Machine Learning CS229/STATS229

Instructors: Tengyu Ma and Chris Re

1. Administrivia

cs229.stanford.edu

2. Topics Covered in This Course

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

- 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

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

Other Information on Course Website
cs229.stanford.edu

- Piazza:
  - technical and logistical question (anonymous or non-anonymous, private or public)
  - to find study groups friends
  - all announcement
- Videos on canvas
- Course calendar: office hours and deadlines
- Section (not Fri section) vs office hour
- Gradescope
  - you will receive invite after Axess enrollment within 24hrs
- Late days policy
- FAQ

1. Administrivia

cs229.stanford.edu

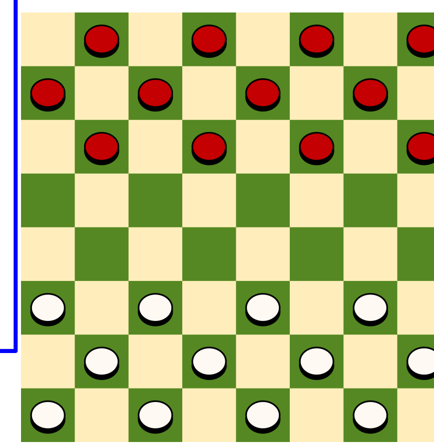
2. Topics Covered in This Course

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

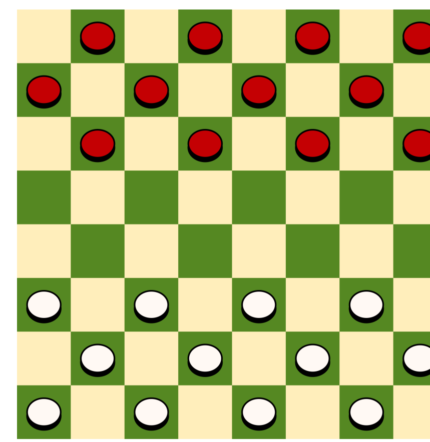


Image from Tom Mitchell’s homepage

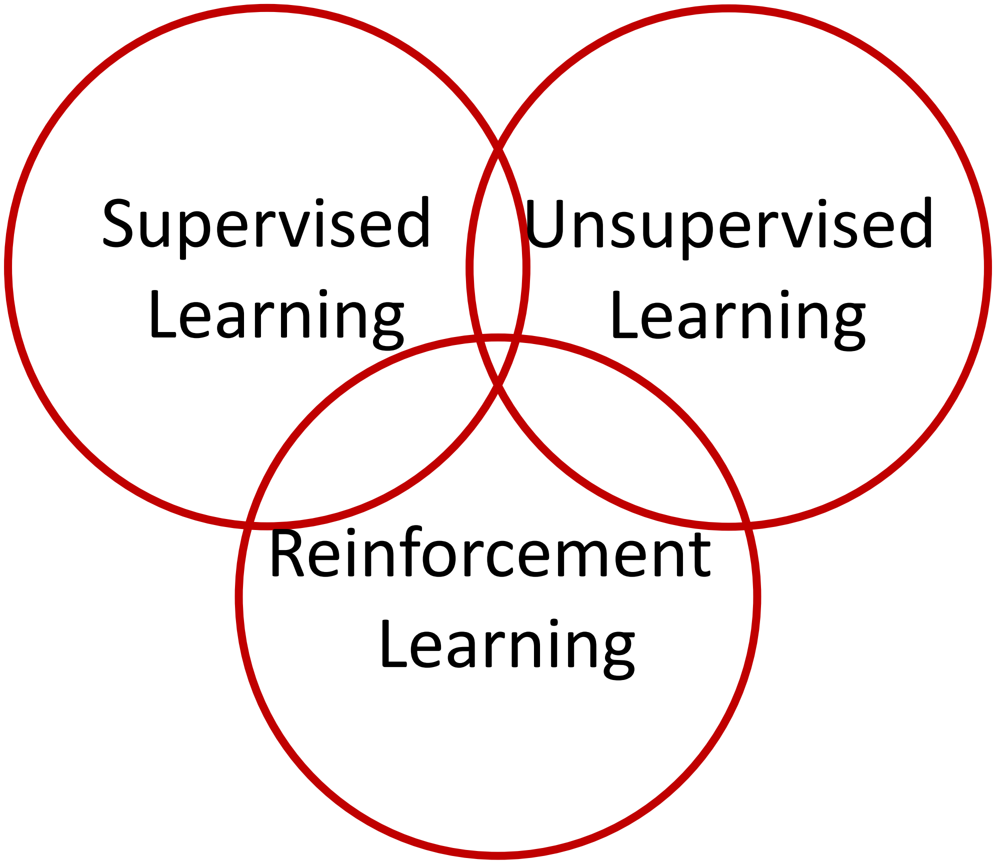
Taxonomy of Machine Learning (A Simplistic View Based on Tasks)

Supervised Learning

Unsupervised Learning

Reinforcement Learning

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



can also be viewed as tools/methods

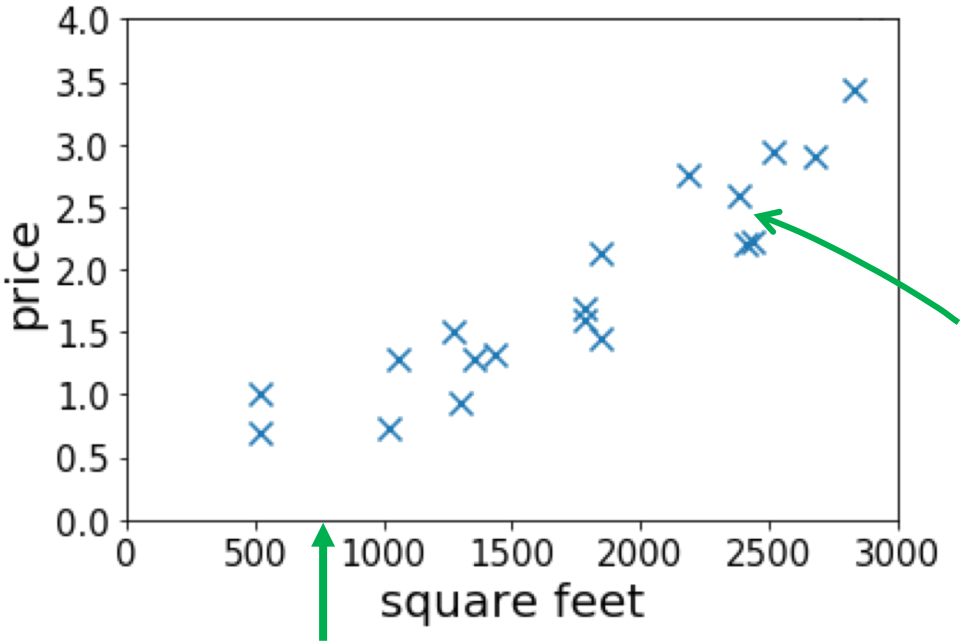
Supervised Learning

Housing Price Prediction

- Given: a dataset that contains n samples

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

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



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

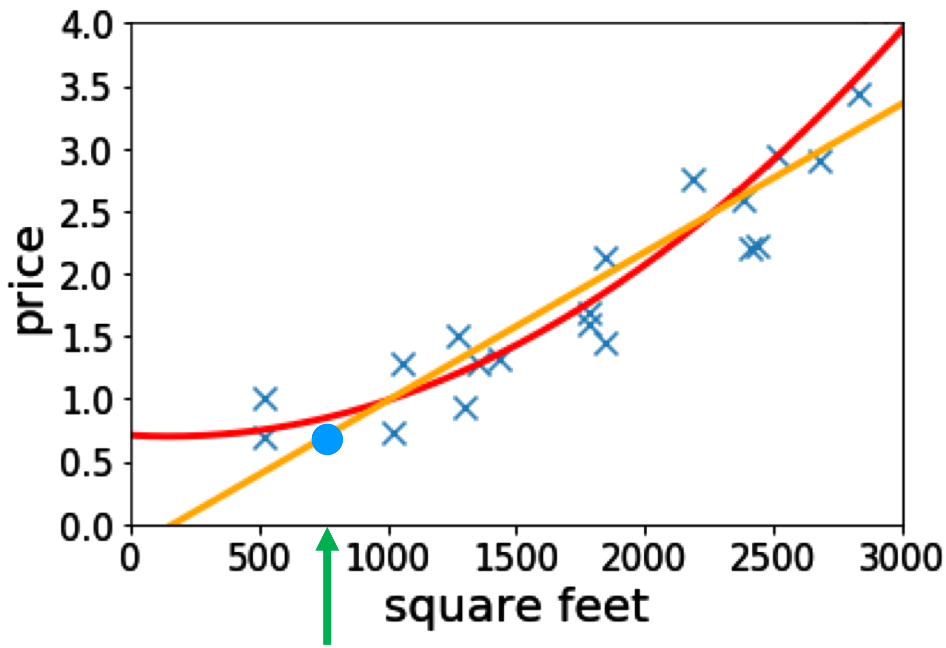
$x = 800$
$y = ?$

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

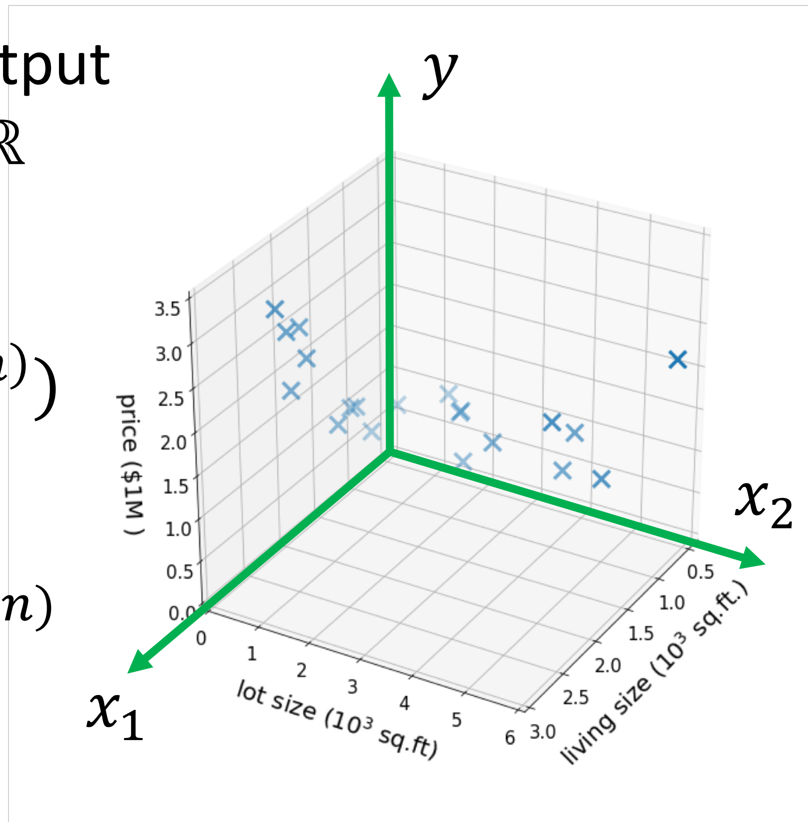
- Suppose we also know the lot size
- Task: find a function that maps

$$\underbrace{\text{(size, lot size)}}_{\substack{\text{features/input}\\ x \in \mathbb{R}^2}} \rightarrow \underbrace{\text{price}}_{\substack{\text{label/output}\\ y \in \mathbb{R}}}$$

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

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

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



High-dimensional Features

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

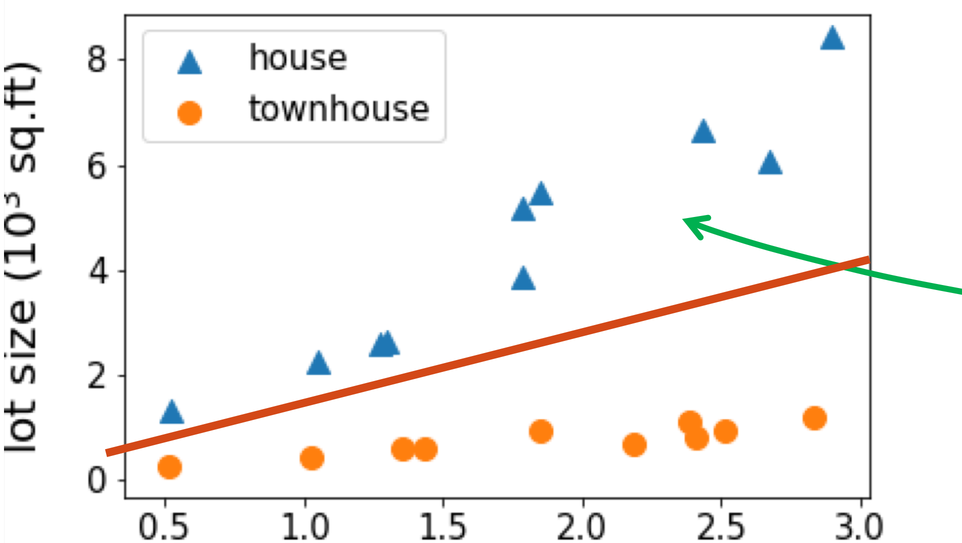
$$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ \vdots \\ \vdots \\ x_d \end{bmatrix} \begin{array}{l} \text{--- living size} \\ \text{--- lot size} \\ \text{--- \# floors} \\ \text{--- condition} \\ \text{--- zip code} \\ \qquad\quad \vdots \\ \\ \end{array} \longrightarrow \quad y \text{ --- price}$$

- Lecture 6-7: infinite dimensional features
- Lecture 10-11: 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) $\rightarrow$ house or townhouse?



$y =$ house or townhouse?

Lecture 3&4: classification

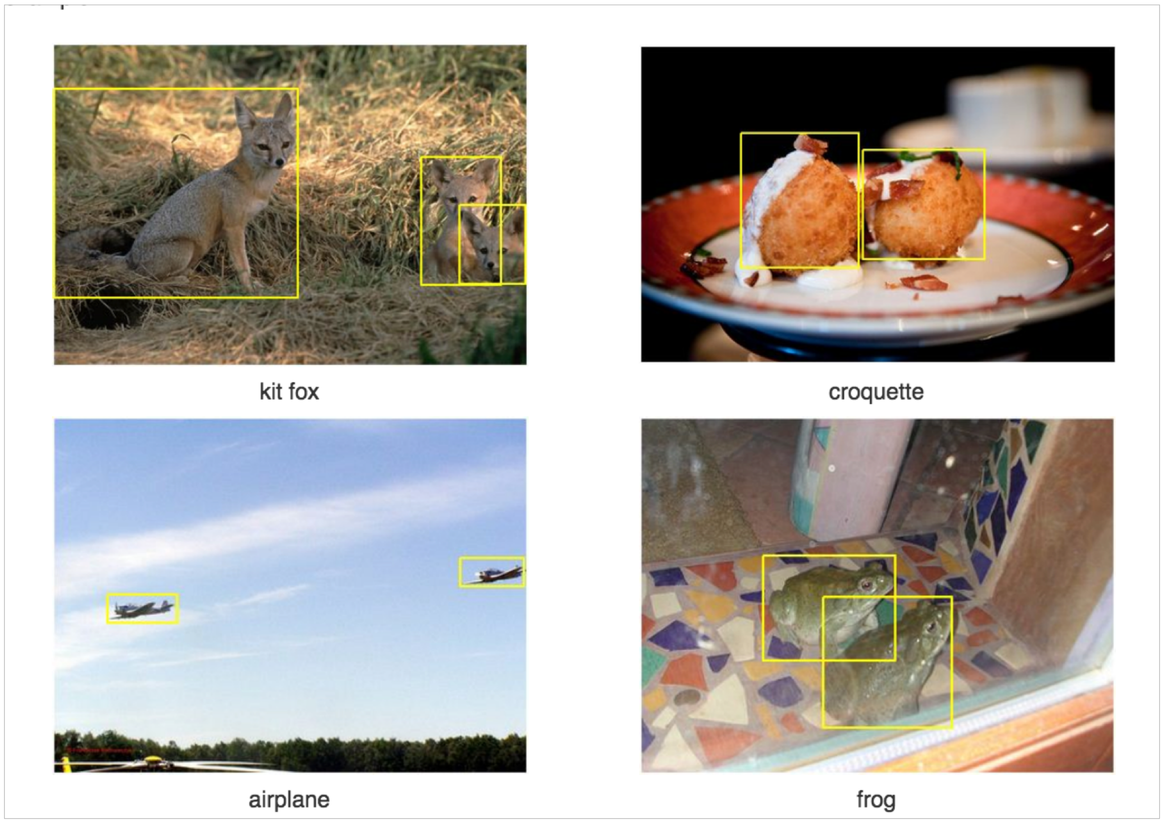
Supervised Learning in Computer Vision

- 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

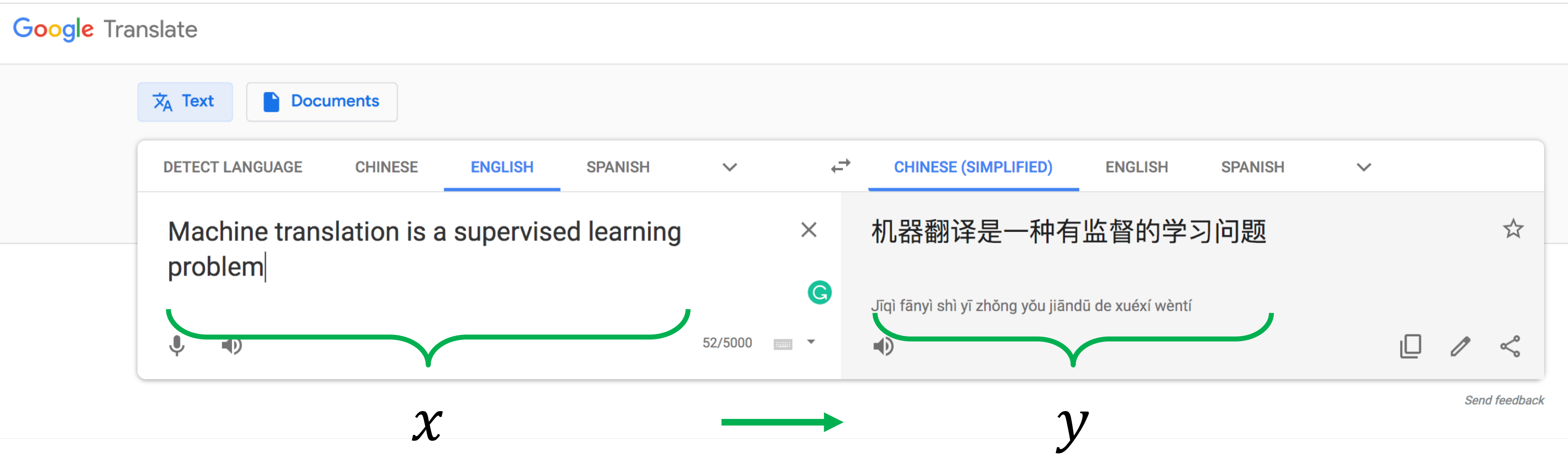
- Object localization and detection
  - x = raw pixels of the image, y = the bounding boxes



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

Supervised Learning in Natural Language Processing

- Machine translation



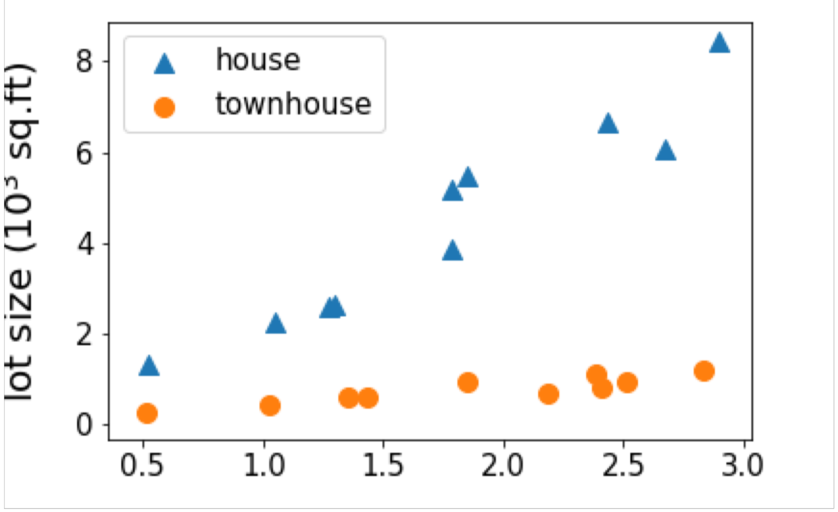
- Note: this course only covers the basic and fundamental techniques of supervised learning (which are not enough for solving hard vision or NLP problems.)
- CS224N and CS231N would be more suitable if you are interested in the particular applications

Unsupervised Learning

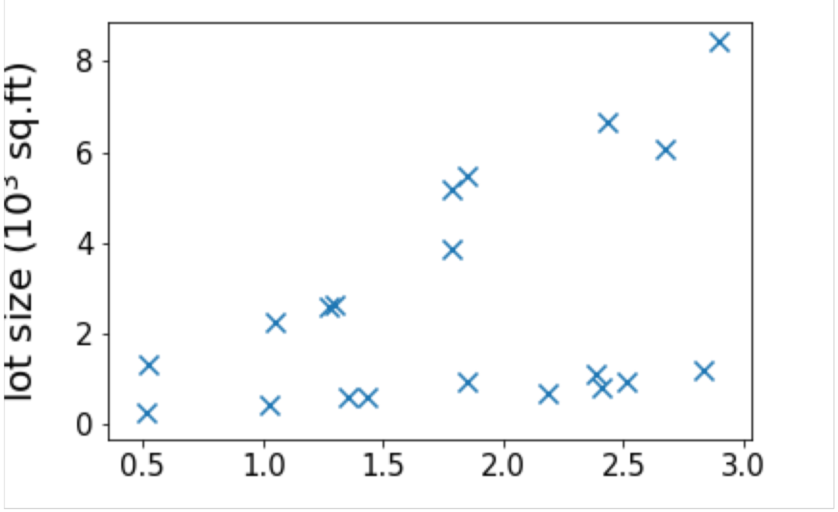
Unsupervised Learning

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

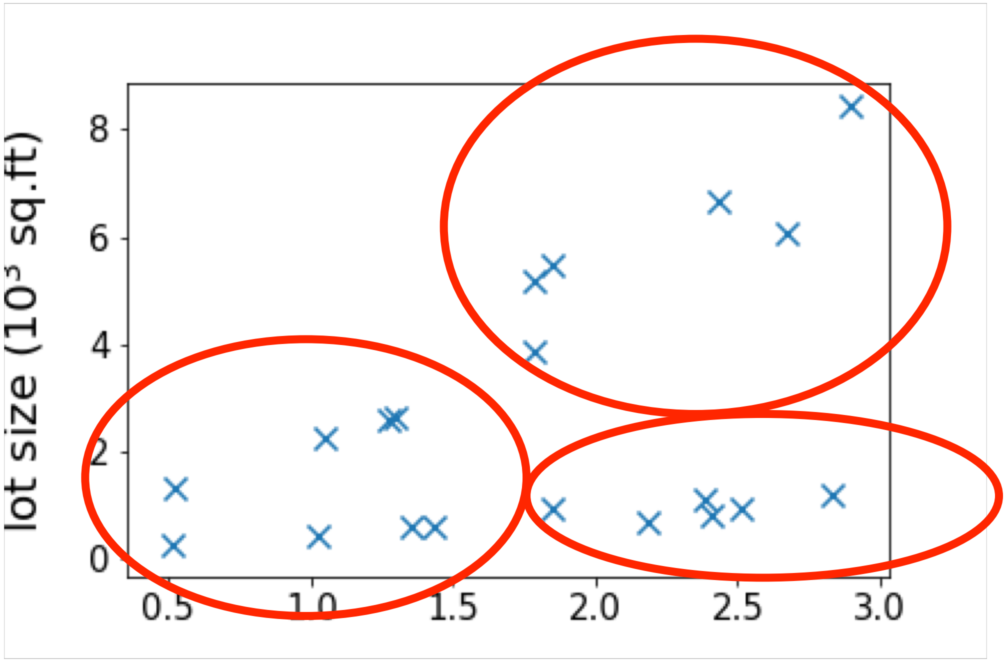
supervised



unsupervised

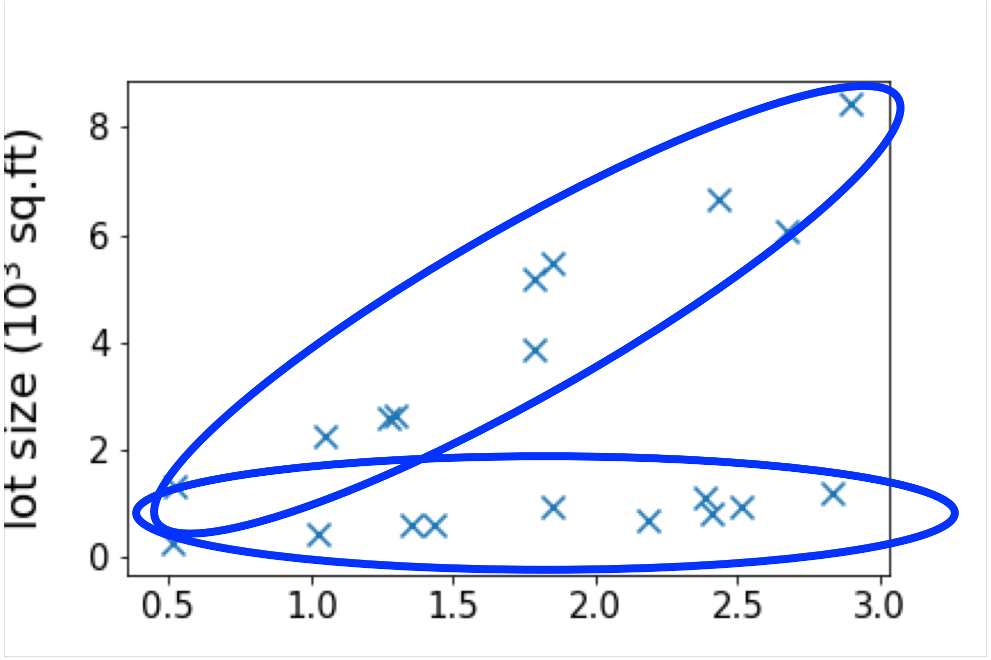


Clustering

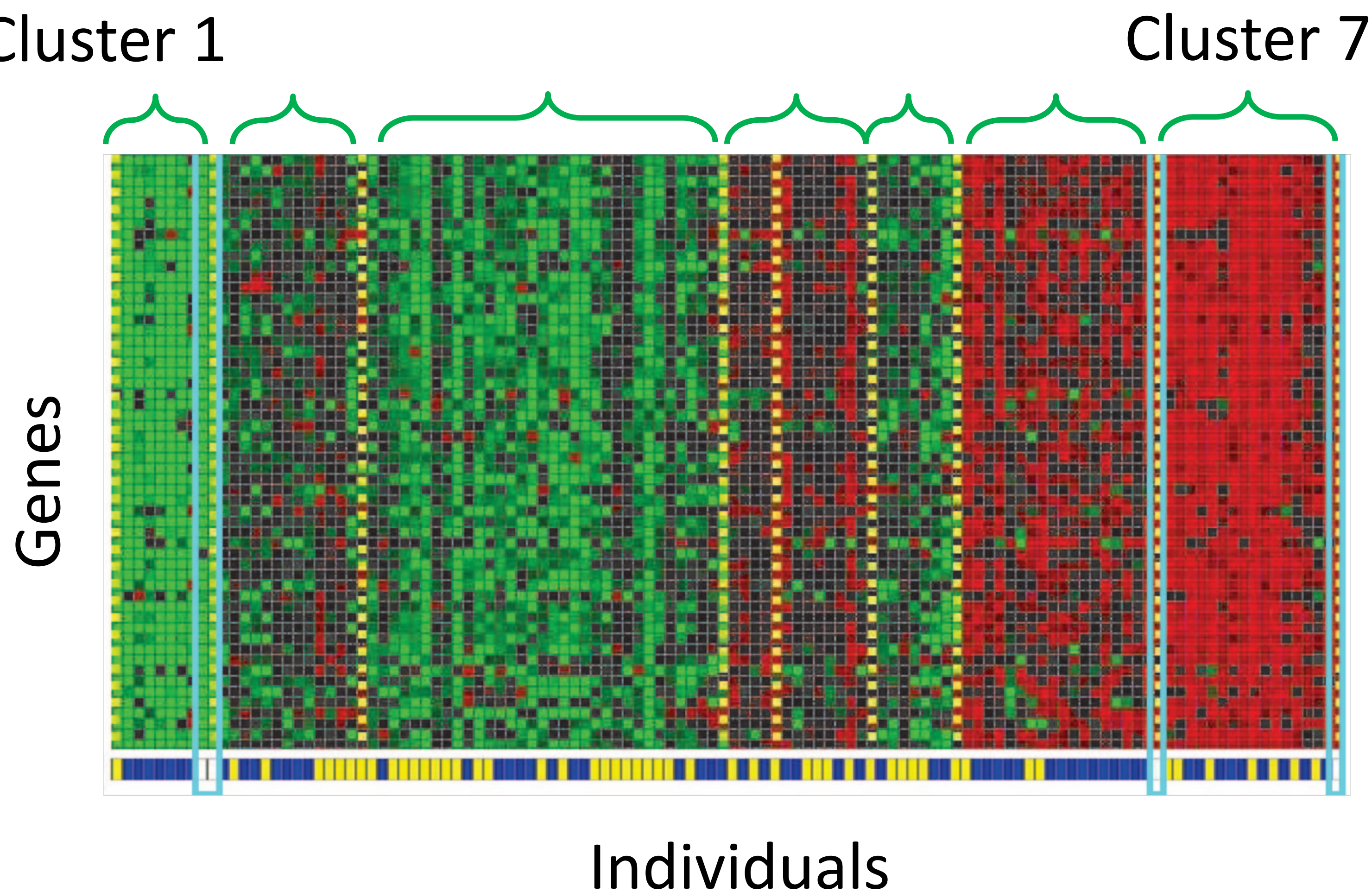


Clustering

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

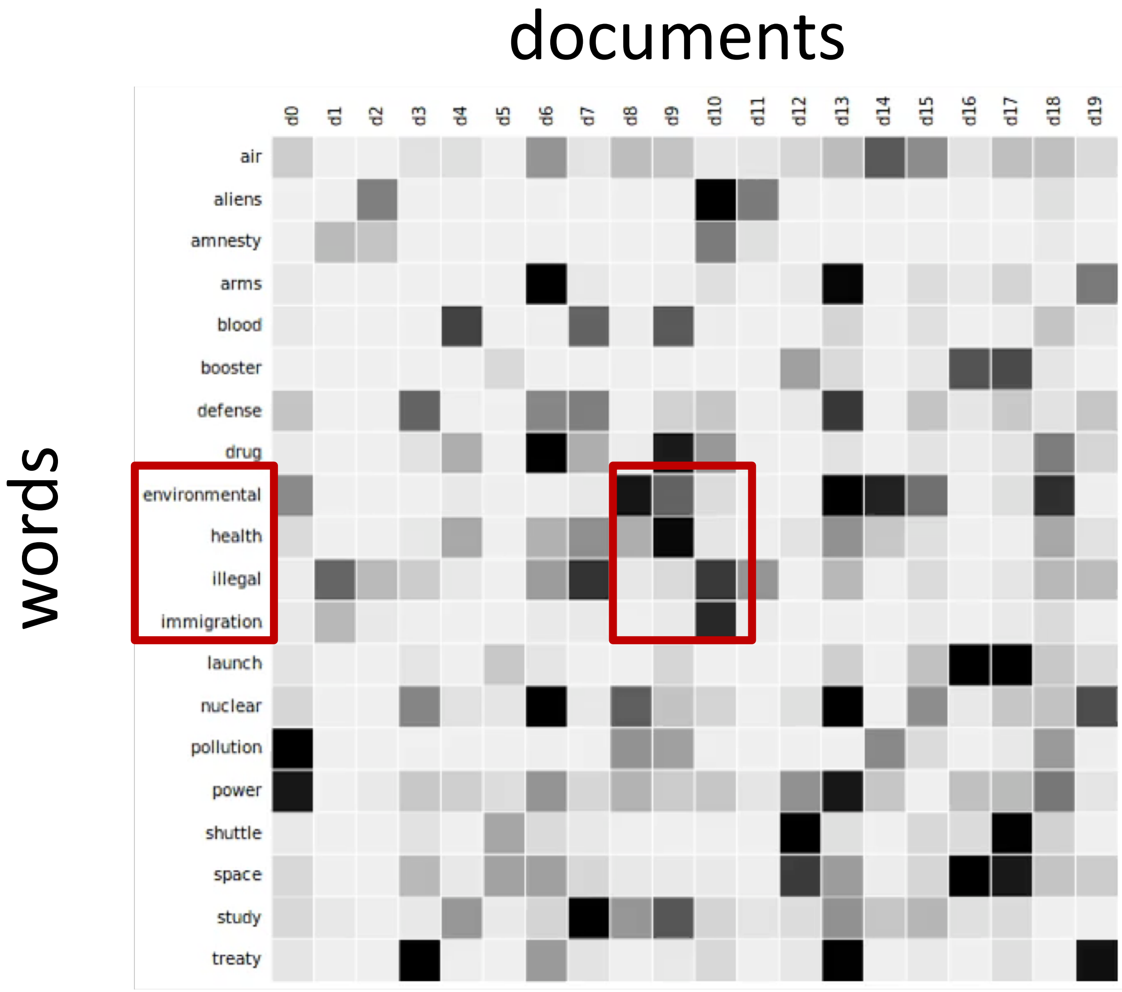


Clustering Genes



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 (tools used in LSA)

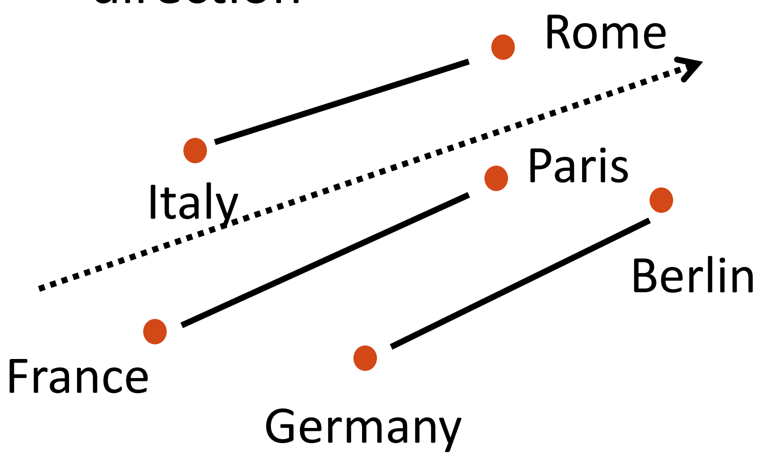
Image credit: https://commons.wikimedia.org/wiki/File:Topic_detection_in_a_document-word_matrix.gif

Word Embeddings

Unlabeled dataset

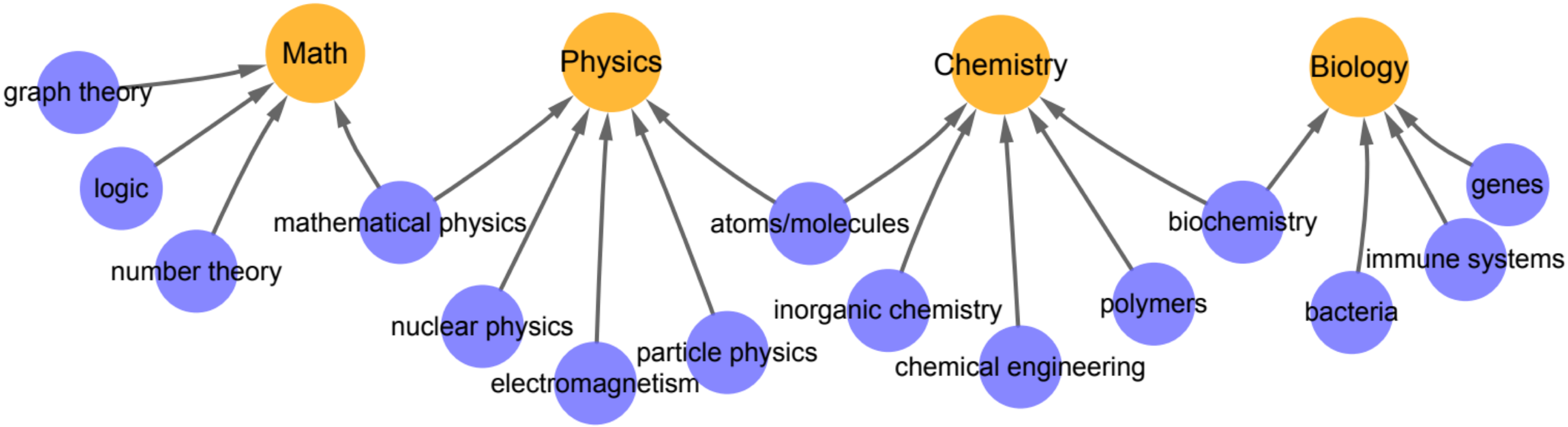
Represent words by vectors

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



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

Clustering Words with Similar Meanings (Hierarchically)

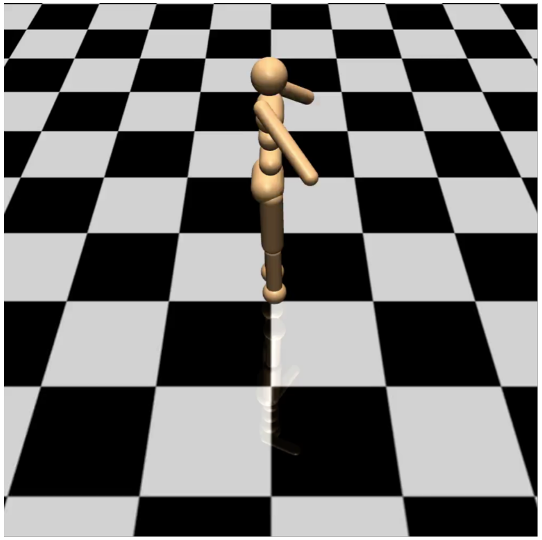


	logic deductive propositional semantics	graph subgraph bipartite vertex	boson massless particle higgs	polyester polypropylene resins epoxy	acids amino biosynthesis peptide
tag	*logic*	*graph theory*	*particle physics*	*polymer*	*biochemistry*

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

Reinforcement Learning

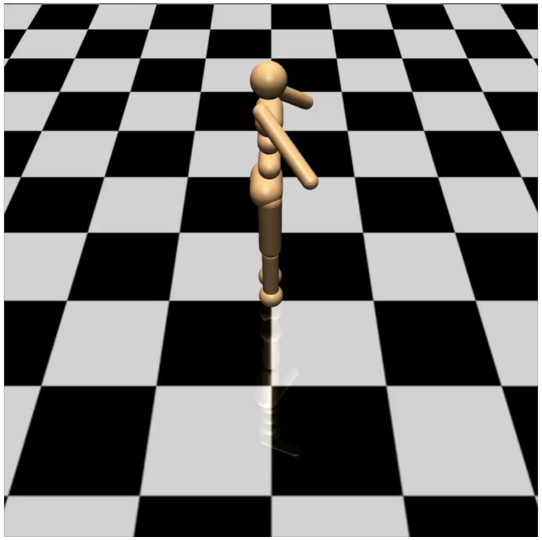
learning to walk to the right



Iteration 10

[Luo-Xu-Li-Tian-Darrell-M.'18]

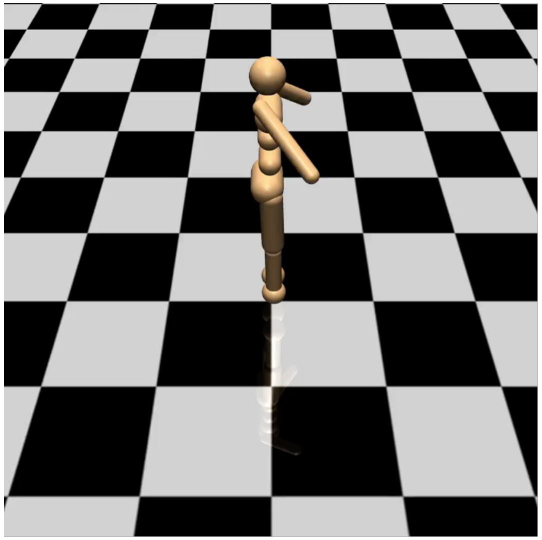
learning to walk to the right



Iteration 20

[Luo-Xu-Li-Tian-Darrell-M.'18]

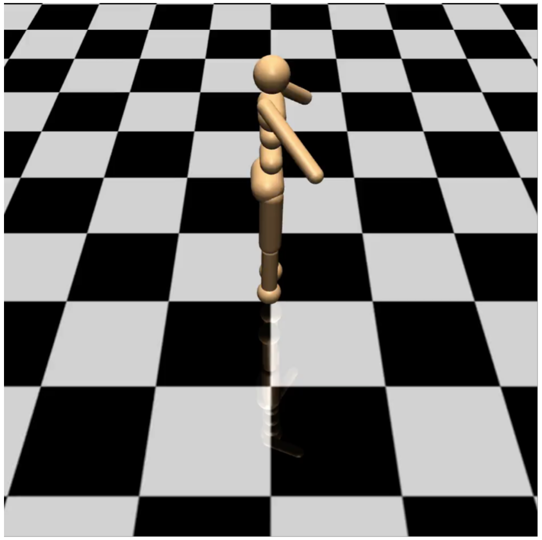
learning to walk to the right



Iteration 80

[Luo-Xu-Li-Tian-Darrell-M.'18]

learning to walk to the right



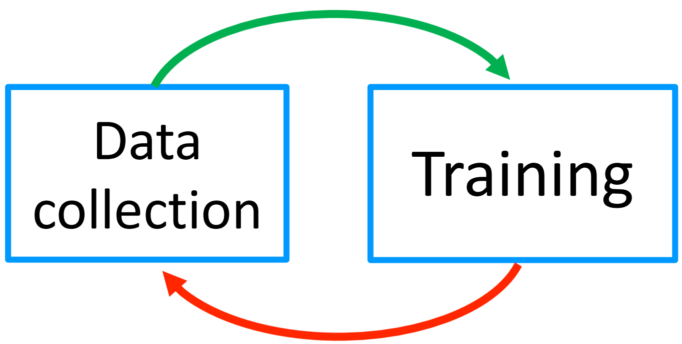
Iteration 210

[Luo-Xu-Li-Tian-Darrell-M.'18]

Reinforcement Learning

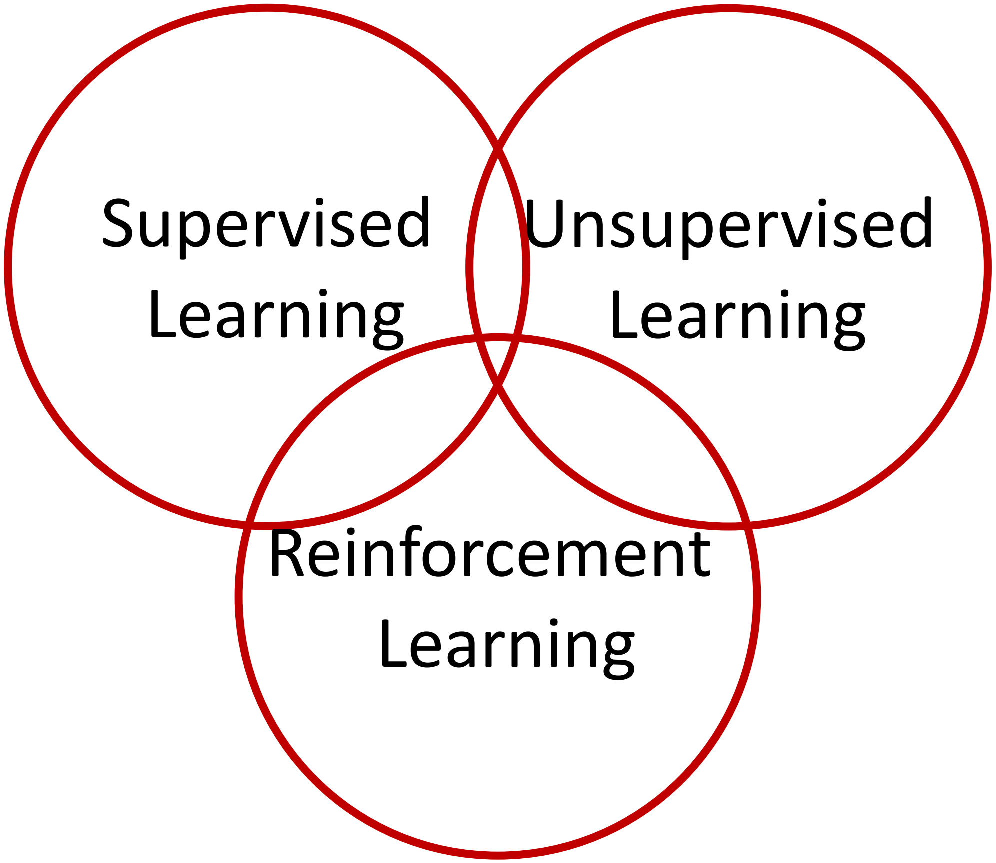
- The algorithm can collect data interactively

Try the strategy and collect feedbacks



Improve the strategy based on the feedbacks

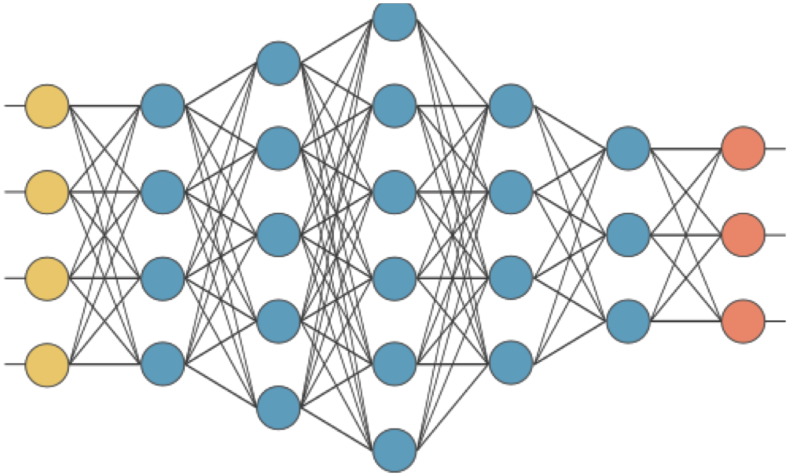
Taxonomy of Machine Learning (A Simplistic View Based on Tasks)



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

Questions?

Thank you!