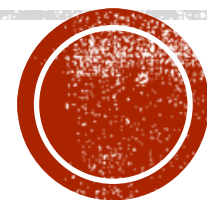




Machine Learning

CS229 / STATS229



Instructors: Moses Charikar, Tengyu Ma, and Chris Re

**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


Differences From Previous Quarters

- Everything will be online --- lectures, Friday and discussion sections, office hours, discussions between students
 - We strongly encourage you to study with others students
 - Technology: Zoom, Slack, ...
- Enrollments increased by 2X in the last two weeks;
Overloaded CAs
- Course project is optional
- Homework can be submitted in pairs
- Take-home exams



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 of the two 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 (Optional)

- 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

- Piazza:
 - All announcements and questions (unless you would only reach out to a subset of course staff)
 - [For logistical questions, please take a look at course FAQ first](#)
 - Finding study groups friends
 - If you enrolled in the class but do not have access to Piazza, it should come within a day. If it has been more than that, send Kush an email
- Slack workspace
- Videos on canvas
- Course calendar: office hours and deadlines
- Gradescope
- Late days policy
- [FAQ on the course website](#)

Teaching Assistants



Course Coordinator
Swati Dube Batra



Course Advisor
Anand Avati



Co-Head TA
Kush Khosla



Co-Head TA
Michael Zhu



Paul Caron



Yining Chen



Taide Ding



Qijia Jiang



Fereshte Khani



Akshay Smit



Guanzhi Wang



Jingbo Yang



Victor Zhang

1. Administrivia

cs229.stanford.edu

2. Topics Covered in This Course

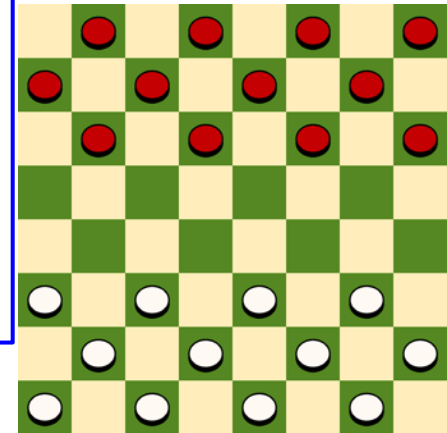
Definition of Machine Learning

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**



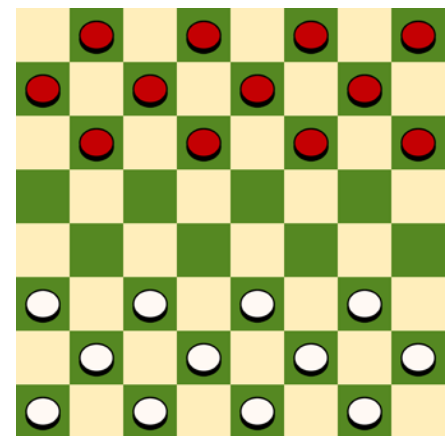
Definition of Machine Learning

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



Taxonomy of Machine Learning

(A Simplistic View Based on Tasks)

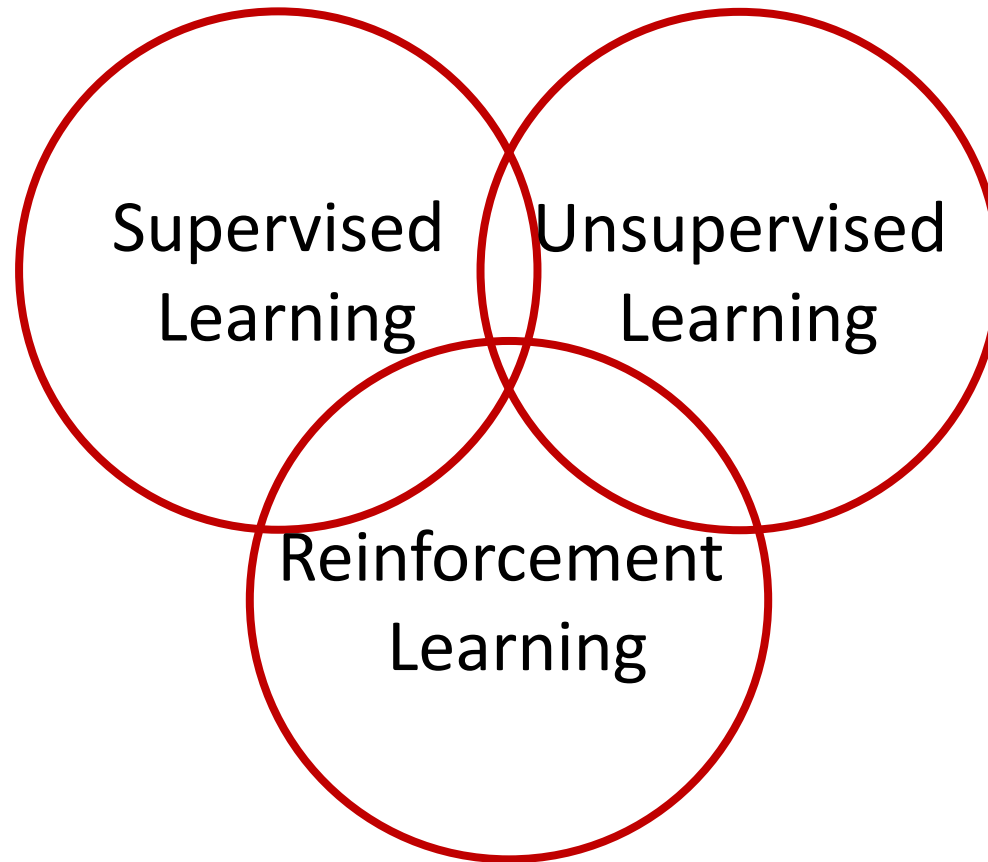


Supervised
Learning

Unsupervised
Learning

Reinforcement
Learning

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



can also be viewed as tools/methods

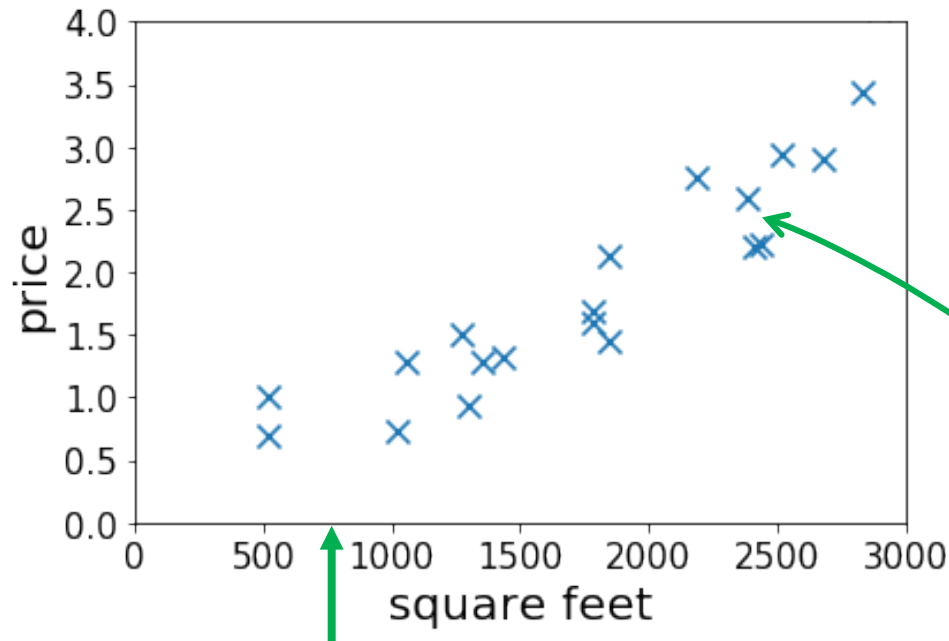
Supervised Learning

Housing Price Prediction

- Given: a dataset that contains n samples

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

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



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

$$x = 800$$

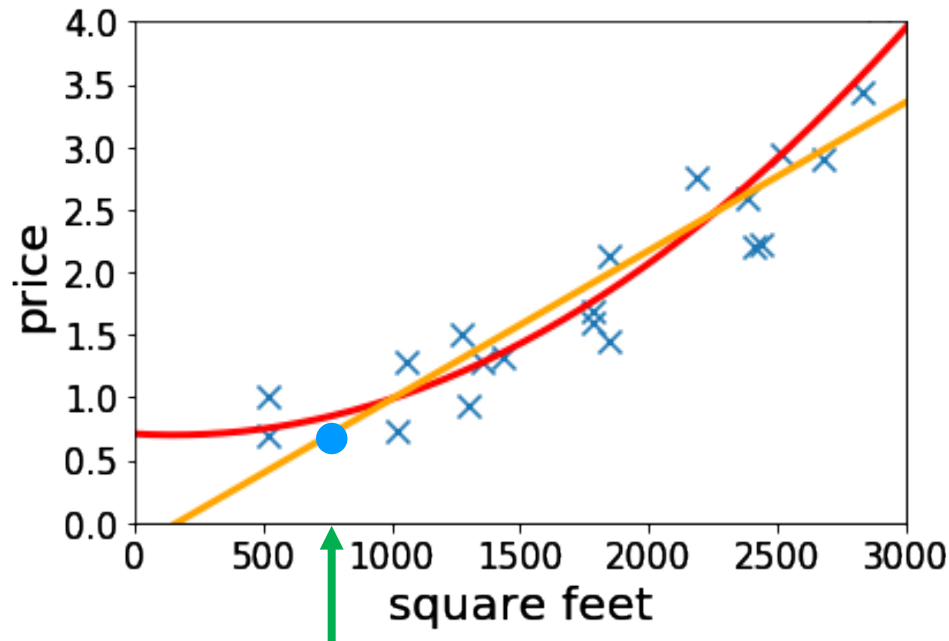
$$y = ?$$

Housing Price Prediction

- Given: a dataset that contains n samples

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

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



- Lecture 2&3: fitting $x = 800$
 $y = ?$

More Features

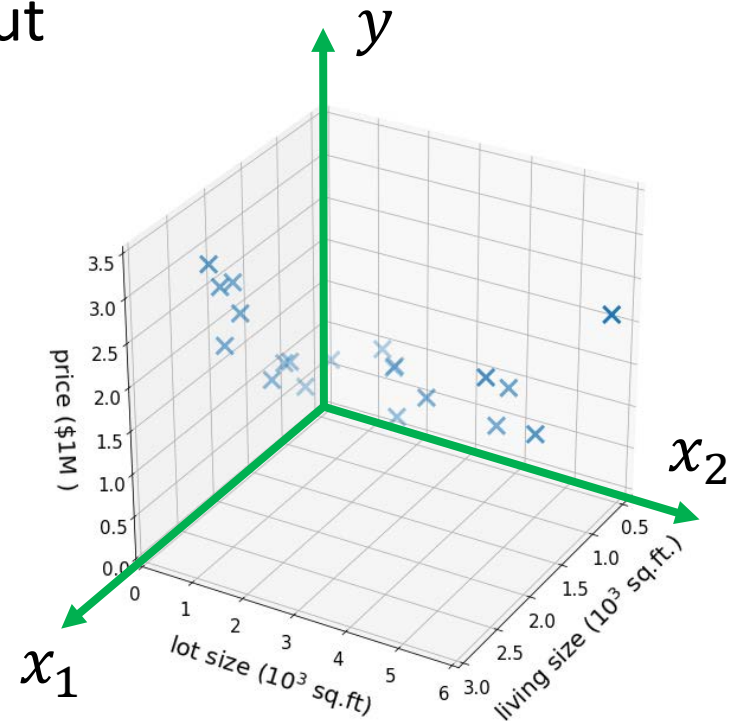
- Suppose we also know the lot size
- Task: find a function that maps

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

- Dataset: $(x^{(1)}, y^{(1)}), \dots, (x^{(n)}, y^{(n)})$

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

- “Supervision” refers to $y^{(1)}, \dots, y^{(n)}$



High-dimensional Features

➤ $x \in \mathbb{R}^d$ for large d

➤ E.g.,

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

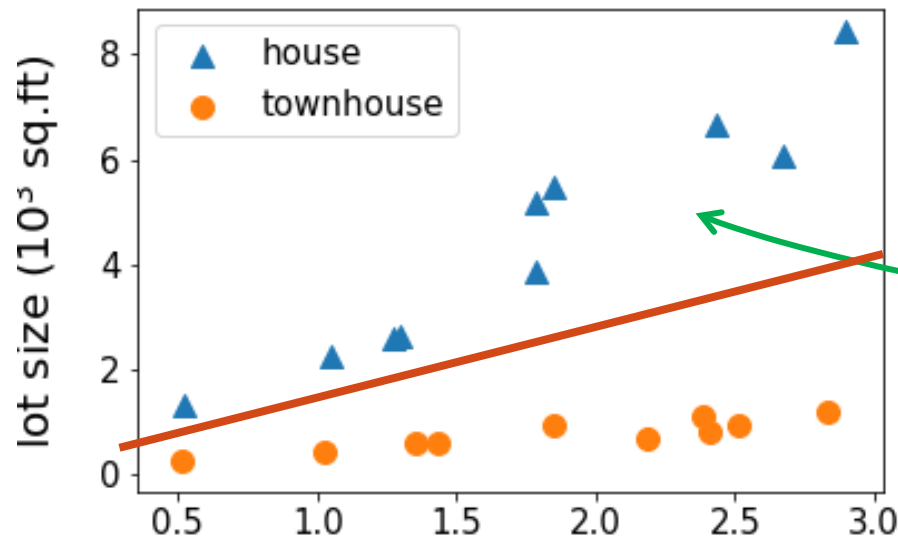
➤ Lecture 6-7: infinite dimensional features

➤ 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

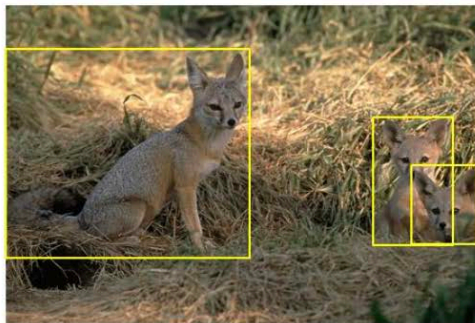
➤ Image Classification

➤ x = raw pixels of the image, y = the main object

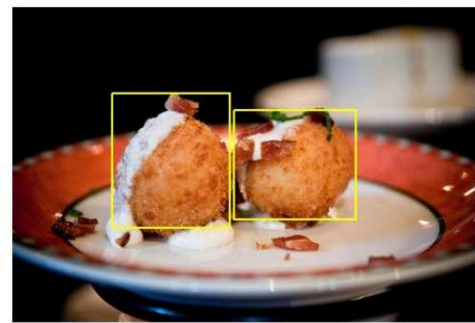


Supervised Learning in Computer Vision

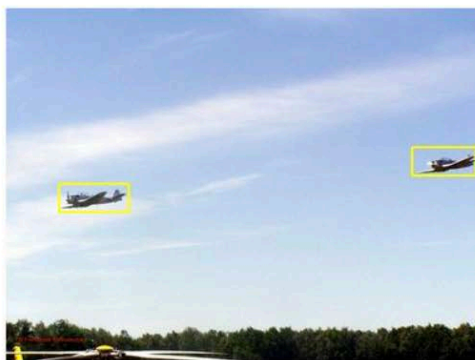
- Object localization and detection
 - x = raw pixels of the image, y = the bounding boxes



kit fox



croquette



airplane

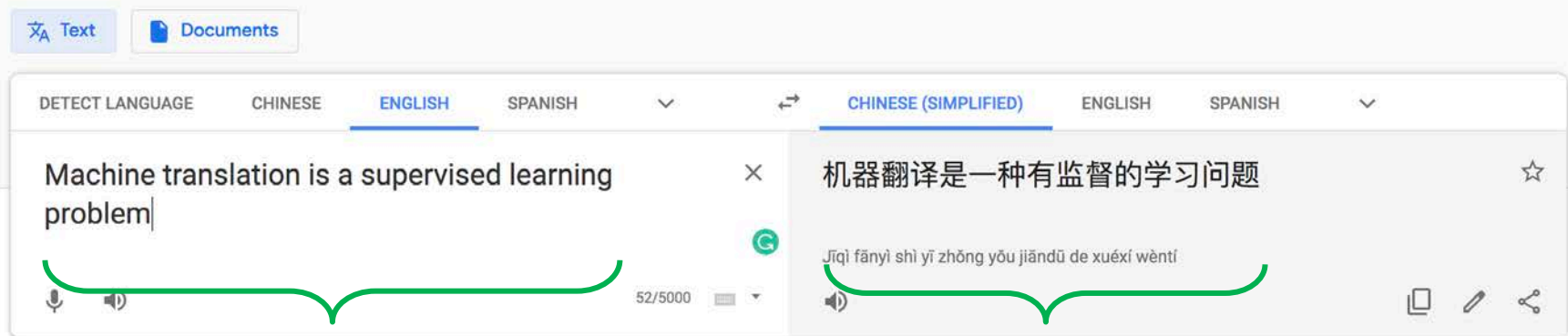


frog

Supervised Learning in Natural Language Processing

➤ Machine translation

Google Translate



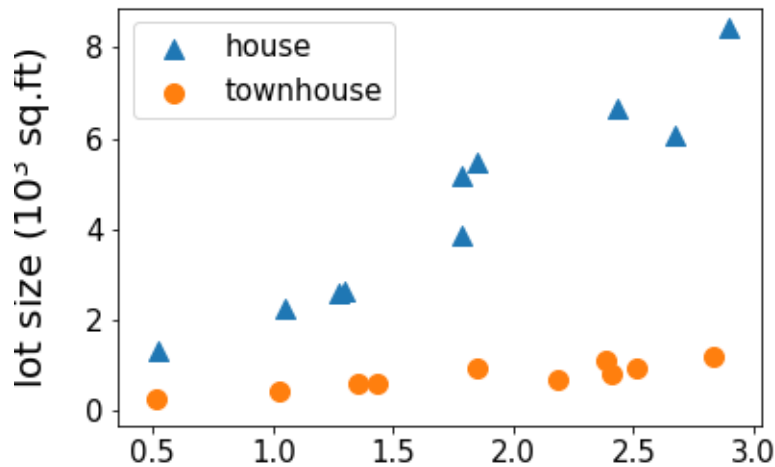
- **Note:** this course only covers the basic and fundamental techniques of supervised learning (which are not enough for solving hard vision or NLP problems.)
- CS224N and CS231N would be more suitable if you are interested in the particular applications

Unsupervised Learning

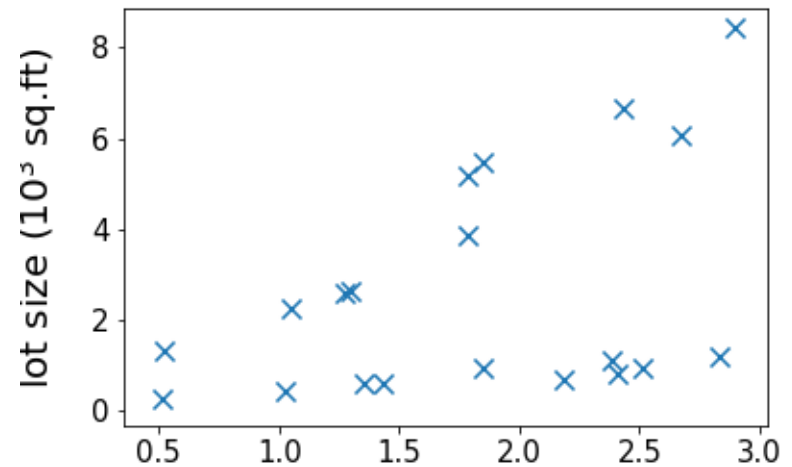
Unsupervised Learning

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

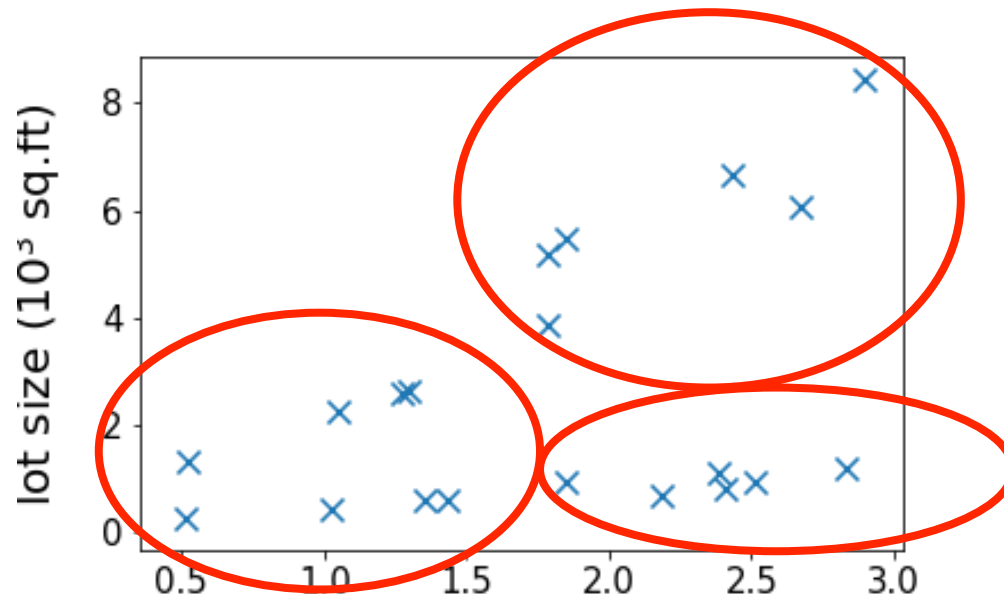
supervised



unsupervised

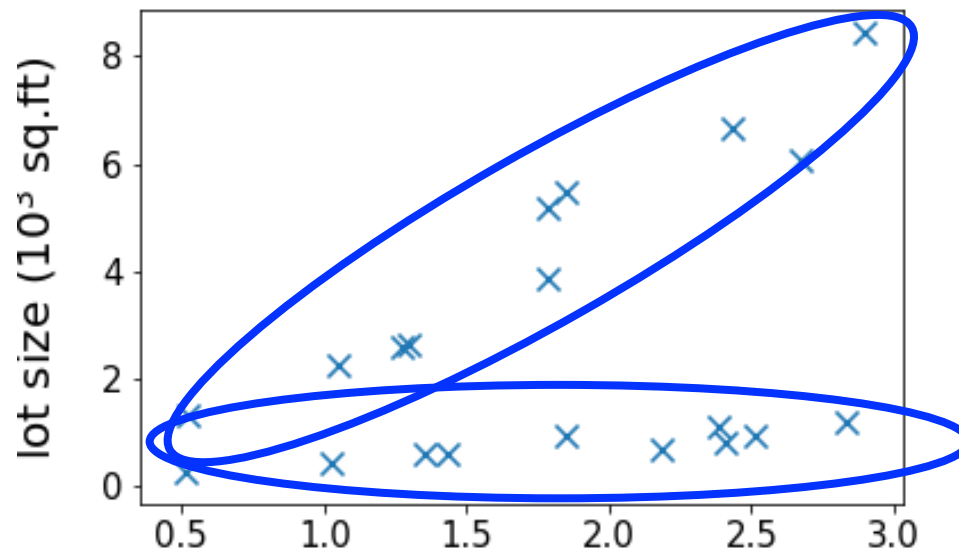


Clustering



Clustering

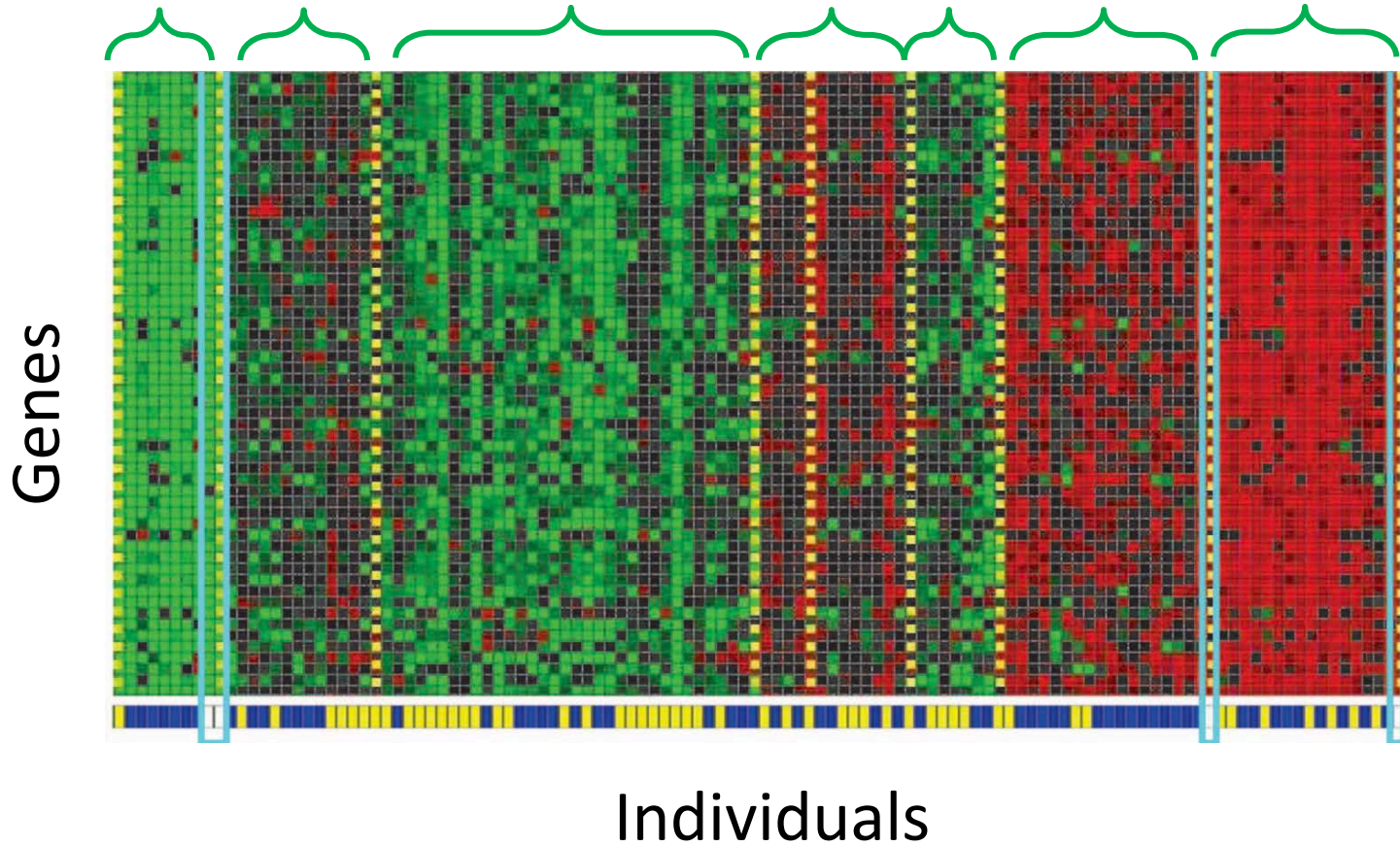
- [Lecture 12&13](#): k-mean clustering, mixture of Gaussians



Clustering Genes

Cluster 1

Cluster 7



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)

documents

words



- 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



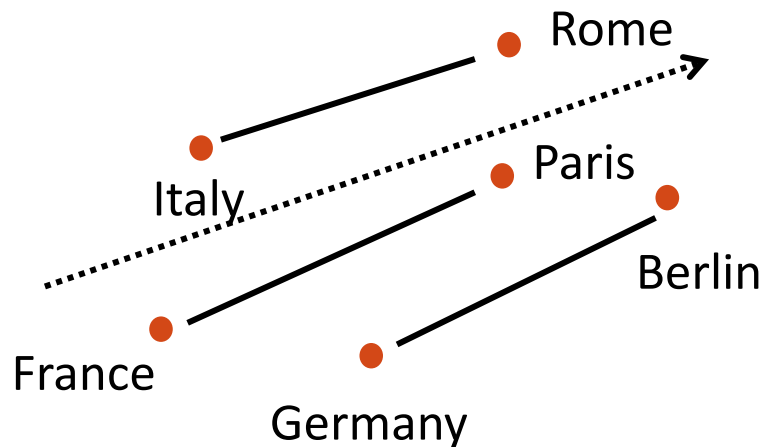
Word Embeddings



Unlabeled dataset

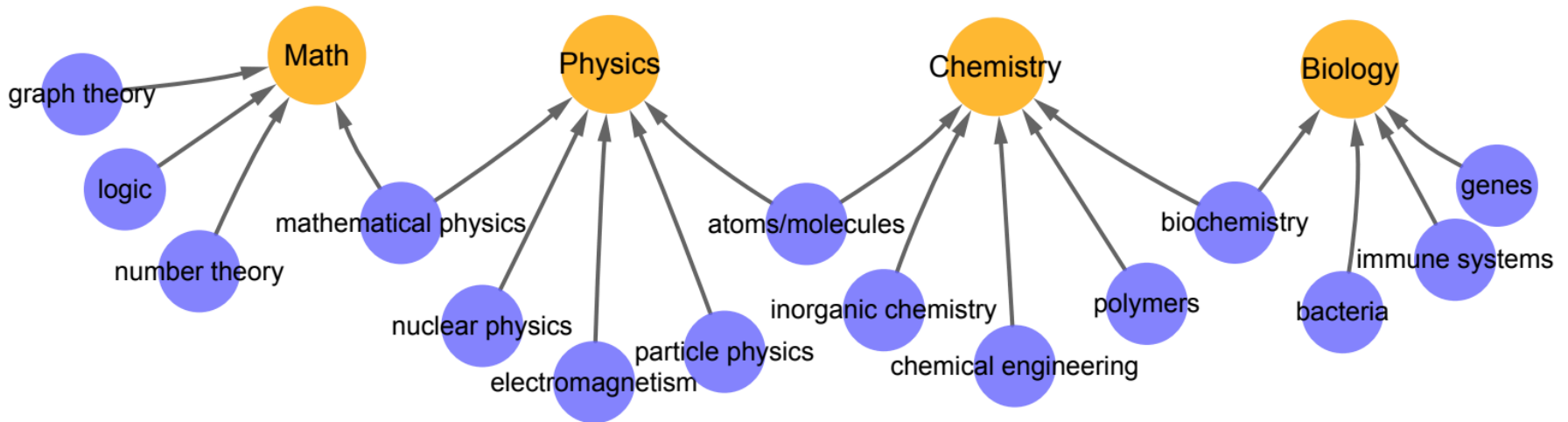
Represent words by vectors

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



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

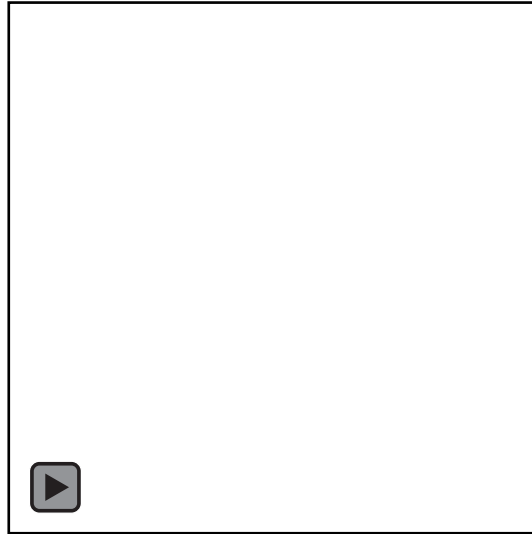
Clustering Words with Similar Meanings (Hierarchically)



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

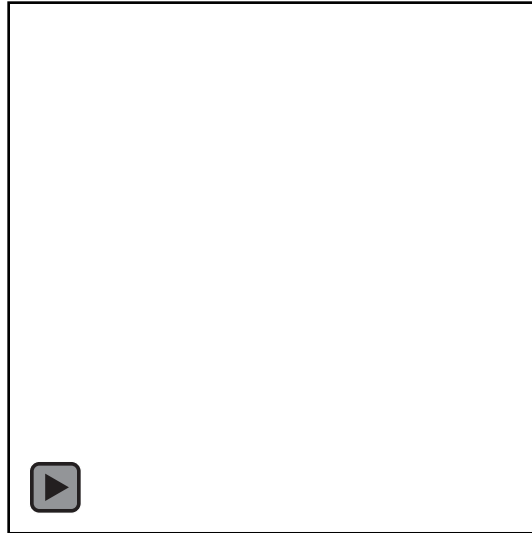
Reinforcement Learning

learning to walk to the right



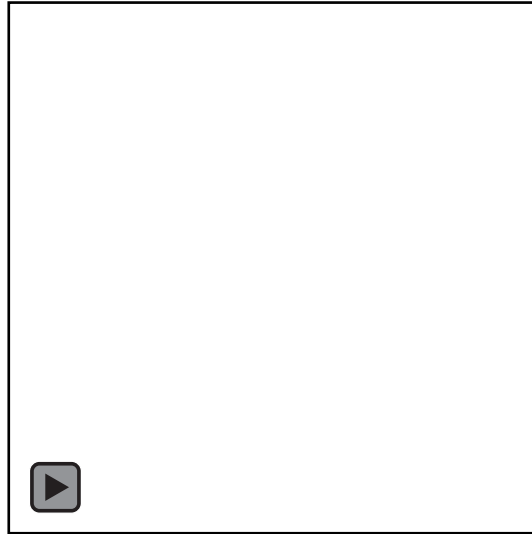
Iteration 10

learning to walk to the right



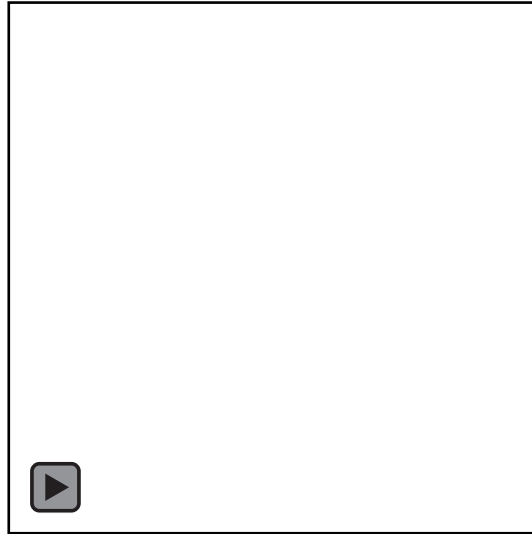
Iteration 20

learning to walk to the right



Iteration 80

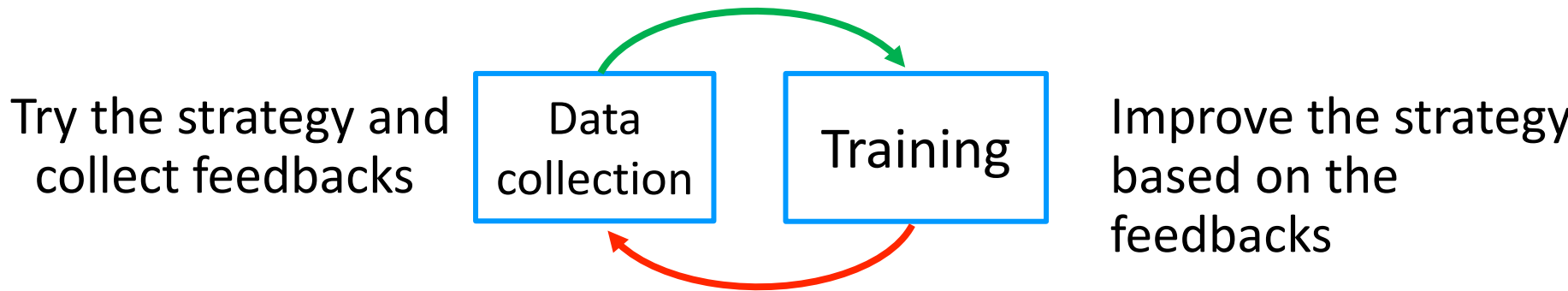
learning to walk to the right



Iteration 210

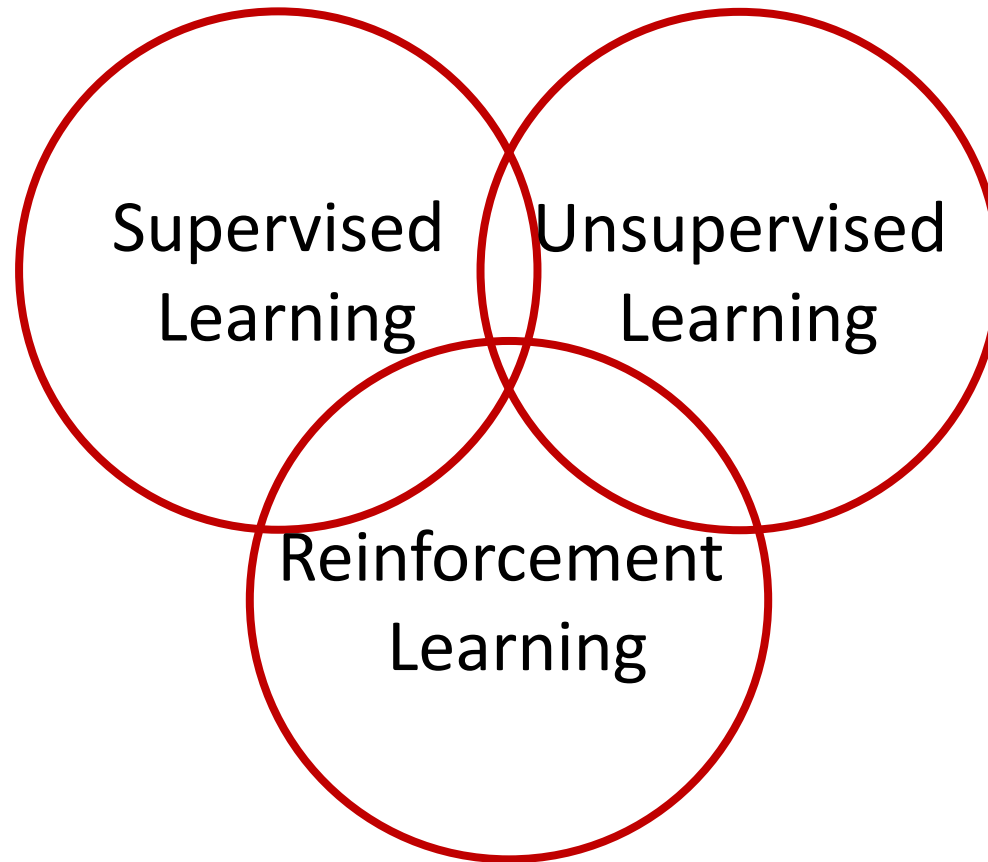
Reinforcement Learning

- The algorithm can collect data interactively



Taxonomy of Machine Learning

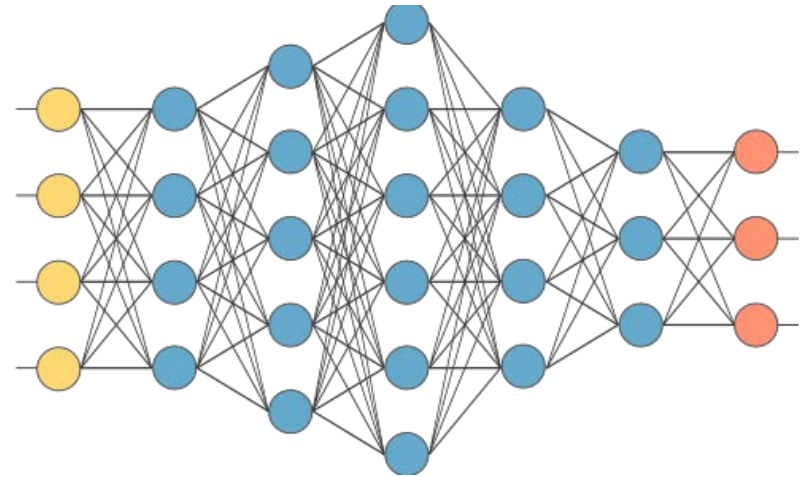
(A Simplistic View Based on Tasks)



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

Questions?

Thank you!