

Machine Learning

CS360

Seating Charts

 SEATING ASSIGNMENTS

How has your experience been with the Seating Assignments tool? [Send us your Feedback!](#)

Viewing seat assignments for: ▾

To get started, see the [Quick-Start guide](#) or review the [Instructor Knowledgebase article](#).



Select a section / cohort from the dropdown menu above.

About us

Instructor: Li Guo lg154@nyu.edu

Recitation instructor: Chloe Song yunjie.song@nyu.edu

Learning assistant: Yijie(Hellen) Wang yt2267@nyu.edu

What is machine learning?

computers machine(algos) react programming inputs generated
things modeling predictions application machines
historical algorithm training different
train based using fun learn future
output tendency know learn teaching
tendency involves
code computer use predict model
actually direct output made answer
big new input huge given
deepen learning analysis logic
multivariate gives statistics create
expect pattern paradigm find days feed
tensorflow stats prediction generative performance
result

Machine Learning

in:spam

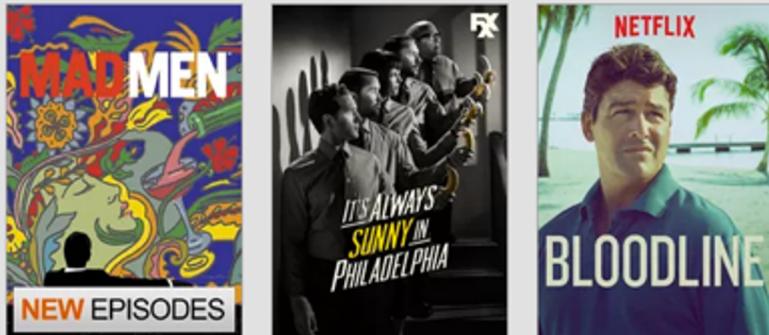
The conversation has been unmarked as spam and moved to the Inbox. [Learn more](#) [Undo](#)

1-12 of 12

Delete all spam messages now (messages that have been in Spam more than 30 days will be automatically deleted)

<input type="checkbox"/>		Linda Smith	College Job - Earn \$795/week working part-time - Dear Student, I would like to offer you an opportunity for earning extra money. Get Paid To t	Mar 28
<input type="checkbox"/>		Mark Putney MNM Partners.	San Francisco - Still the Most Expensive City for New Renters - Click here to view this message in a browser window. Benjamin, It's A Whole	Mar 27

Top Picks for Benjamin



Everything from E-mail spam filters to Netflix recommendation engines & Amazon suggested products all rely on machines crunching incomprehensible amounts of data in order to better “learn” your preferences!

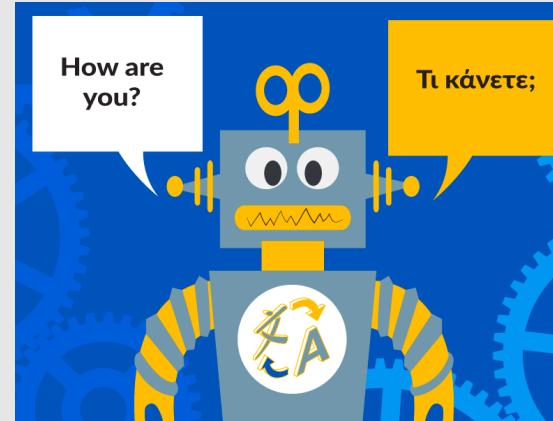
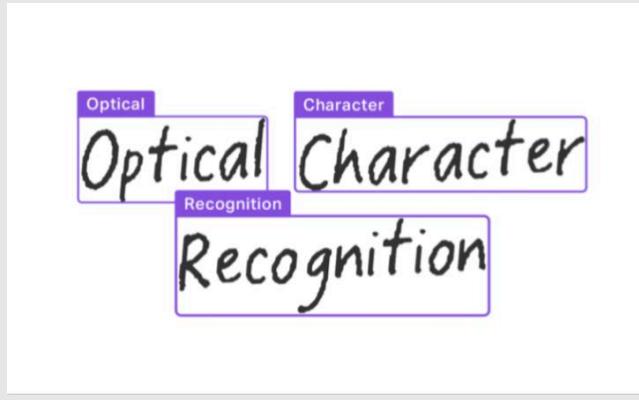
Frequently Bought Together



Some of these items ship sooner than the others. Show details

- This item: Vanity Fair Everyday, 400 Count \$8.51 (\$0.02 / Count) [Add-on Item](#)
- Glad Tall Kitchen Drawstring Trash Bags, 13 Gallon, 90 Count \$14.98 (\$0.17 / Count) [Add-on Item](#)
- Lysol Disinfecting Wipes, Lemon and Lime Blossom, 240 Count \$9.97 (\$0.04 / Count) [Add-on Item](#)

Machine Learning



Machine Learning



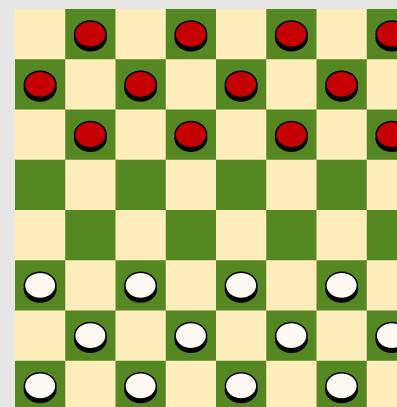
Machine Learning

“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.”

Tom Mitchell(1998)

Experience(Data): games played by the program

Performance measure: winning rate



A new kind of programming

Write a computer program with **explicit rules** to follow:

```
If email contains V!agr@  
    Then mark is-spam  
If email contains ...  
If email contains ...
```

Traditional Programs

Write a computer program to **learn from examples**:

```
Try to classify some emails;  
Change self to reduce errors;  
Repeat;
```

Machine Learning Programs

Machine Learning

“Machine learning is a field of computer science that uses statistical techniques to give computer system the ability to “learn” with data, without being explicitly programmed.”

Arthur Samuel(1958)

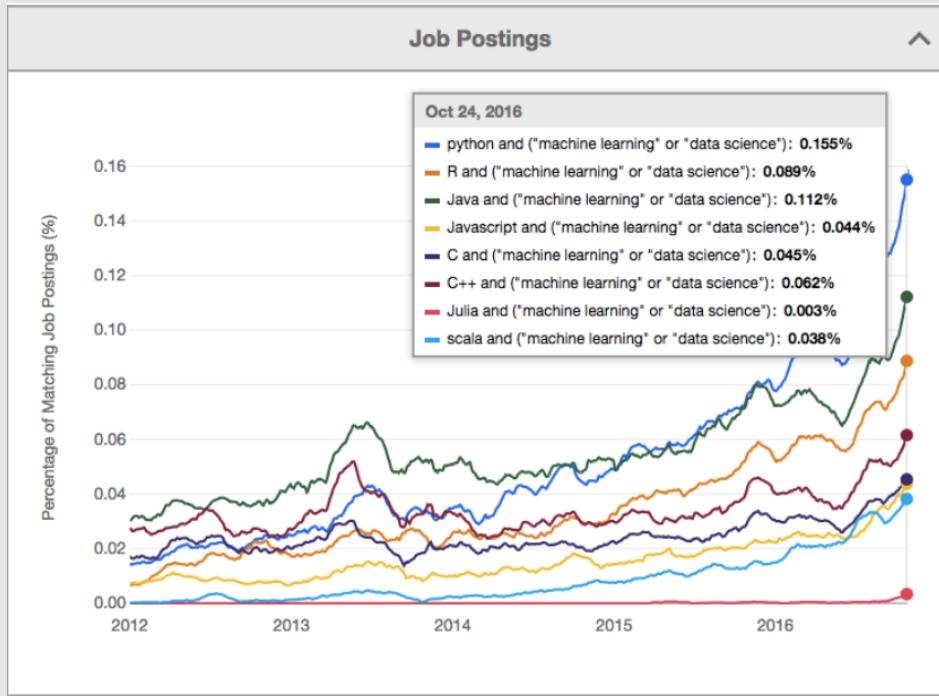
Understand Machine Learning from Memes



don't worry about it if you don't
understand



Why study machine learning



Why study machine learning?

MACHINE LEARNING USE EXAMPLES



THE SELF DRIVING GOOGLE CAR



WEB SEARCH RESULTS



SOCIAL LISTENING APPLICATIONS



MARKET PRICING MODELS



TEXT BASED SENTIMENT ANALYSIS



FRAUD DETECTION



PATTERN RECOGNITION



CREDIT SCORING



PREDICTION OF SUCCESS AND FAILURE



ONLINE RECOMMENDATION OR OFFERS ON BIG ECOMMERCE SITES (AMAZON, NETFLIX)

Big data empowers machine learning



Course organization

Teaching Mode

- Jan 26th – Mar 6th: Mixed Mode
- Mar 6th – May 14th: In Person



WeChat group



Anonymous
feedbacks

Logistics

NYU classes for course related materials, slides, assignment

Office hours

Li Guo: Wed 9:30-10:30PM 1157, Thu 2:30-3:30PM 28-117 Shinmay

Yunjie Song: Tue 3:00-4:00PM 1162, Fri 2:30-3:30PM 28-102 Shinmay

Hellen: TBD

Prerequisites

- Probability and statistics
- Calculus (differential and integral)
- Intro to Computer Programming (Python)
- Desirable
 - Linear Algebra
 - Multivariable Calculus
 - Data structure

Reference Materials

- Andrew Ng's Machine Learning and Deep Learning courses (Stanford/Coursera)
- James G., Witten D., Hastie T. Tibshirani R., In Introduction to Statistical Learning
- Hastie T., Tibshirani R., Friedman J., The Elements of Statistical Learning
- Bishop C., Pattern Recognition and Machine Learning

Course Evaluation

- Homework assignments: 25%
- Quizzes: 35%
- Final project: 40%

Final project:

- Team of ≤ 3
- Solve a real world problem using ML
 - Try to find a problem that interests you.
 - You need to get the data yourself
 - Do not use problems from a stats course, it will put you at a disadvantage.
- Deliverables: Final presentation & a paper to summarize your work
- [final project guidelines](#)

Some Examples

Examples from last year

Potential topics:

- Theory
- General Machine Learning
- Final and Commerce
- Computer vision
- Natural Language Processing
- Reinforcement Learning
- Covid-19

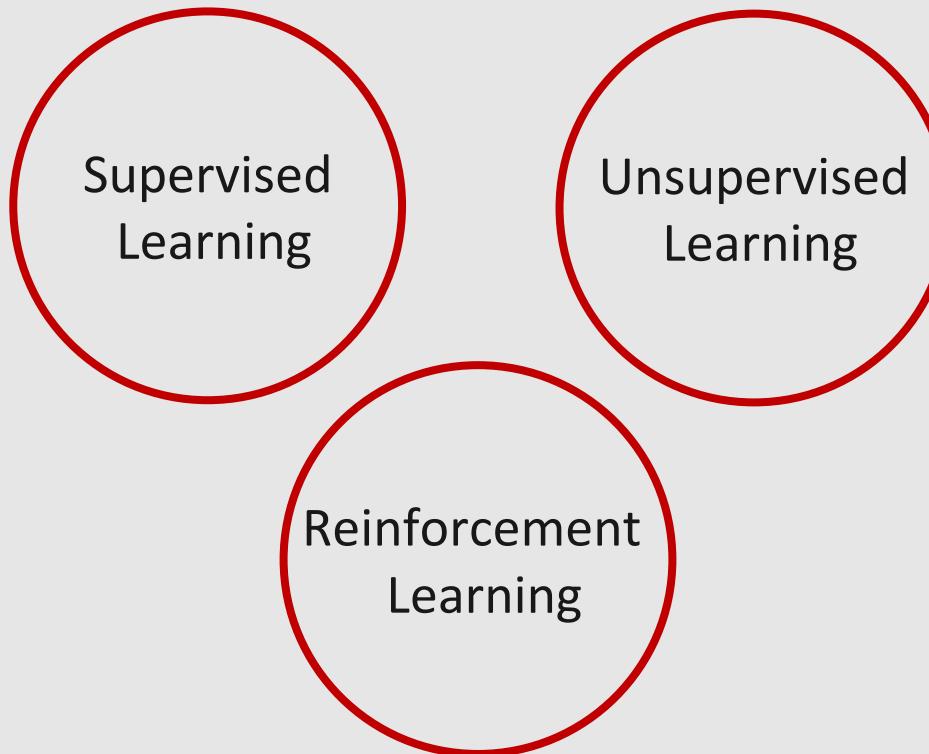
Key dates (subject to change)

- 2021-3 : Group formation due
- 2021-3: Project proposal
- 2021-5 Final paper due
- 2021-5: Final presentation due

Learning Map

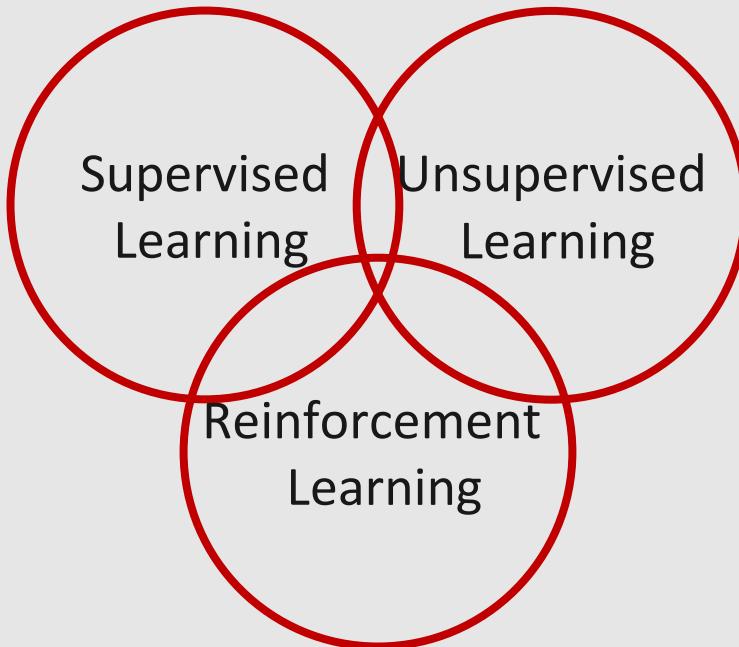
Taxonomy of Machine Learning

(A Simplistic View Based on Tasks)

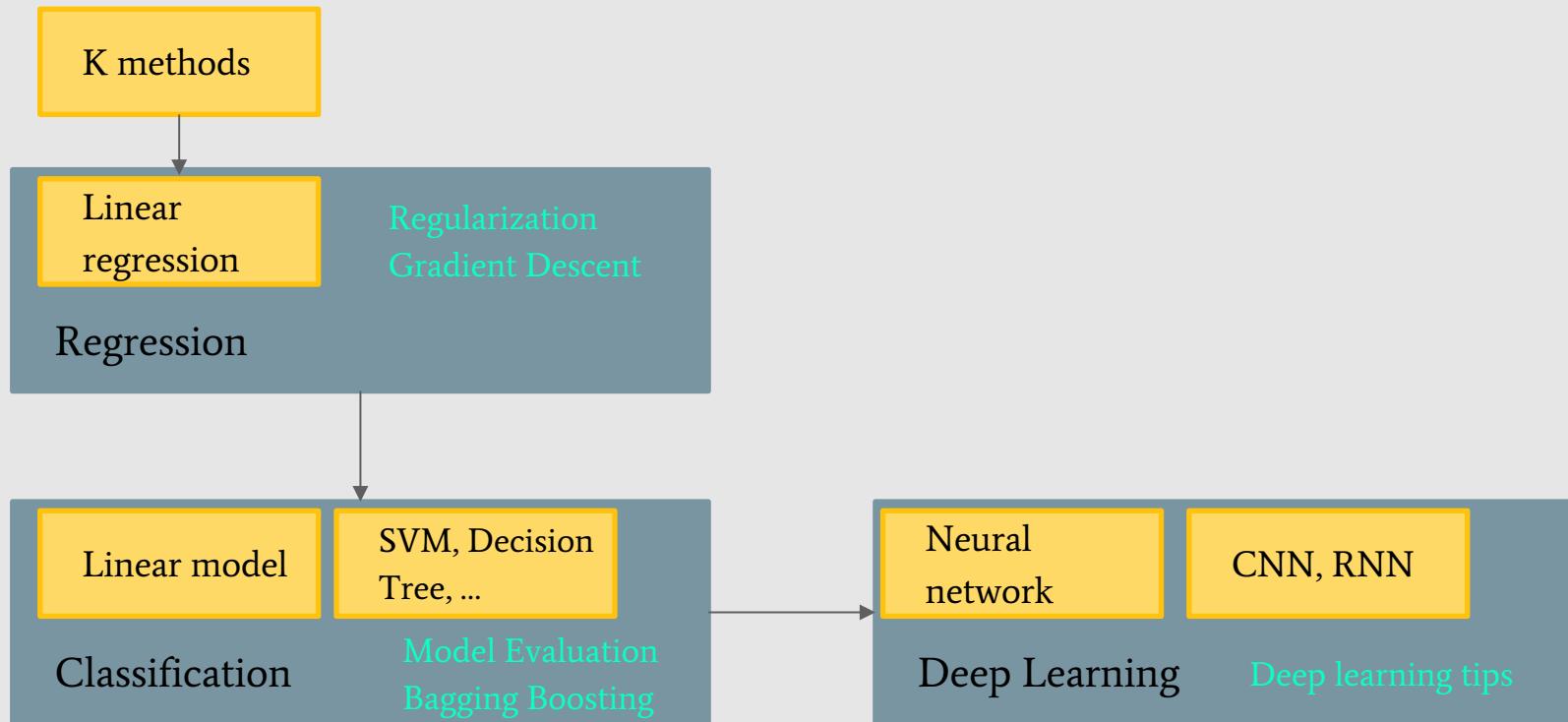


Taxonomy of Machine Learning

(A Simplistic View Based on Tasks)

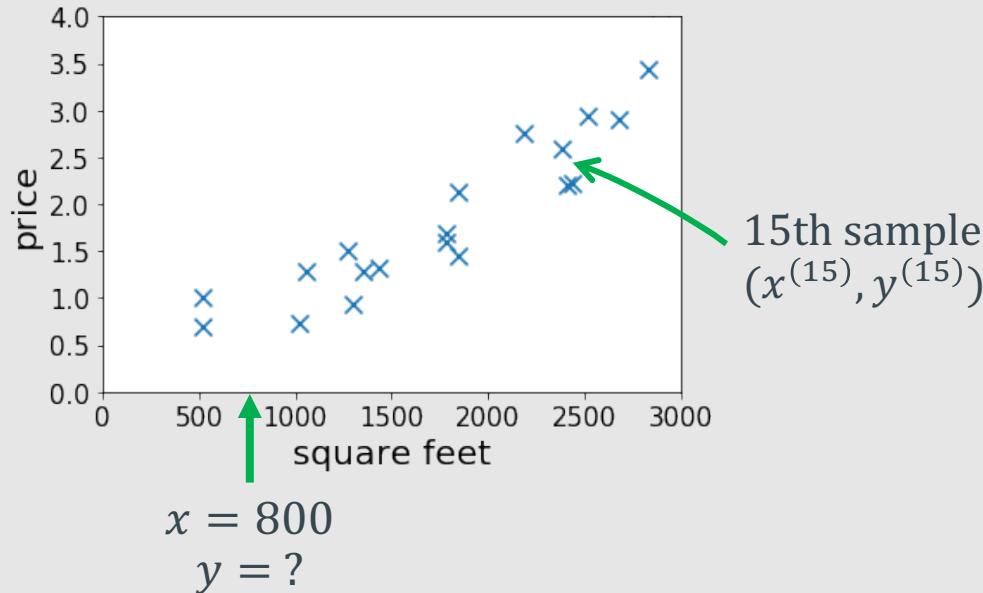


Learning Map



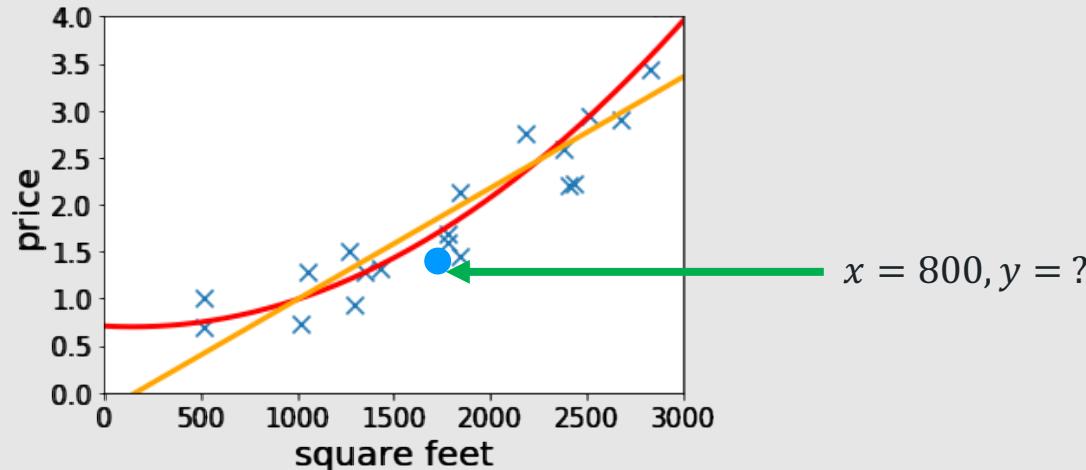
Housing Price Prediction

- 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?



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?



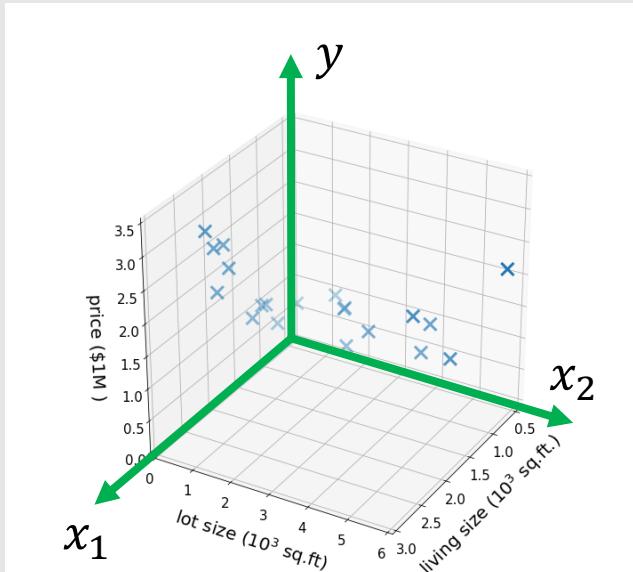
- fitting linear/quadratic functions to the dataset

More Features

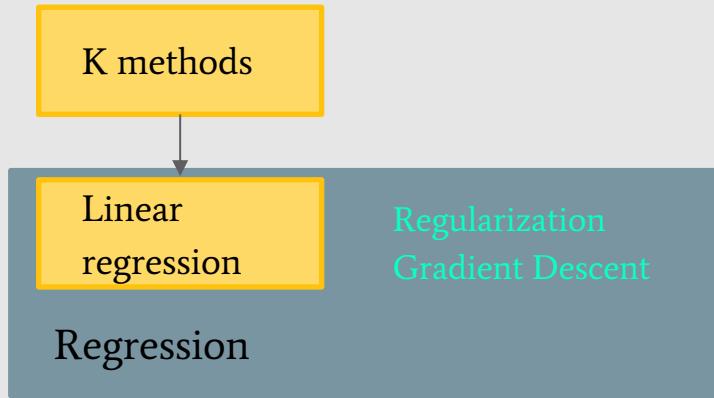
- Suppose we also know the lot size
- Task: find a function that maps

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

- Dataset: $(x^{(1)}, y^{(1)}), \dots, (x^{(n)}, y^{(n)})$
- where $x^{(i)} = (x_1^{(i)}, x_2^{(i)})$
- “Supervision” refers to $y^{(1)}, \dots, y^{(n)}$

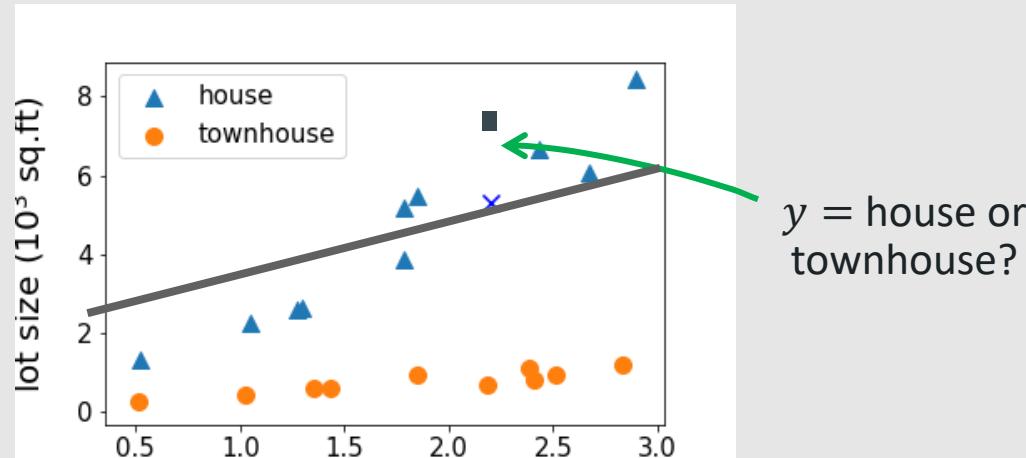


Learning Map



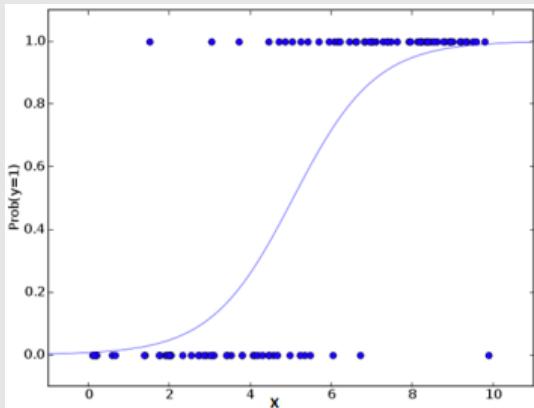
Classification

- 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
 $(\text{size}, \text{lot size}) \rightarrow \text{house or townhouse?}$

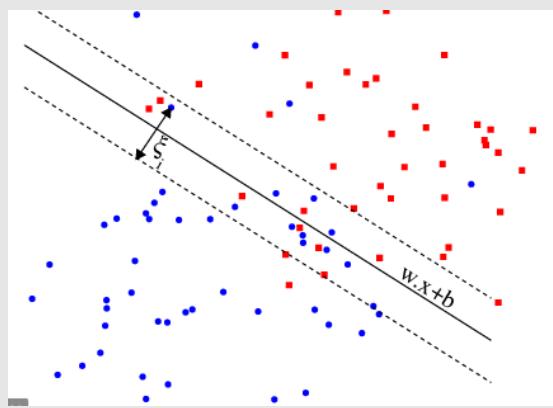


Classification

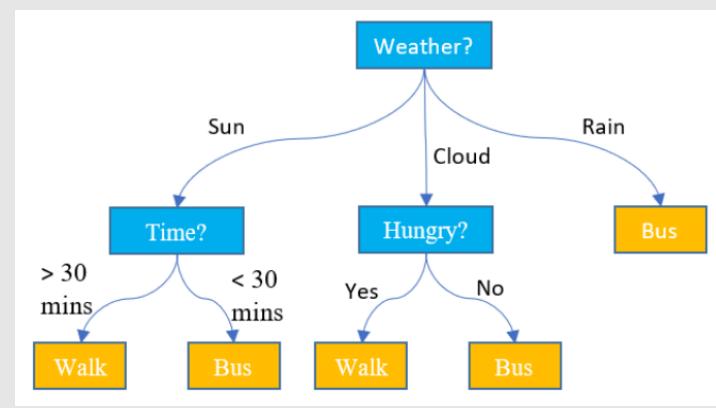
Logistic regression



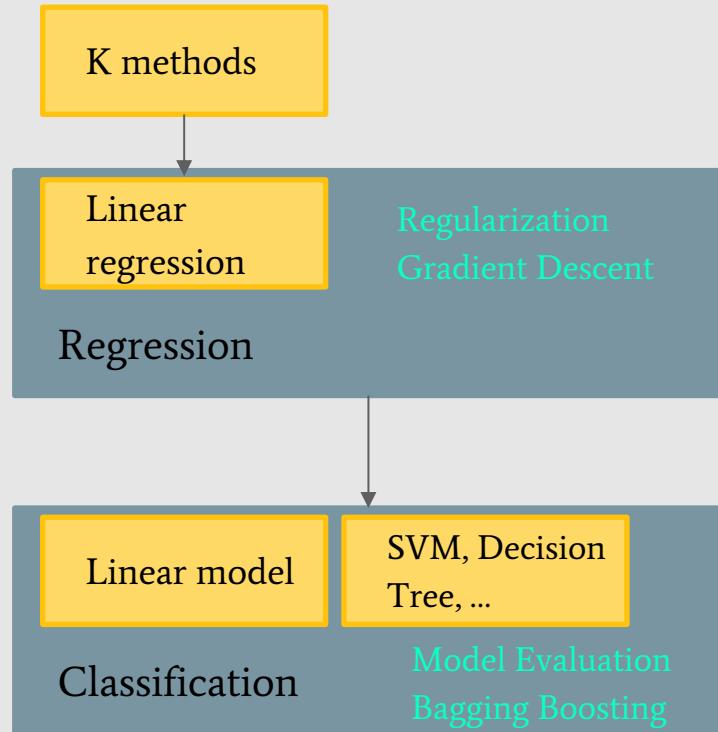
SVM



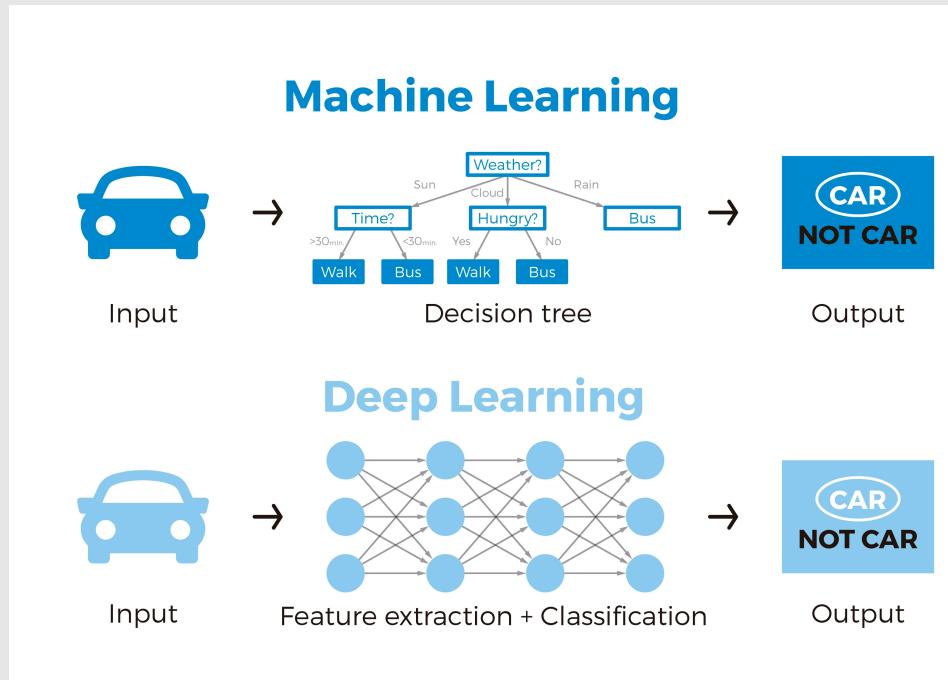
Decision Trees



Learning Map



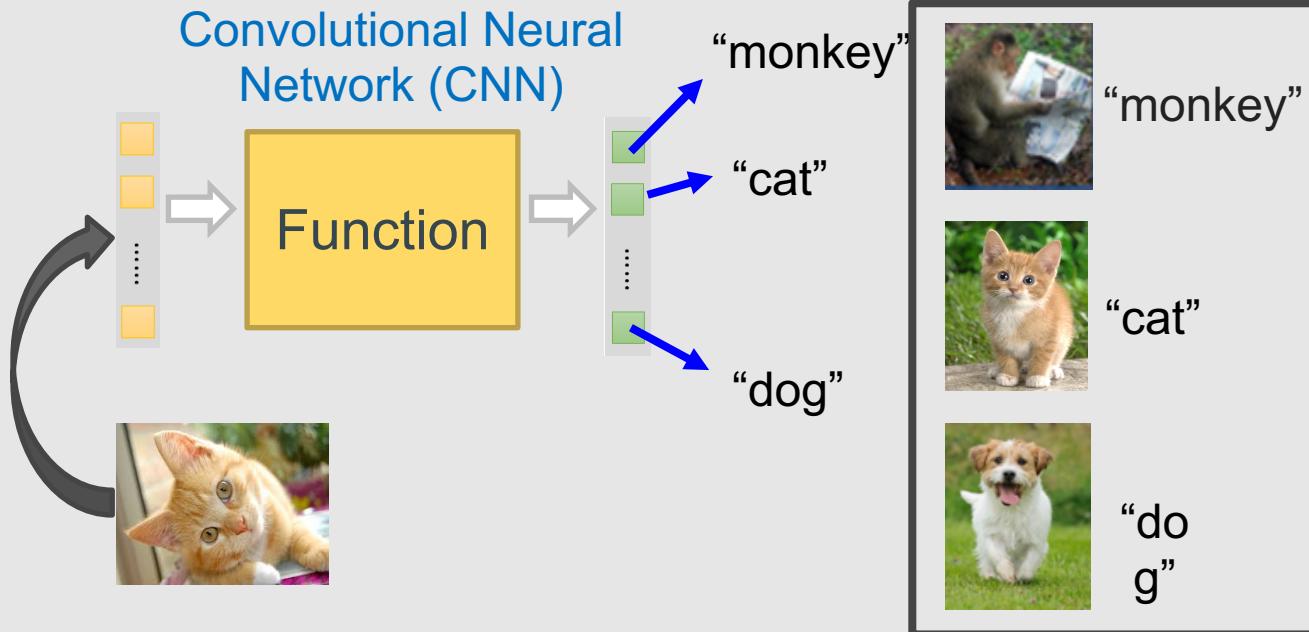
Neural Network



Classification - Deep Learning

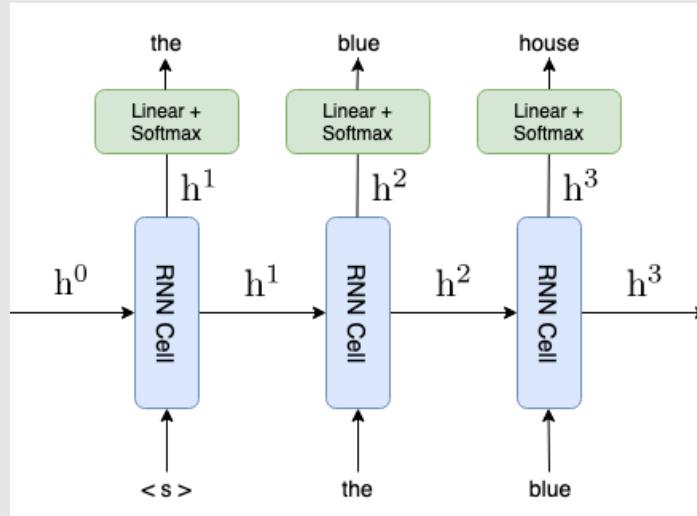
Image Recognition

Training Data



Dealing with sequential data

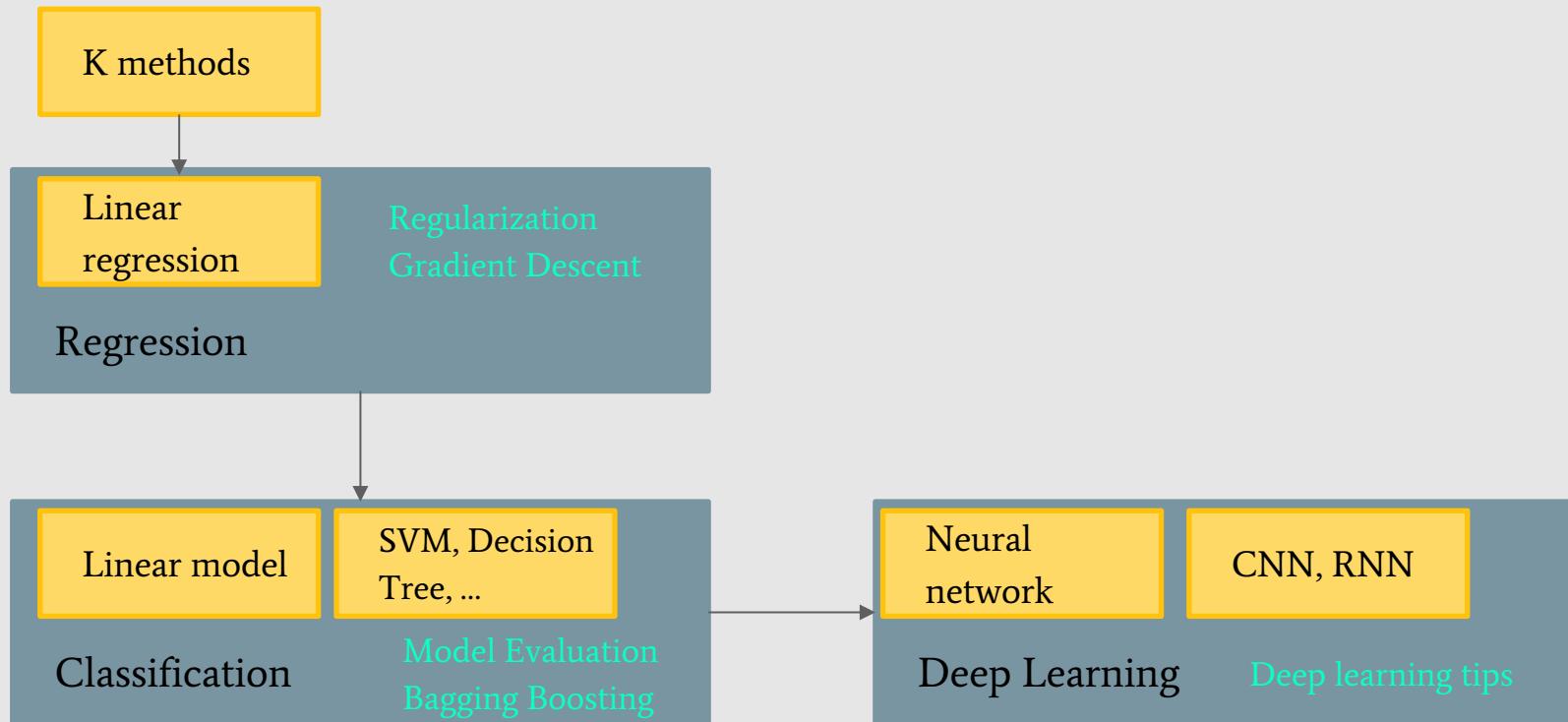
Recurrent neural network



Natural language process

Times series data

Learning Map



For more information please check the syllabus