

This class will not be about:

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

# Two parts of the class

## Supervised learning algorithms

- ▶ regression
- ▶ classification

## Unsupervised learning

- ▶ looking for structure without predictive goal