COMP 551 – Applied Machine Learning
Lecture 1: Introduction

Instructor: Joelle Pineau (*jpineau@cs.mcgill.ca)*

Class web page: *www.cs.mcgill.ca/~jpineau/comp551*

Outline for today

- Overview of the syllabus
- Summary of course content
- Broad introduction to Machine Learning (ML)
- Examples of ML applications

Course objectives

- To develop an understanding of the fundamental concepts of ML.
  - Algorithms, models, practices.

- **To emphasize good methods and practices for effective deployment of real systems.**

- To acquire hands-on experience with basic tools, algorithms and datasets.

About you

177 registered, primarily from:

- Computer Science (**approx. 50%**)
- Math, Statistics, Biostats, Epidemiology, Information Studies
- Electrical, Biomedical, Software, Mechanical, Mining Engineering

… and a few from:

- Physics, Biology, Neuroscience, Cognitive science, Economics.
- Interuniversity transfers.

…and new this year:

- Music, Political Science, History, Human genetics, Chemical Eng.

Approx. 10% PhD, 30% Masters, 60% Bachelors candidates.

About me

- What have I done?
  - B.A.Sc. in Engineering (U.Waterloo) 1993 - 1998
  - Ph.D. in Robotics (Carnegie Mellon U.) 1998 – 2003
  - Assistant / Associate Prof at McGill 2004 - …
- Co-director of the Reasoning and Learning Lab.
- What kind of research do I do?
  - Machine learning (reinforcement learning, deep learning, online learning, representation learning, …)
  - Planning and decision-making
  - Robotics
  - Personalized medicine and health-care

The rest of the teaching team

<u>Associate instructor:</u>

Dr. Herke van Hoof, *herke.vanhoof@mail.mcgill.ca*
Office hours: T.b.d., MC104N

<u>Teaching assistants:</u>

Philip Amortila, *philip.amortila@mail.mcgill.ca*
Office hours: T.b.d.

Christopher Glasz, *christopher.glasz@mail.mcgill.ca*
Office hours: T.b.d.

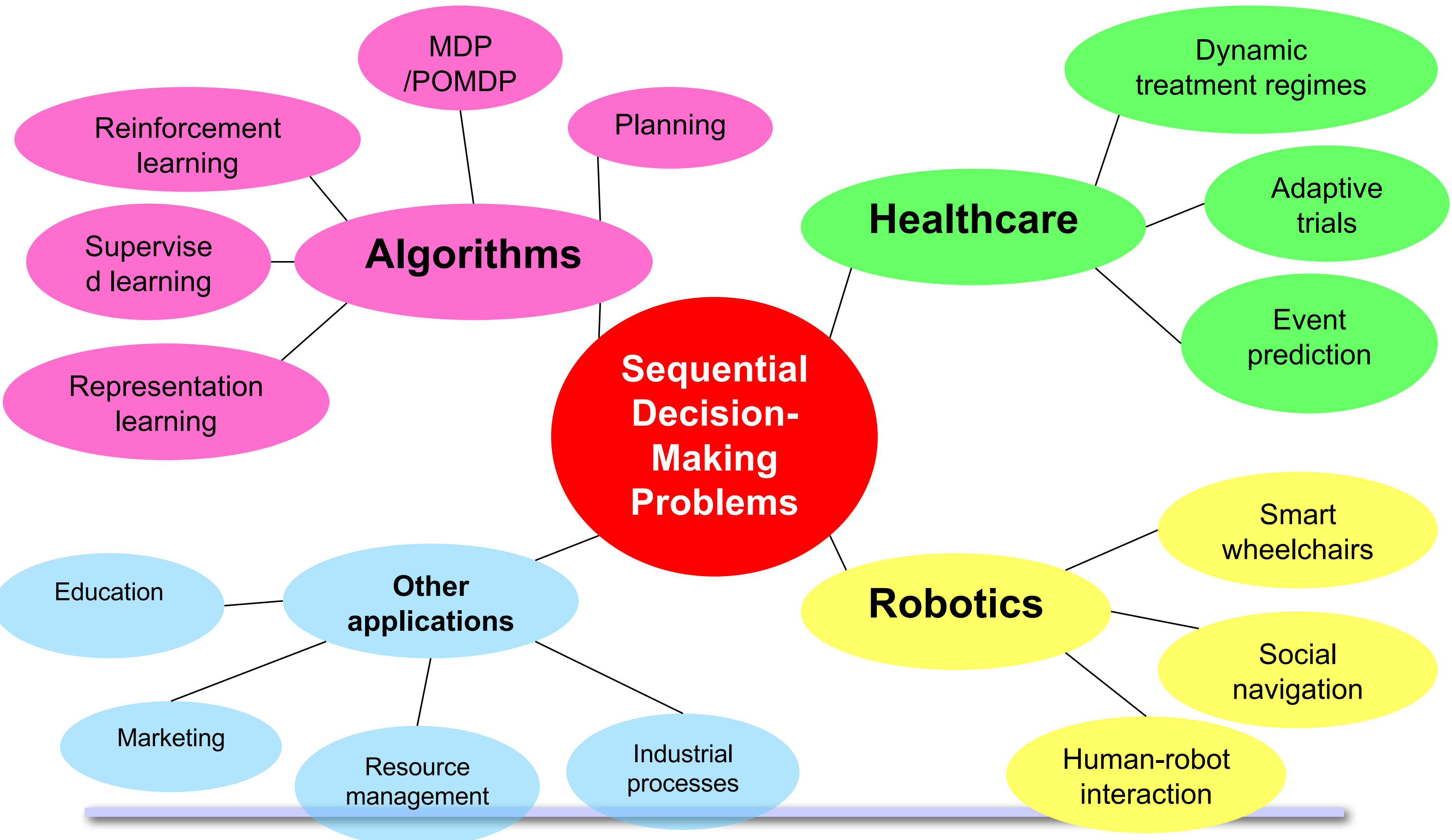
Harsh Satija, *harsh.satija@mail.mcgill.ca*
Office hours: Mondays 10:30-11:30am, MC106

Koustiv Sinha, *koustuv.sinha@mail.mcgill.ca*
Office hours: Wednesdays 4-5pm, TR3110

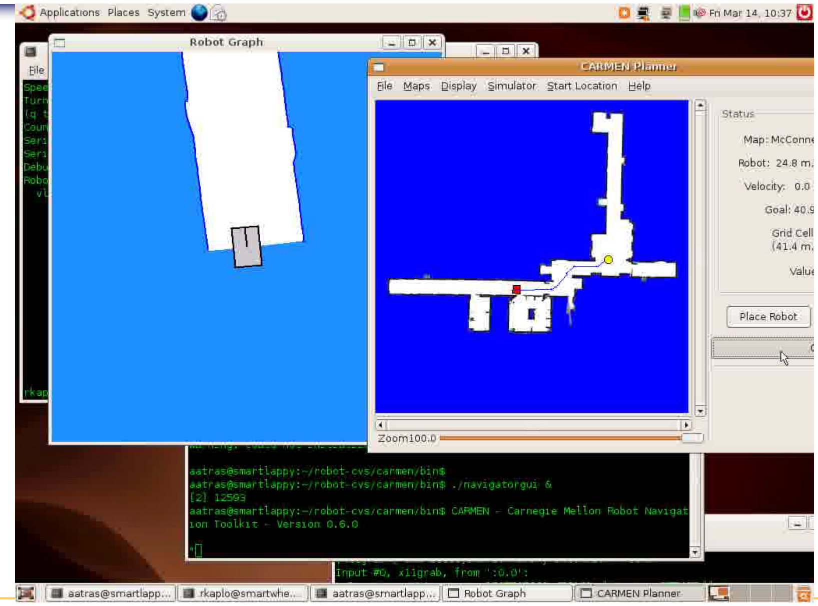
Matthew Smith, *matthew.smith5@mail.mcgill.ca*
Office hours: T.b.d.

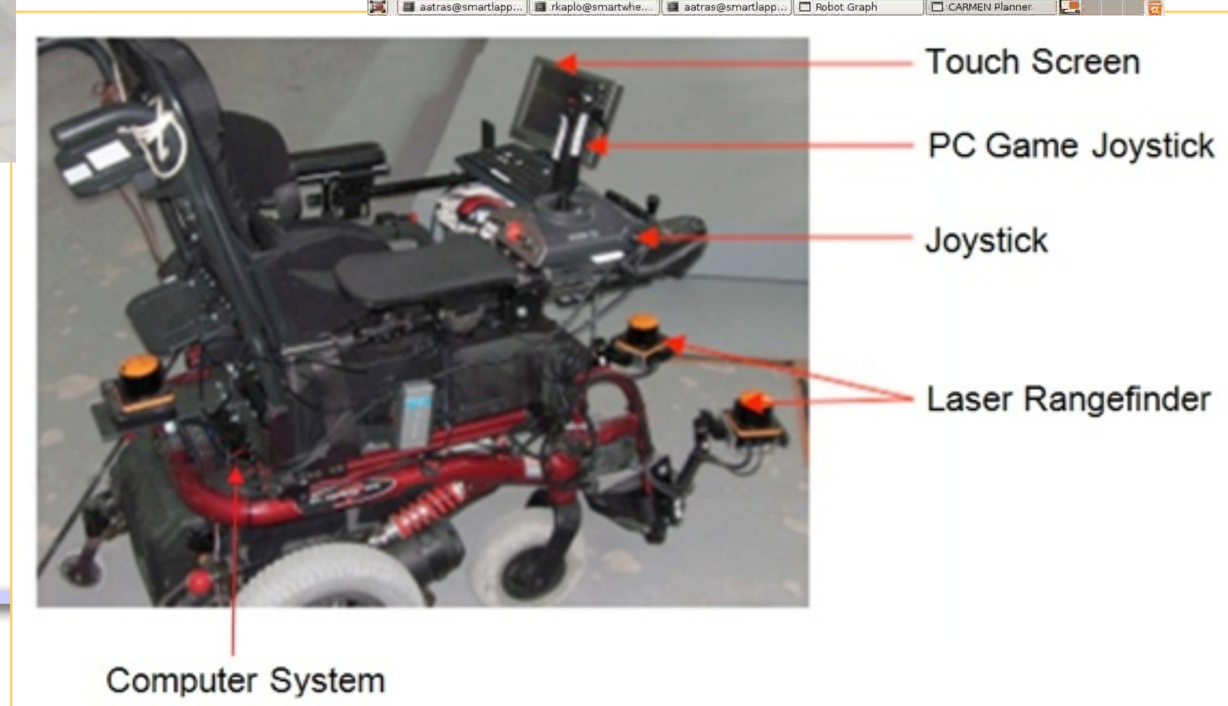
Sanjay Thakur, *sanjay.thakur@mail.mcgill.ca*
Office hours: Fridays 10:30-11:30am, MC105

Research areas in my lab



From the lab to the real world

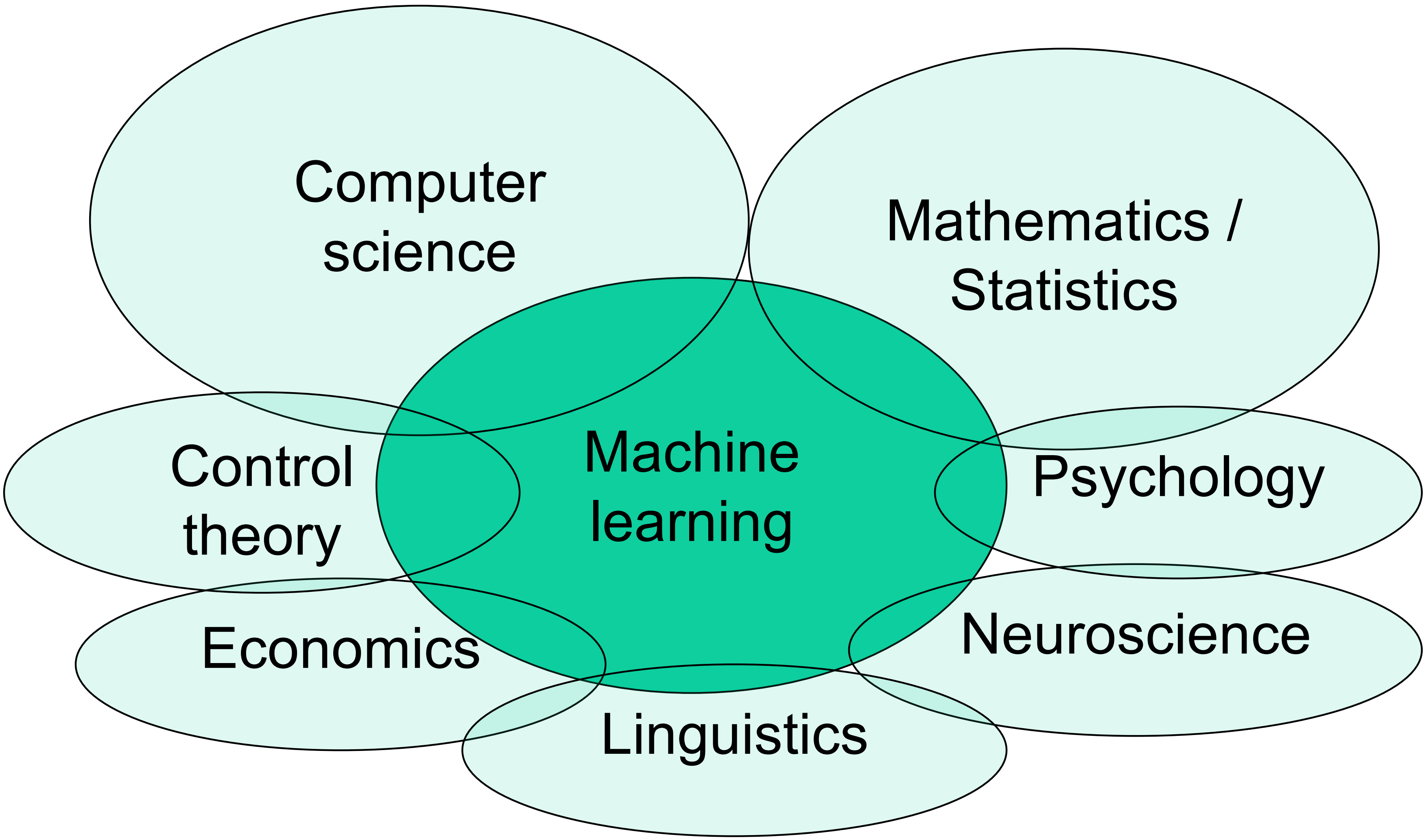




Sample publication – See http://www.cs.mcgill.ca/~jpineau/publications.html

- A.M.S. Barreto, D. Precup, J. Pineau. "**Practical Kernel-Based Reinforcement Learning**". Journal of Machine Learning Research. 17(67):1:70. 2016.
- B. Wang, J. Pineau. "**Generalized Dictionary for Multitask Learning with Boosting**". International Joint Conference on Artificial Intelligence (IJCAI). 2016.
- C. Zhou, B.B. Balle, J. Pineau. "**Learning Time Series Models for Pedestrian Motion Prediction**". International Conference on Robotics and Automation (ICRA). 2016.
- I.V. Serban, A. Sordoni, Y. Bengio, A. Courville, J. Pineau. "**Building End-To-End Dialogue Systems Using Generative Hierarchical Neural Network Models**". American Association for Artificial Intelligence (AAAI). 2016.
- B. Wang, B.B. Balle, J. Pineau. "**Multitask Generalized Eigenvalue Program**". American Association for Artificial Intelligence (AAAI). 2016.
- A.M.S. Barreto, R.L. Beirigo, J. Pineau, D. Precup. "**Incremental Stochastic Factorization for On-line Reinforcement Learning**". American Association for Artificial Intelligence (AAAI). 2016.
- S.M. Shortreed, E.B. Laber, J. Pineau, S.A. Murphy. "**Imputing Missing Data from Sequential Multiple Assignment Randomized Trials**". Book chapter. Adaptive Treatment Strategies in Practice: Planning Trials and Analyzing Data for Personalized Medicine. M.R. Kosorok and E.E.M. Moodie (eds). 2016.
- R. Vincent, J. Pineau. "**Practical reinforcement learning in dynamic treatment regimes**". Book chapter. Adaptive Treatment Strategies in Practice: Planning Trials and Analyzing Data for Personalized Medicine. M.R. Kosorok and E.E.M. Moodie (eds). 2016.

About machine learning

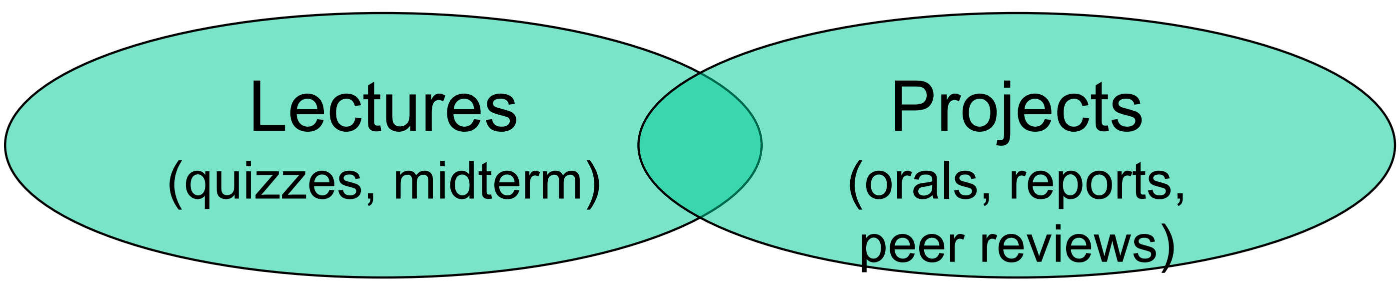


About the course: Tentative list of topics

- Linear regression.
- Linear classification.
- Performance evaluation, overfitting, cross-validation, bias-variance analysis, error estimation.
- Dataset analysis.
- Naive Bayes.
- Decision and regression trees.
- Support vector machines.
- Neural networks.
- Deep learning.
- Unsupervised learning and clustering.
- Feature selection.
- Dimensionality reduction.
- Regularization.
- Data structures and Map-Reduce.
- Ensemble methods.
- Cost-sensitive learning.
- Online / streaming data.
- Time-series analysis.
- Semi-supervised learning.
- Recommendation systems.
- Applications.

About the course

- During class:
  - Primarily lectures, with some seminars, paper discussions, problem-solving.
- Outside of class:
  - 4 optional tutorial sessions.
  - Complete 4 projects, online quizzes, peer review work of colleagues, review your notes, read papers, watch videos.



IMPORTANT! These target <u>different</u>, but complementary, knowledge & skills!

About the course

- **Prerequisites**:
  - Knowledge of a programming language (Matlab, R are ok; Python is best.)
  - Knowledge of probabilities/statistics (e.g. MATH-323, ECSE-305).
  - Knowledge of calculus and linear algebra.
  - Some AI background is recommended (e.g. COMP-424, ECSE-526) but not required.
- **Anterequisites**:
  - If you took COMP-652 before 2014, you CANNOT take COMP-551.
  - However taking COMP-652 during/after Winter 2014 is ok (course was re-designed to avoid overlap).

About the course

Evaluation:

- Weekly quizzes and exercises (5%)
- One in-class midterm (35%)
- Four data analysis case studies (projects) + peer reviews (60%)

Coursework policy:

All course work should be submitted online (details to be given in class), by 11:59pm, on the assigned due date. Late work will be subject to a 30% penalty, and can be submitted up to 1 week after the deadline.

No make-up quizzes or midterm will be given.

About the course

- Four case studies (projects):
  1. Machine learning task #1. (Dataset curation) 10%
  2. Supervised learning task #2. (Text classification) 15%
  3. Supervised learning task #3. (Image classification) 15%
  4. Final project. (Imposed topic; variety of datasets) 20%

- Format:
  – Projects will be submitted as written report + working code/data.
  – Final project will involve a short oral presentation.

- Work to be done in teams of 3. Work with different people for each project.

About the course

- I will not be using the classroom recording system.
- **My advice**: Do not to register for two courses in same time block. **Plan on attending every class**.
- Slides and projects will be available on the class website.
- MyCourses is available for discussions and finding project teams.
- We will use MyCourses for quizzes.
- We will use *https://cmt3.research.microsoft.com/* for project reports and peer-reviews.

Course material

No mandatory textbook, but a few good textbooks are recommended on the syllabus (some freely available online).

- Shalev-Schwartz & Ben-David. *Understanding Machine Learning*. Cambridge University Press. 2014.
    - *More theoretical.*
- Hastie, Tibshirani & Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd Edition. Springer. 2009.
    - *More mathematical.*
- Bishop. *Pattern Recognition and Machine Learning*. Springer. 2007.
    - *More practical, more accessible.*
- Goodfellow, Bengio &Courville. *Deep Learning*. MIT Press. 2016.
    - *For neural networks and deep learning modules.*

Software tools

- **Many** software packages are available, including broad ML libraries in Java, C++, Python and others.
- Many advanced packages for specialized algorithms.
- Strong push in the community to make software freely available.

Expectations

The courses is intended for **hard-working, technically skilled, highly motivated students**.

- Take notes during class. Do the readings. Review the slides.
- Participate in discussions and sessions. Ask questions.
- Respect the coursework policy.

Participants are expected to show **initiative, creativity, scientific rigour, critical thinking, and good communication skills**.

- Be prepared to work hard on the projects. Work well in a team.
- Provide constructive feedback in peer-reviews.

Read this carefully

- **Some of the course work will be individual, other components can be completed in groups.** It is the responsibility of each student to understand the policy for each work, and ask questions of the instructor if this is not clear.

- It is the responsibility of each student to **carefully acknowledge all sources** (papers, code, books, websites, individual communications) using appropriate referencing style when submitting work.

- **We will use automated systems to detect possible cases of text or software plagiarism.** Cases that warrant further investigation will be referred to the university disciplinary officers. Students who have concerns about how to properly use and acknowledge third-party software should consult a McGill librarian or the TAs.

Questions?

What is machine learning?

- A definition (by Tom Mitchell):

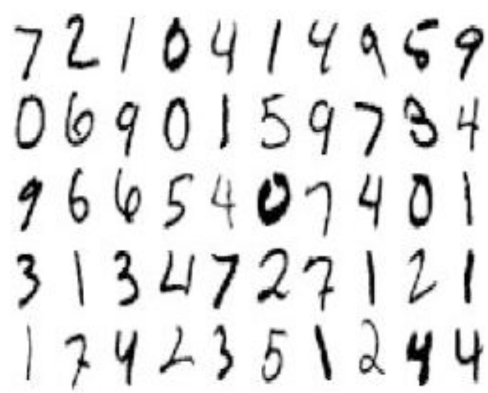
  “How can we build computer systems that automatically improve with experience, and what are the fundamental laws that govern all learning processes?”

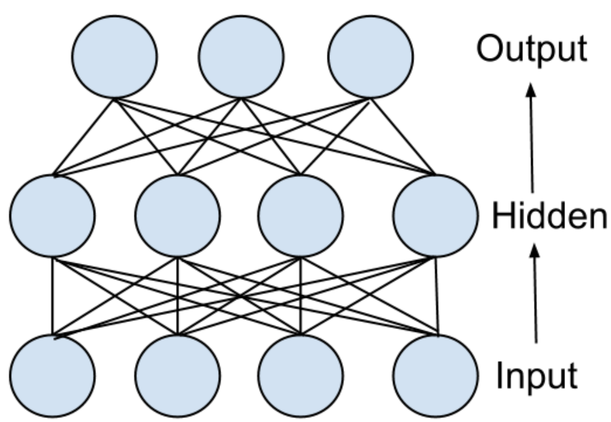
- More technically:

  "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"

Case study #1: Optimal character recognition

- Handwritten digit recognition: >99% accuracy (on a large dataset).



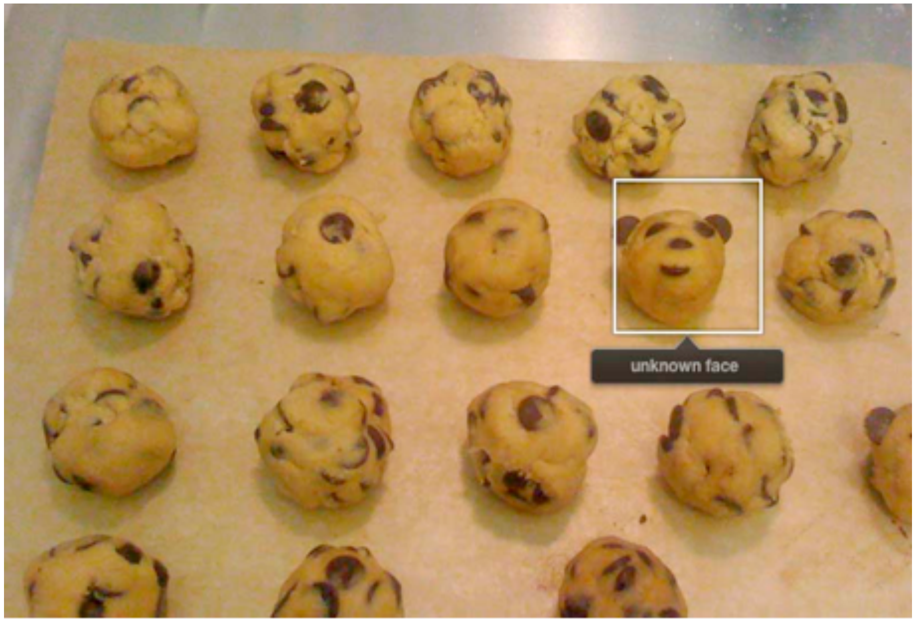


Boxes represent the weights into a hidden node in a neural network learner.

Case study #2: Computer Vision

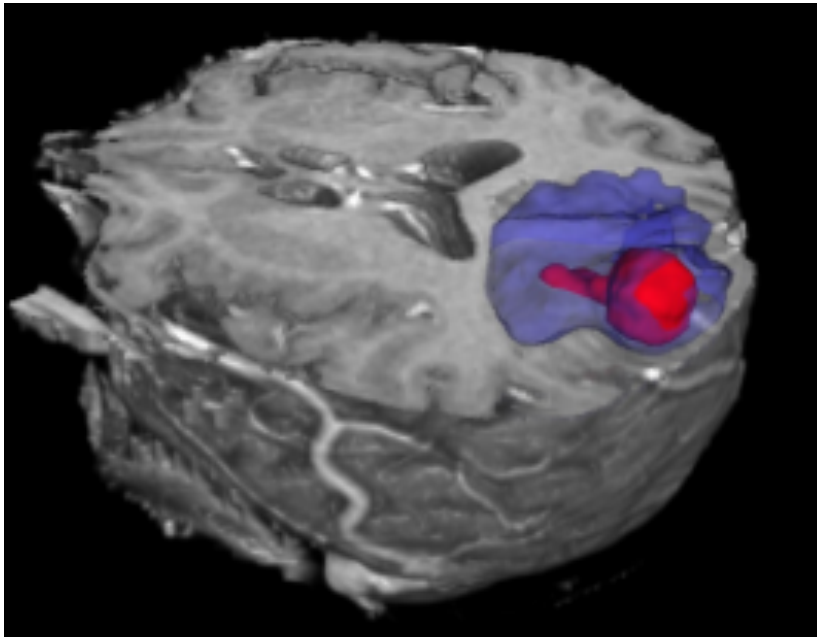
- Face recognition.

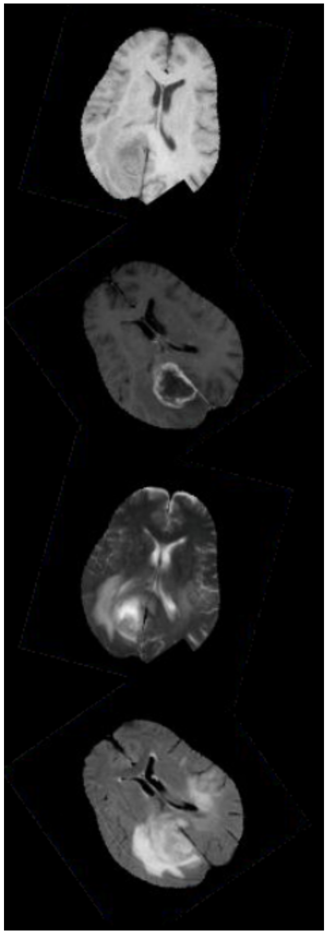
Not always perfect!



Case study #2: Computer Vision

- Voxel-level tumour segmentation





Case study #3: Text analysis

- Learning a language model:

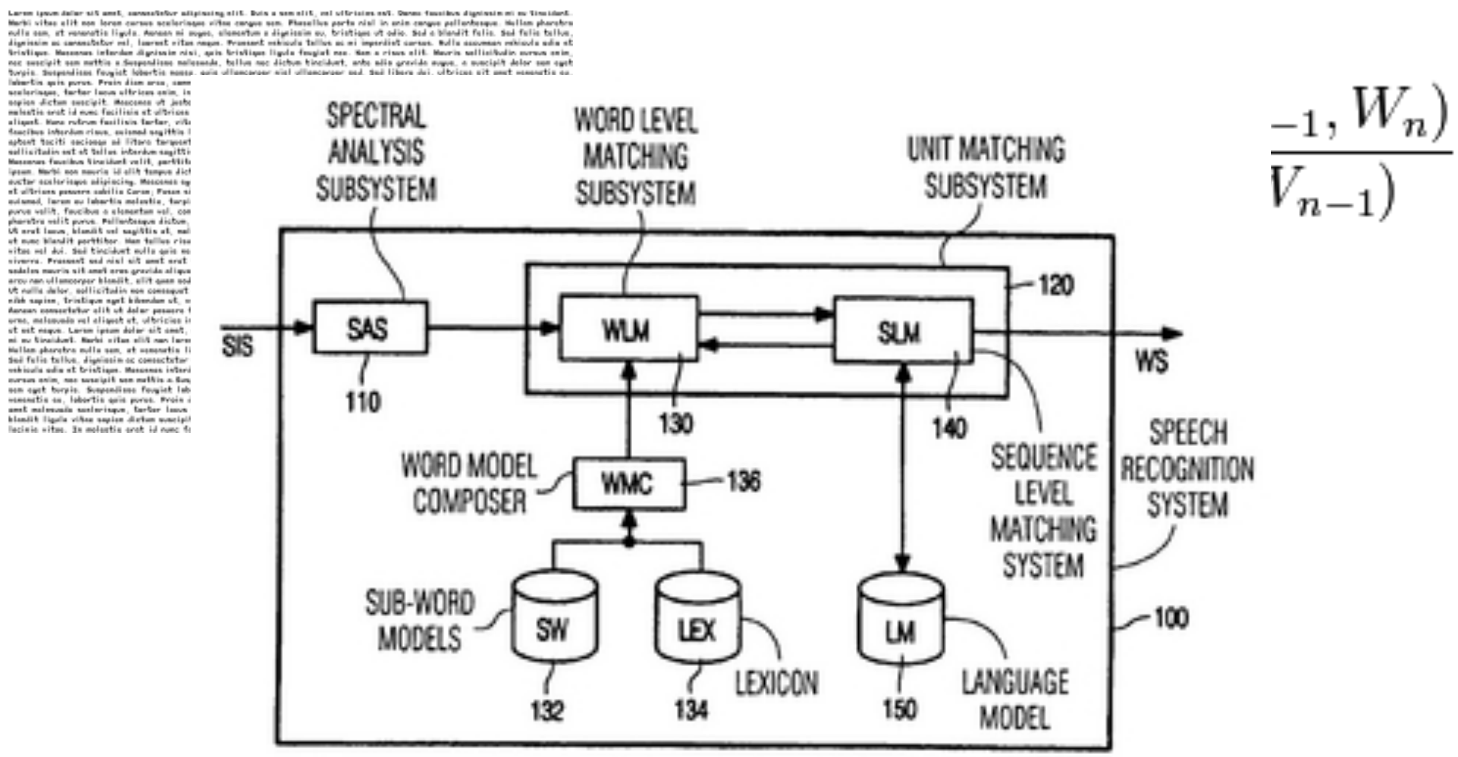
Text corpus → Statistical language model

$$P(W_n|W_{n-1}) = \frac{P(W_{n-1}, W_n)}{P(W_{n-1})}$$

Case study #3: Text analysis

- Learning a language model:

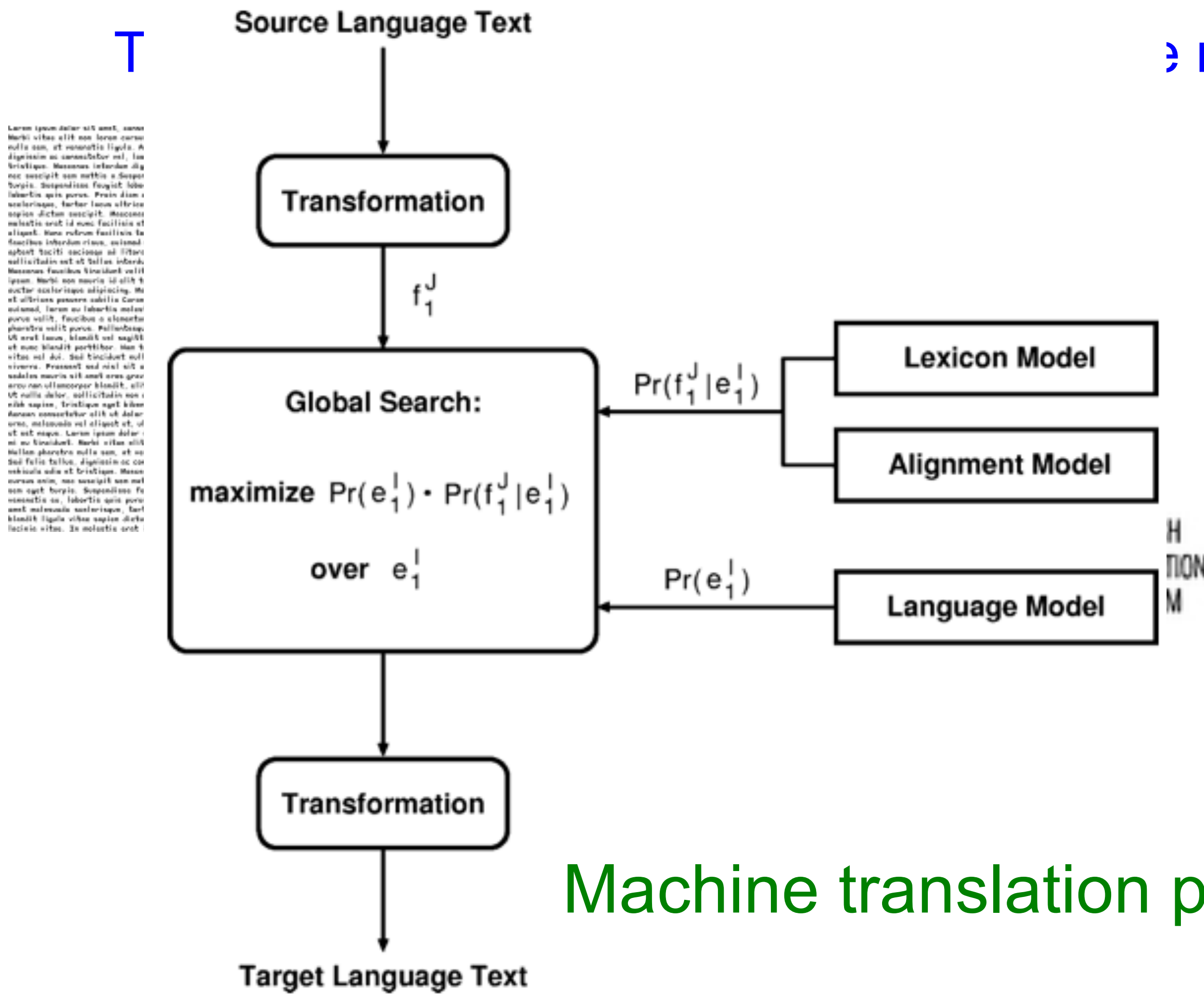
Text corpus ⟶ Statistical language model



Speech recognition pipeline

Case study #3: Text analysis

- Learning a language model:



Machine translation pipeline

Case study #3: Text analysis

- From vision input to text output:

“Two pizzas sitting on top of a stove top oven”

“A group of young people playing a game of frisbee”

Case study #3: Text analysis

- Still some work to do!

"A refrigerator filled with lots of food and drinks"

"A yellow school bus parked in a parking lot"

Case study #4: The Netflix Prize

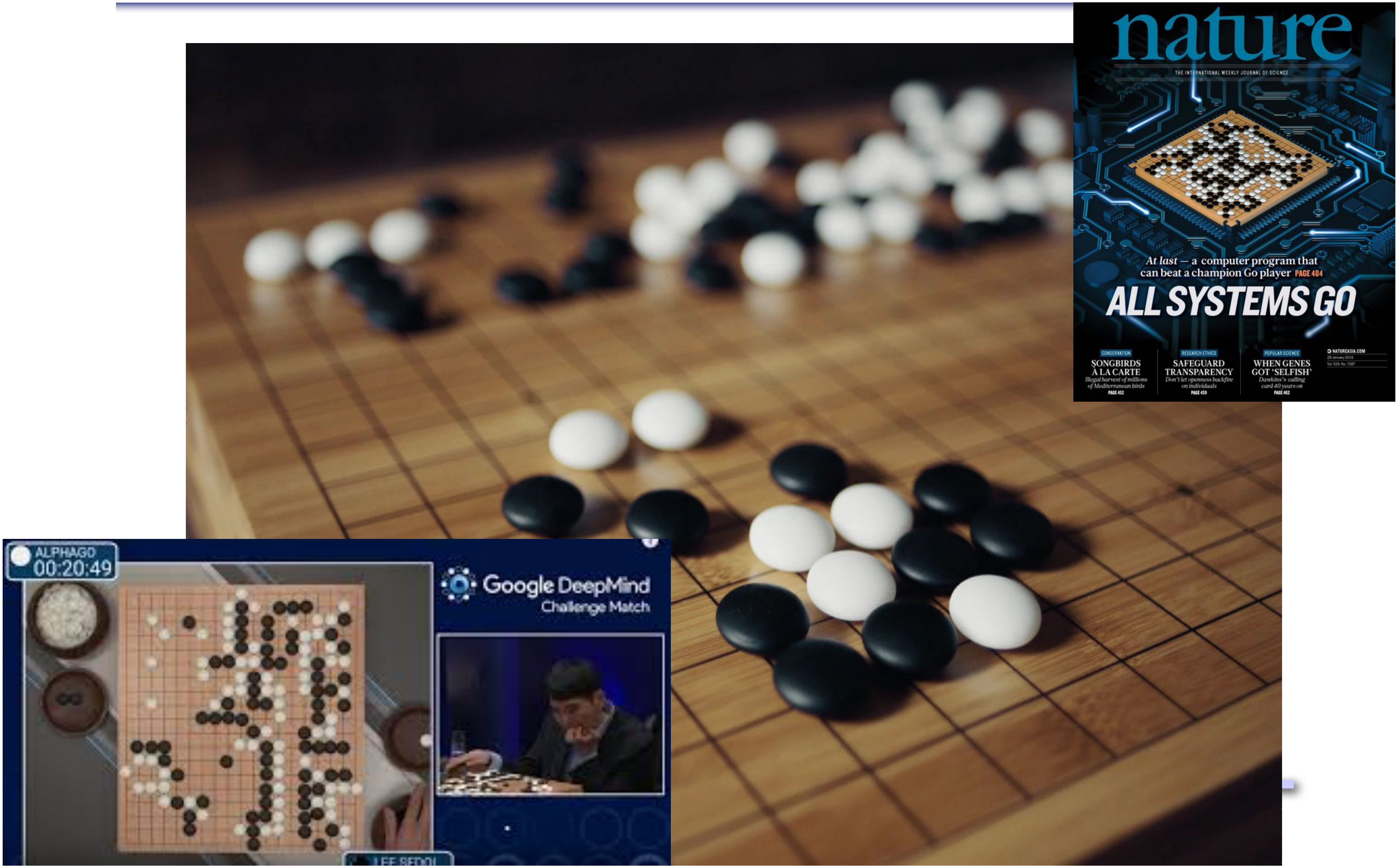
Task: Improve Netflix's recommendation system by 10%.

Training data: 10^8 movie ratings, to build the ML algorithm.

Test set: 1.5×10^6 ratings to evaluate final performance.

Quiz set: 1.5×10^6 ratings to calculate leaderboard scores.

Case study #5: Playing games



Types of machine learning

- Supervised learning
  - Classification
  - Regression
- Unsupervised learning
- Reinforcement learning

Terminology

- Columns are called **input variables** or **features** or **attributes**.
- The columns we are trying to predict (outcome and time) are called **output variables** or **targets**.
- A row in the table is called a **training example** or **instance**.
- The whole table is called a **data set**.

tumor size	texture	perimeter	. . .	outcome	time
18.02	27.6	117.5		N	31
17.99	10.38	122.8		N	61
20.29	14.34	135.1		R	27
. . .					

Supervised learning - Classification

Goal: Learning a function for a categorical output.

E.g.: Spam filtering. The output ("Spam?") is binary.

	Sender in address book?	Header keyword	**Word 1**	**Word 2**	…	**Spam?**
x1	Yes	Schedule	Hi	Profesor	…	No
x2	Yes	meeting	Joelle	I	…	No
x3	No	urgent	Unsecured	Business	…	Yes
x4	No	offer	Hello	I	…	Yes
x5	No	cash	We'll	Help	…	Yes
x6	No	comp-551	Dear	Professor	…	No
…						

Supervised learning - Regression

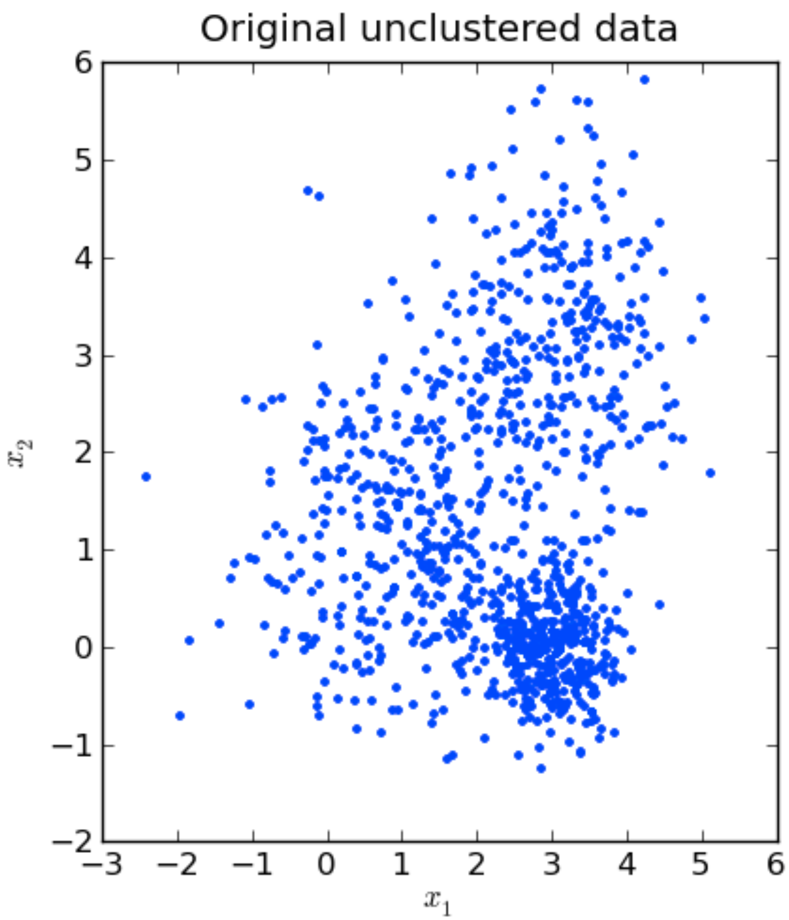
<u>Goal</u>: Learning a function for a continuous output.

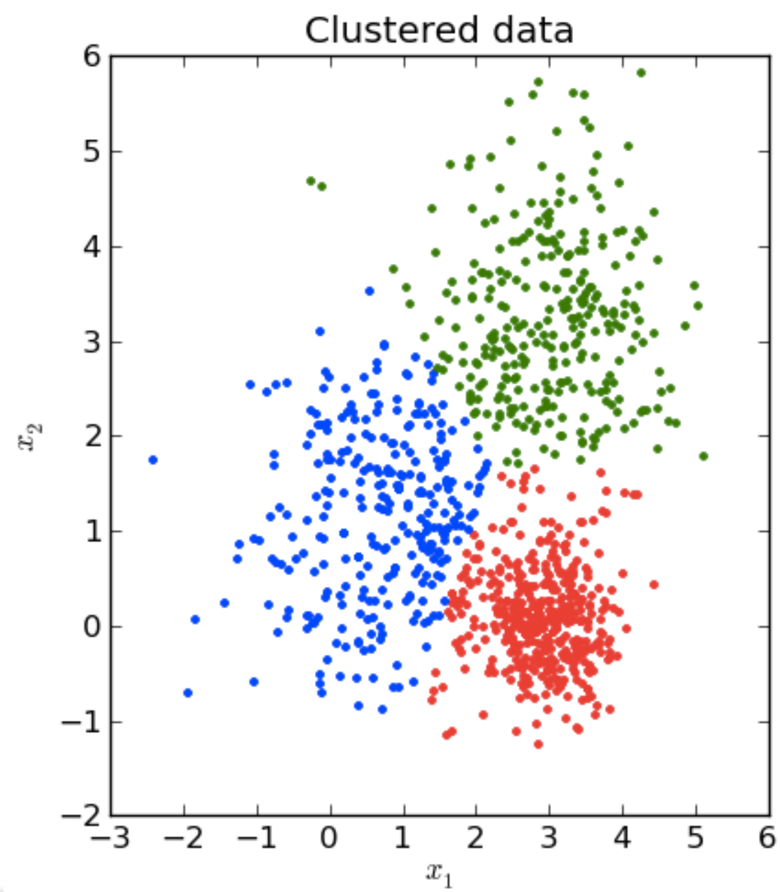
E.g.: Self-driving car speed control. The output ("speed") is continuous.

Unsupervised learning

Goal: Learning a function over the input alone.

E.g. Organizing data into classes. Inferring distances between data points.

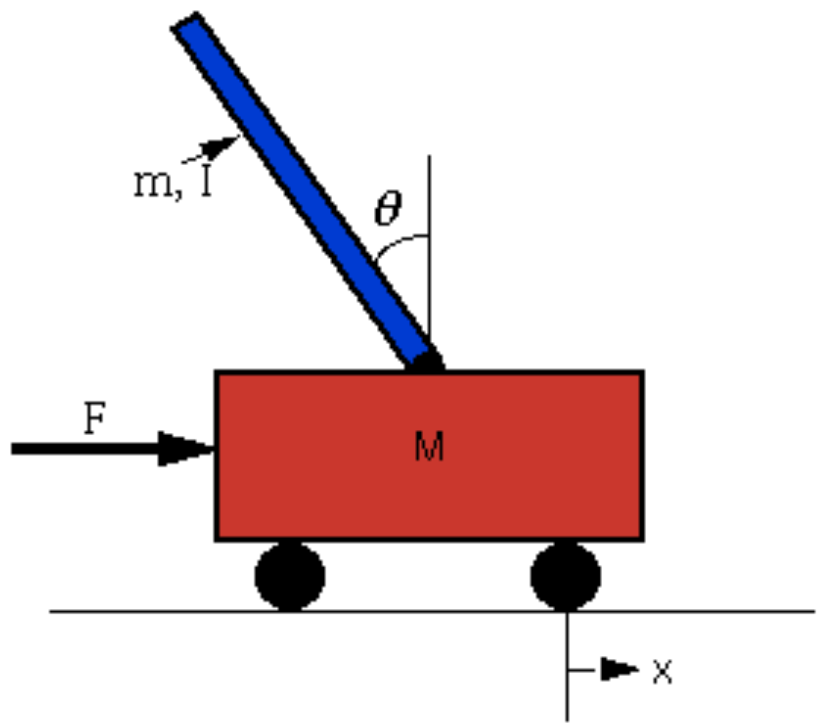




Reinforcement learning

Goal: Learning a sequence of actions that optimizes costs/rewards.

E.g.: Balancing an inverted pendulum.



ML today

- Currently the method of choice for many applications:
  - Speech recognition
  - Computer vision
  - Robot control
  - Fraud detection

  and growing…
- Why so many applications?

ML today

- Currently the method of choice for many applications:
  - Speech recognition
  - Computer vision
  - Robot control
  - Fraud detection

  and growing…
- Why so many applications?
  - Increase in number of sensors/devices ➔ We have loads of data!
  - Increase in computer speed and memory ➔ We can process the data!
  - Better ML algorithms and software for easy deployment.
  - Increasing demand for customized solutions (e.g personalized news).

Harvard Business Review

SEARCH

THE MAGAZINE | BLOGS | AUDIO & VIDEO | BOOKS | CASES | WEBINARS

Guest | limited access

Register today and save 20%* off your first order! Details

THE MAGAZINE

October 2012

ARTICLE PREVIEW To read the full article: Sign in or Register for free. HBR Subscribers activate your free archive access »

Data Scientist: The Sexiest Job of the 21st Century

by Thomas H. Davenport and D.J. Patil

Comments (87)

RELATED

Executive Summary

ALSO AVAILABLE

- Buy PDF

Research challenge: Big data

- Old-style $O(n^2)$ algorithms simply won't work.
- Fitting the data on a single machine may not be feasible. Work from a stream of examples (process every example only once.)
- Must distribute computations across multiple machines.

E.g. Predicting which ad is interesting (from John Langford)

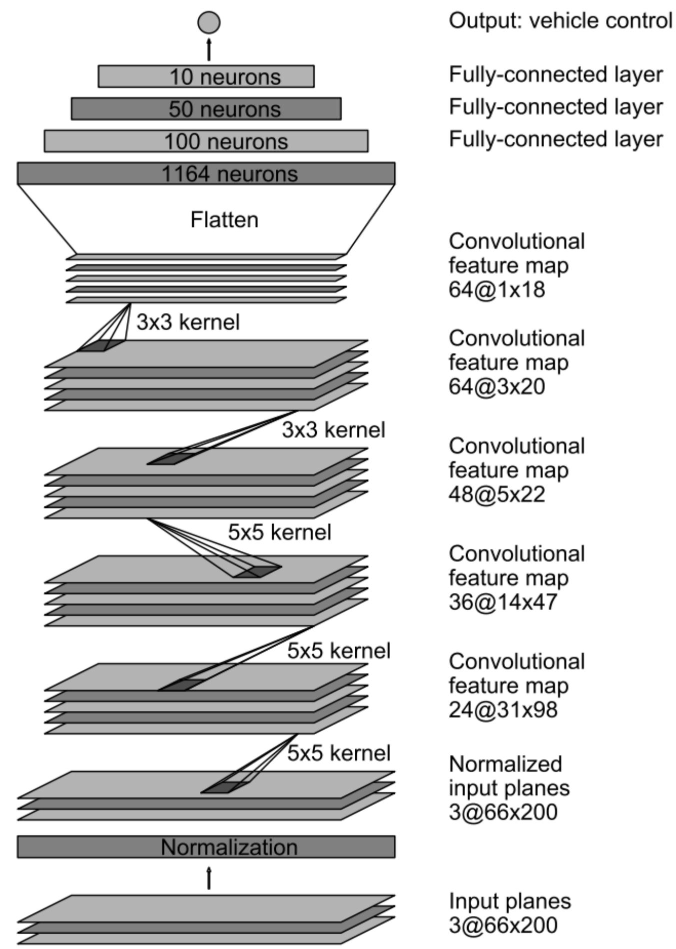
- 2.1TB sparse features
- 17B examples
- 16M parameters
- 1K computation nodes

Research challenge: End-to-end learning

- From raw features => high-order decision.

E.g.

- Single characters => Text classification
  - https://arxiv.org/abs/1509.01626
- Pixels => Steering angle for autonomous driving
  - https://arxiv.org/pdf/1604.07316v1.pdf



Lots of other (inter-disciplinary) challenges

- Many open questions about algorithmic methods and theoretical characterization.
  - Inferring the right representation / model.
  - Exploration vs Exploitation
- Weakness in evaluation methods.
- Privacy concerns in obtaining and releasing data.
- Many exciting under-explored applications!

DOI:10.1145/2347736.2347755

Tapping into the "folk knowledge" needed to advance machine learning applications.

BY PEDRO DOMINGOS

A Few Useful Things to Know About Machine Learning

MACHINE LEARNING SYSTEMS automatically learn programs from data. This is often a very attractive alternative to manually constructing them, and in the last decade the use of machine learning has spread rapidly throughout computer science and beyond. Machine learning is used in Web search, spam filters

is needed to successfully develop machine learning applications is not readily available in them. As a result, many machine learning projects take much longer than necessary or wind up producing less-than-ideal results. Yet much of this folk knowledge is fairly easy to communicate. This is the purpose of this article.

» key insights

- Machine learning algorithms can figure out how to perform important tasks by generalizing from examples. This is

Final comments

- Come to class! Come prepared!

- For next class:
  - (Must) Read this paper:
    - *http://homes.cs.washington.edu/~pedrod/papers/cacm12.pdf*
  - (If necessary) Review basic algebra, probability, statistics.
    - Ch.1-2 of Bishop.
    - *http://www.cs.mcgill.ca/~dprecup/courses/ML/Materials/prob-review.pdf*
    - *http://www.cs.mcgill.ca/~dprecup/courses/ML/Materials/linalg-review.pdf*
    - Many online resources.
  - (Optional) Read Chap.1-2 of Bishop, Ch. 1 of Hastie et al. or Ch.2 of Shalev-Schwartz et al.