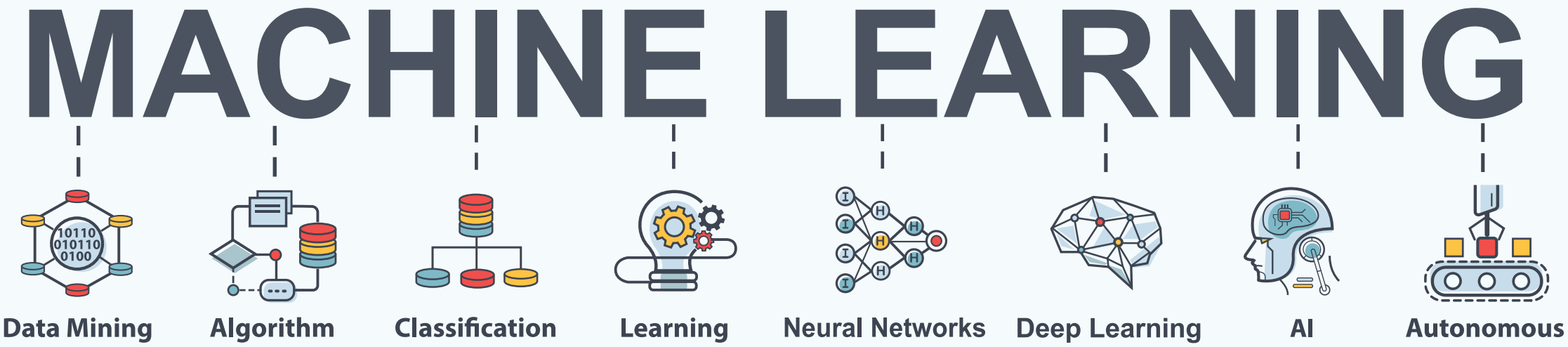


CP8318/CPS803
Lecture 1

Instructor: Nariman Farsad

* Some of the slides in this deck are adopted from Stanford Machine Learning course.

Why Machine Learning is Important?

Andrew Ng

“AI is the new electricity!”

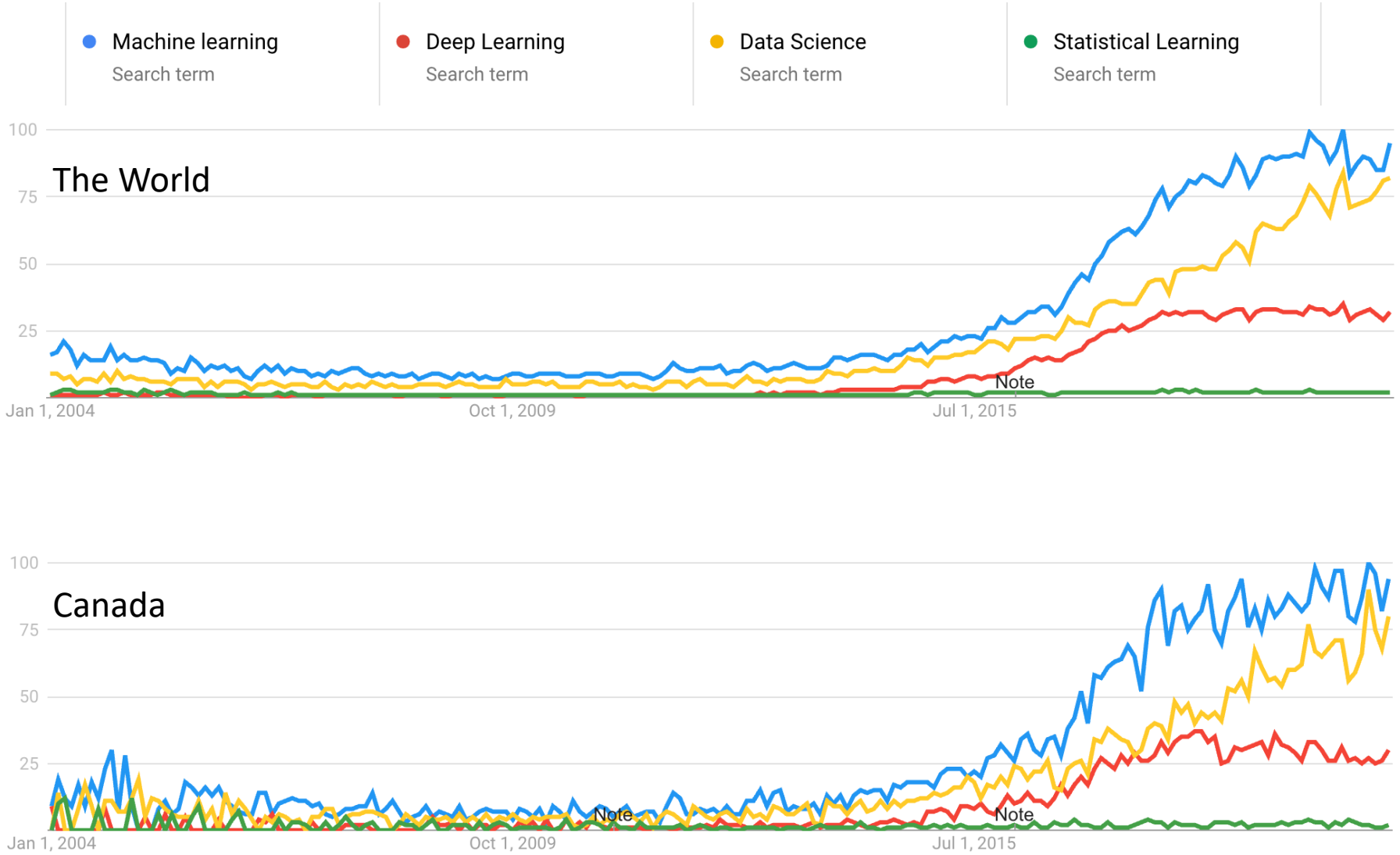
Electricity transformed many industries: transportation, manufacturing, healthcare, communications and more

AI will also bring about a big transformation

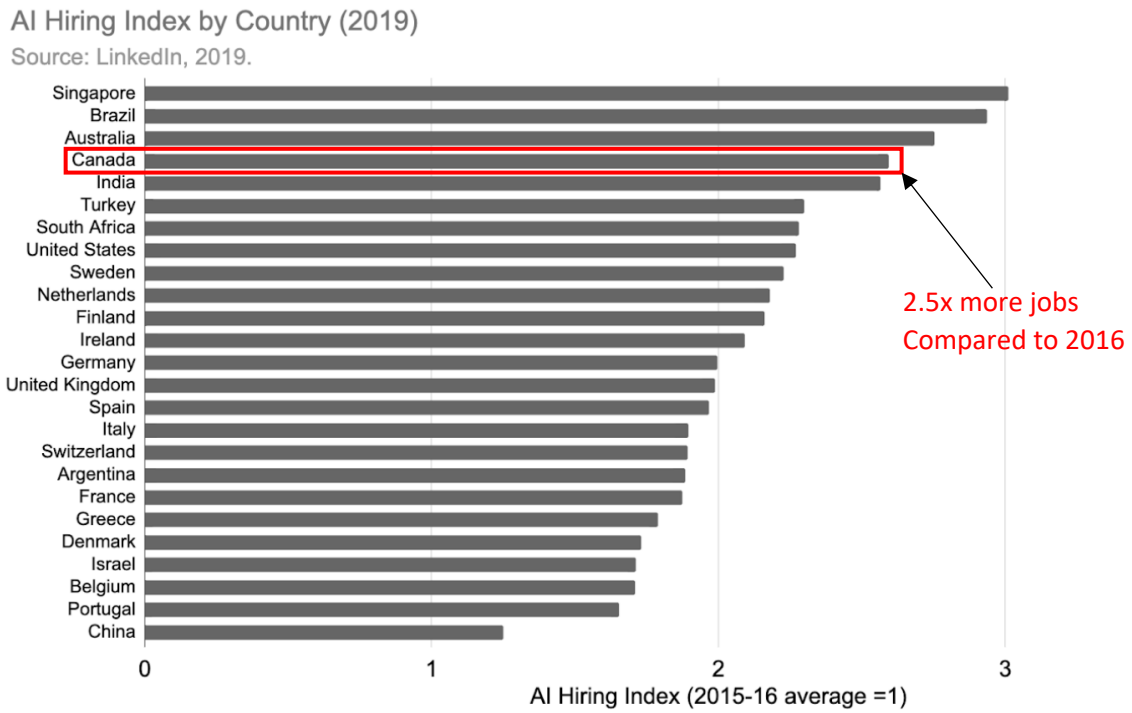
<u>AI versus Machine Learning</u>

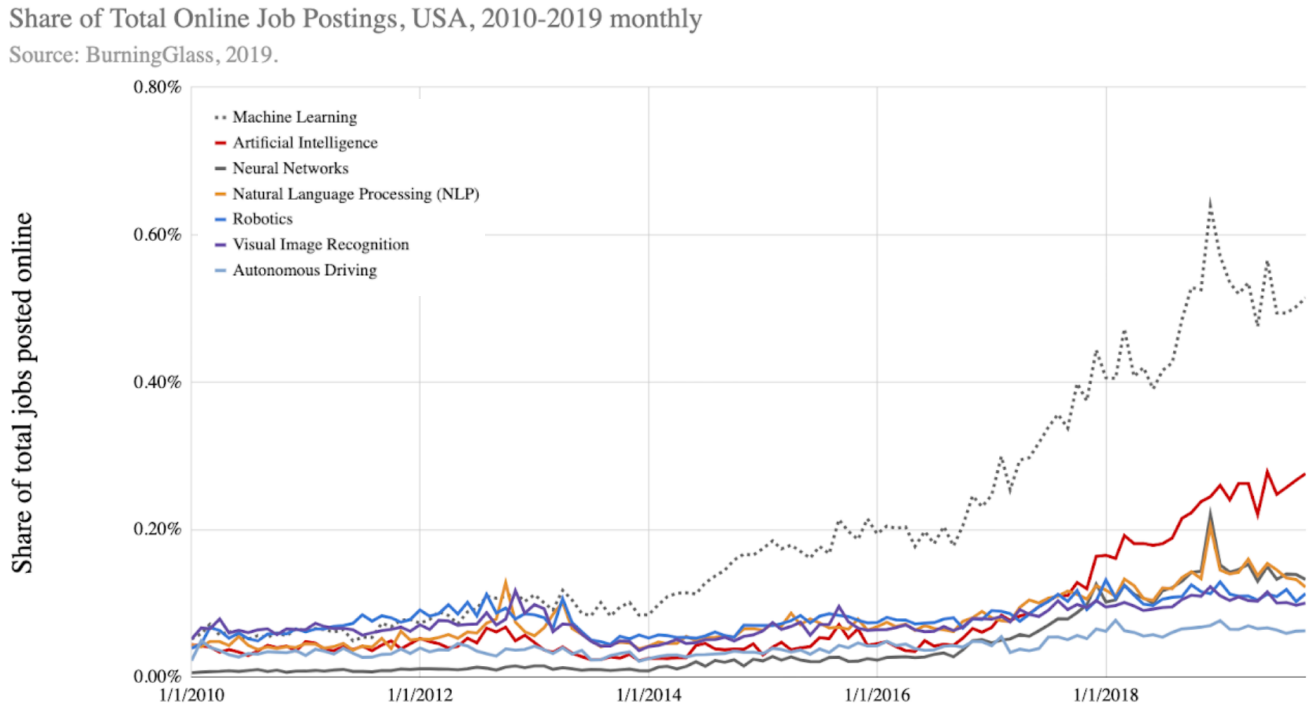
- Definition of AI is vague, and many agree that it changes with time
- Machine learning has a clear definition as we will see and is the most important part of AI (at least currently)

Google Trends



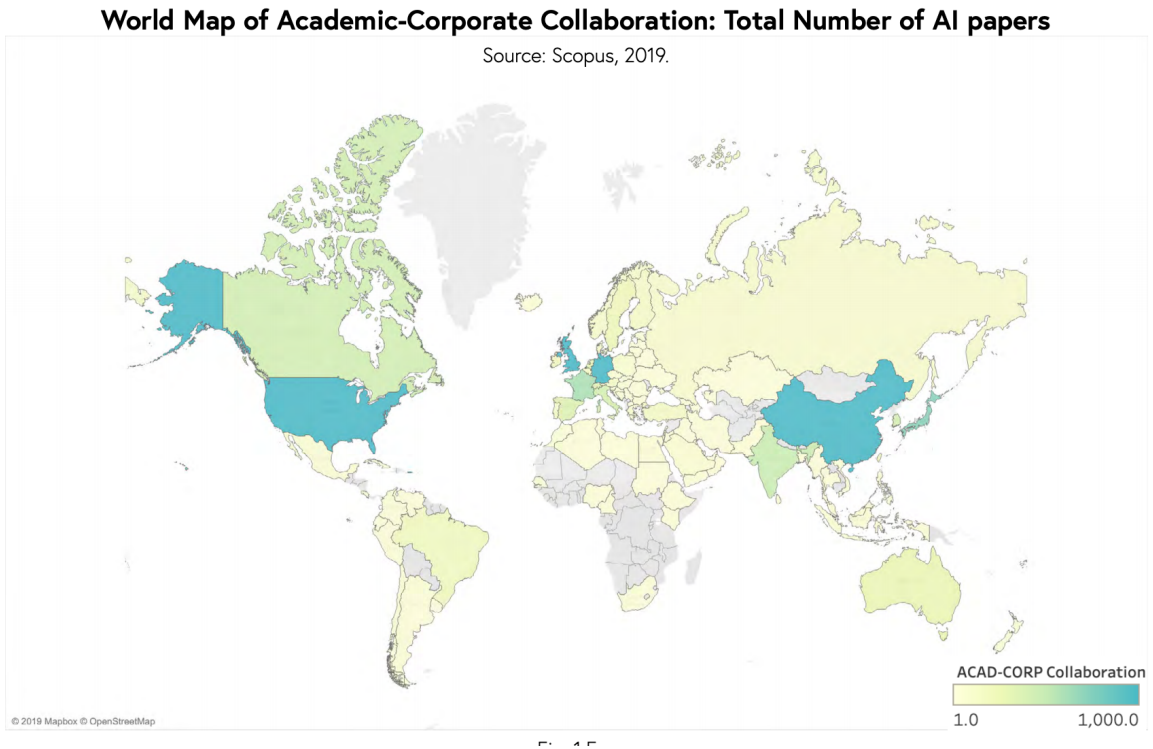
Jobs in Machine Learning and AI

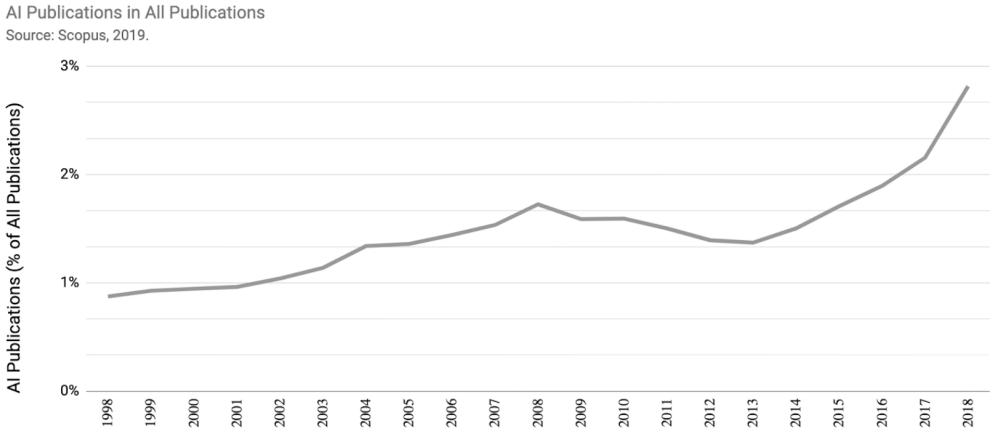


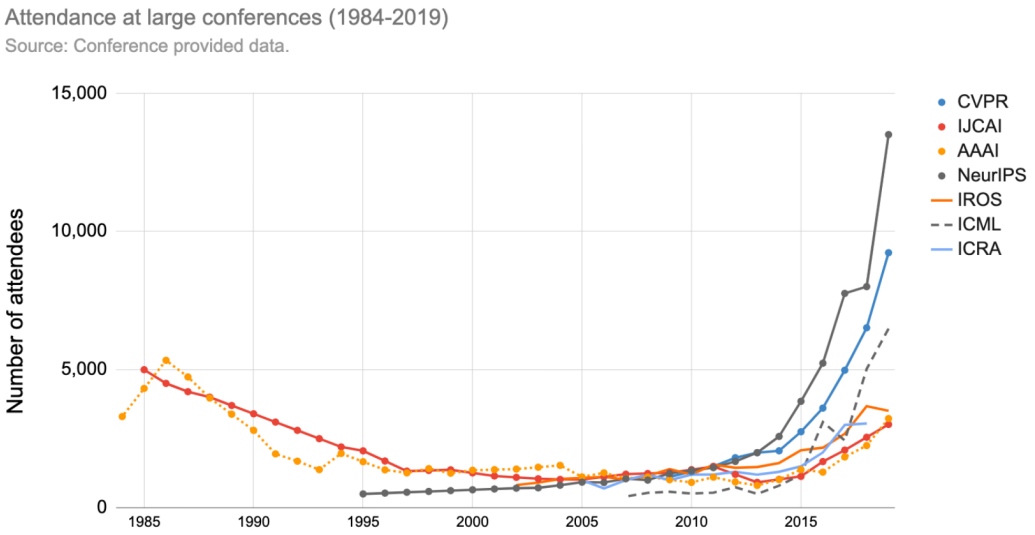


Source: https://hai.stanford.edu/sites/default/files/ai_index_2019_report.pdf

Research in Machine Learning and AI







Source: https://hai.stanford.edu/sites/default/files/ai_index_2019_report.pdf

Poll: Why are you taking this course?

A. To learn to apply machine learning to different problems.

B. To become an expert in machine learning or do research in this field.

C. I was just curious what is the big deal with machine learning.

Poll: Why are you taking this course?

A. To learn to apply machine learning to different problems.

B. To become an expert in machine learning or do research in this field.

C. I was just curious what is the big deal with machine learning.

Poll: How was the video/audio quality so far?

A. Good.

B. Not good, but I can follow all the time.

C. Not great, but I can follow for majority of time.

D. Not good, I can’t see or hear most of the time.

E. I don’t see or hear anything!

Today's Agenda

1. Administrative
   - Please read the course outline in D2L carefully
   - Course website: http://narimanfarsad.com/cps803/index.html

2. Course overview and introduction to ML

Teaching Staff

Instructor

Nariman Farsad

Teaching Assistants

Kayvan Tirdad

Research:
Machine Learning & Deep Learning

Matthew Kowal

Research:
Deep Learning & Computer Vision

Soosan Naderi Mighan

Research:
Information security, Big Data and IoT, Blockchain

Alex Dela Cruz

Research:
Deep Learning, Natural Language Processing, Health Care

Ian Robinson

Research:
Neurofeedback and Deep Learning

Pre-requisite

- Linear algebra (MTH 108)
  - matrix multiplication, eigenvector
- Multivariable Calculus (MTH 207)
  - Partial derivative, gradients, Jacobian Matrix
- Probability (Unwritten pre-requisite)
  - distribution, random variable, expectation, conditional probability, variance, density
- Basic programming (in Python)
- We will review these during first 3 weeks
  - Good Resource: http://narimanfarsad.com/cps803/background.html

This is a mathematically intense course.
But that's why it's exciting and rewarding!

Differences From Previous Years

- We will be using course notes of Andrew Ng at Stanford
- Everything will be online --- lectures, office hours, discussions between students
  - We strongly encourage you to study with other students
  - Technology: D2L discussion boards, Zoom, Slack, Discord, WhatsApp …
- Enrollments increased by ~1.5x compared to last year
  - About 180 students in class

Course Evaluation

- Three assignment each 15% (total 45%)
  - have theoretical (math) and practical (programming) questions
  - Are very intensive, start as soon as they are released, or you can't finish them
  - It is fine to discuss the problems with your classmates, but must write your own solutions
  - CP8318 students get extra questions for each assignment

- Final Project (5% for proposal and 50% for final submission)
  - CPS803 will work in groups of 4 (randomly assigned)
    - Each group is assigned a TA, who will be their mentor for the projects throughout the semester
  - CP8318 can work individually or in groups of up to 4
    - The instructor will mentor for the projects throughout the semester
  - Final submission evaluation based on TA feedbacks, groupmates feedbacks, code, report, video

Final Project

- Since this is most of your grade, the project you do must be “significant”
  - See this page for more details: http://narimanfarsad.com/cps803/project.html

- Some potential topic areas:

  - Athletics & Sensing Devices
  - Audio & Music
  - Computer Vision
  - Finance & Commerce
  - General Machine Learning
  - Life Sciences
  - Natural Language
  - Physical Sciences
  - Theory
  - Reinforcement Learning
  - Covid-19

Online Lecture Structures

- Three hours is too long for lectures on machine learning (even if in person)
- Every week 2PM – 4PM lectures over zoom (synchronous lectures)
  - Lectures will be recorded and released by end of the Saturday after the Friday lecture
- Every week from 4PM – 5PM breakout rooms for work on final project
- Every week there will be more lectures and/or video materials released for learning (asynchronous lectures)
  - Will be released at the same time as recorded video of synchronous lectures

The Last Lecture

- For the last lecture in we plan to have guest lecture(s) from industry.

- I will announce who are the speaker(s) by the end of October.

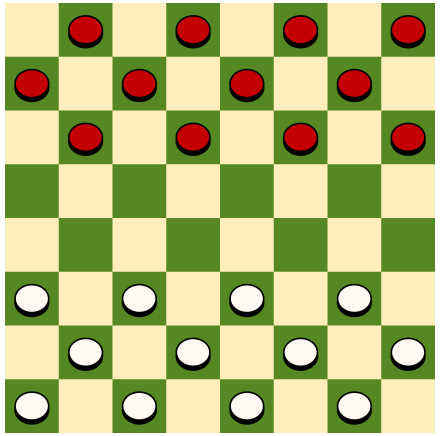
Introduction to Machine Learning

Definition of Machine Learning

Arthur Samuel (1959): Machine Learning is the field of study that gives the computer the ability to learn without being explicitly programmed.

A. L. Samuel*

Some Studies in Machine Learning
Using the Game of Checkers. II—Recent Progress



Photos from Wikipedia

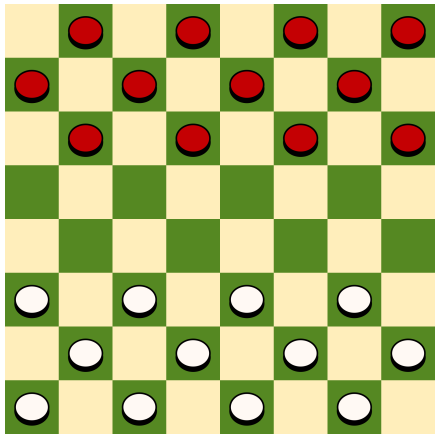
Definition of Machine Learning

Tom Mitchell (1998): 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.

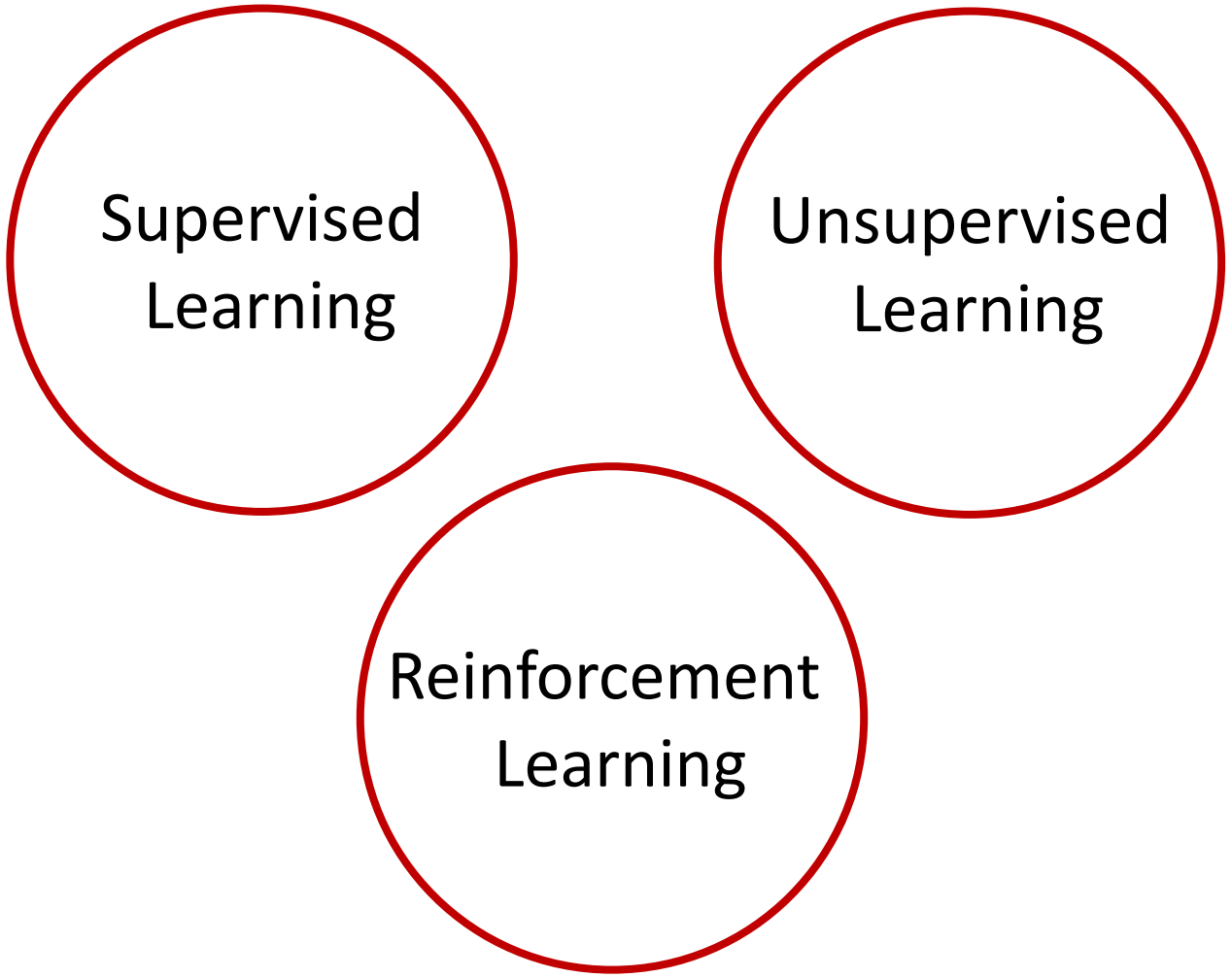
Image from Tom Mitchell's homepage

Experience (data): games played by the program (with itself)

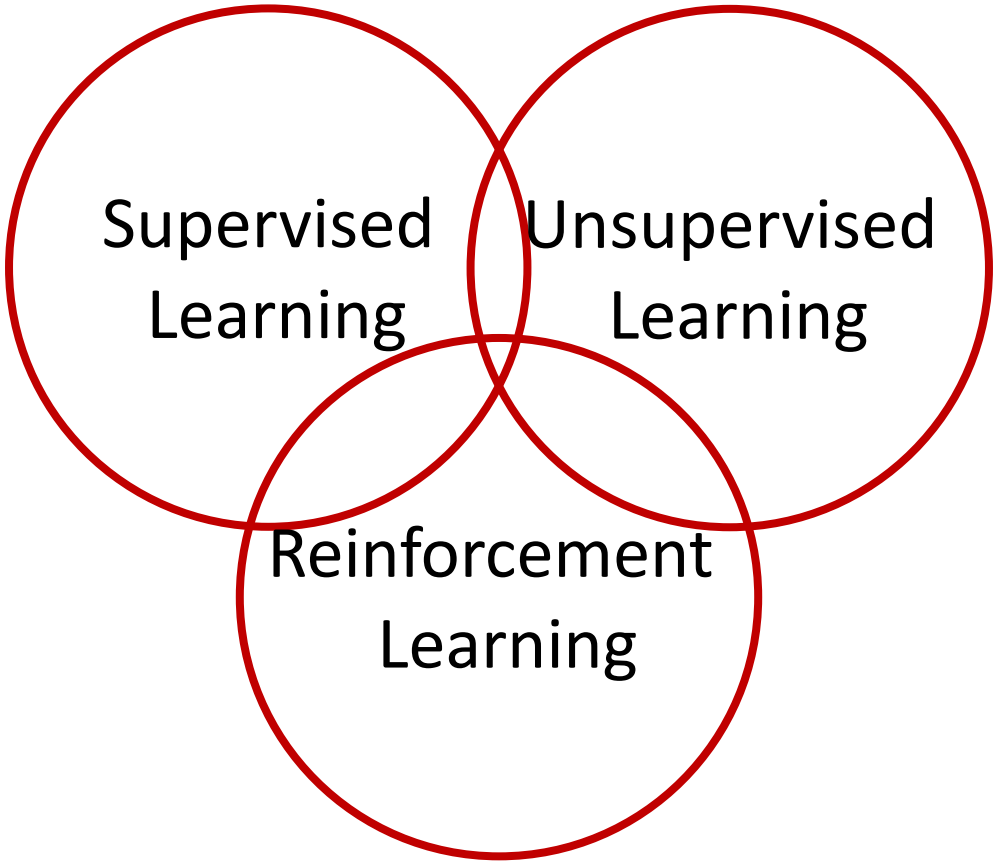
Performance measure: winning rate



Taxonomy of Machine Learning (Different Learning Approaches)



Taxonomy of Machine Learning (Tools and Methods Overlap)



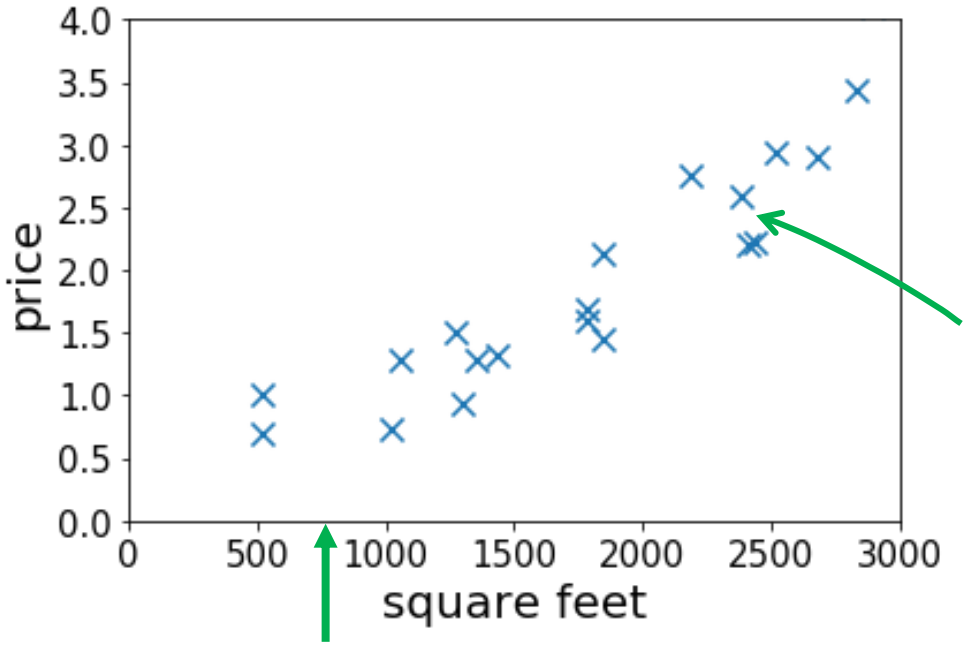
Supervised Learning

Housing Price Prediction

- Given: a dataset that contains n samples

$$(x^{(1)}, y^{(1)}), \ldots (x^{(n)}, y^{(n)})$$

- Task: if a residence has x square feet, predict its price?



15th sample
$(x^{(15)}, y^{(15)})$

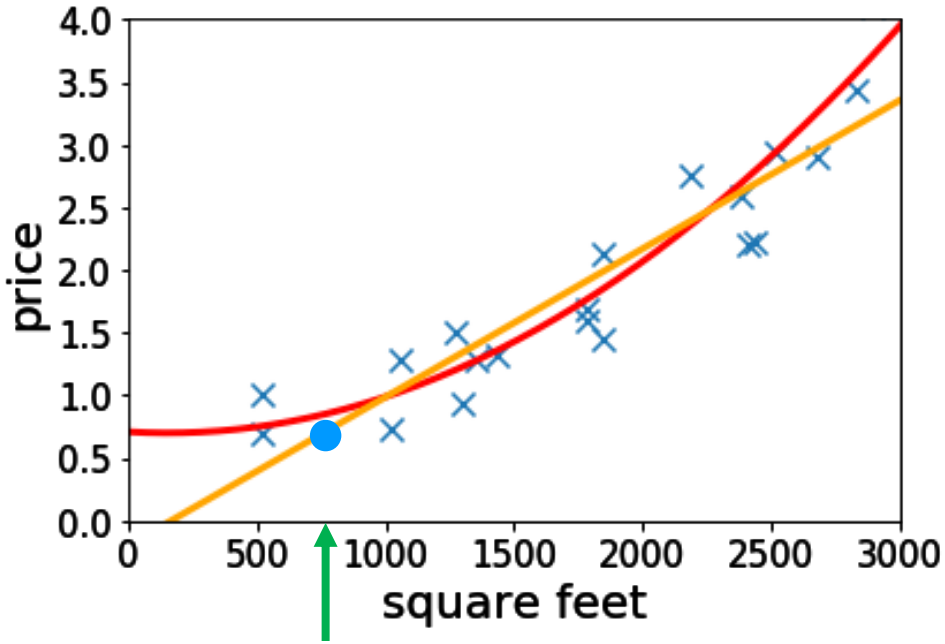
$x = 800$
$y = ?$

Housing Price Prediction

- Given: a dataset that contains n samples

$$\left(x^{(1)}, y^{(1)}\right), \ldots \left(x^{(n)}, y^{(n)}\right)$$

- Task: if a residence has x square feet, predict its price?



$x = 800$

$y = ?$

- Lecture 2&3: fitting linear/qaudratic functions to the dataset

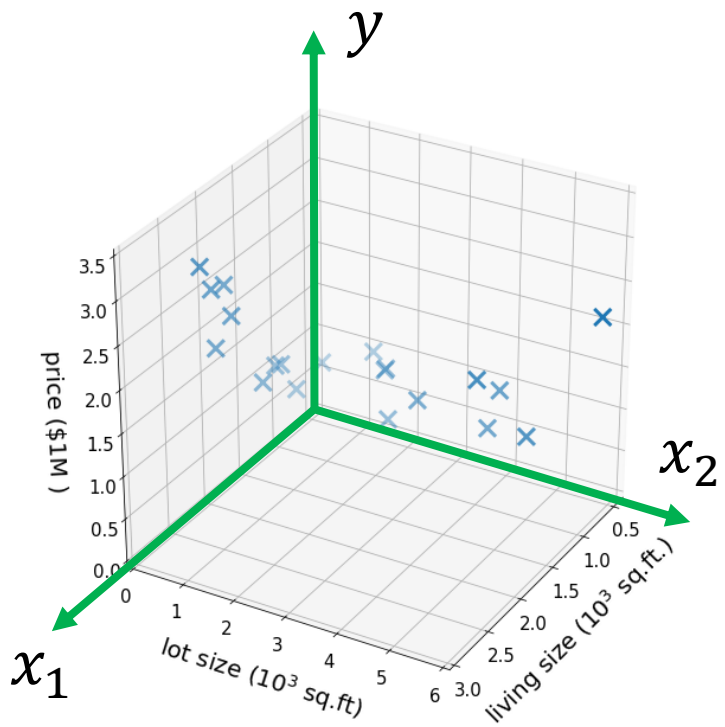
More Features

- Suppose we also know the lot size
- Task: find a function that maps

(size, lot size) → price

features/input $\mathbf{x} \in \mathbb{R}^2$

label/output $y \in \mathbb{R}$



- Dataset: $\left(\mathbf{x}^{(1)}, y^{(1)}\right), \ldots, \left(\mathbf{x}^{(n)}, y^{(n)}\right)$

where $\mathbf{x}^{(i)} = (x_1^{(i)}, x_2^{(i)})$

- "Supervision" refers to $y^{(1)}, \ldots, y^{(n)}$

High-dimensional Features

- $\mathbf{x} \in \mathbb{R}^d$ for large d
- E.g.,

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ \vdots \\ \vdots \\ x_d \end{bmatrix}$$

--- living size
--- lot size
--- # floors
--- condition
--- zip code
⋮

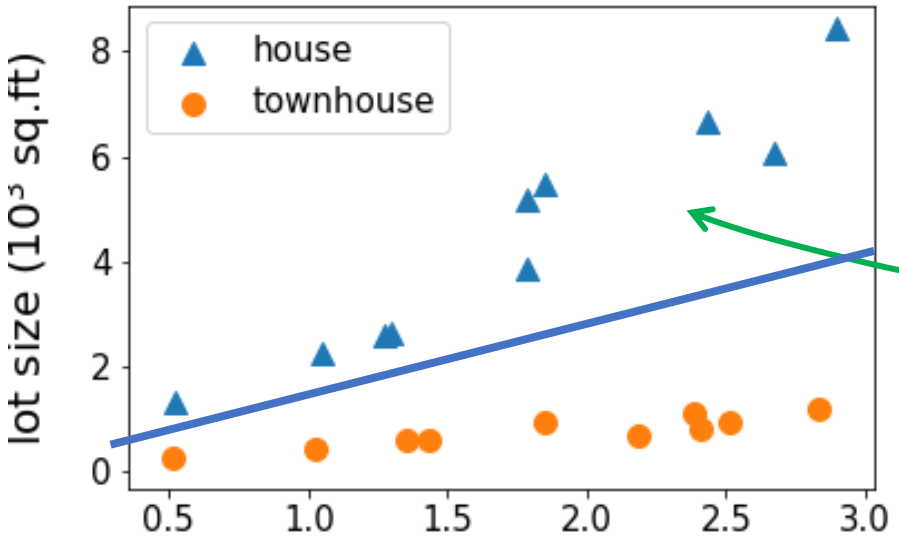
→ y --- price

- Lecture 5-6: infinite dimensional features
- Lecture 8-9: select features based on the data

Regression vs Classification

- regression: if $y \in \mathbb{R}$ is a continuous variable
  - e.g., price prediction
- classification: the label is a discrete variable
  - e.g., the task of predicting the types of residence

(size, lot size) → house or townhouse?



y = house or townhouse?

Lecture 3&4: classification

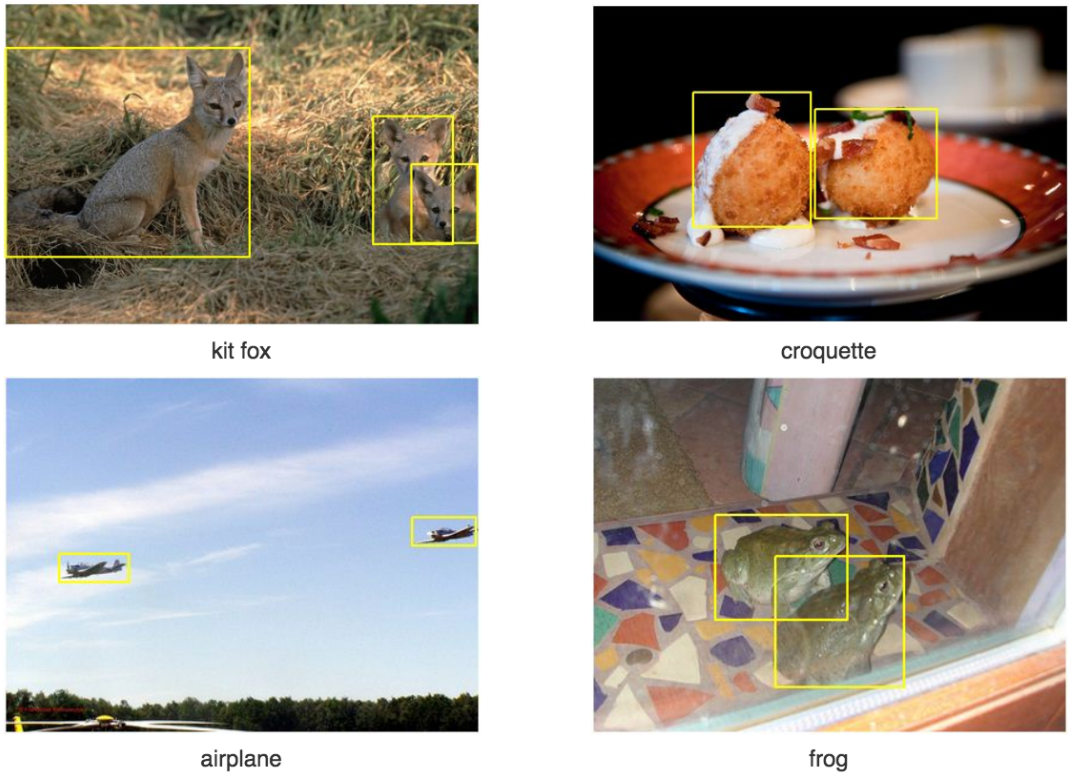
Supervised Learning in Computer Vision

- Image Classification
  - $\mathbf{x}$ = raw pixels of the image, y = the main object

ImageNet Large Scale Visual Recognition Challenge. Russakovsky et al.'2015

Supervised Learning in Computer Vision

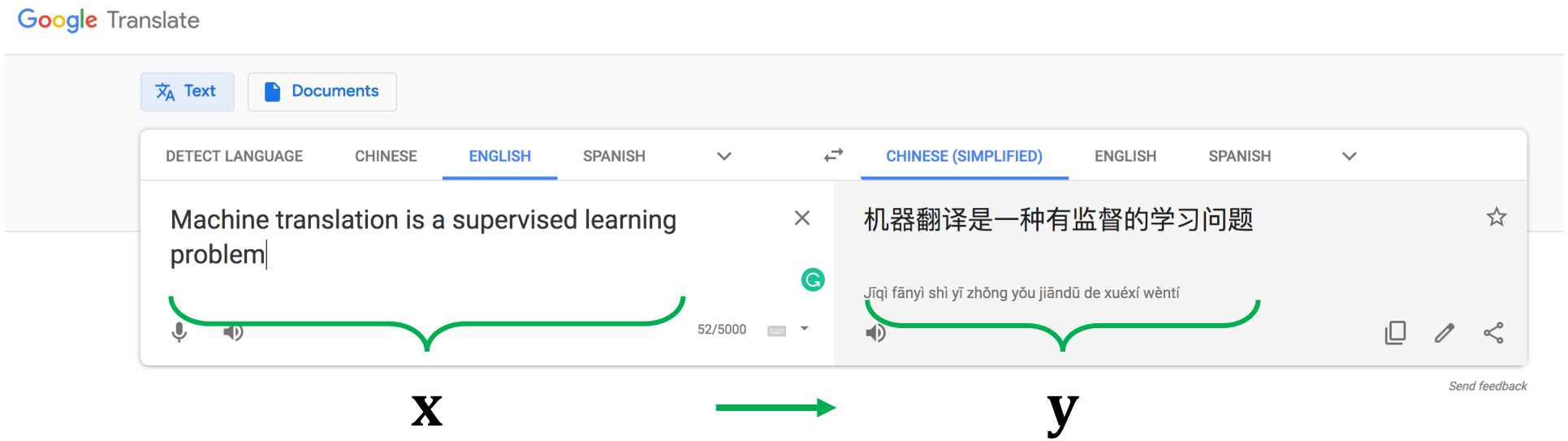
- Object localization and detection
  - $\mathbf{x}$ = raw pixels of the image, $\mathbf{y}$ = the bounding boxes



ImageNet Large Scale Visual Recognition Challenge. Russakovsky et al.'2015

Supervised Learning in Natural Language Processing

- Machine translation



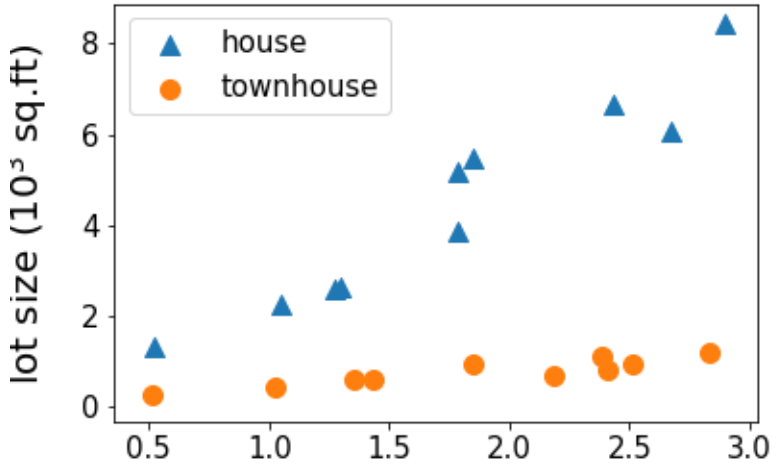
- Note: this course only covers the basic and fundamental techniques of supervised learning
- This is not enough for solving hard vision or NLP problems. We have other courses for that CPS870, CPS843.

Unsupervised Learning

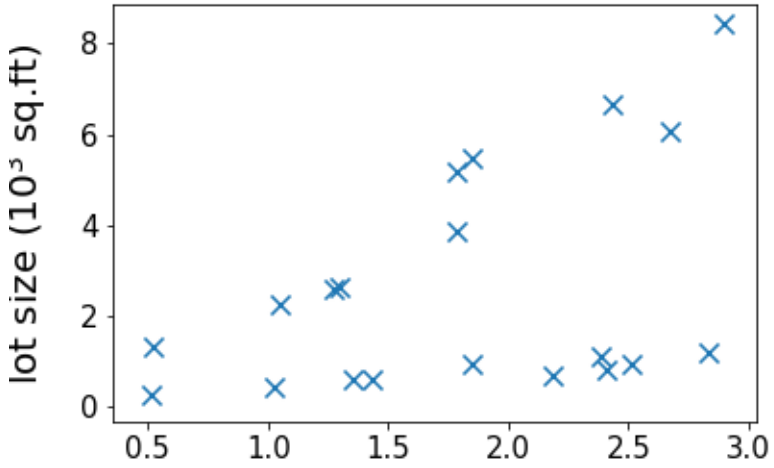
Unsupervised Learning

- Dataset contains no labels: $\mathbf{x}^{(1)}, \ldots \mathbf{x}^{(n)}$
- Goal (vaguely-posed): to find interesting structures in the data

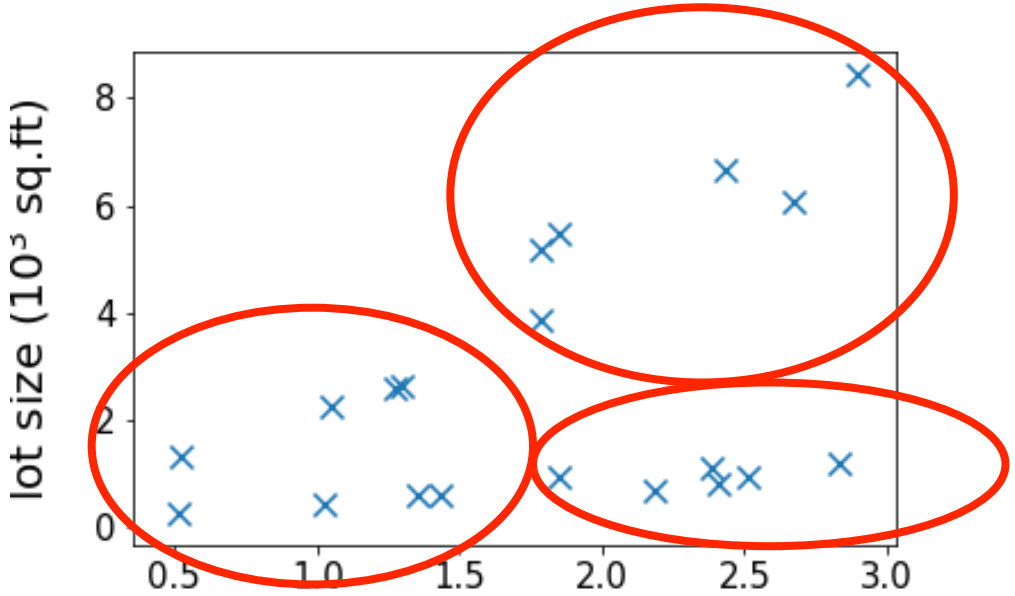
supervised



unsupervised

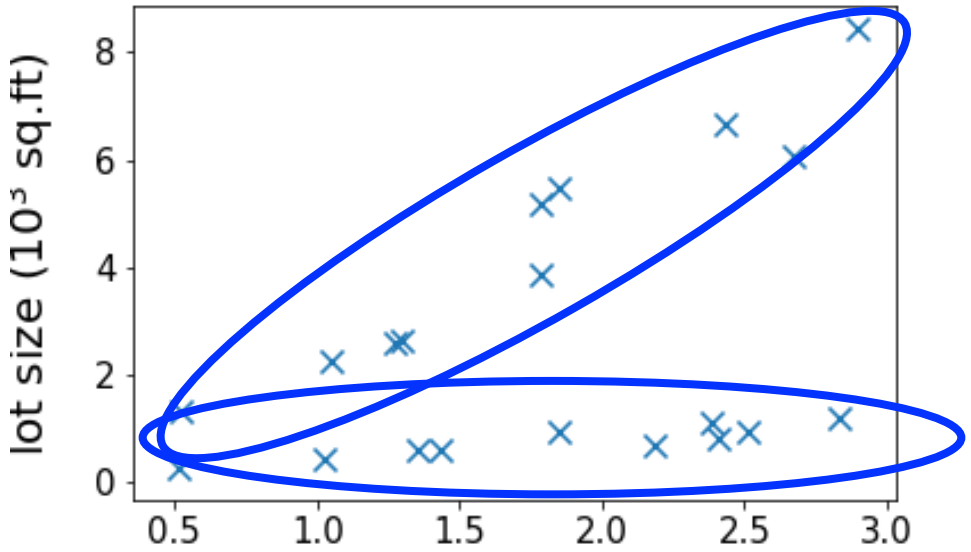


Clustering

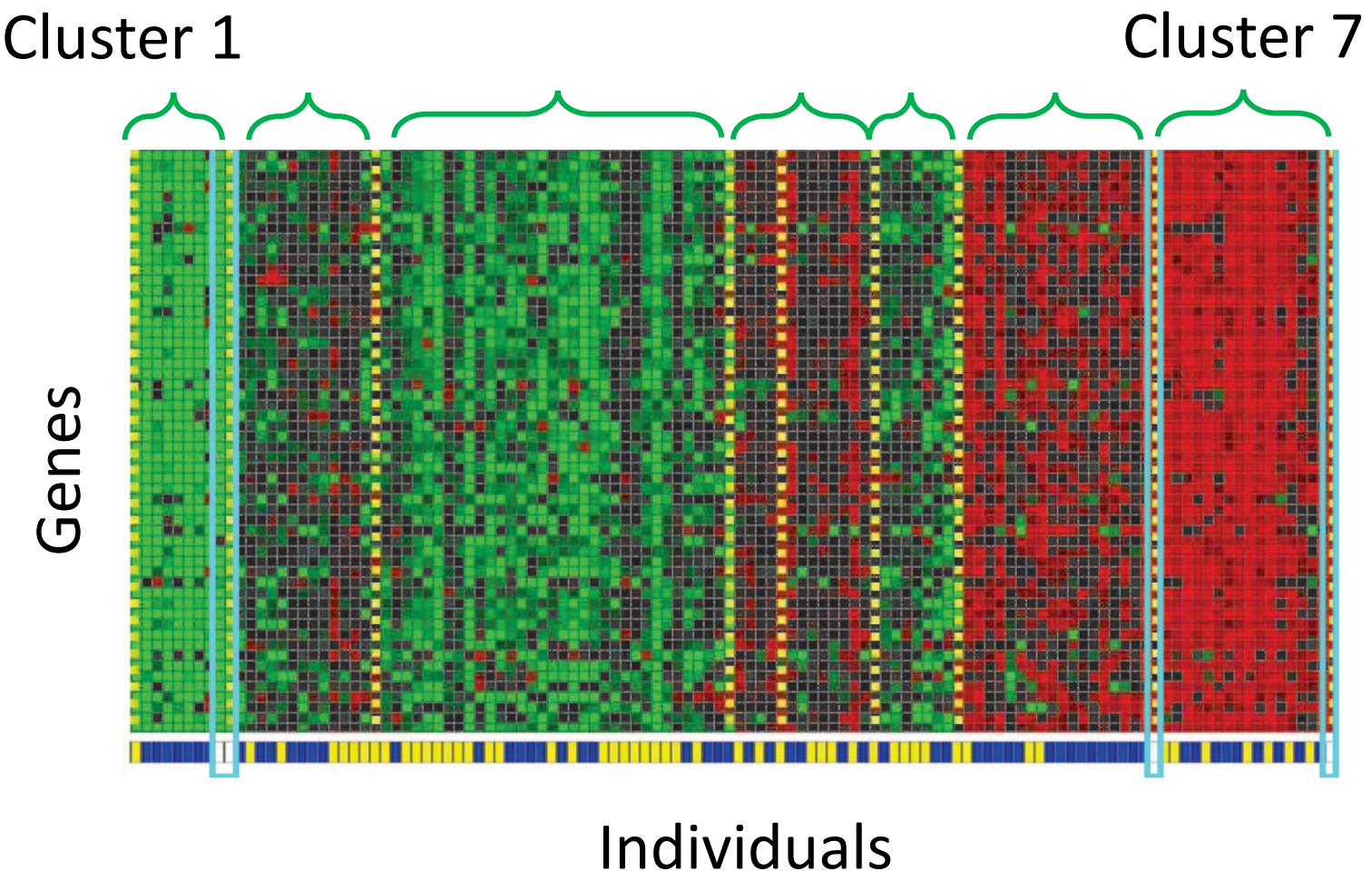


Clustering

- Lecture 8: k-mean clustering, mixture of Gaussians

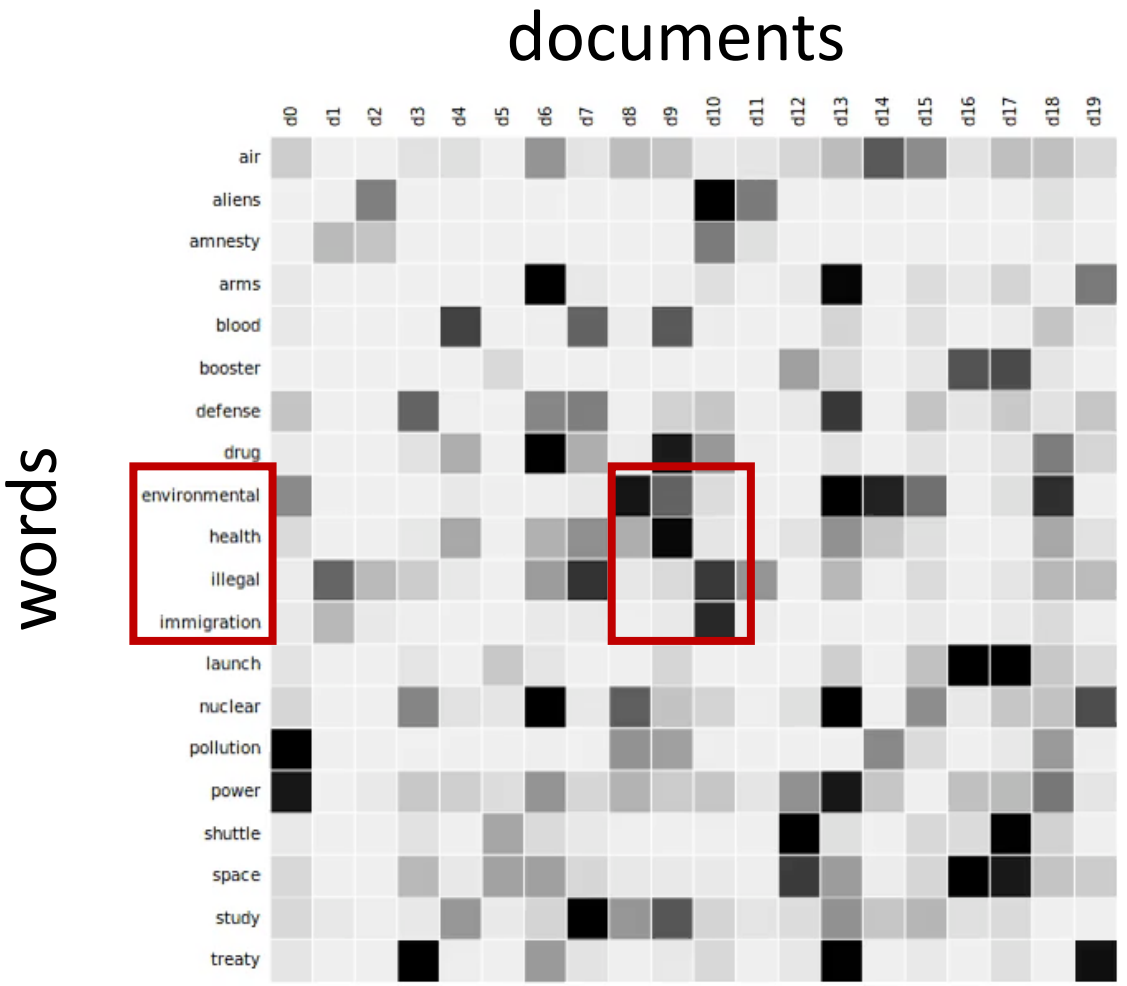


Clustering Genes



Identifying Regulatory Mechanisms using Individual Variation Reveals Key Role for Chromatin Modification. [Su-In Lee, Dana Pe'er, Aimee M. Dudley, George M. Church and Daphne Koller. ’06]

Latent Semantic Analysis (LSA)



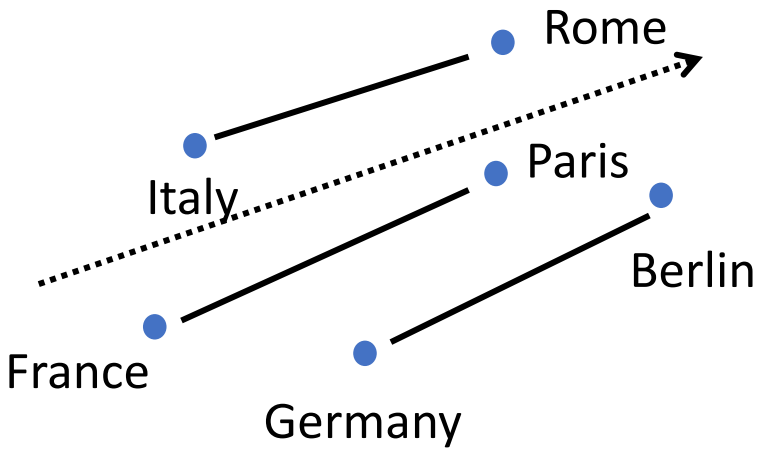
- Lecture 9: principal component analysis (tools used in LSA)

Image credit: https://commons.wikimedia.org/wiki/File:Topic_detection_in_a_document-word_matrix.gif

Word Embeddings

Represent words by vectors

- word $\xrightarrow{\text{encode}}$ vector
- relation $\xrightarrow{\text{encode}}$ direction

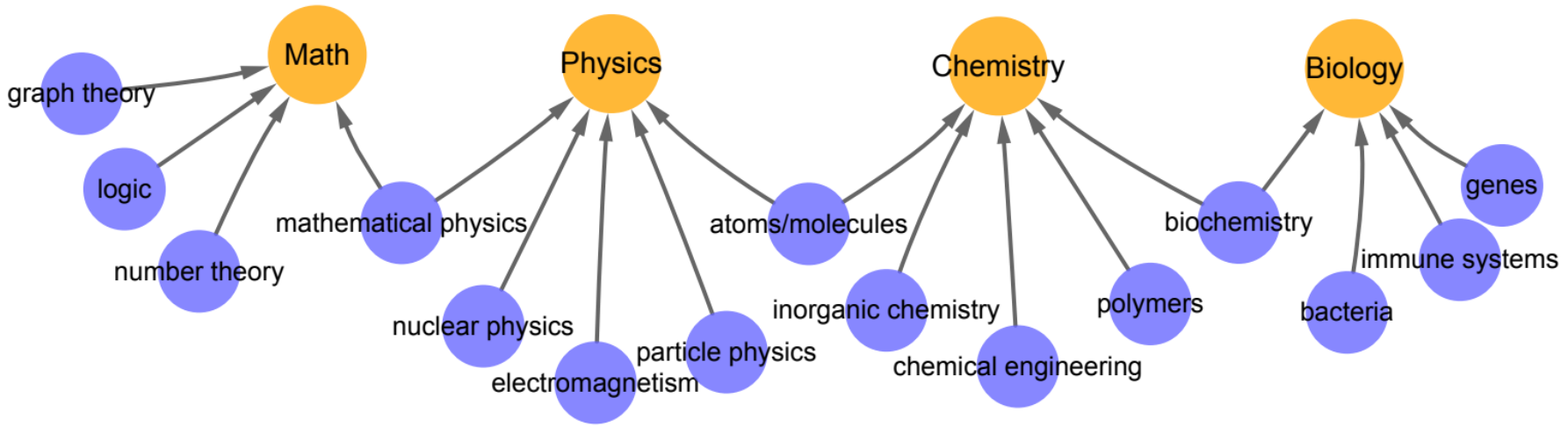


Paris – France + Italy = Rome

Unlabeled dataset

Word2vec [Mikolov et al'13]
GloVe [Pennington et al'14]

Hierarchical Clustering of Words



	logic deductive propositional semantics	graph subgraph bipartite vertex	boson massless particle higgs	polyester polypropylene resins epoxy	acids amino biosynthesis peptide
tag	*logic*	*graph theory*	*particle physics*	*polymer*	*biochemistry*

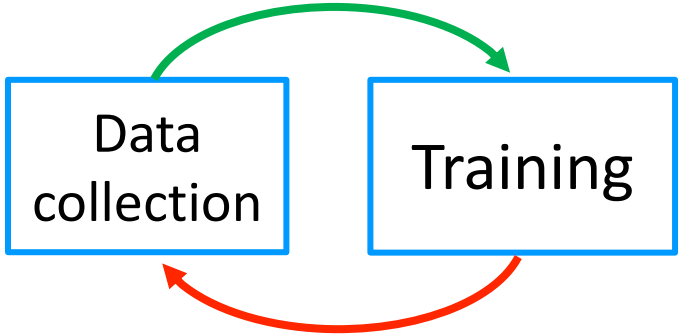
[Arora-Ge-Liang-M.-Risteski, TACL'17,18]

Reinforcement Learning

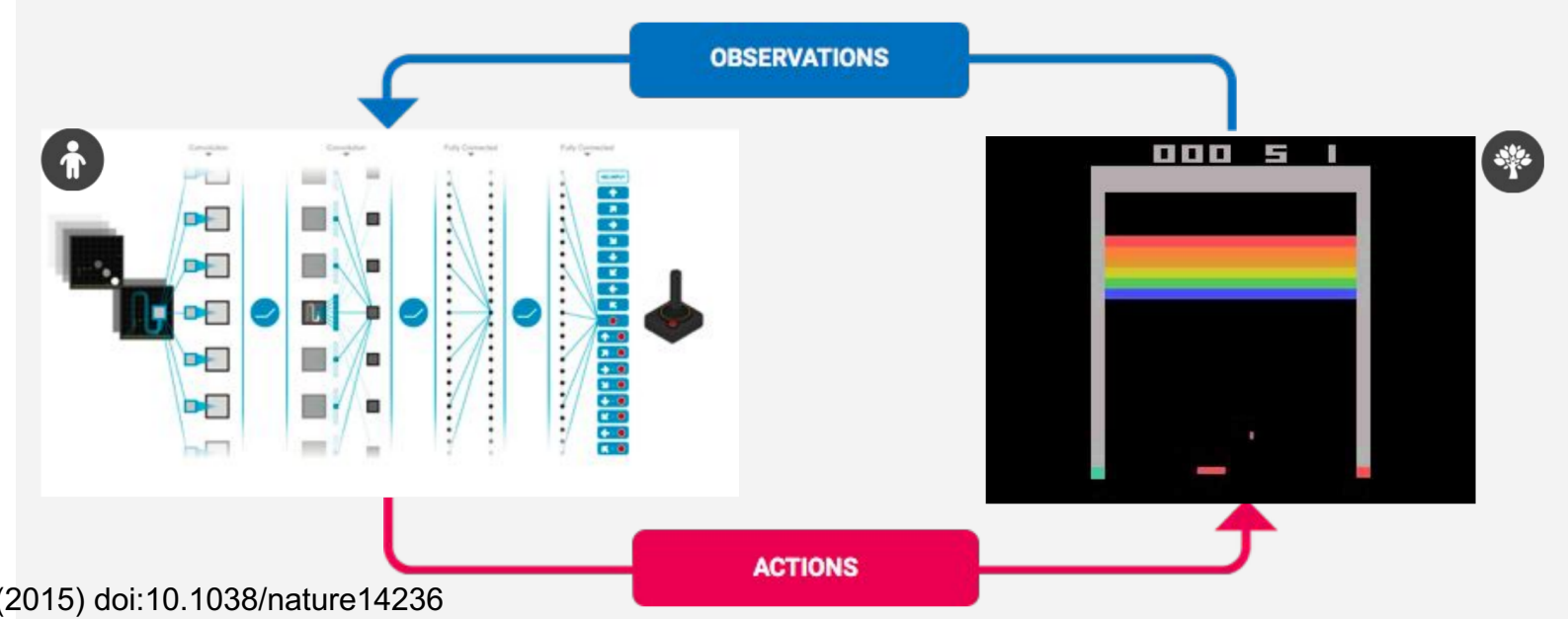
Reinforcement Learning

- The algorithm can collect data interactively

Try the strategy and collect feedbacks



Improve the strategy based on the feedbacks



V Mnih *et al. Nature* **518**, 529-533 (2015) doi:10.1038/nature14236

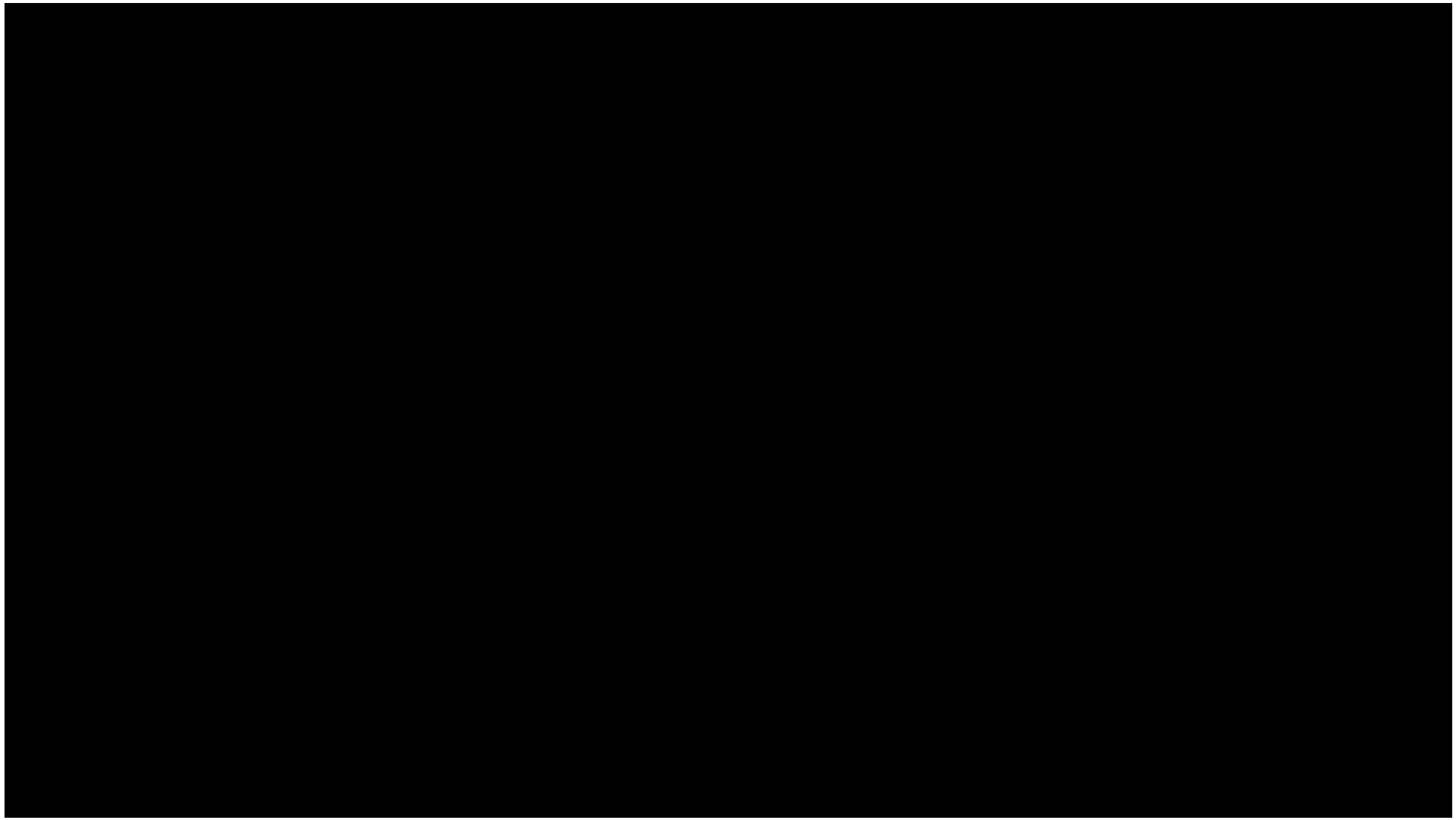
Learning to Play

V Mnih *et al. Nature* **518**, 529-533 (2015) doi:10.1038/nature14236

Stanford Autonomous Helicopter

- Two controllers
- Can be flown in many ways
- How do we learn to fly?

Stanford Autonomous Helicopter Using RL



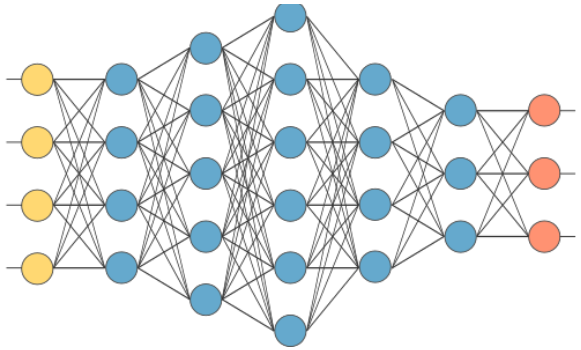
Source: http://heli.stanford.edu/icml2008/

Like to Learn More About RL?

- Lecture 11 will cover the very basics of RL
- Please take CPS824/CP8319: Reinforcement Learning in the Winter 2021 to lean more

Other Tools/Topics In This Course

- Deep learning basics



- Introduction to learning theory
  - Bias variance tradeoff
  - Feature selection
  - ML advice
- Broader aspects of ML
  - Robustness/fairness