Machine Learning

Chapter 1: Introduction

Fall 2023

Instructor: Xiaodong Gu

Today

History

Why study machine learning?

What is machine learning?

- Key elements

How machine learning works?

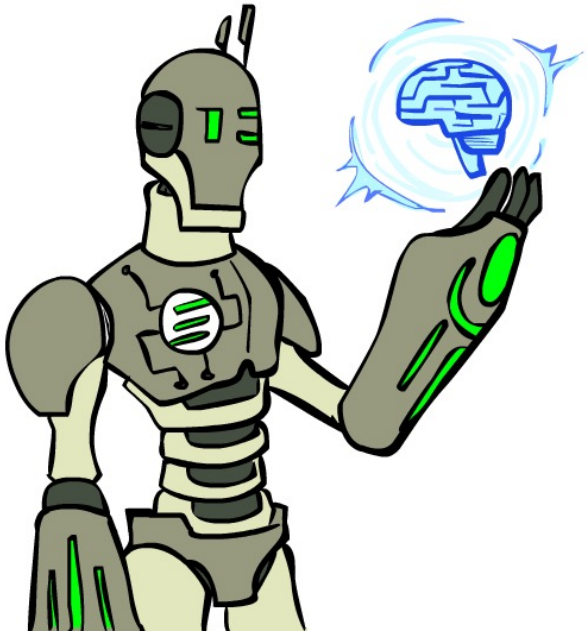
- A generic pipeline
- A working example

Overview of machine learning algorithms

- Categories
- Typical algorithms

Important ML concepts that we'll use throughout the semester

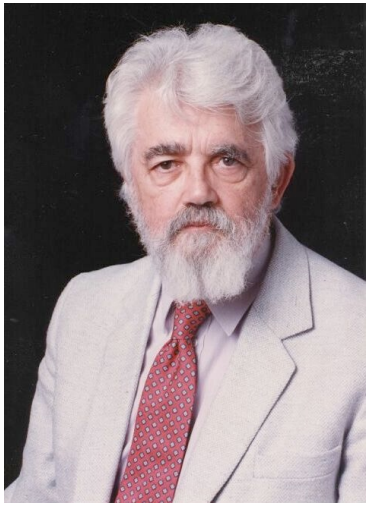
- Generalization、overfitting、regularization



What is Artificial Intelligence (AI) ?

What is intelligence ?

- Capacity for logic, understanding, self-awareness, learning, reasoning, planning , creativity, critical thinking, and problem-solving.

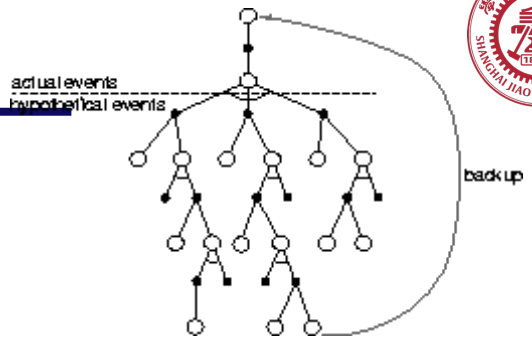


John McCarthy
one of the founders of the discipline of AI

What is artificial intelligence ?

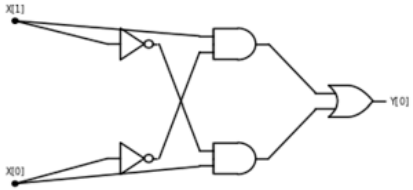
- It is the science and engineering of making intelligent machines, especially intelligent computer programs.

A Brief History of AI



1940—1950s: Early days

- 1943: McCulloch & Pitts: **Boolean circuit** model of brain
- 1950: Turing's "Computing Machinery and Intelligence"

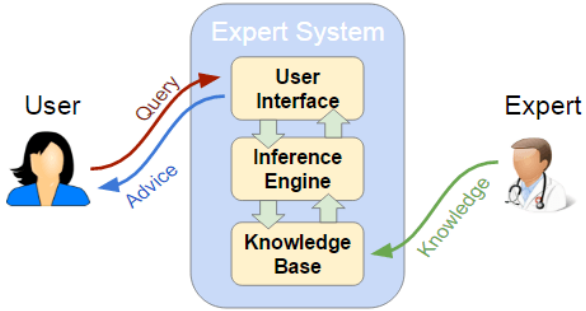


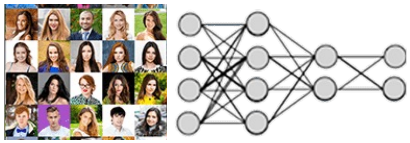
1950—1960s: Excitement: Look, Ma, no hands!

- 1950s: Early AI **programs**, including Samuel's checkers program, Newell & Simon's **Logic** Theorist, Gelernter's Geometry Engine
- 1956: Dartmouth meeting: "Artificial Intelligence" adopted

1970—80s: Knowledge-based approaches

- 1969—79: Early development of **knowledge**-based systems
- 1980—88: Expert systems industry booms
- 1988—93: Expert systems industry busts: "AI Winter"



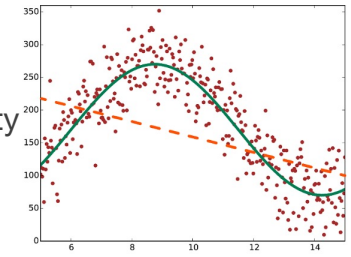


2012—: Deep learning

- 2012: ImageNet & AlexNet
- **Deep** Neural Networks

1990—: Statistical approaches

- **Data**-driven machine learning algorithms
- Resurgence of **probability**, focus on uncertainty
- General increase in technical depth
- Agents and learning systems… "AI Spring"?



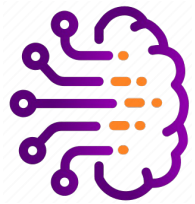
Artificial Intelligence vs. Machine Learning

Artificial Intelligence

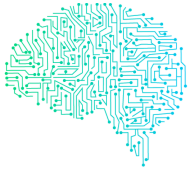
Computer systems that perform tasks that would usually require human intelligence

Machine Learning

Statistical techniques that learn from a series of inputs and outputs



Deep Learning



Machine learning algorithms that enable self learning of data representations

History of Machine Learning

1950-1960s

Logical Inference

The creation of "Machine Learning"
Turing test
Neural networks

1970-1980s

Knowledge Based Learning

Assign human knowledge to computer systems

1980-1990s

Inductive and Statistical Learning

Learning from examples
BP Neural Networks
Statistical Learning

2012-

Deep Learning

Deep Neural Networks
AlphaGo

Why Machine Learning?

Traditional Programming vs Machine Learning

Traditional Programming

Write a computer program with explicit instructions to follow

Machine Learning

Write a computer program to automatically learn from examples

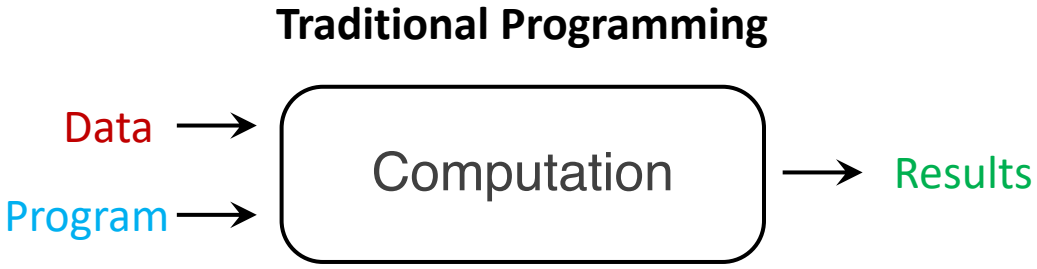
```
if email contains “发票”
  then mark is-spam;
if email contains …
if email contains …
```

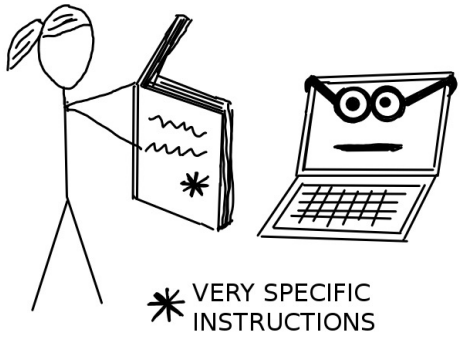
So many decision rules…

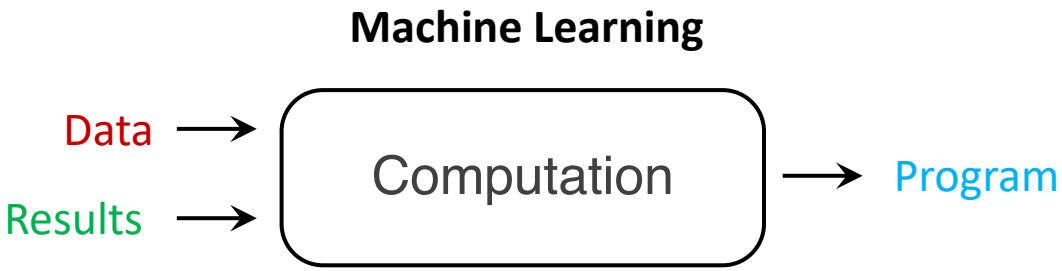
Magic!

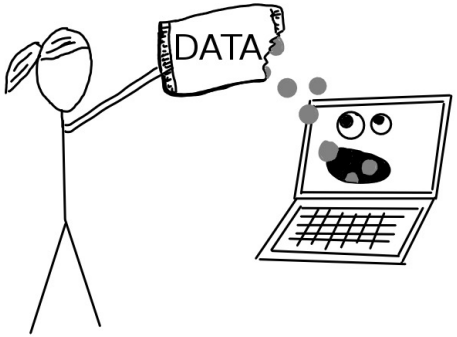
Machine Learning = “The ability to learn without being programmed”

Traditional Programming vs Machine Learning



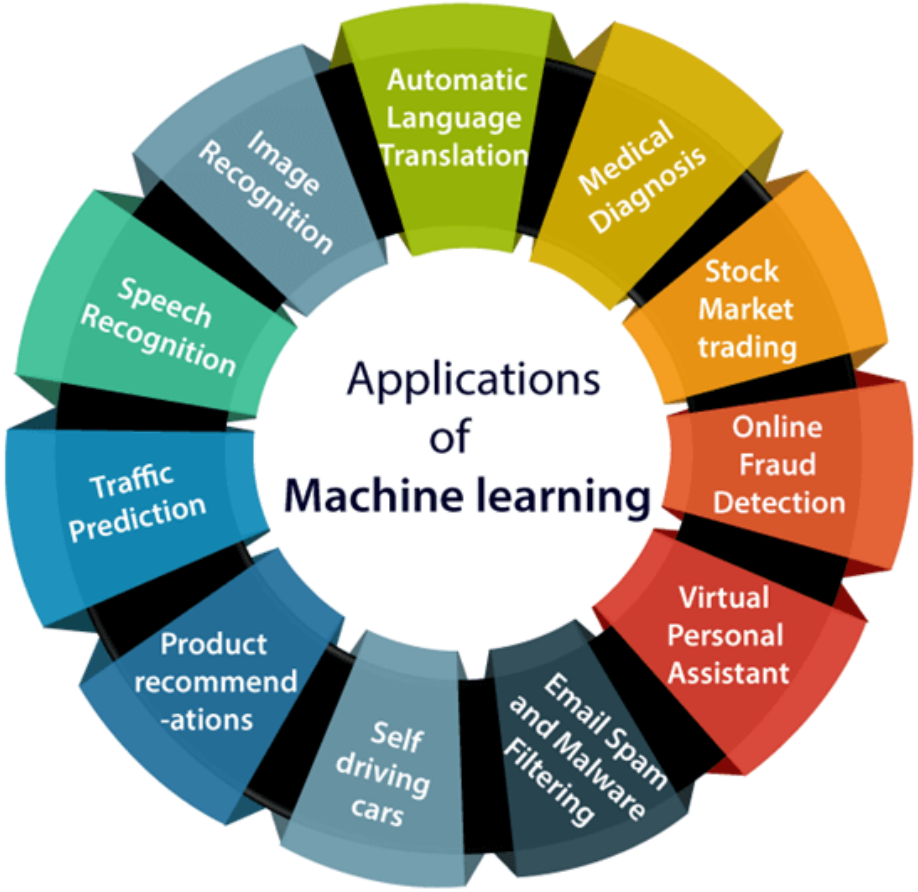






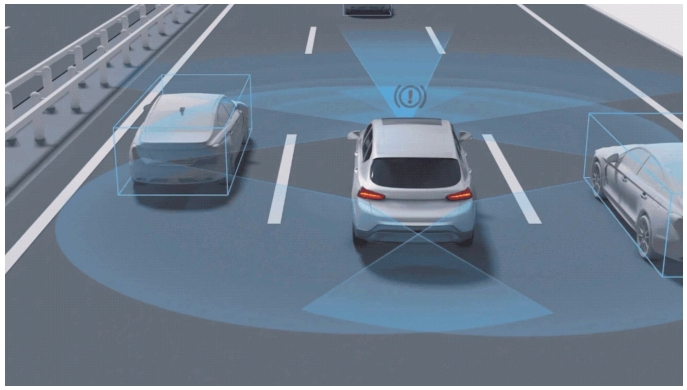
Machine Learning Applications

Machine learning is very widely used in:



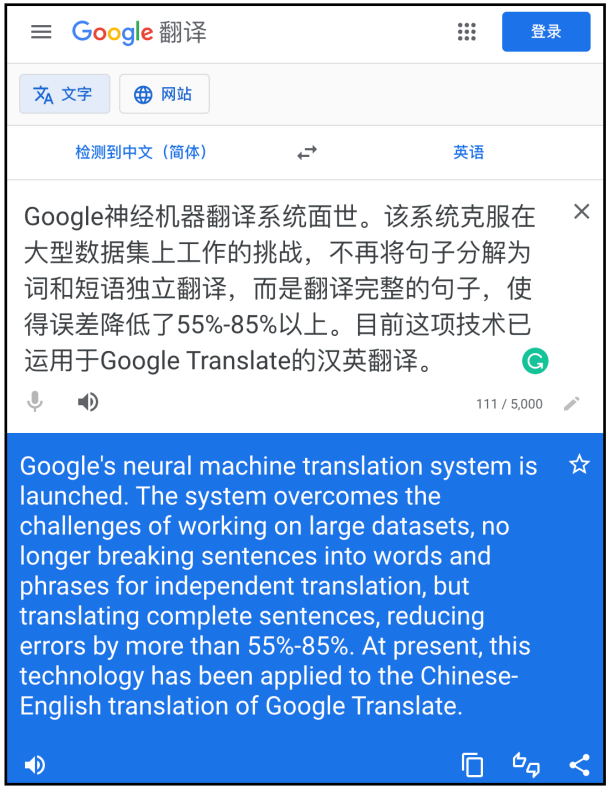
Machine Learning Applications

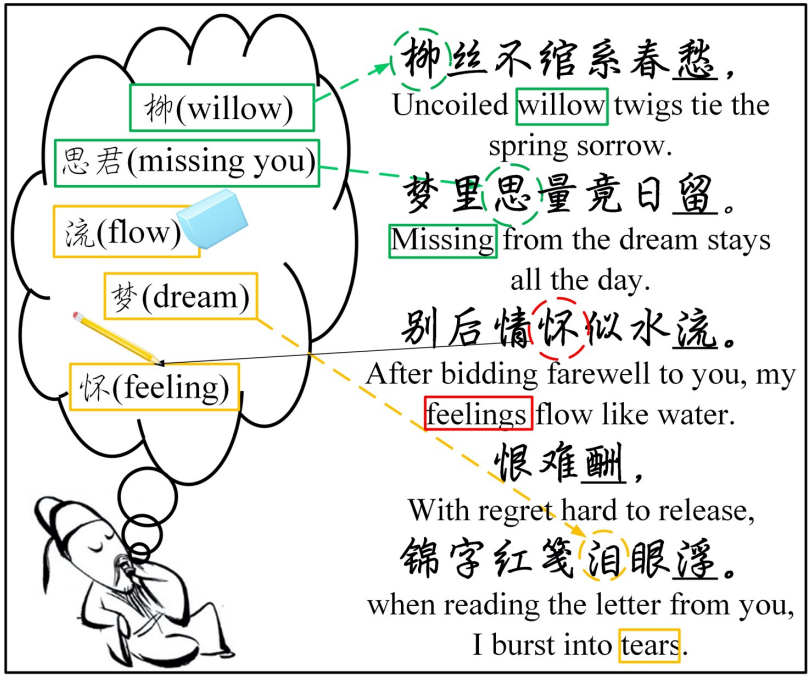
Computer Vision: object detection, semantic segmentation, diagnoses, self-driving, etc.



Machine Learning Applications

NLP: translation, QA, summary, article, etc.



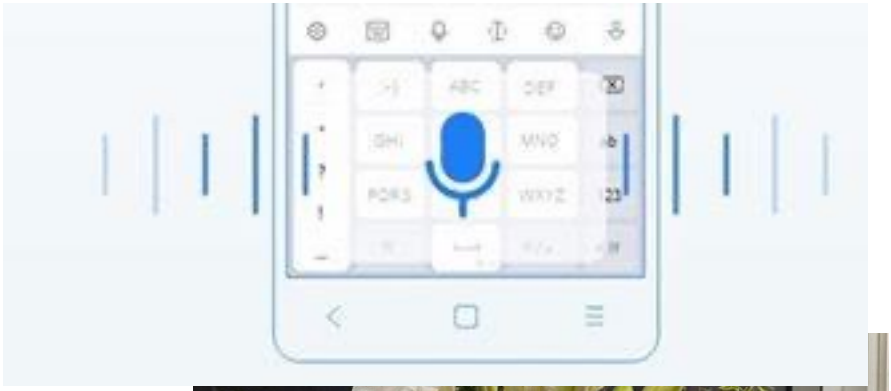


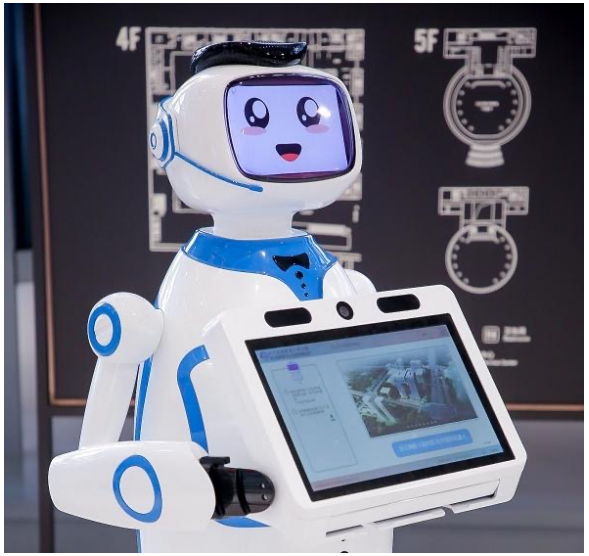
Chinese poem generation. Can you distinguish?

Yi et al. Chinese Poetry Generation with a Working Memory Model. IJCAI'18

Machine Learning Applications

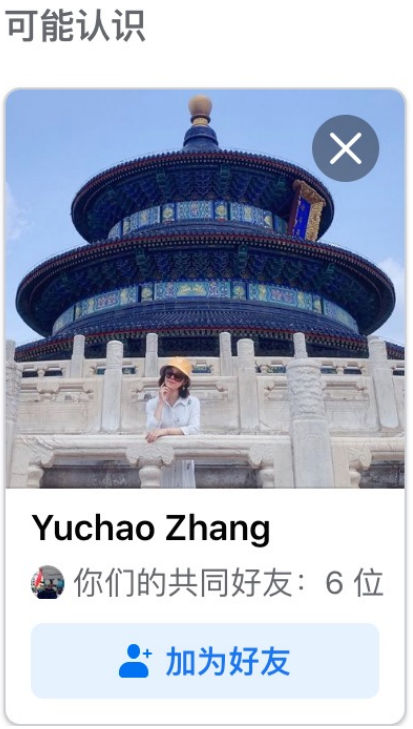
Speech Recognition & Conversational Assistants





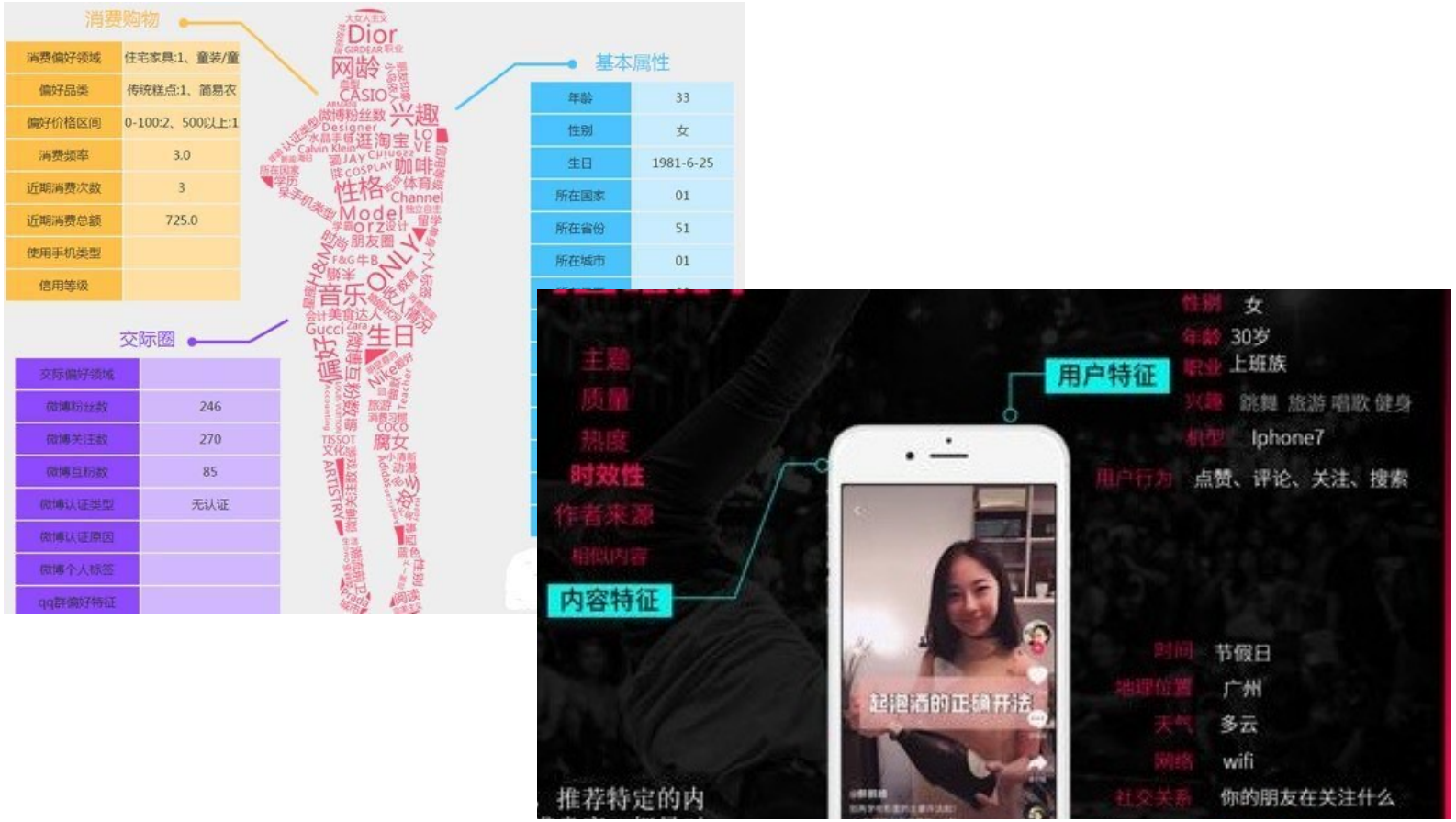
Machine Learning Applications

Recommender Systems



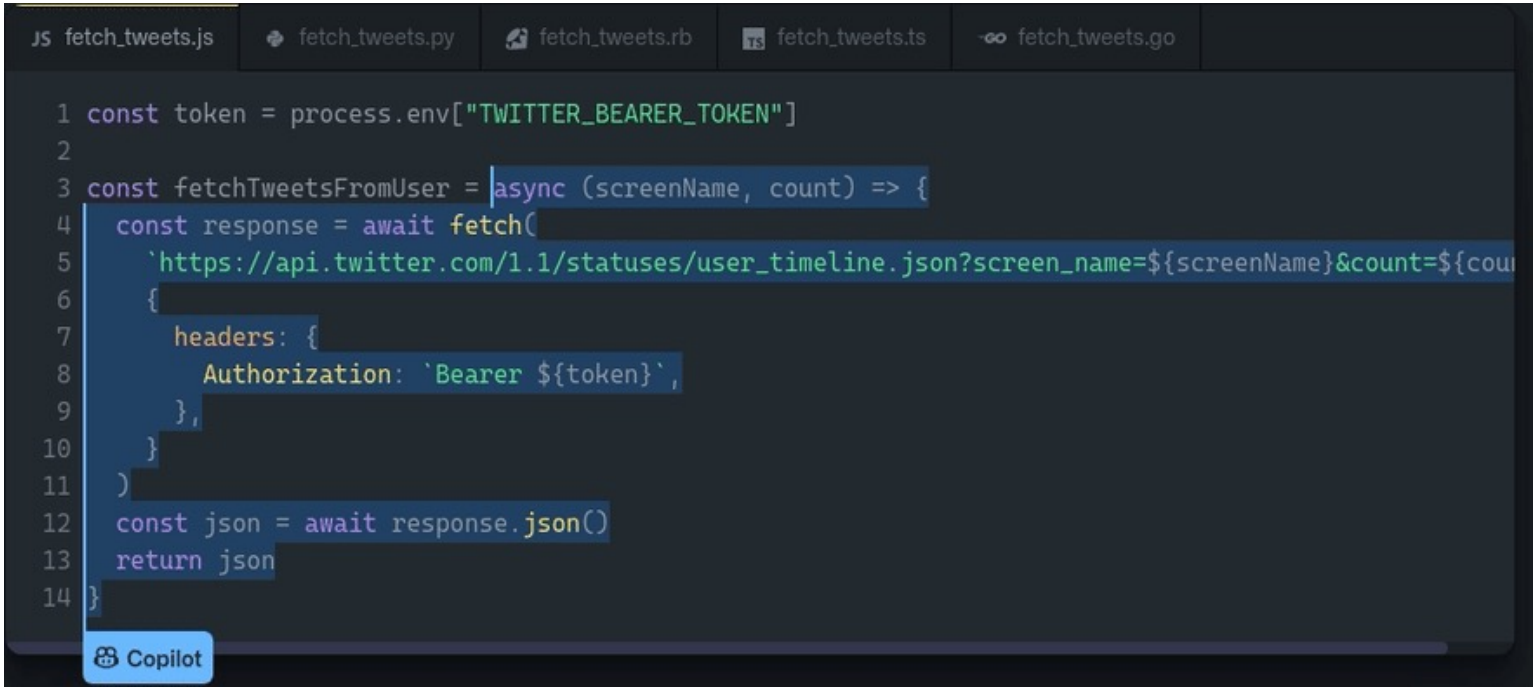
Machine Learning Applications

User Profiling



Machine Learning Applications

AI Programmer



Machine Learning Applications

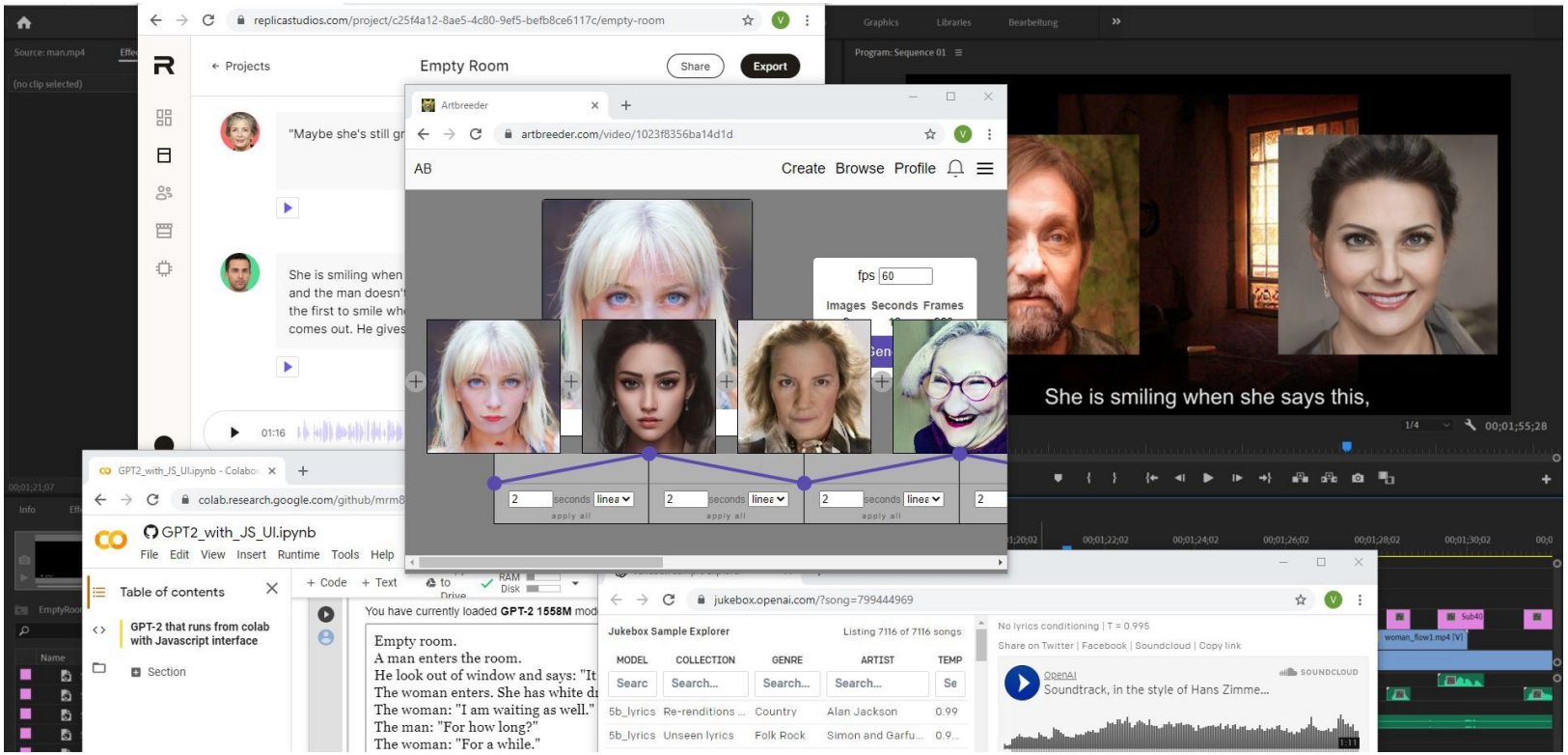
AI Painting

https://openai.com/dall-e-2/

Original: *Girl With a Pearl Earring* by Johannes Vermeer

Machine Learning Applications

AI Movie Maker



英伟达真假发布会：
https://www.bilibili.com/video/BV1kb4y1U7Re?spm_id_from=333.337.search-card.all.click

What Is Machine Learning?

What Is Machine Learning?

Human Learning

Knowledge Imparting

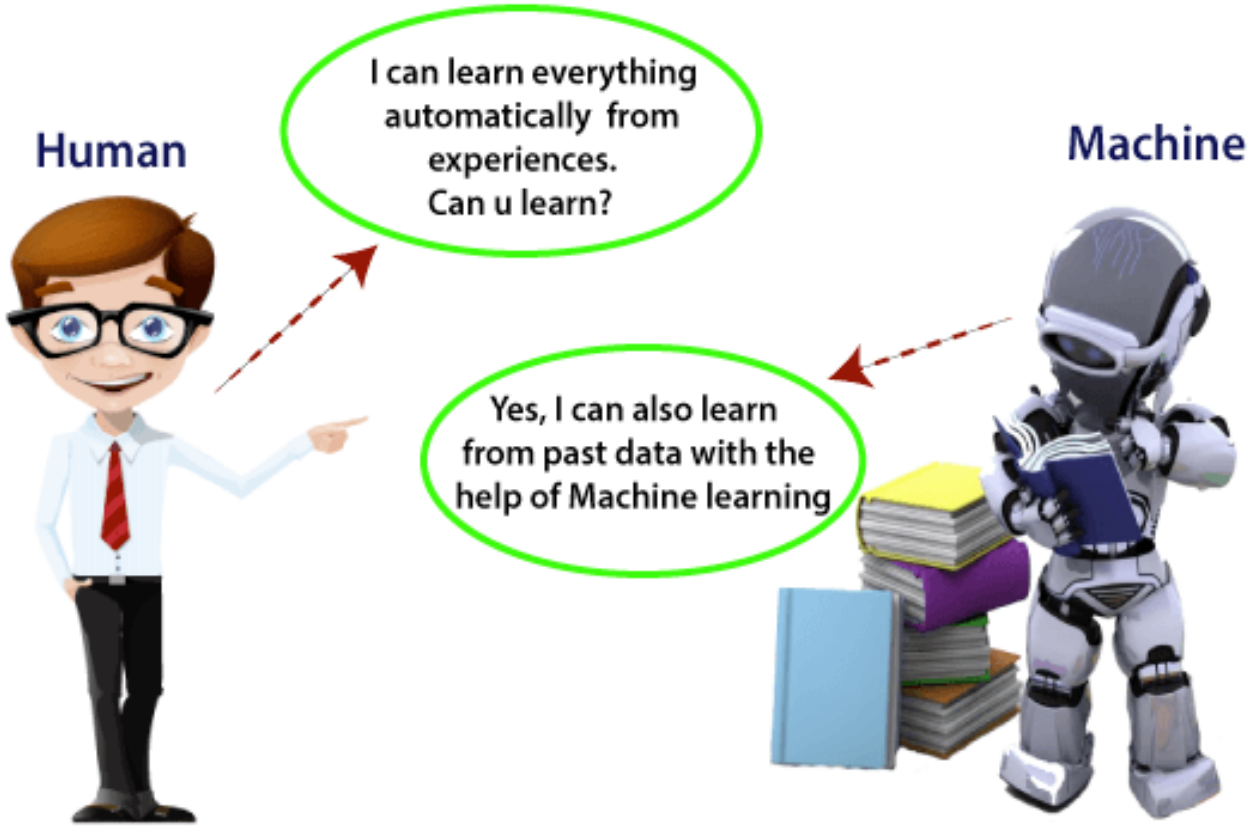
Practice

Exam

Summarize &Amends

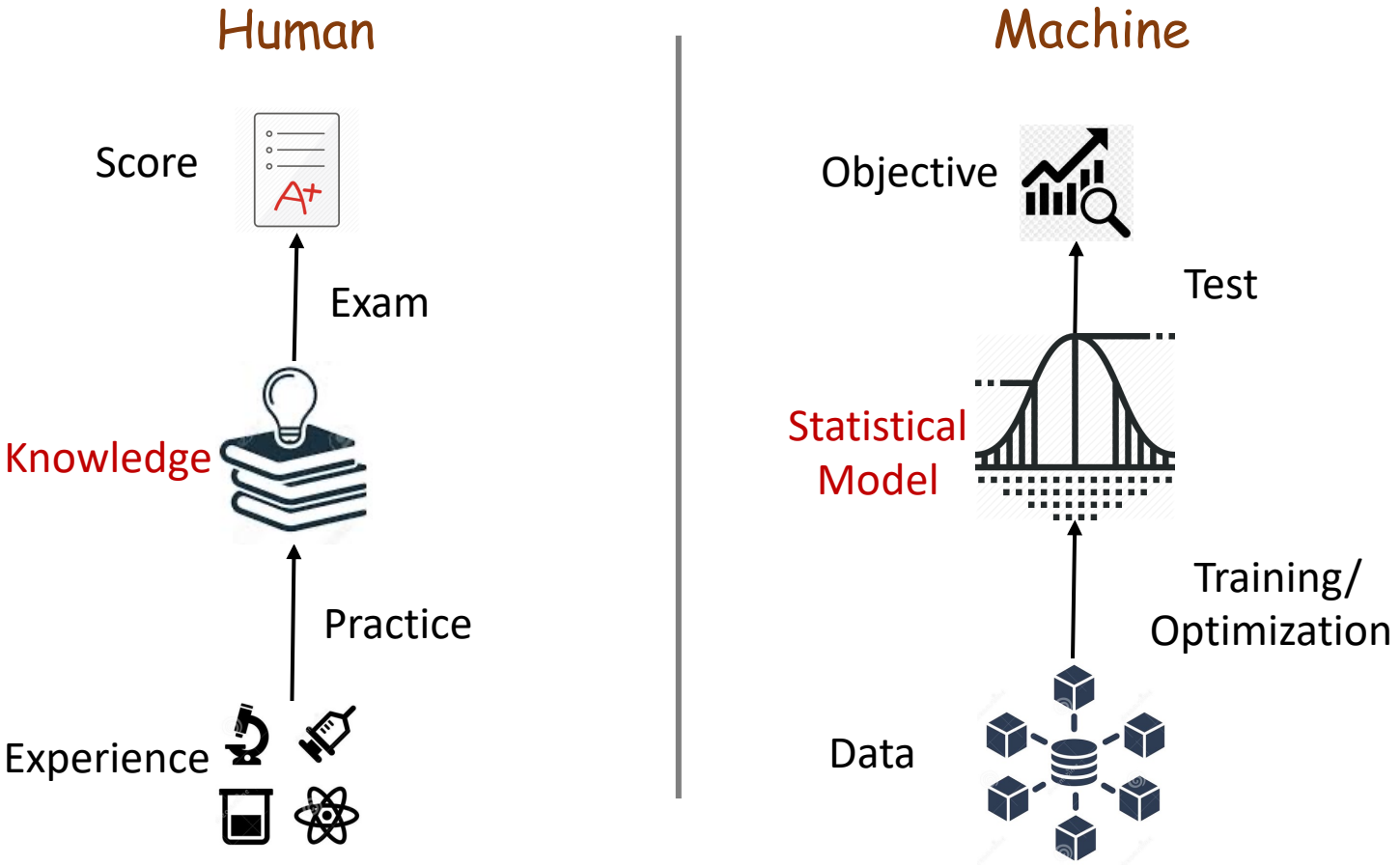
A Simple Analogy

Human Learning vs. Machine Learning



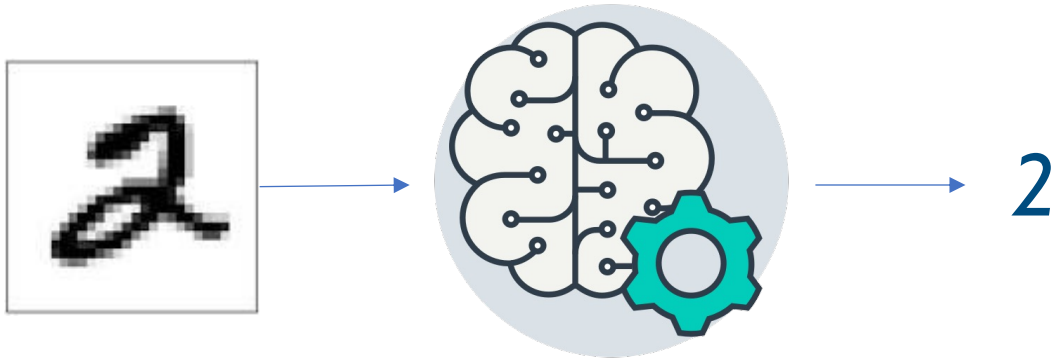
A Simple Analogy

Human Learning vs. Machine Learning

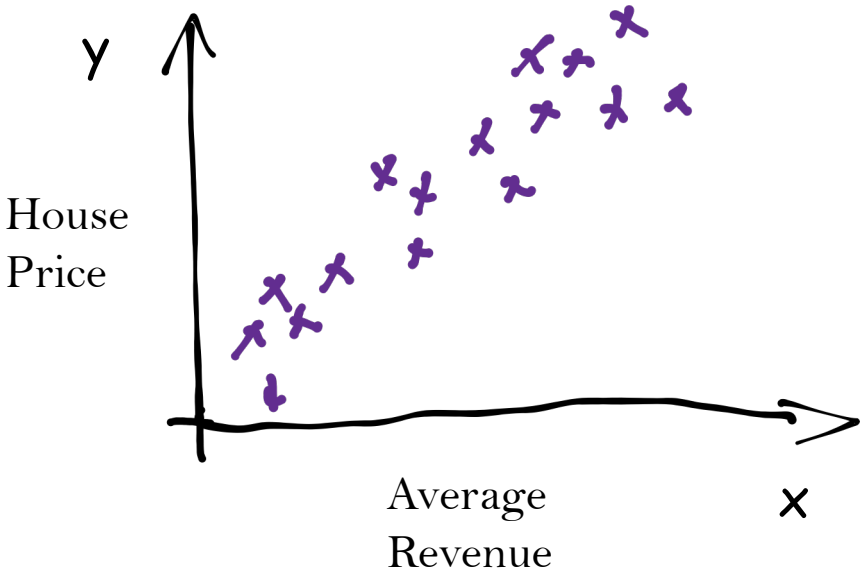


So, what exactly is machine learning?

A subset of artificial intelligence which uses **statistical methods** to enable machine to **improve** towards an **objective** using **data** without requiring explicit programming by human.



Key Elements of Machine Learning



Elements:

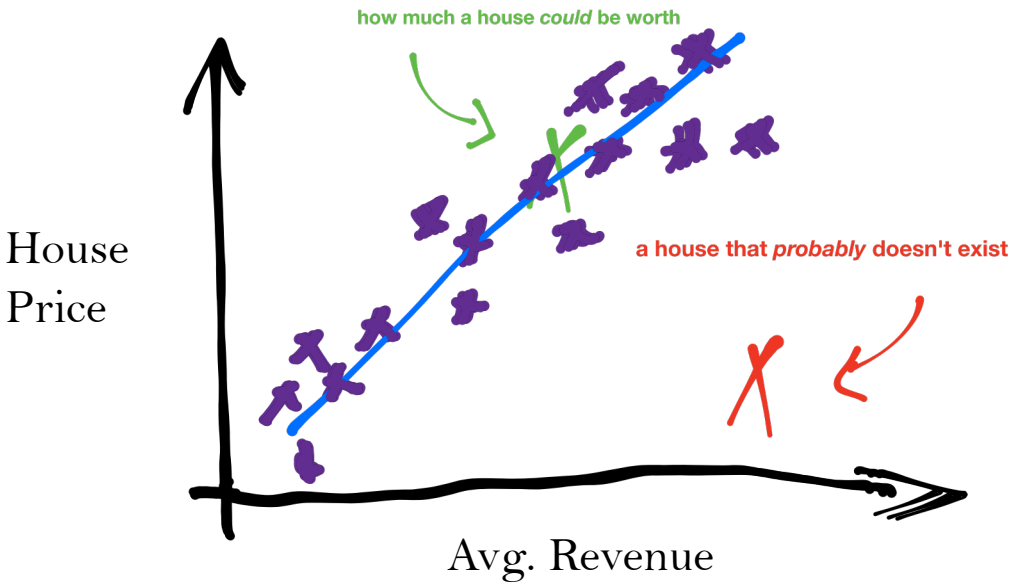
#1 Data (Experience)

$D = \{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \ldots, (x^{(N)}, y^{(N)})\}$

Key Elements of Machine Learning

We **cannot** learn from data

We **can** learn from data + *hypothesis*



What hypothesis do we make about this data?

Elements:

#1 Data (Experience)

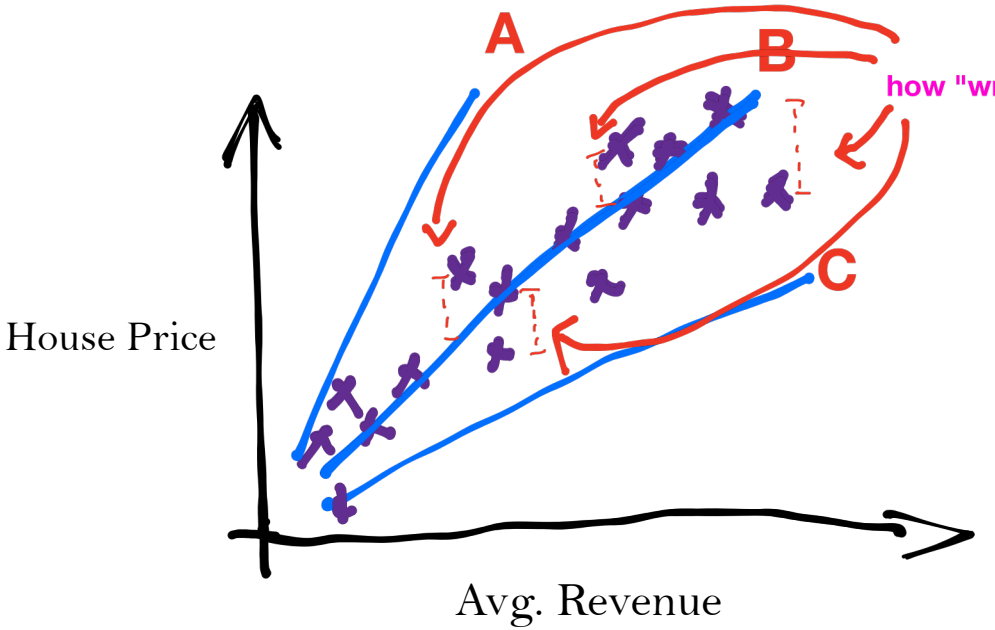
#2 Model(Hypothesis)

$$\hat{y} = g(x|\theta) + \epsilon$$

e.g. $\hat{y} = \boldsymbol{w}x + \boldsymbol{b}$

where θ (e.g., W and b) denotes the parameters

Key Elements of Machine Learning



Elements:

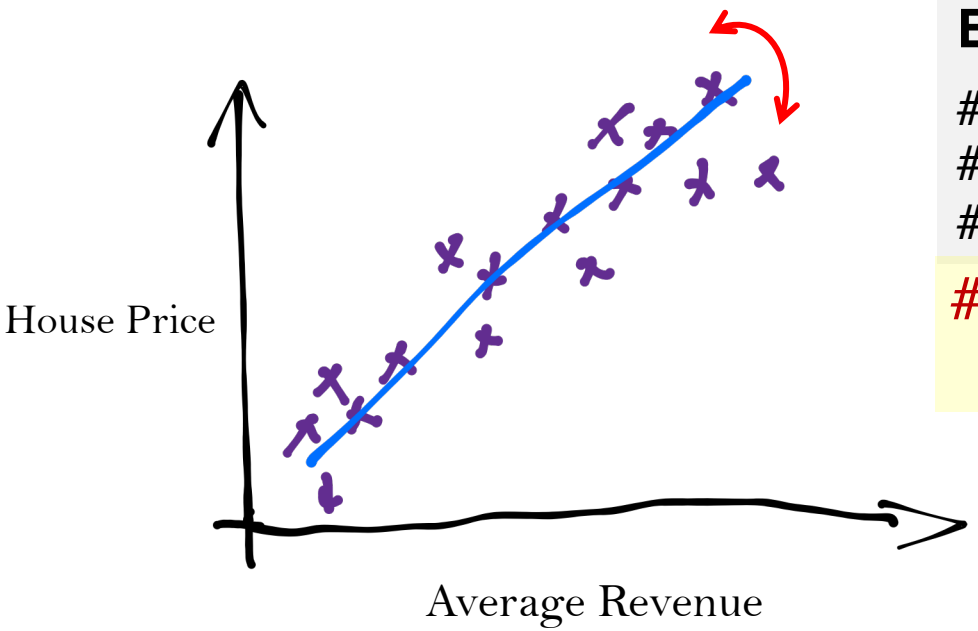
#1 Data (Experience)
#2 Model (Hypothesis)

#3 Loss Function (Objective)

$$\mathcal{L}(\theta|\mathcal{D}) = \frac{1}{2}\sum_{n=1}^{N}[wx_n + b - y_n]^2$$

How to evaluate a model? (performance criterion)

Key Elements of Machine Learning

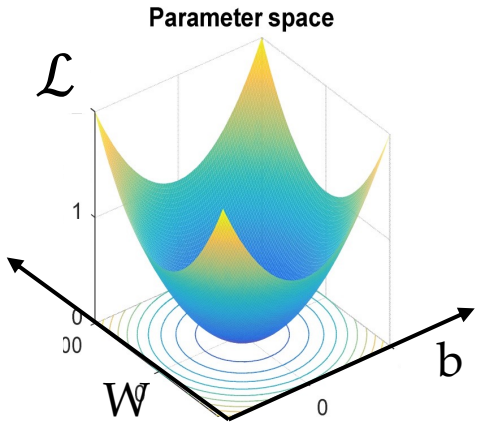


Elements:

#1 Data (Experience)
#2 Model (Hypothesis)
#3 Loss Function (Objective)

#4 Optimization Algorithm
(Improvement)

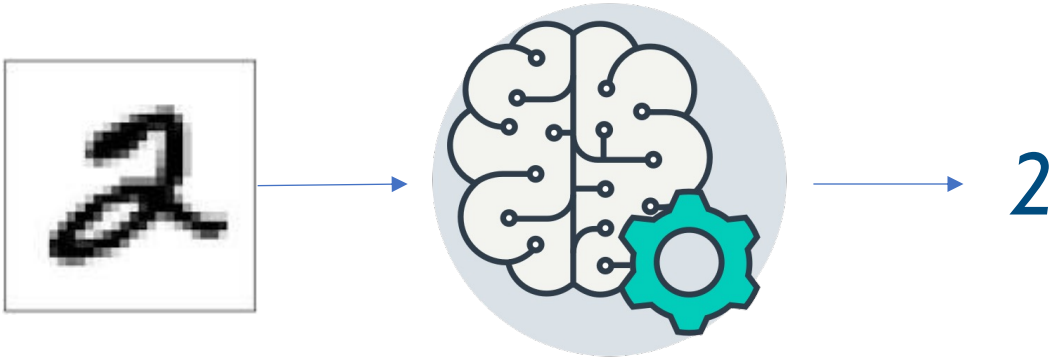
$$\theta^* = \underset{\theta}{\operatorname{argmin}}\, \mathcal{L}(\theta|\mathcal{D})$$



How to find the optimal model?

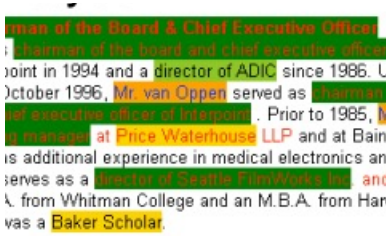
So, what exactly is machine learning ?

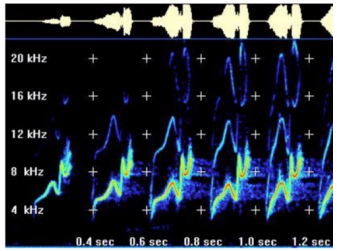
A subset of artificial intelligence which uses statistical methods (model) to enable machine to improve (optimization) towards an objective (loss function) using (data) without requiring explicit programming by human.

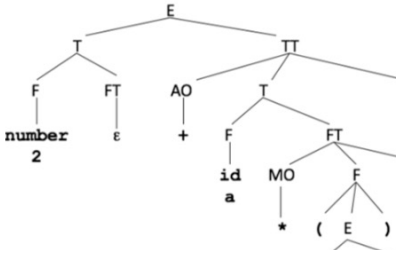


Generalized to other data

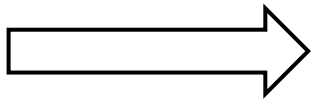
What about other data structures?

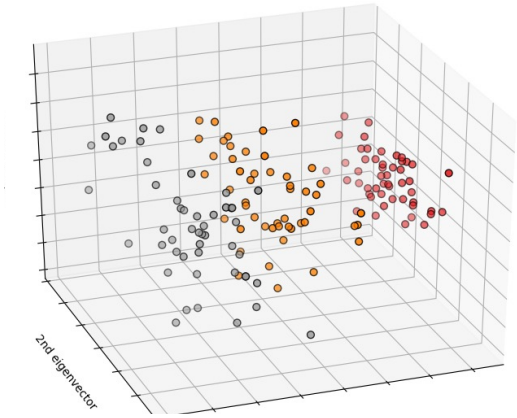






Vectorization/
Feature Extraction

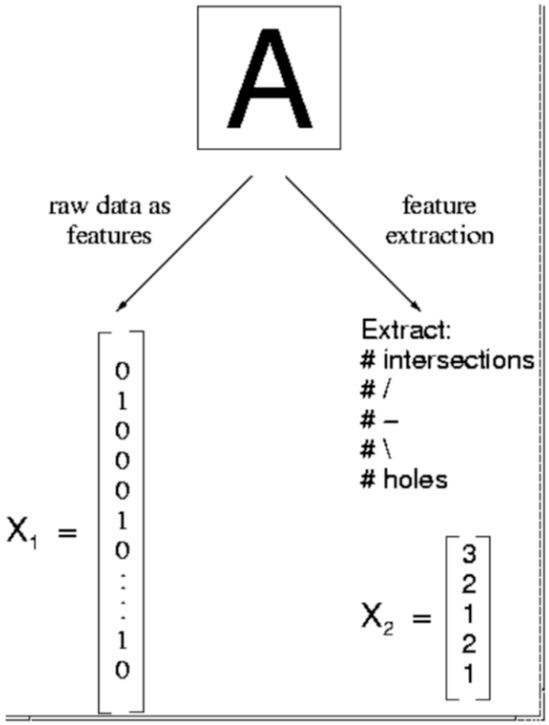


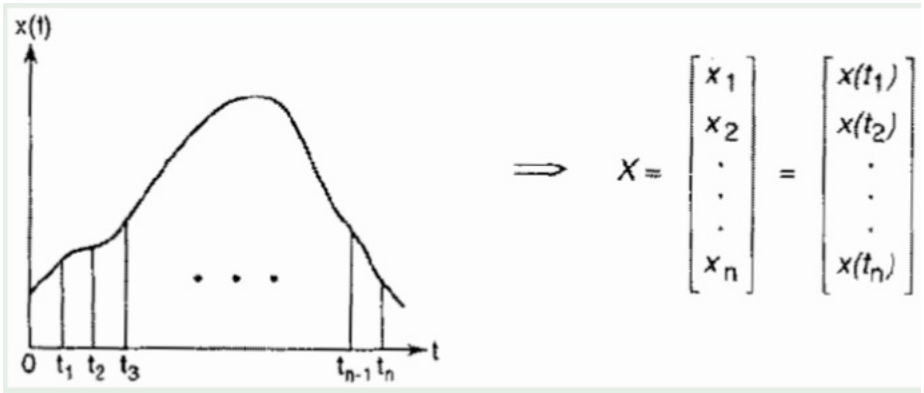


Data Points in High-Dimensional Vector Space

Generalized to other data

Vectorization/Feature Extraction





Quiz

以下人类学习行为分别对应哪种机器学习要素？

1. 查文献
2. 考试
3. 错题总结
4. 挂科

How machine learning works?

"Hello-World" for machine learning

Recall: Key Elements of Machine Learning

- **Data** (Experience)
- **Model** (Hypothesis)
- **Loss Function** (Objective)
- **Optimization Algorithm** (Improve)

How to Approach a Machine Learning Problem?

① consider the nature of available data **D**

- how much amount of data can you obtain? how would it cost (in time, computation, human efforts)?

② select a representation for the input **X**

- data preprocessing, feature extraction, etc.

③ choose a set of possible models **H** (hypothesis space)

- set of functions h: X → Y

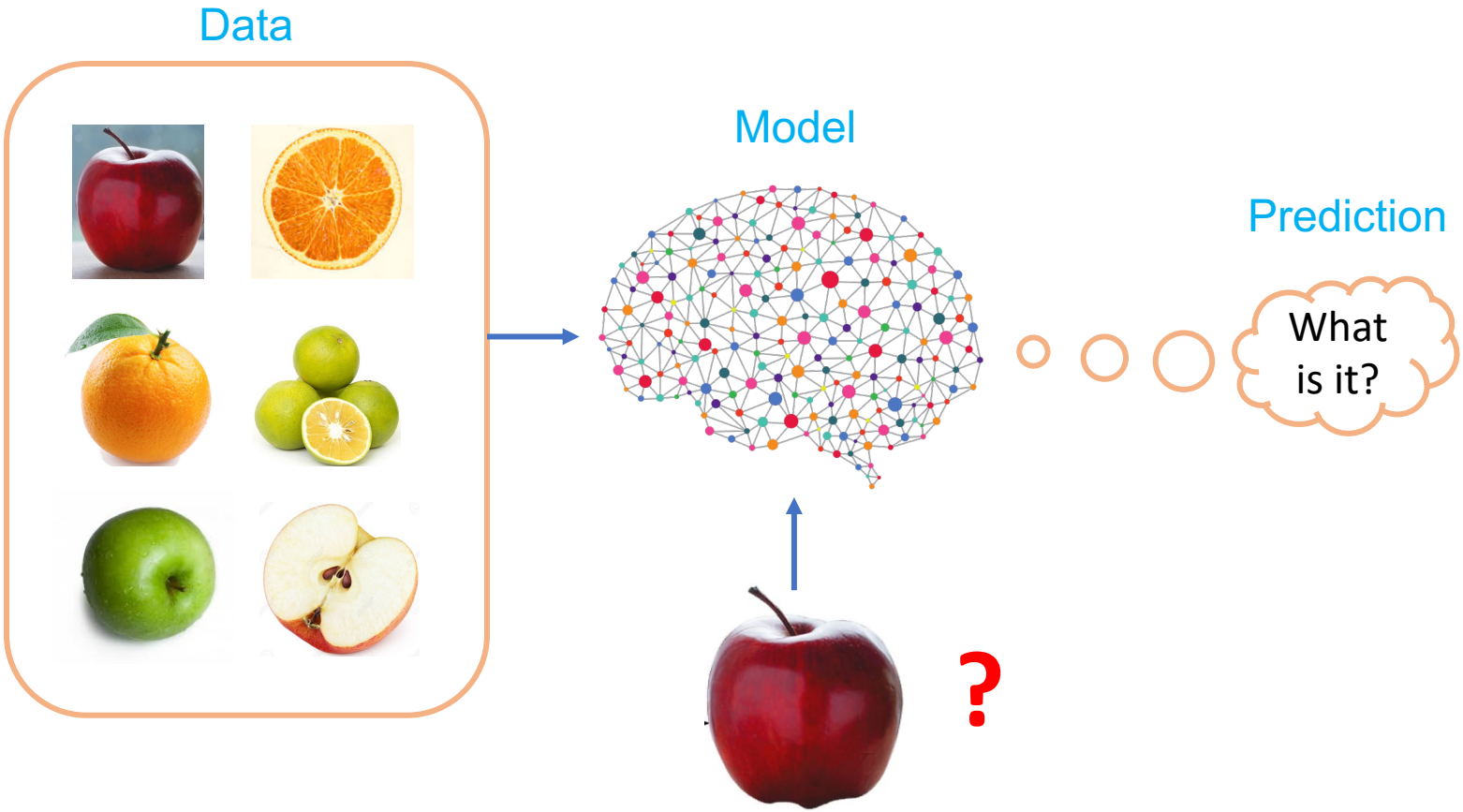
④ choose the performance measure **P** (loss function)

⑤ choose or design a learning algorithm

- for using examples (**E**) to converge on a member of **H** that optimizes **P**

A Working Example

Task: Automatic Fruit Classification



Data Collection

Create training samples

- Tell me which type a specific fruit belongs to

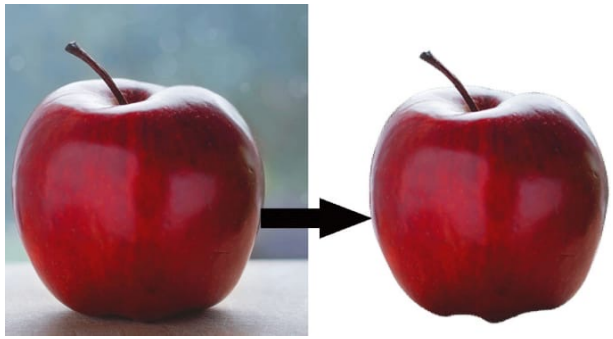
Sample	Label
	Apple
	Orange
	Apple
...	...

Pre-processing

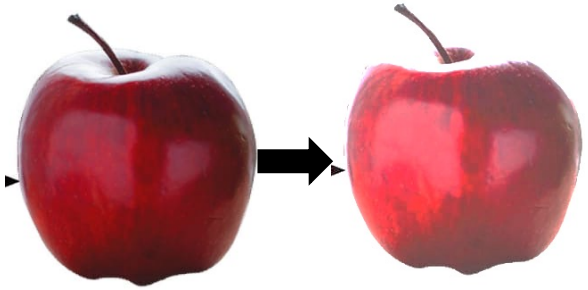
Different types of fruit differ in size, lightness, position, etc.

Needs pre-processing!

Example:



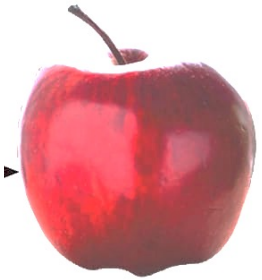
remove the background



adjust for light level

Feature Extraction

Consider each fruit as a point in some feature space



Feature:

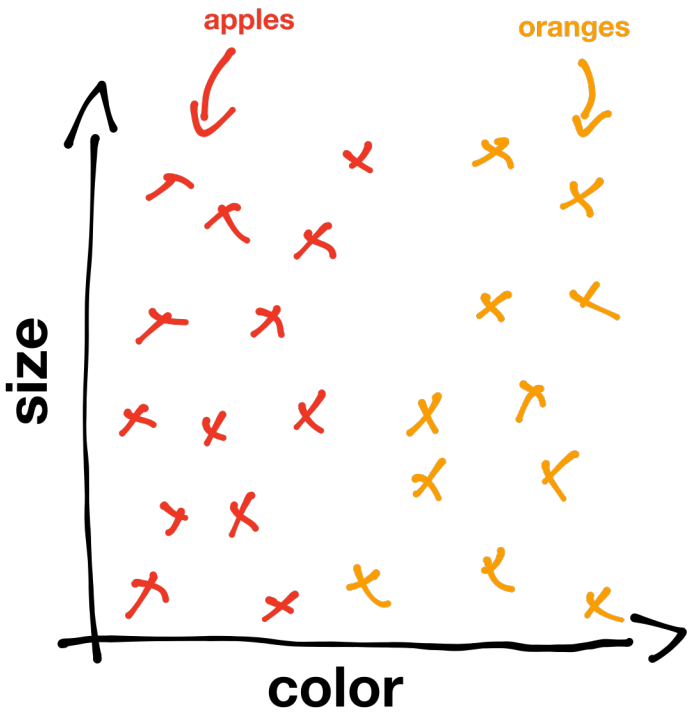
Color: red
Shape: round
Size: 3
...

$$\begin{bmatrix} 1 \\ 0 \\ 3 \end{bmatrix}$$

Feature:

Color: orange
Shape: round
Size: 2
...

$$\begin{bmatrix} 2 \\ 0 \\ 2 \end{bmatrix}$$



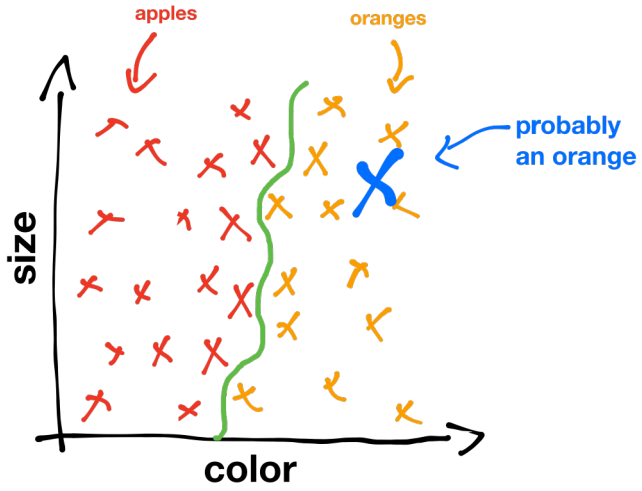
Learning (Training)

Partition the feature space into 2 regions, one for each type of fruit

- decision boundary

Test and Evaluation

How to handle new images?



Classifying a new image using the trained classifier .

How to measure the classifier performance?

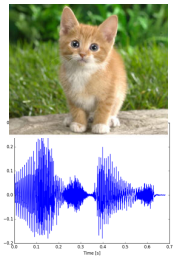
- classification error rate
  - % patterns that are assigned to the wrong category
- other aspects may be important too
  - e.g. computational complexity
  - e.g. user-friendliness

Pipeline of an ML System

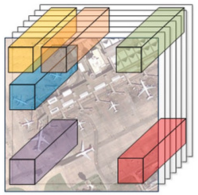
Data Preparation

Training

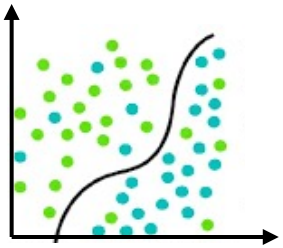
Evaluation



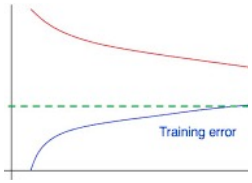
Data Collection



Feature Extraction

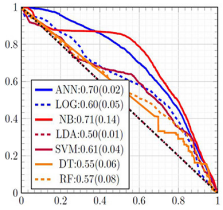


Train model



Validation

Prediction

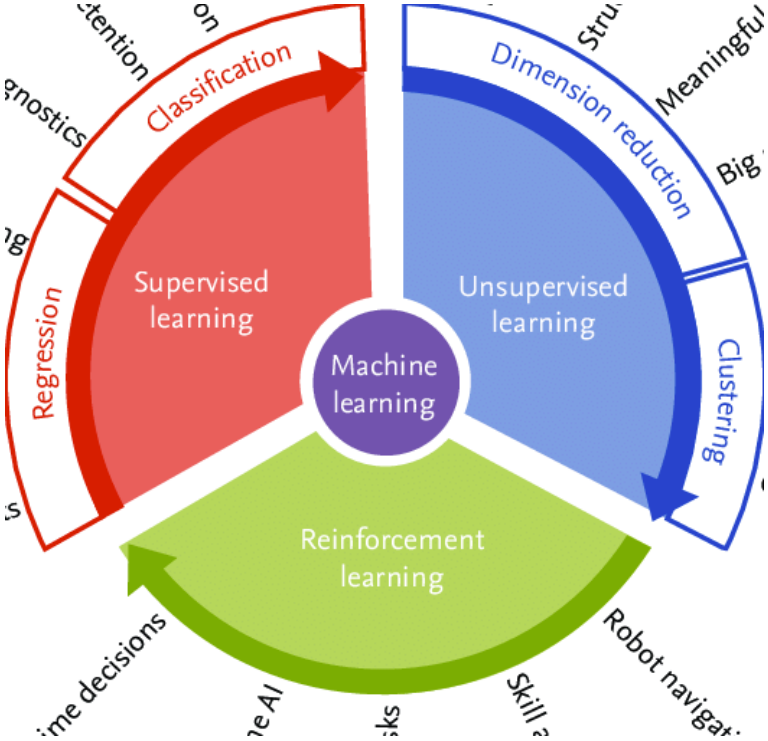


Evaluation

Overview of Machine Learning Algorithms

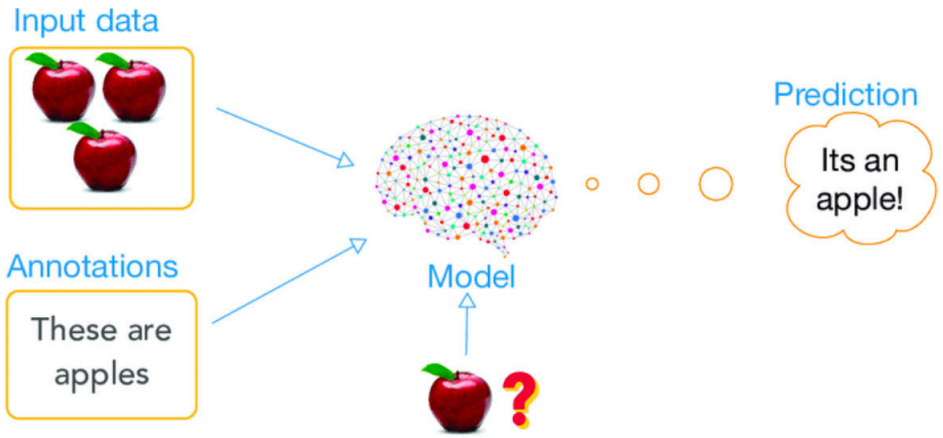
Categories of Machine Learning Algorithms

- **Supervised Learning**
- **Unsupervised Learning**
- **Reinforcement Learning**



Supervised Learning

- The learner is provided with a set of inputs together with the corresponding desired outputs.



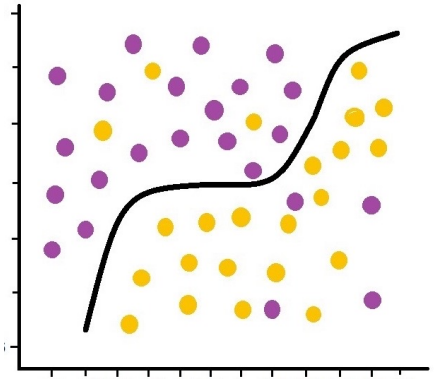
- has a "teacher"

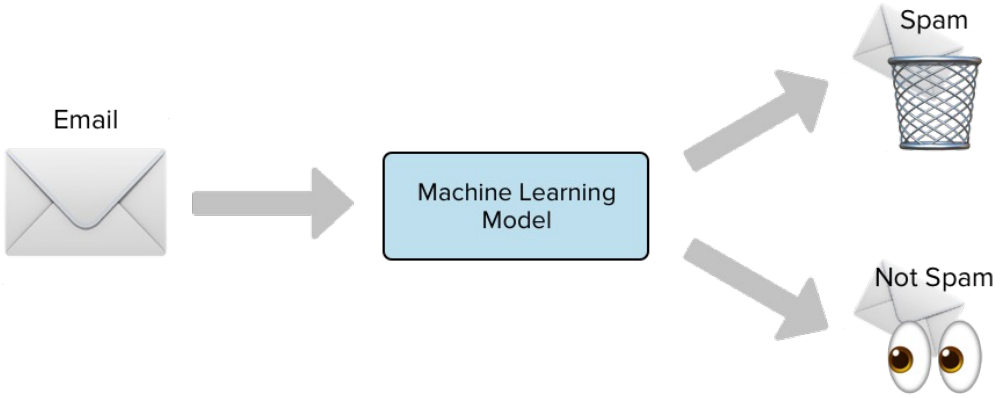
Example

- teaching kids to recognize different animals.
- graded examinations with correct answers provided.

Supervised Learning

Classification

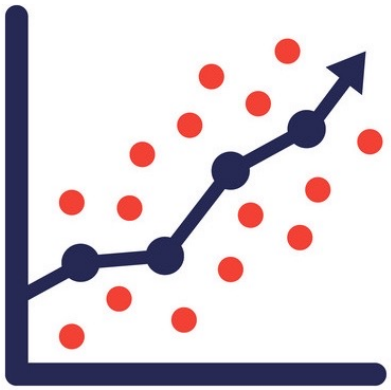


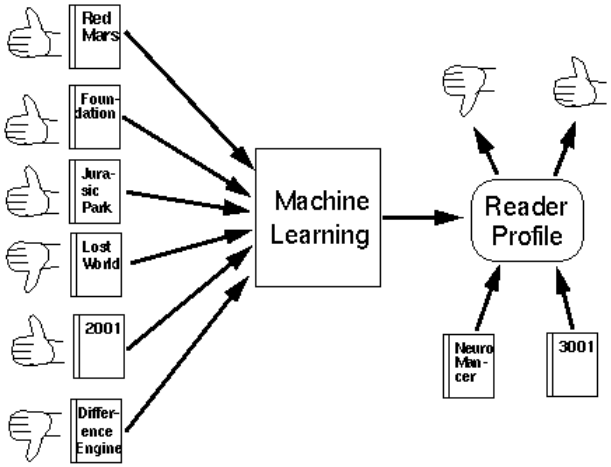


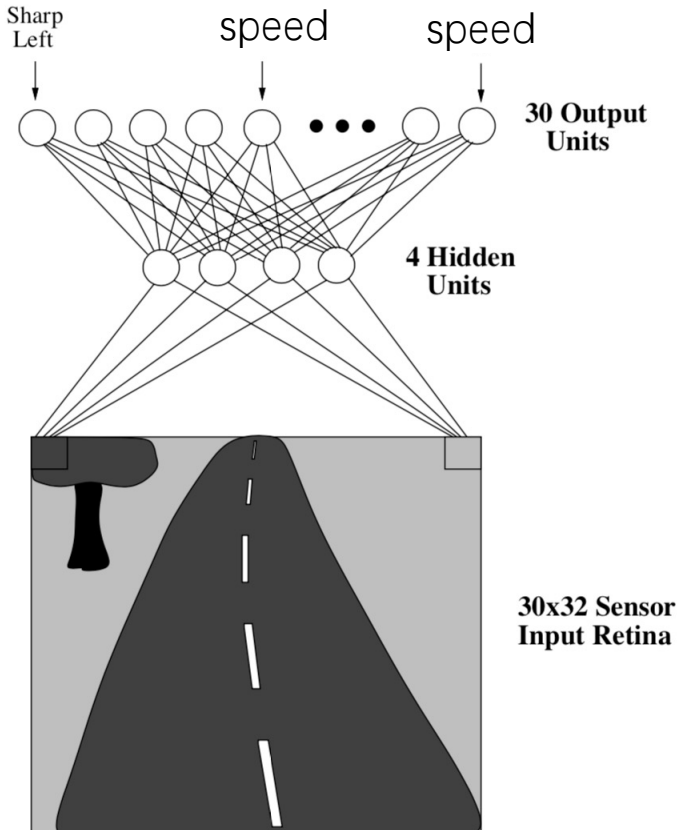
Sentiment analysis

Supervised Learning

Regression

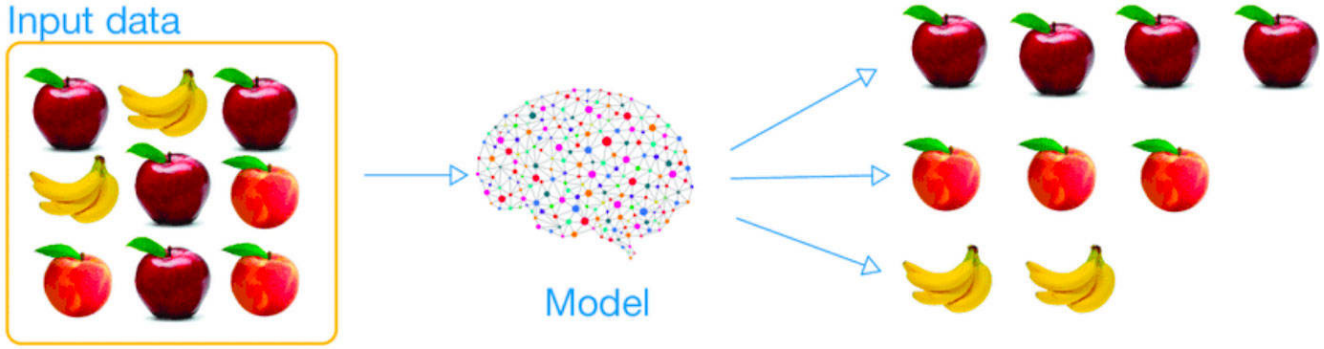






Unsupervised Learning

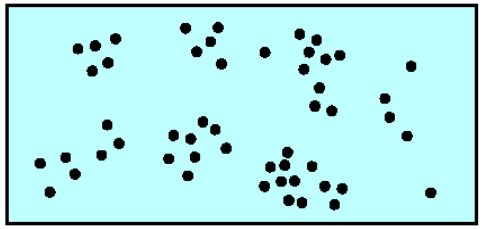
Training examples as input patterns, with no associated output.



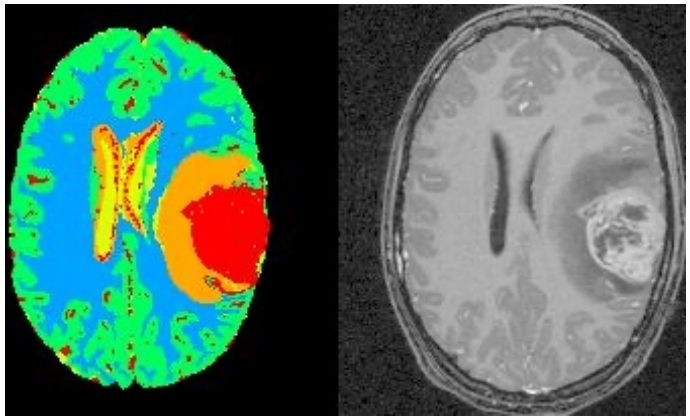
- no teacher
- similarity measure exists to detect groupings / clusterings

Unsupervised Learning

Clustering



- in the early stages of an investigation, it may be helpful to perform exploratory data analysis to gain some insight into the nature or structure of the data.

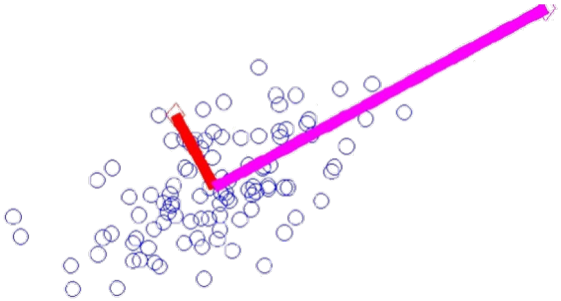


Unsupervised Learning

Dimensionality Reduction and Feature Selection

- find features or preprocess existing features for the subsequent pattern classification problem (supervised learning)

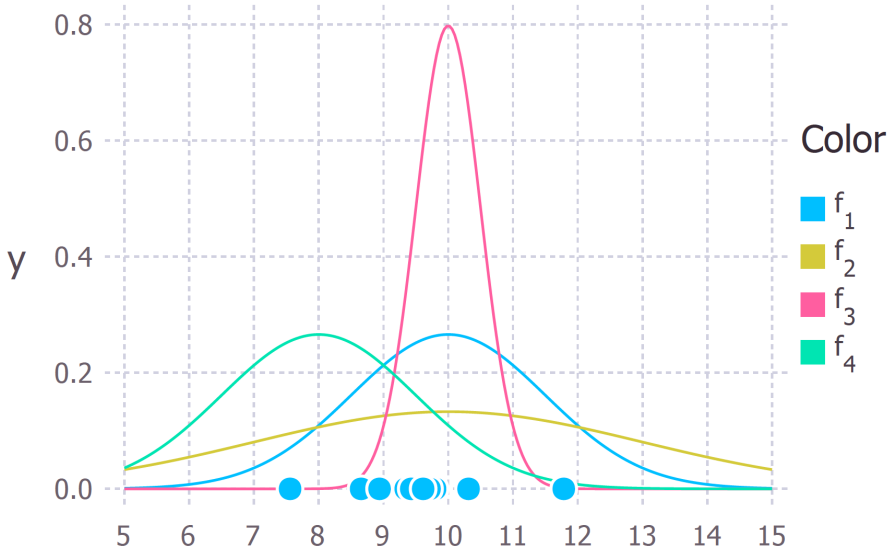
Principal component analysis (PCA)



Example (eigenface)

Unsupervised Learning

Probability Estimation



Applications

- Density estimation
- Data generation
- Abnormal Detection

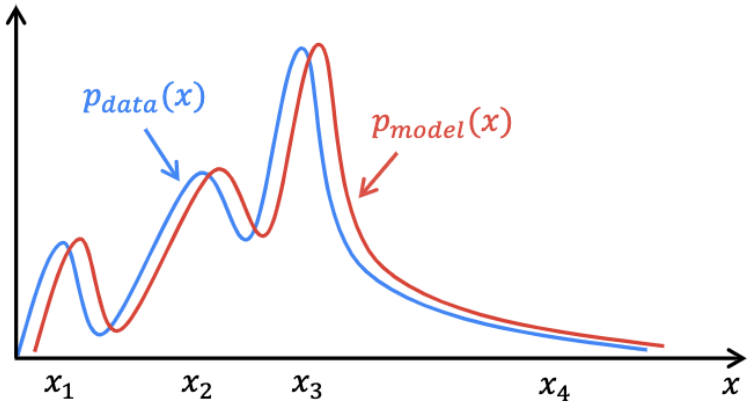
Unsupervised Learning

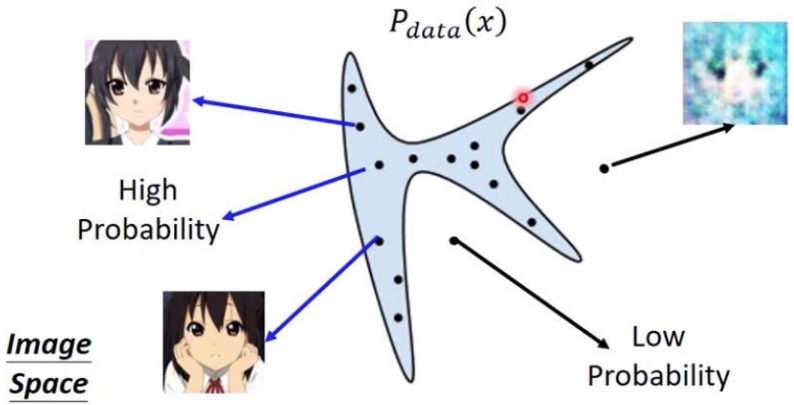
Data Generation

Distribution of images generated by the model

The goal of the generative model is to find a $p_{model}(x)$ that approximates $p_{data}(x)$ well.

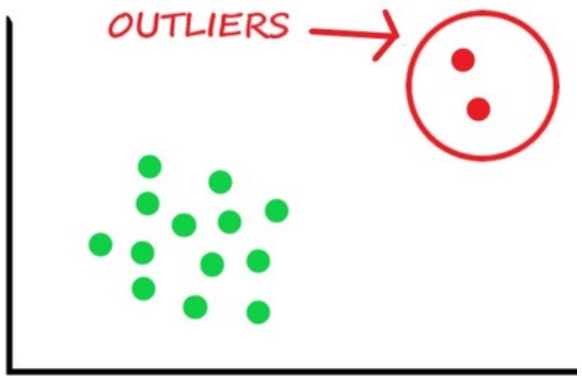
Distribution of actual images





Unsupervised Learning

Outlier Detection – find the least likely observations from a dataset.



Example: network intrusion detection

NETWORK SECURITY FUNCTION	INGRESS	EGRESS	EAST-WEST
Web Application Firewall (WAF)	✔	-	**
Intrusion Detection/ Prevention (IDS/IPS)	✔	✔	✔
AntiVirus Detection/ Blocking	✔	✔	**
URL/FQDN Filtering (includes Explicit and Category based profiles)	-	✔	✔
Data Loss Prevention (DLP)	-	✔	✔
Layer7 DoS	✔	-	✔
Malicious IP Blocking	✔	✔	-
GeoIP Blocking	✔	-	**
Threat Packet Captures	✔	✔	✔

Supervised vs. Unsupervised Learning

Supervised Learning

Data: (x, y)
x is data, y is label

Goal: Learn function to map
$x \rightarrow y$

Examples: Classification, regression, object detection, semantic segmentation, etc.

Unsupervised Learning

Data: x
x is data, no label

Goal: Learn the hidden or underlying structure of the data.

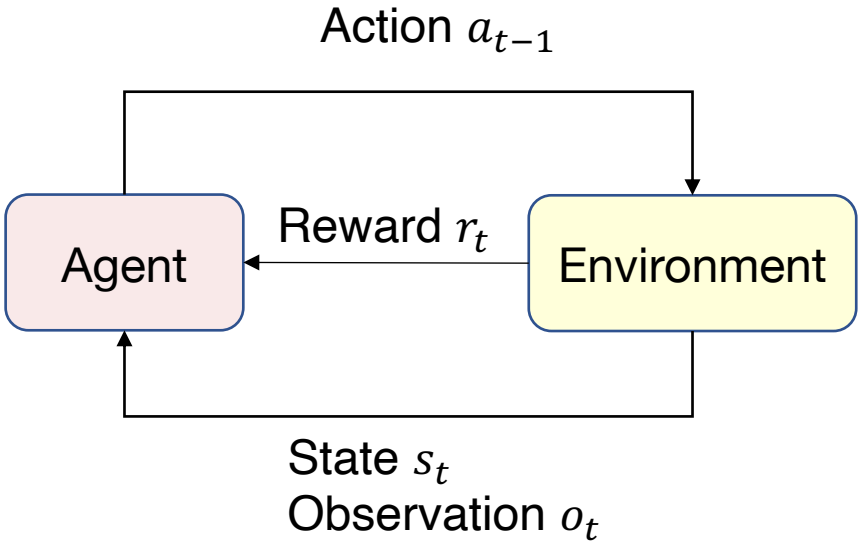
Examples: Clustering, dimensionality reduction, probability estimation, etc.

Reinforcement Learning

- Learning from **interacting** with an **environment.**

pain

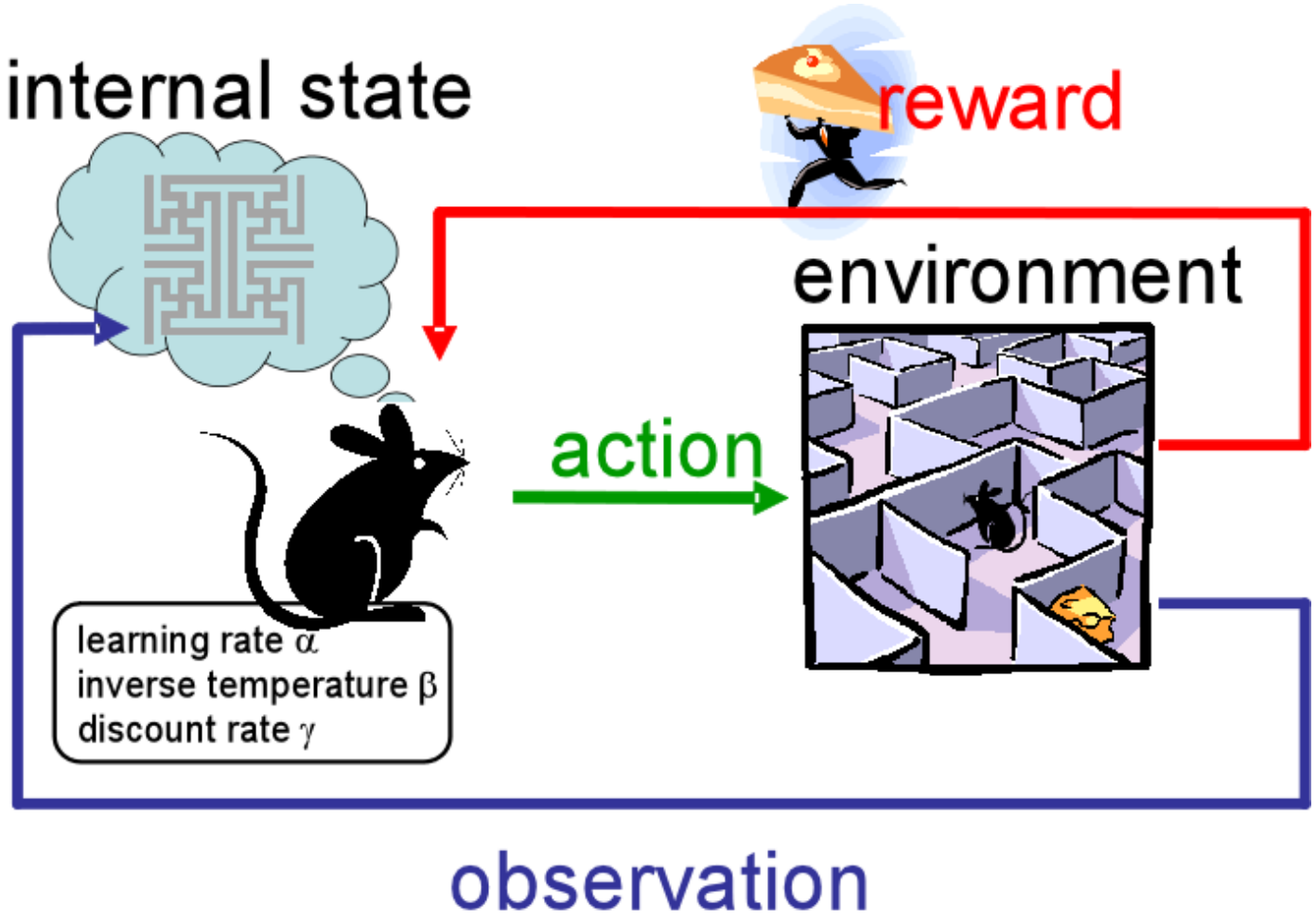
compliment



Learning a mapping from states to actions to maximize long-term reward.

Example: graded examinations with only overall scores but no correct answers

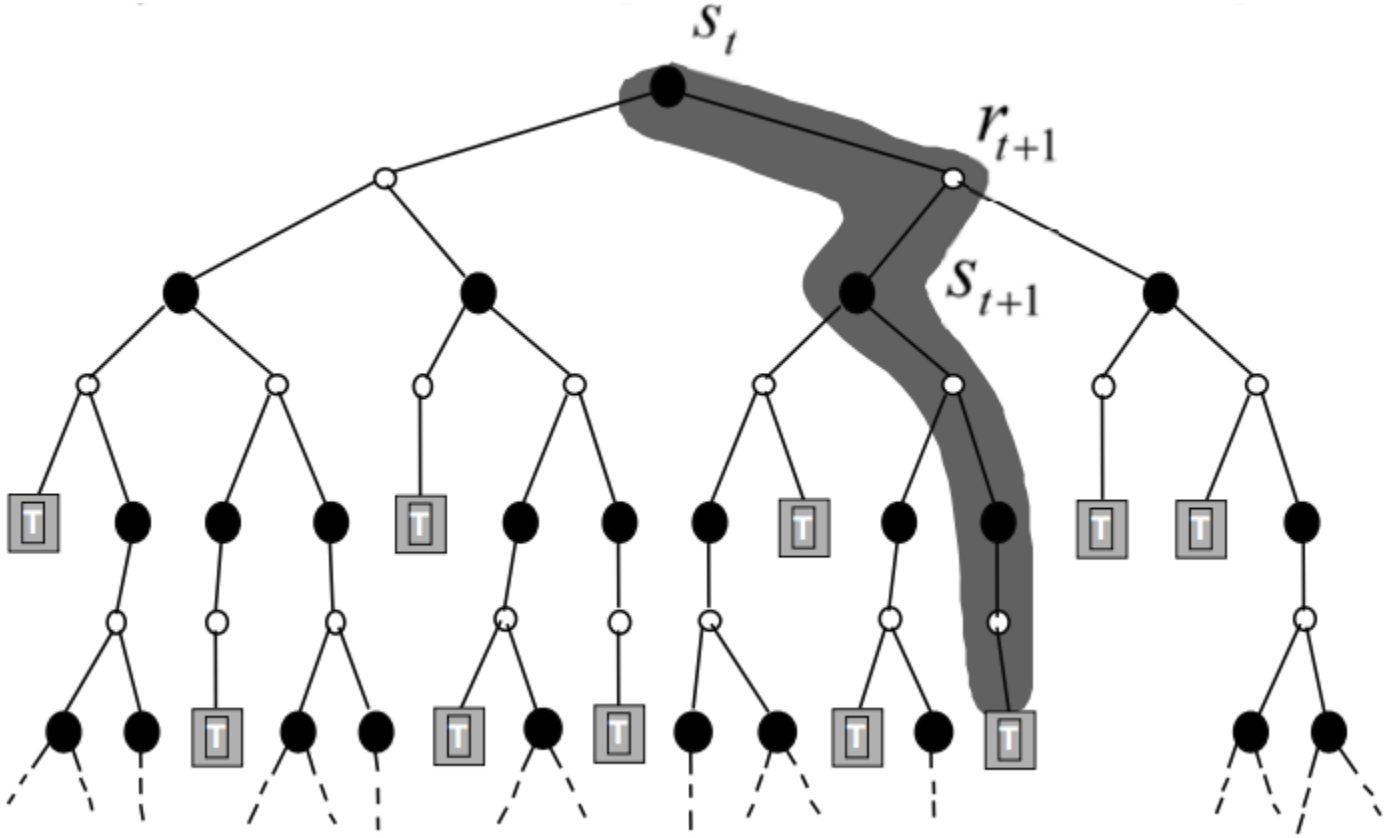
Example


internal state
reward
environment
action
learning rate α
inverse temperature β
discount rate γ
observation

Monte-Carlo Reinforcement Learning

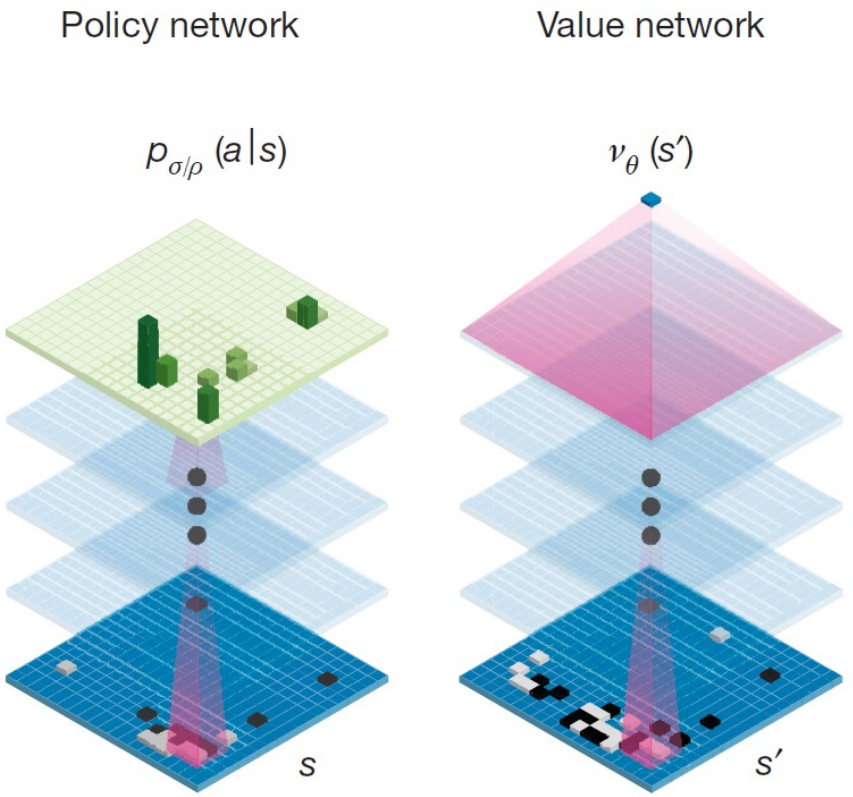
$$v(s_t) \leftarrow v(s_t) + \alpha[R_t - v(s_t)]$$

R_t is the actual long-term return following state s_t in a sampled trajectory



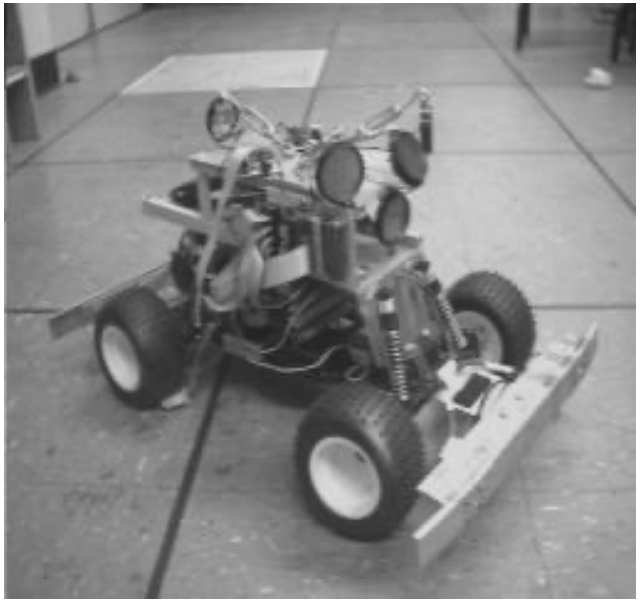
Reinforcement Learning

Example: AlphaGo



Reinforcement Learning

Example: Robots



- **input**: sensory information and reinforcement signal.
- **output**: avoid negative reward and try to have positive reward (obstacle avoidance, wall following, etc.)

Important ML concepts that we'll use throughout the semester

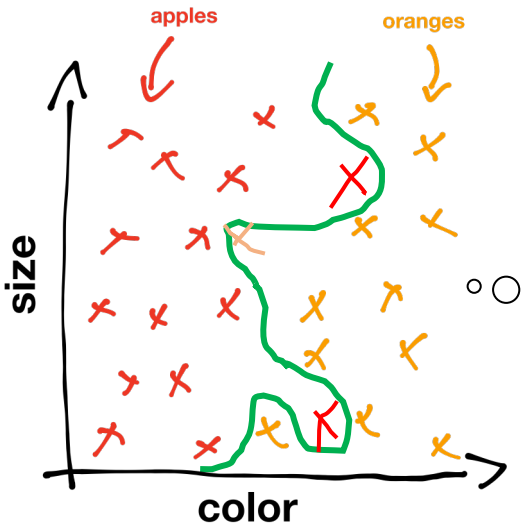
Generalization, Overfitting, Cross Validation, etc

Generalization

Will the classifier work for unseen fruits?

Is this an apple?

The issue of generalization: whether ML models are encouraged to learn generic patterns or simply remember details?

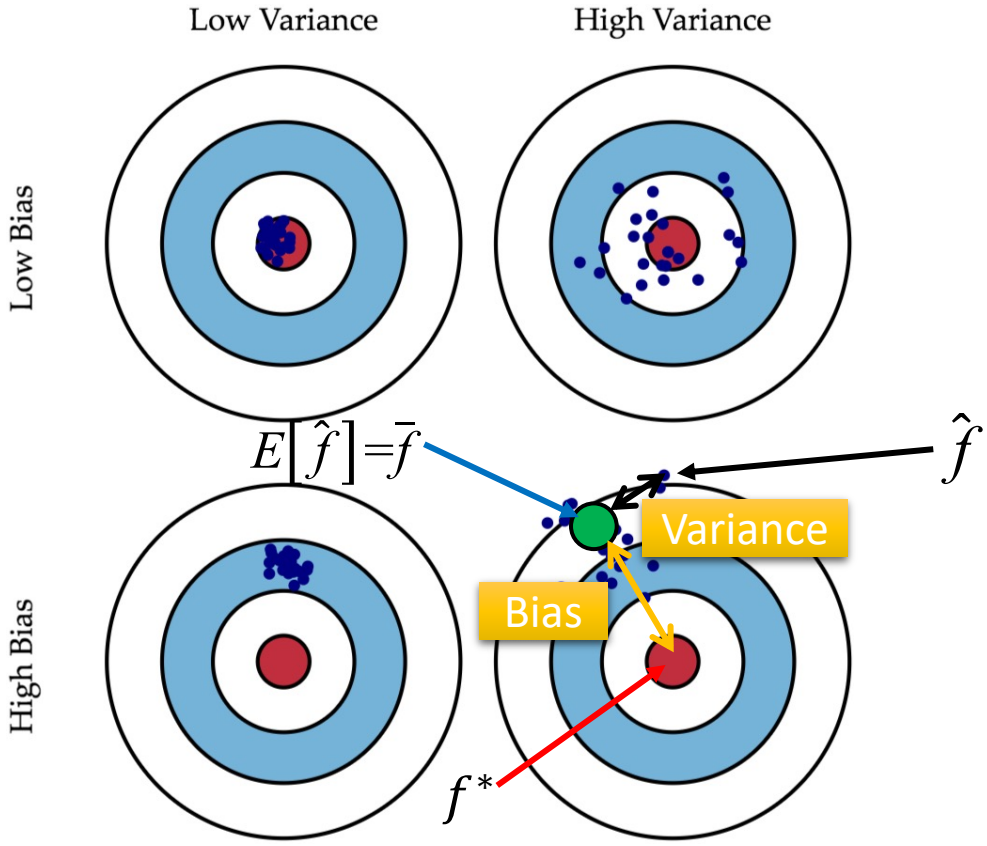


Which model is better?

Bias and Variance

Bias – how close does the assumed model fit for the observed data?

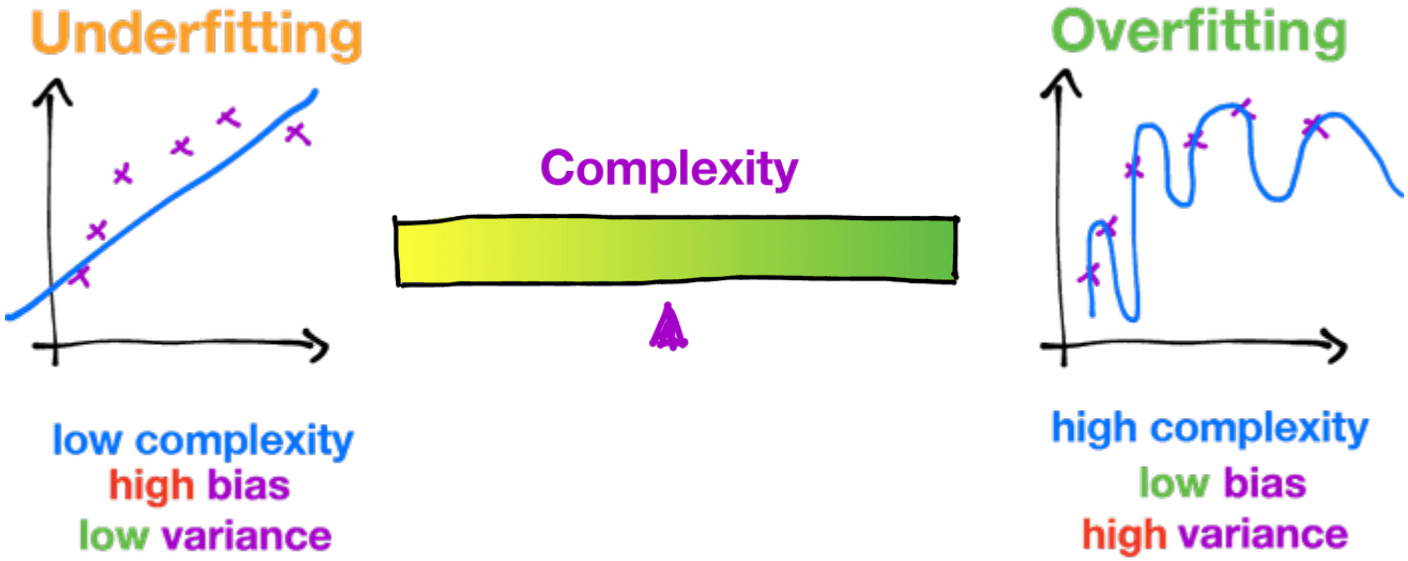
Variance – how complex (e.g., freedom) the assumed model is?



The Bias-Variance Dilemma

Bias – how close does the assumed model fit for the observed data?

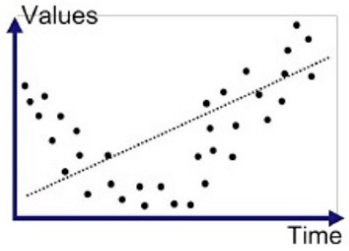
Variance – how complex (e.g., freedom) the assumed model is?



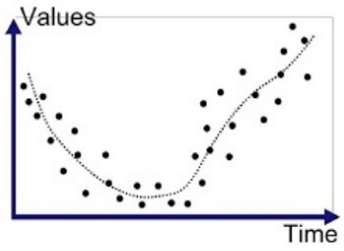
Overfitting

Underfitting – the model isn't complex enough to capture the real knowledge, the assumption may not hold.

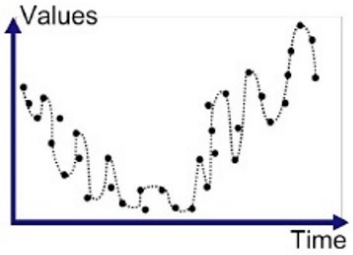
Overfitting – the model is too complex and thus describes the details of data (e.g., random noise) instead of underlying knowledge



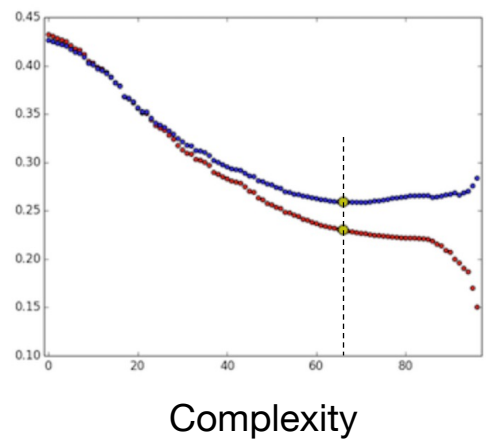
Underfitted



Good Fit/Robust



Overfitted



Overfitting

A real-world example (driving test)

1. 机动车遇到行人时应 A. 快速通过 B. **减速通过** C. 正常行驶
2. 在如图所示的场景中，驾驶员应 A. 停车等待 B **减速让行** C.
3. 遇到下图所示的情况，下面哪个操作正确？ A.
4. …

- 题中有"**应**"字样的选B,
- 有"**操作**"字样的选C,
- 最长的题目选D,
- 看图题有女孩的选A, 没人的选C

三短一长选最长

非正常行驶的场合，应选择谨慎的选项

Prevent Overfitting

Selecting models with appropriate complexity

Widely used approaches (to mitigate overfitting)

- Increase training data
- Regularization (penalizing model complexity)
- Hold-out & cross validation (unseen data to ensure generalization)
- Early stopping
- Prior knowledge (e.g., Bayesian prior)
- …

Prevent Overfitting

Fit the overall distribution instead of the training set.

- To estimate the generalization error, we need data unseen during model training.
- Data Splitting (Hold-Out):
  - Training set (e.g., 50%) — 平时练习
  - Validation set (e.g., 25%) — 模拟考试
  - Test set (e.g., 25%) — 高考!

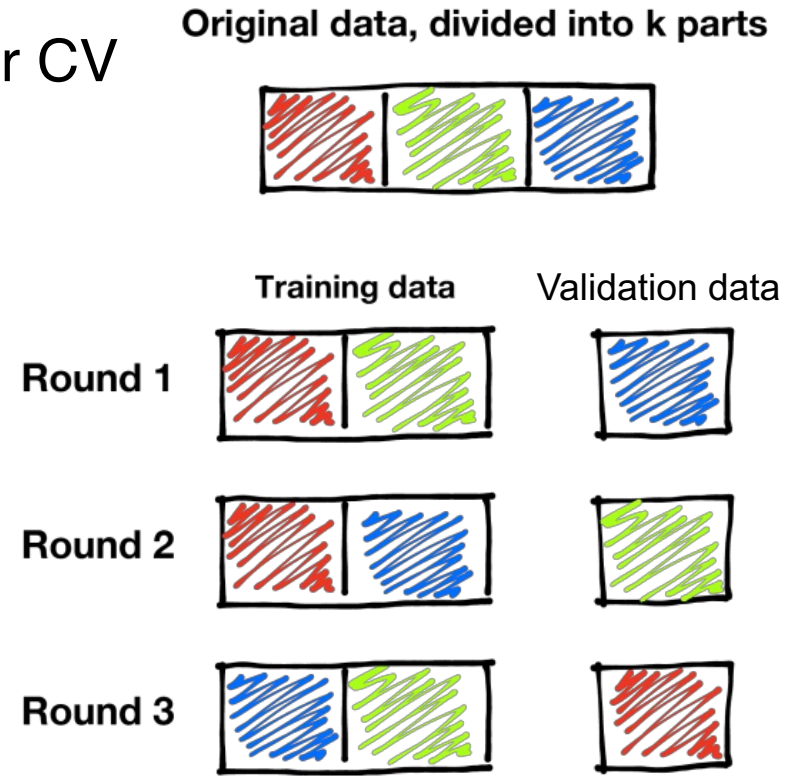
In general, design a good loss function to indicate the performance of the predictive model over the whole-data distribution instead of the training data can help achieve compromise between simplicity and complexity of the model structure.

Cross Validation

What if we want to make the most of the data?

(e.g., when the dataset is small)

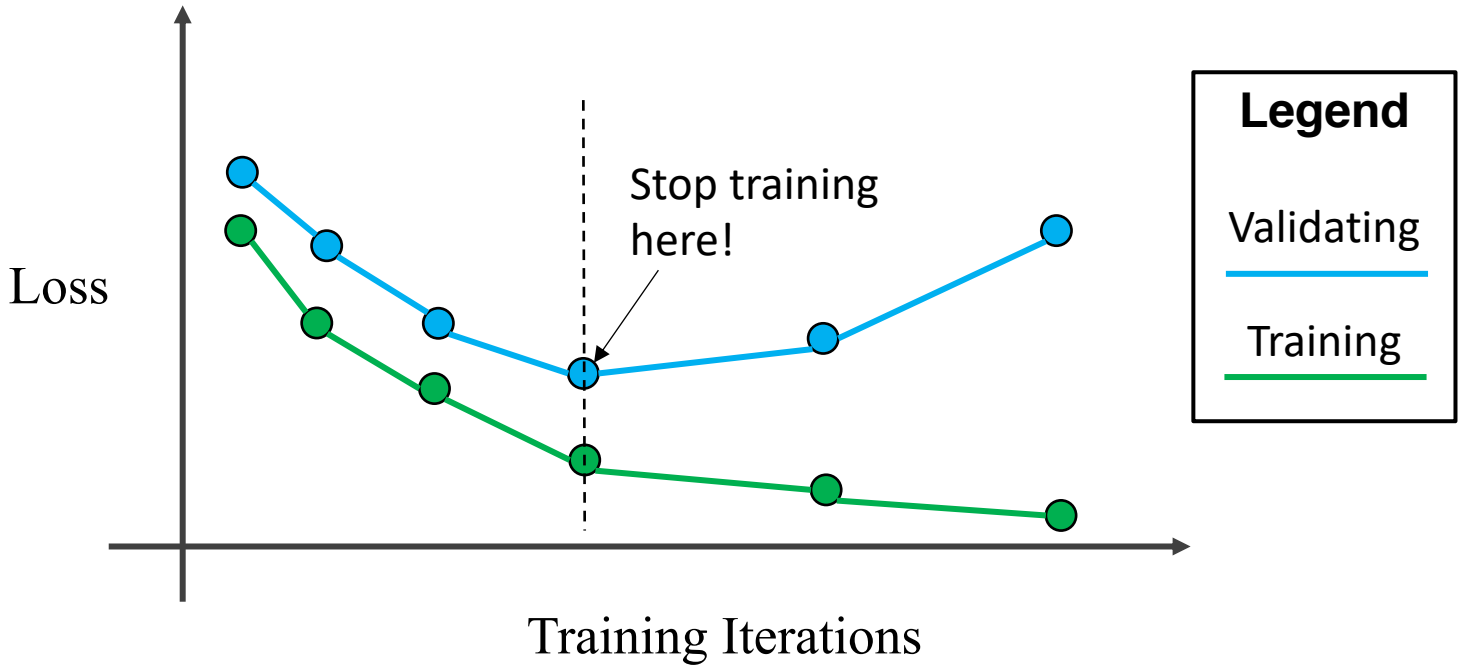
- K-folder CV



- Allow for more train-test splits
- A better indication of how well your model performs on unseen data
- Takes more computational power and time

Early Stop

- Stop training before we have a chance to overfit



Regularization

- Add a **penalty term** of the parameters to prevent the model from overfitting the training data.

$$L(\mathbf{W}, \lambda_1, \lambda_2) = ||y - \mathbf{W}X||^2 + \lambda_2||w||_2^2 + \lambda_1||w||_1$$

Penalize model complexity in the loss function

No free lunch theorem

- There is no universally best model.
- Different types of models have to be developed to suit the nature of the data in real applications.

What's Next

Linear Regression:

- a simple and well-known machine learning algorithm.
- Similar to the least squares method (最小二乘法)