Machine Learning CS229/STATS229

Instructors: Moses Charikar and Chris Ré

Hope everyone stays safe and healthy in these difficult times!

1. Administrivia

cs229.stanford.edu

(you may need to refresh to see the latest version)

2. Topics Covered in This Course

Who we are


Course Advisor
Swati Dube Batra


Course Coordinator
Amelie Byun


Head TA
Ian Tullis


Moses Charikar


Chris Ré


Anand Avati


Daniel Do


Jeff Z. HaoChen


Qijia Jiang


Soyeon Jung


Yao Liu


Akshay Smit


Angelica


Christopher Wolff

- We have wonderful course coordinators (Swati and Amelie). They are your resource for **any** admin decision
- We have fantastic TAs! Please be kind and generous with them!

Pre-requisite

- Probability (CS109 or STAT 116)
  - distribution, random variable, expectation, conditional probability, variance, density
- Linear algebra (Math 104, Math 113, or CS205)
  - matrix multiplication
  - eigenvector
- Basic programming (in Python)
- Will be reviewed in Friday sections (recorded)

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

Honor Code

Do's

- form study groups (with arbitrary number of people); discuss and work on homework problems in groups
- write down the solutions independently
- write down the names of people with whom you've discussed the homework
- read the longer description on the course website

Don'ts

- copy, refer to, or look at any official or unofficial previous years' solutions in preparing the answers

Honor Code for Submission In Pairs

- Students submitting in a pair act as one unit
  - may share resources (such as notes) with each other and write the solutions together
- Both students should fully understand all the answers in their submission
- Each student in the pair must understand the solution well enough in order to reconstruct it by him/herself

Course Project

- We encourage you to form a group of 1-3 people
  - same criterion for 1-3 people
- More information and previous course projects can be found on course website
- List of potential topics

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

Other Information on Course Website
cs229.stanford.edu

- Ed:
  - All announcements and questions (unless you would only reach out to a subset of course staff)
    - For logistical questions, please look at course FAQ first
  - Finding study groups friends
  - ➢ If you enrolled in the class but do not have access to Ed, it should come within a day. If it has been more than that, send Amelie Byun an email (aebyun@Stanford.edu)
- Nooks: Office Hours
- Videos on canvas: Under Panopto Videos tab (will be uploaded EOD)
- Course calendar & Syllabus for deadlines
- Canvas calendar for office hours/ section/ lecture dates and links
- Gradescope: You will be automatically enrolled in course Gradescope
- Late days policy
- FAQ on the course website

… Course feel …

- This class is almost all “whiteboard” and mathematical
  - We try to be self contained, but there are a diverse set of backgrounds.
  - Please ask questions! When you ask questions, we’re so happy!!
- Some of you will learn from lectures, notes, each other. Find what works for **you.**
- Please be generous with the staff (and yourself!)
  - We’re getting better (we hope) at this virtual experience.
  - We really want to help you learn this material, and that’s why I love this class.

1. Administrivia

cs229.stanford.edu

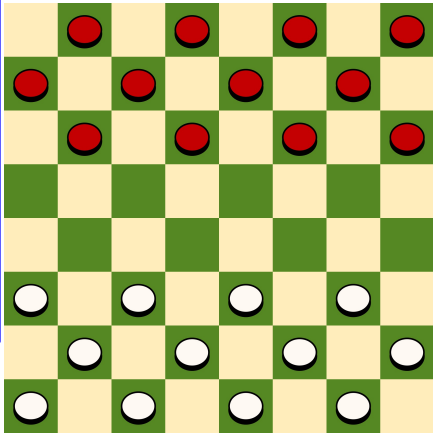
2. Topics Covered in This Course

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.

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

Performance measure: winning rate

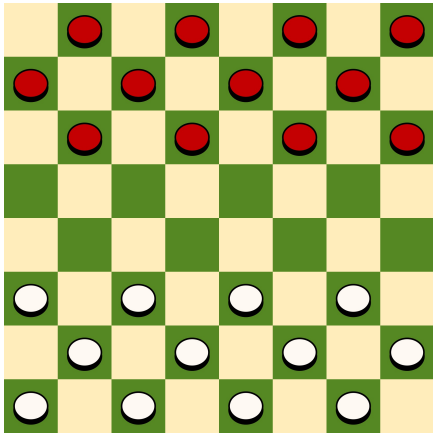
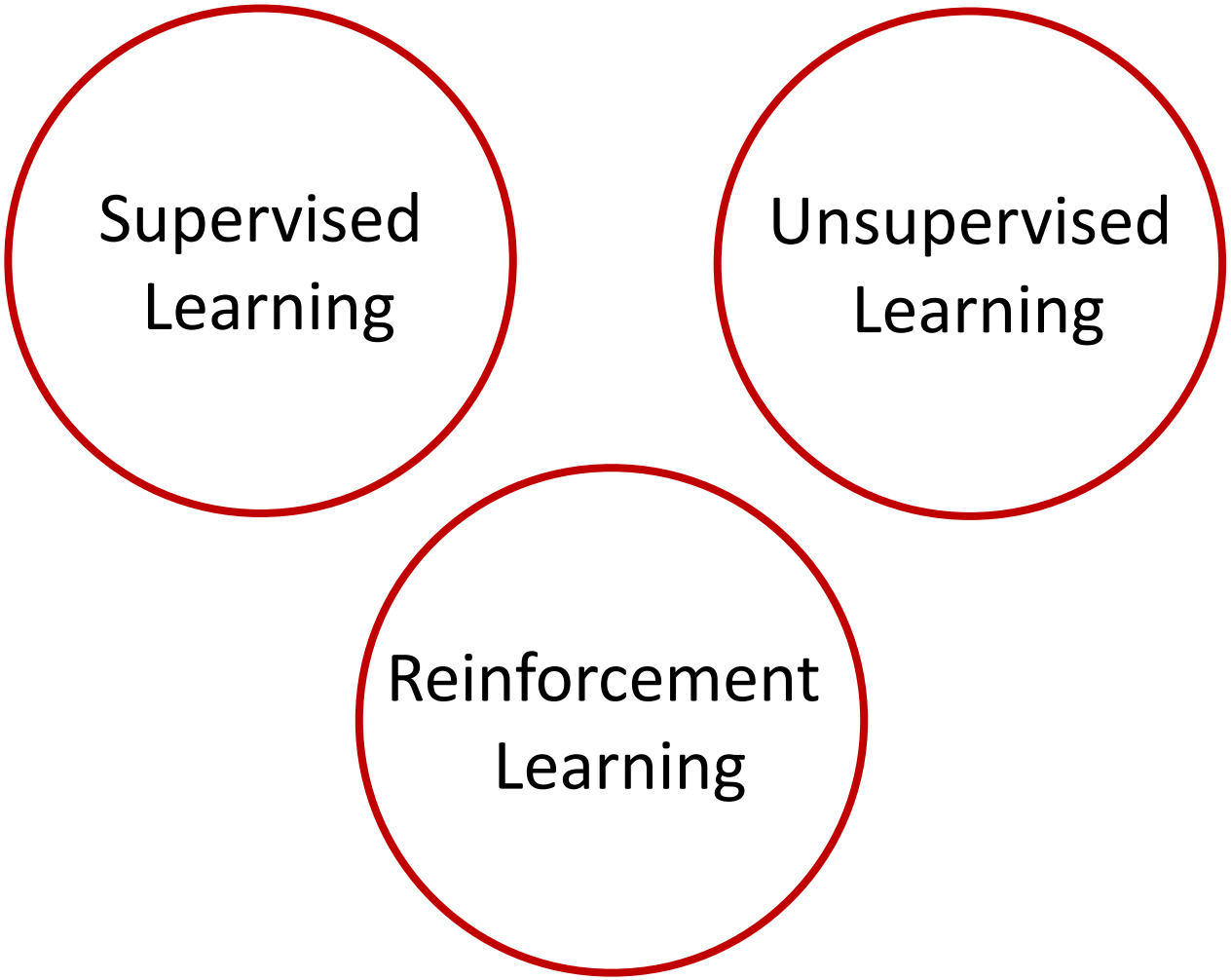
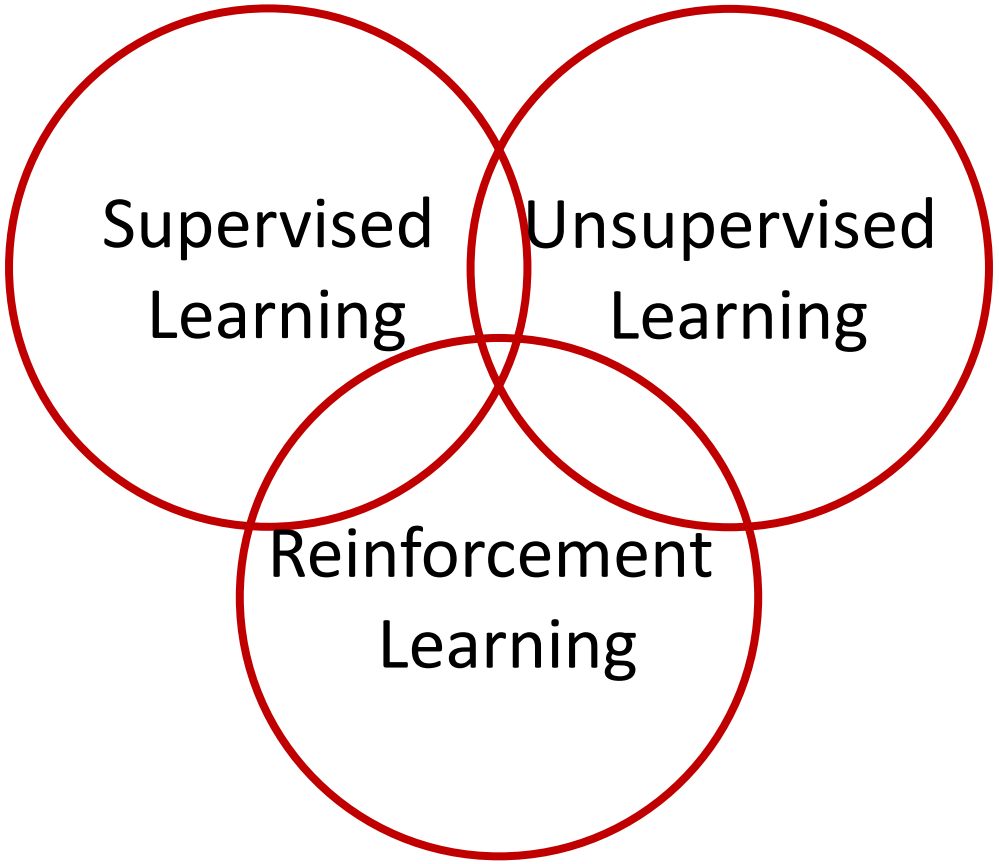


Image from Tom Mitchell's homepage

Taxonomy of Machine Learning (A Simplistic View Based on Tasks)



Taxonomy of Machine Learning (A Simplistic View Based on Tasks)



can also be viewed as tools/methods

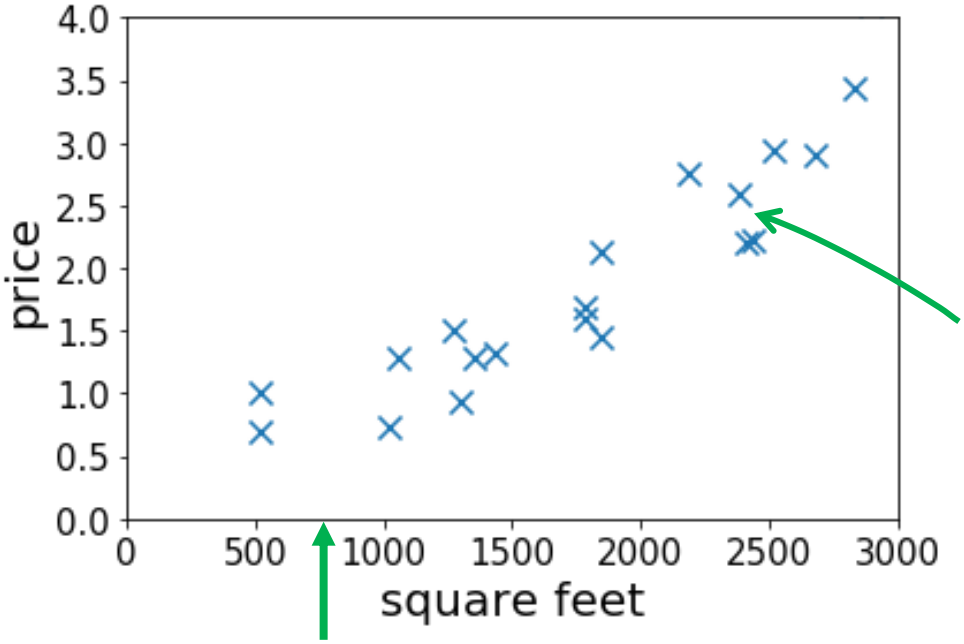
Supervised Learning

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

$x = 800$

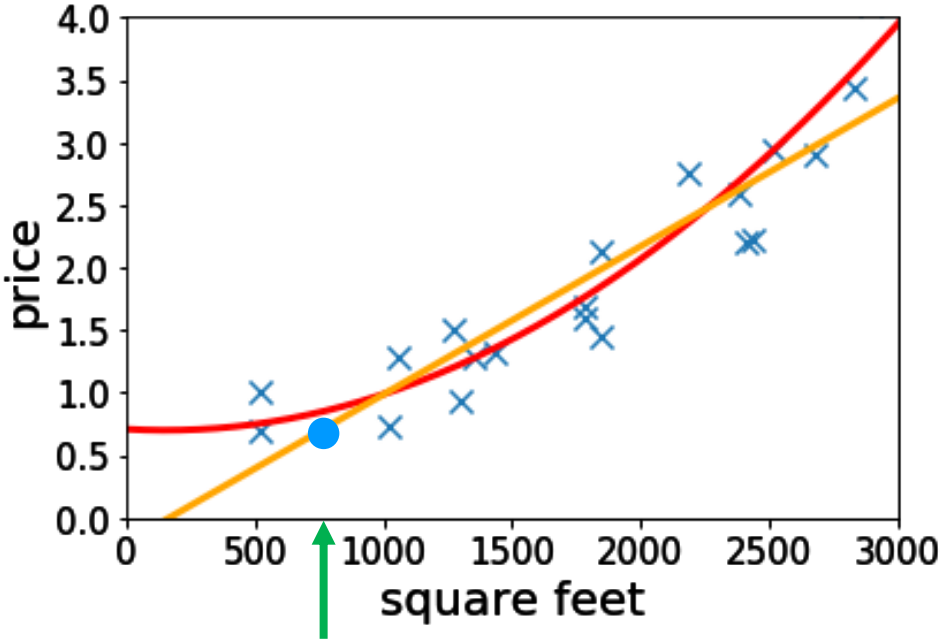
$y = ?$

Housing Price Prediction

- Given: a dataset that contains n samples

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

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



$$x = 800$$

$$y = ?$$

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

More Features

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

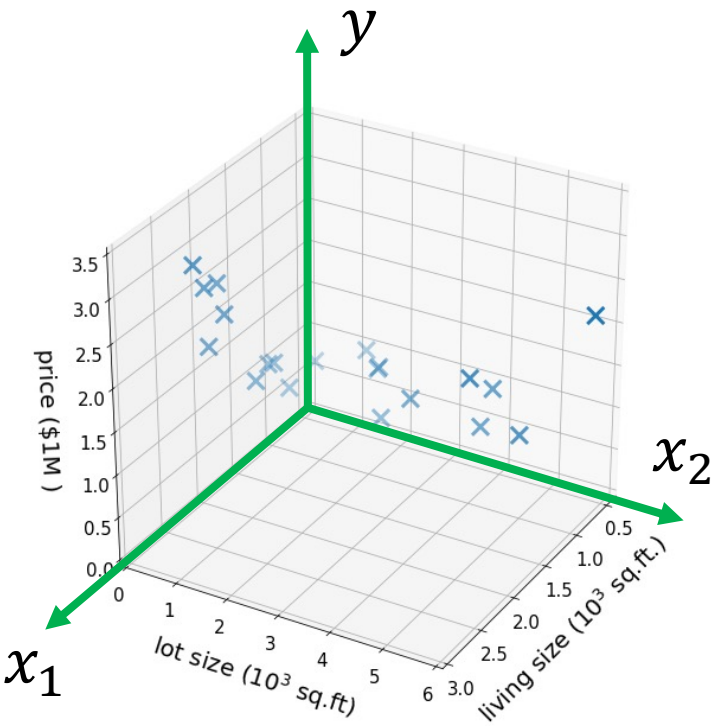
(size, lot size) → price

features/input $x \in \mathbb{R}^2$ label/output $y \in \mathbb{R}$

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

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

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



High-dimensional Features

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

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

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

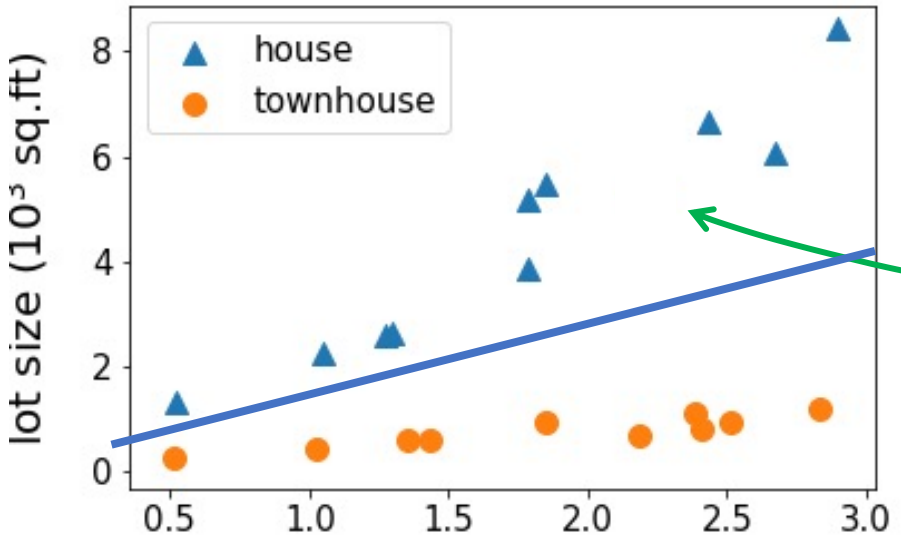
$\longrightarrow$ y --- price

- Lec. 6-7: infinite dimensional features (kernels)
- Lec. 10-11: select features based on data (deep learning)

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

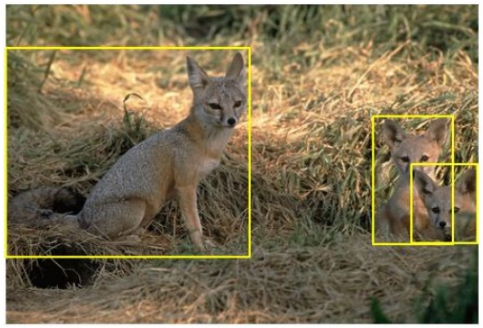
Supervised Learning in Computer Vision

- Image Classification
  - $x =$ raw pixels of the image, $y =$ the main object

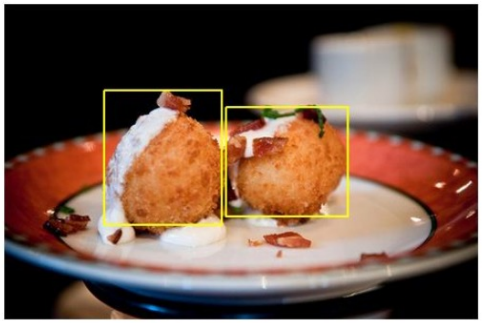
ImageNet Large Scale Visual Recognition Challenge. Russakovsky et al.'2015

Supervised Learning in Computer Vision

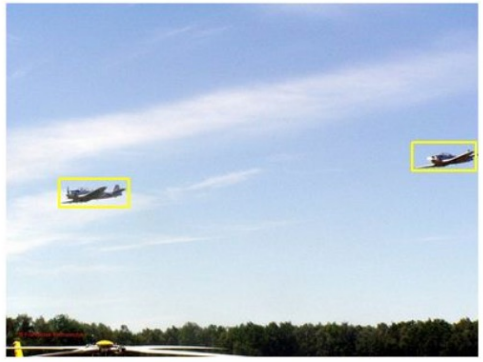
- Object localization and detection
  - $x =$ raw pixels of the image, $y =$ the bounding boxes



kit fox



croquette



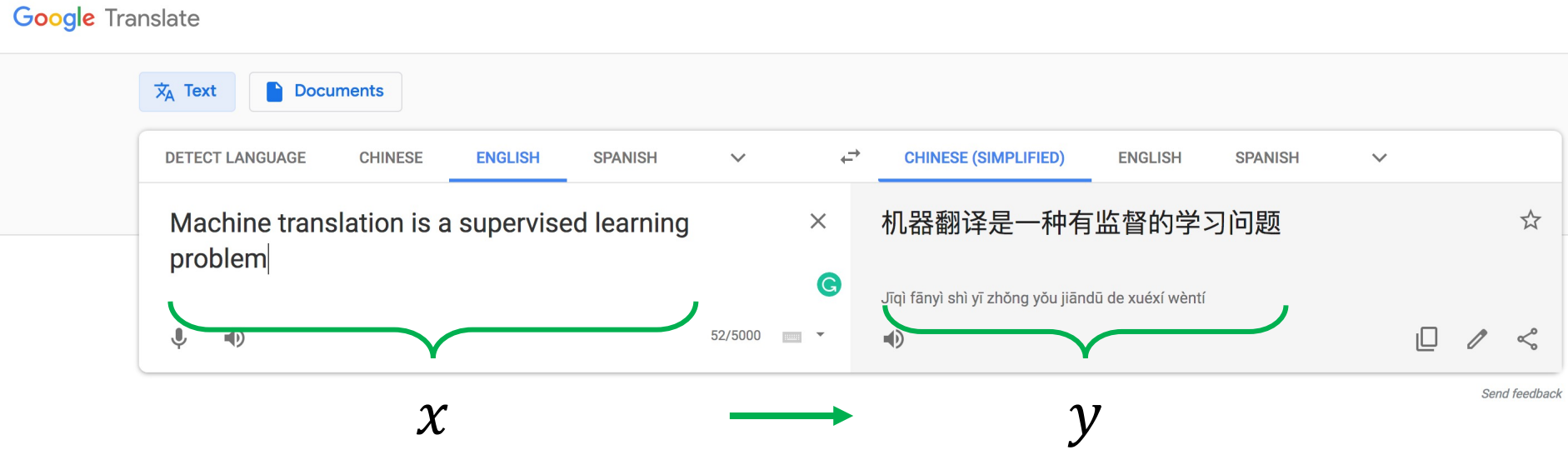
airplane

frog

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

Supervised Learning in Natural Language Processing

- Machine translation



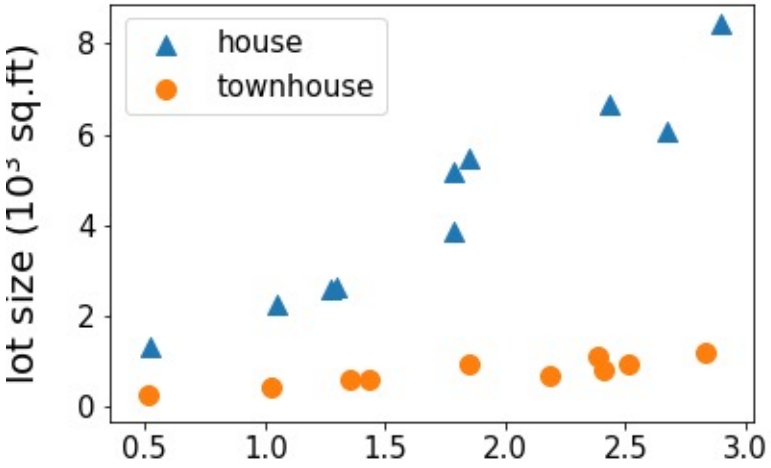
- Note: this course only covers the basic and fundamental techniques of supervised learning (which are not enough for solving hard vision or NLP problems.)
- CS224N and CS231N, if you are interested in the particular applications.

Unsupervised Learning

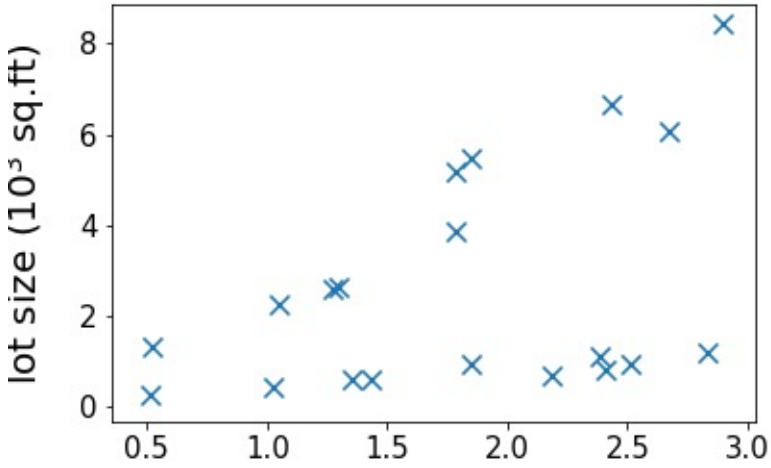
Unsupervised Learning

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

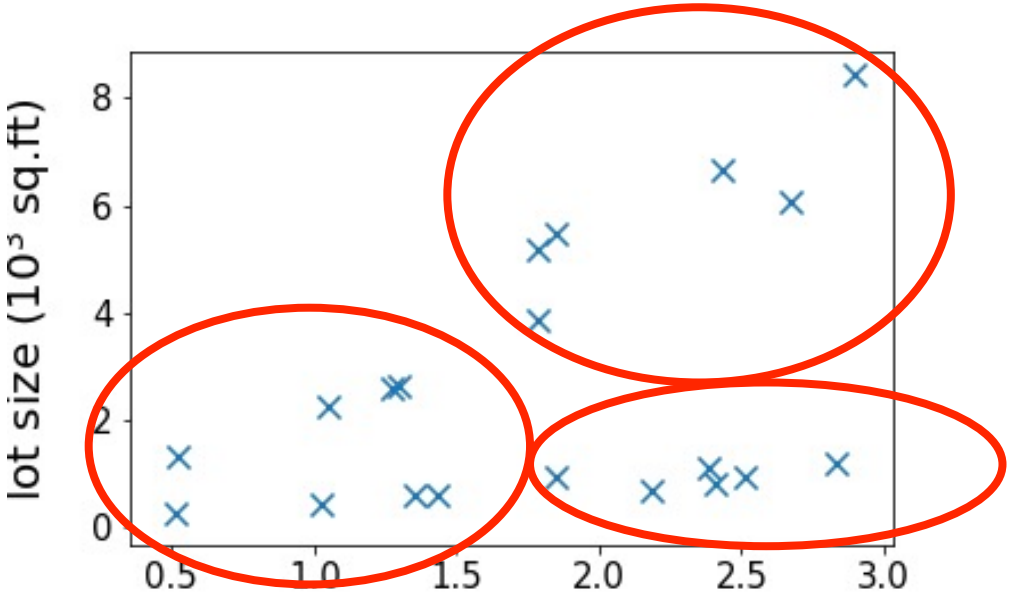
supervised



unsupervised

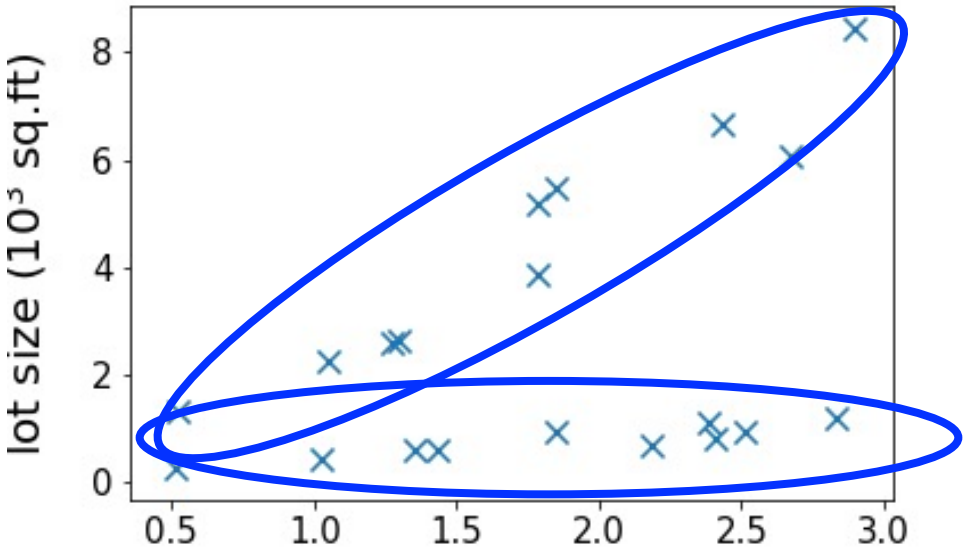


Clustering

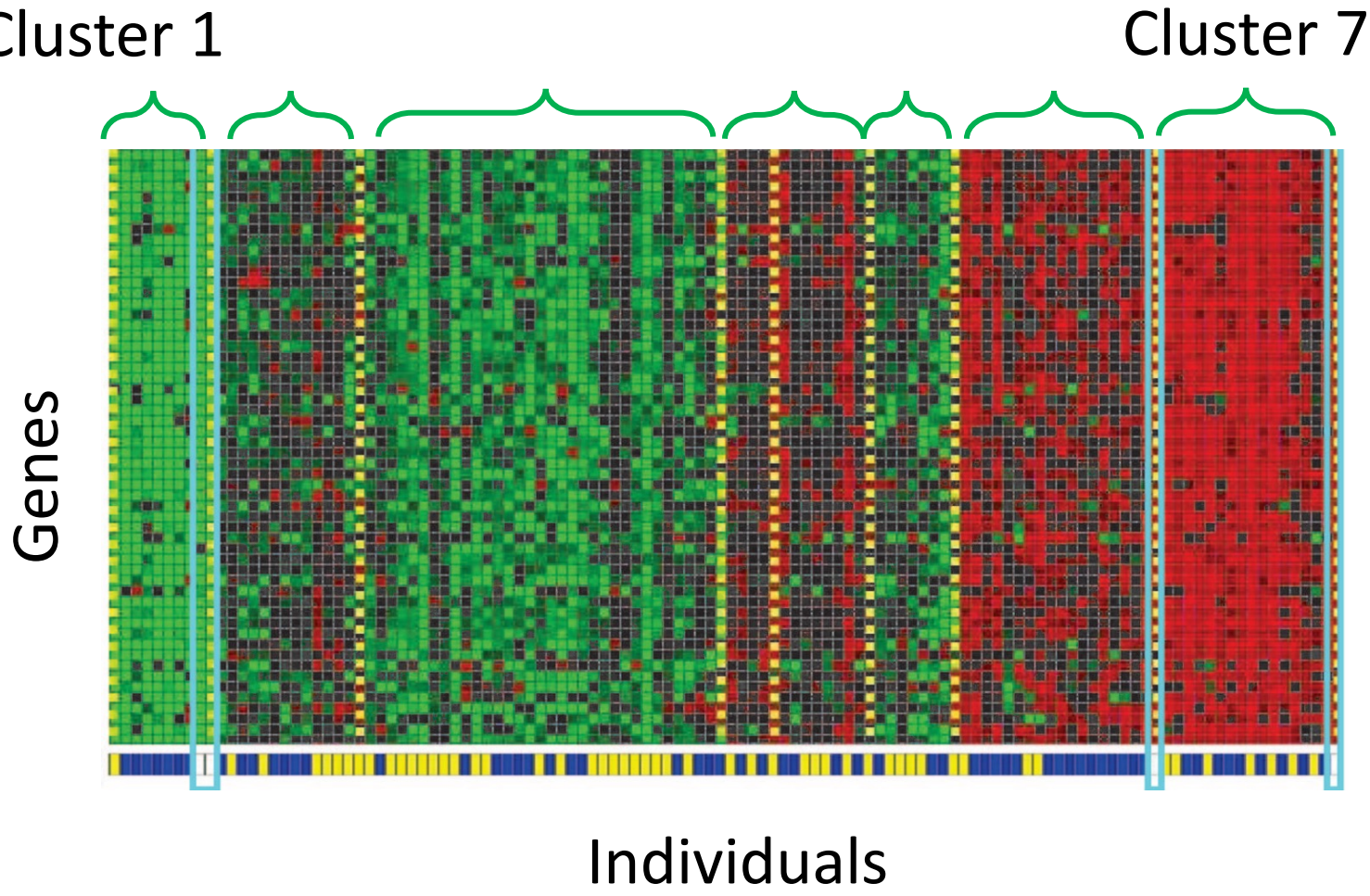


Clustering

- Lecture 12&13: k-mean clustering, mixture of Gaussians

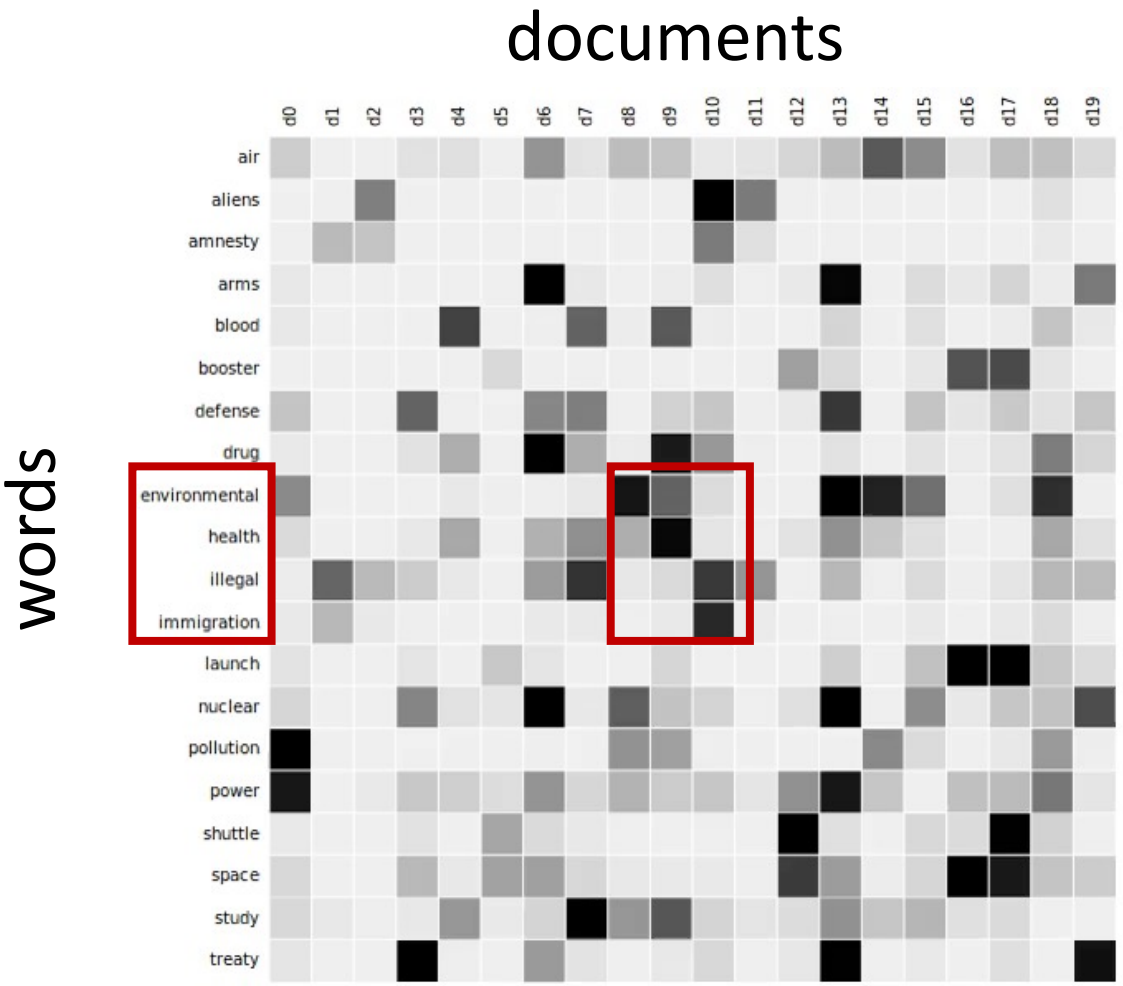


Clustering Genes



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 14: principal component analysis (used in LSA)

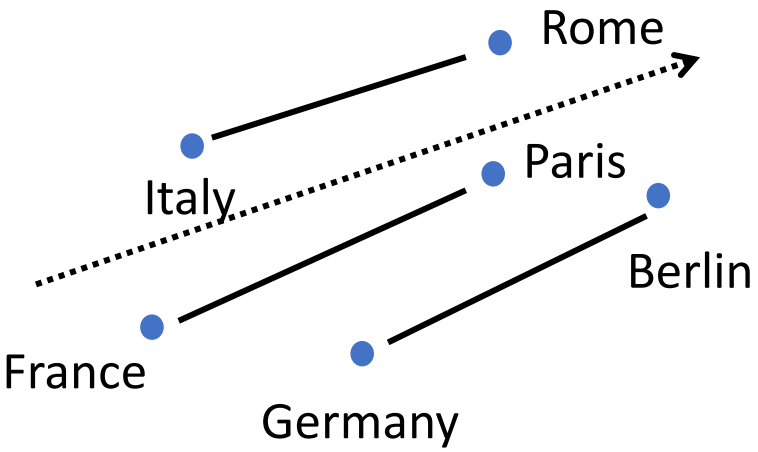
Image credit: https://commons.wikimedia.org/wiki/File:Topic_detection in a document-word matrix.gif

Word Embeddings

Unlabeled dataset

Represent words by vectors

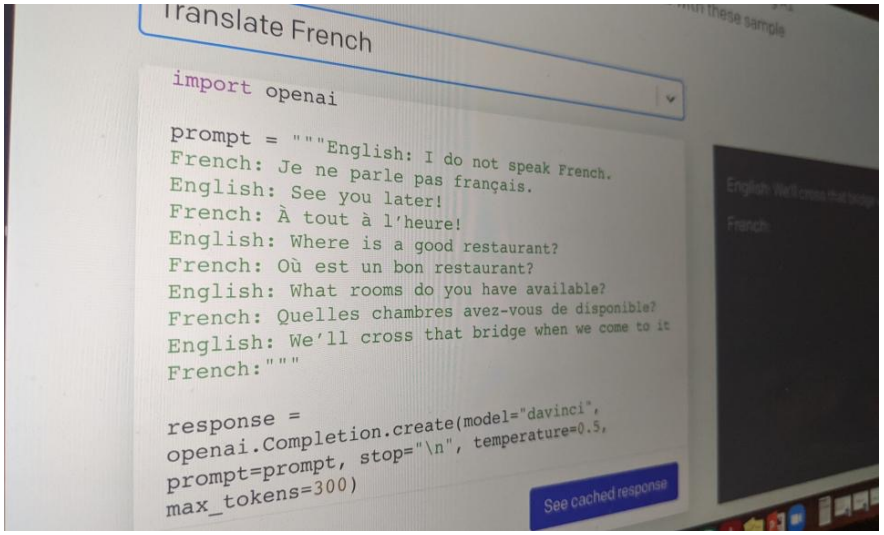
- word $\xrightarrow{\text{encode}}$ vector
- relation $\xrightarrow{\text{encode}}$ direction



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

Language Processing, Machine Learning

How Large Language Models Will Transform Science, Society, and AI

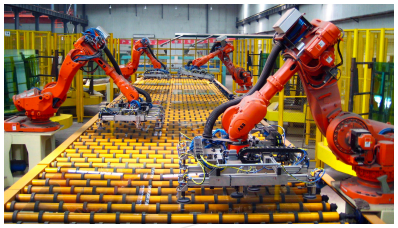


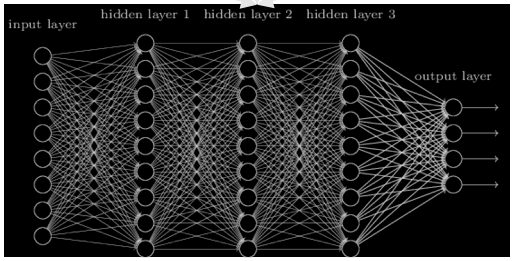
GPT3

- Lecture 16
- (maybe whole course next year!!)

- https://hai.stanford.edu/news/how-large-language-models-will-transform-science-

Software 2.0 is eating Software 1.0





1000x Productivity: Google shrinks language translation code from 500k LoC to 500 lines of **dataflow**.

AI driven by **data—** not **the model**

https://jack-clark.net/2017/10/09/import-ai-63-google-shrinks-language-translation-code-from-500000-to-500-lines-with-ai-only-25-of-surveyed-people-believe-automationbetter-jobs

"Software 2.0", Andrej Karpathy, https://medium.com/@karpathy/software-2-0-a64152b37c35

… you probably used SW2.0 in the last hour…

Overton: A Data System for Monitoring and Improving Machine-Learned Products

Christopher Ré, Apple
Feng Niu, Apple
Pallavi Gudipati, Apple
Charles Srisuwananukorn, Apple

Snorkel DryBell: A Case Study in Deploying Weak Supervision at Industrial Scale

Stephen H. Bach[†] Daniel Rodriguez[‡] Yintao Liu[‡] Chong Luo[‡]
Haidong Shao[‡] Cassandra Xia[‡] Souvik Sen[‡] Alex Ratner[§]
Braden Hancock[§] Houman Alborzi[‡] Rahul Kuchhal[‡] Chris Ré[§] Rob Malkin[‡]
[†]Brown University [‡]Google [§]Stanford University

Migrating a Privacy-Safe Information Extraction System to a Software 2.0 Design

Ying Sheng, Google, Mountain View, CA, USA, yingsheng@google.com
Nguyen Vo, Google, Mountain View, CA, USA, nguyenvo@google.com
James B. Wendt, Google, Mountain View, CA, USA, jwendt@google.com
Sandeep Tata, Google, Mountain View, CA, USA, tata@google.com
Marc Najork, Google, Mountain View, CA, USA, najork@google.com

Leveraging Organizational Resources to Adapt Models to New Data Modalities

Sahaana Suri[†]; Raghuveer Chanda, Neslihan Bulut, Pradyumna Narayana, Yemao Zeng
Peter Bailis[†], Sugato Basu, Girija Narlikar, Christopher Ré[†], Abishek Sethi
Google, Stanford[†]

Lec 15: basic theory of these new systems “Weak Supervision theory”.
Also new course on ML Engineering next year!

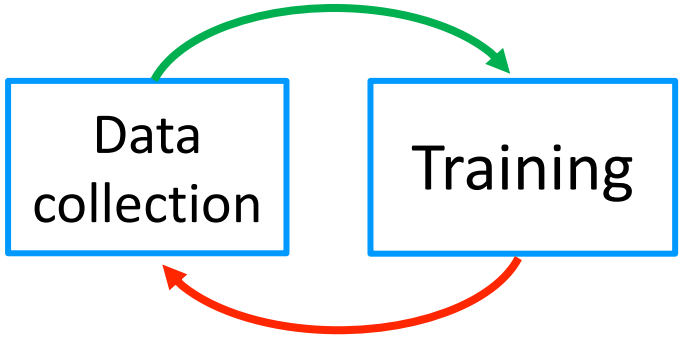
“SW2.0 will add 30 trillion dollars to public equity markets” … this is not trading advice.

REINFORCEMENT LEARNING

Reinforcement Learning

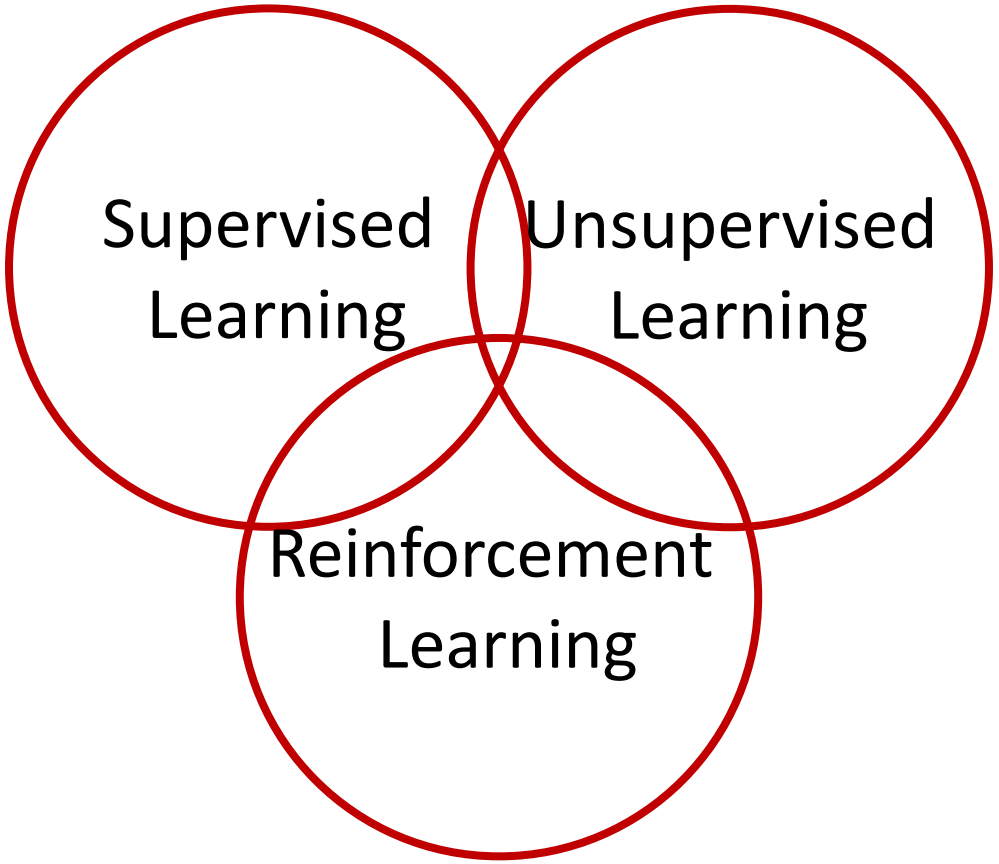
- The algorithm can collect data interactively

Try the strategy and collect feedback



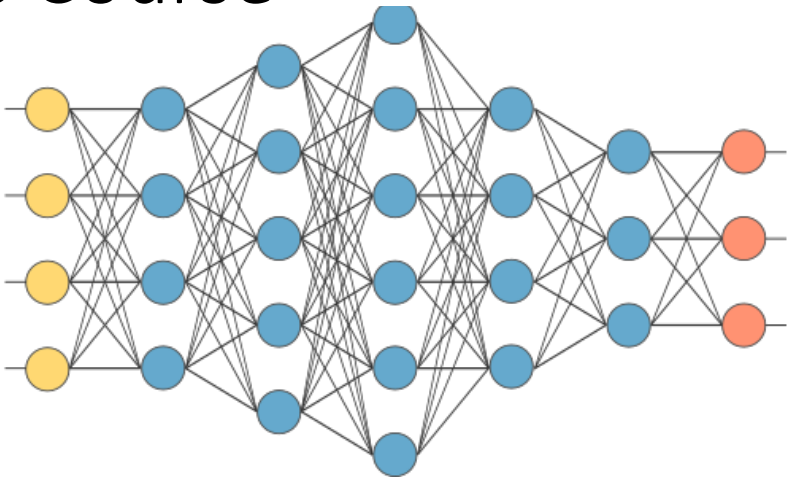
Improve the strategy based on the feedback

Taxonomy of Machine Learning (A Simplistic View Based on Tasks)



can also be viewed as tools/methods

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

Thank
you!