Introduction

Machine Learning & Artificial Intelligence

Chen Yishuai

yschen@bjtu.edu.cn

Network Intelligence Lab of Beijing Jiaotong University

Content

- Definition
- History
- Type
- Application
- Concerns

Content

- Definition
- History
- Type
- Application
- Concerns

Artificial Intelligence

AI

What is Intelligence?

Intelligence

- Ability to acquire and apply knowledge and skills.
  - Understanding: text, vision
  - Communication: text, voice
  - Planning: design, planning
  - Learning: knowledge, imitation

Artificial Intelligence

Make Computers as Smart as Human

Artificial Intelligence

Make people think they are talking to people

Turing Test - Conversation

:

Q: Please write me a sonnet on the subject of the Forth Bridge.
A : Count me out on this one. I never could write poetry.
Q: Add 34957 to 70764.
A: (Pause about 30 seconds and then give as answer) 105621.
Q: Do you play chess?
A: Yes.
Q: I have K at my K1, and no other pieces. You have only K at K6 and R at R1. It is your move. What do you play?
A: (After a pause of 15 seconds) R-R8 mate.

Turing Test (2)

- Question: In the first line of the sonnet which reads “Shall I compare thee to a summer’s day,” would not “a spring day” do as well or better?
- Answer: It wouldn’t scan.
- Question: How about “a winter’s day”? That would scan all right.
- Answer: Yes, but nobody wants to be compared to a winter’s day.
- Question: Would you say Mr. Pickwick reminded you of Christmas?
- Answer: In a way.
- Question: Yet Christmas is a winter’s day, and I don’t think Mr. Pickwick would mind the comparison.
- Answer: I don’t think you’re serious. By a winter’s day one means a typical winter’s day, rather than a special one like Christmas.

Review

- What is AI?
  - Make Computers as Smart as Human
- What is turing test?
  - Make people think they are talking to people

Content

- Definition
- History
- Type
- Application
- Concerns

60 Years of Artificial Intelligence

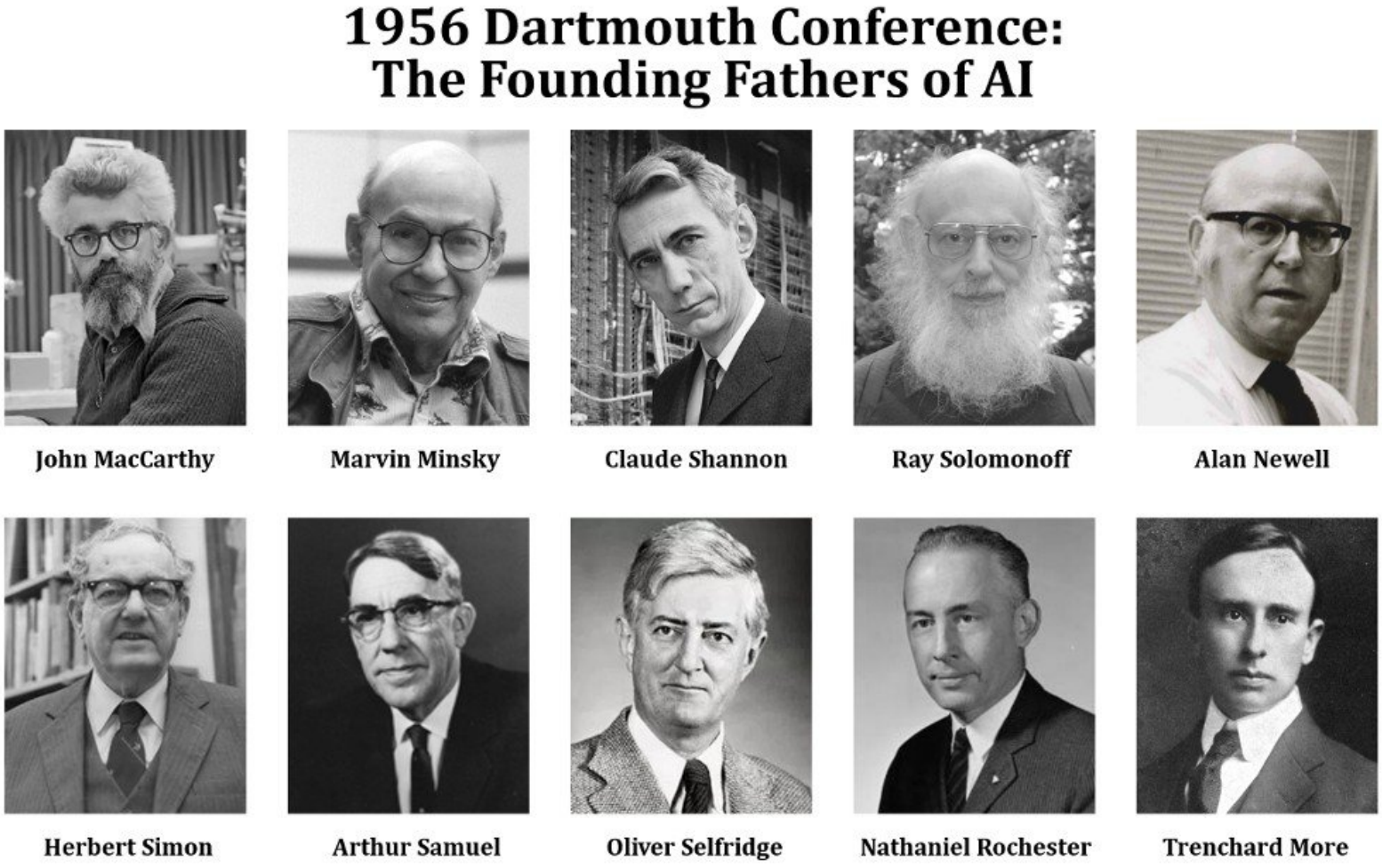
1956 to present

1956-1974

First wave

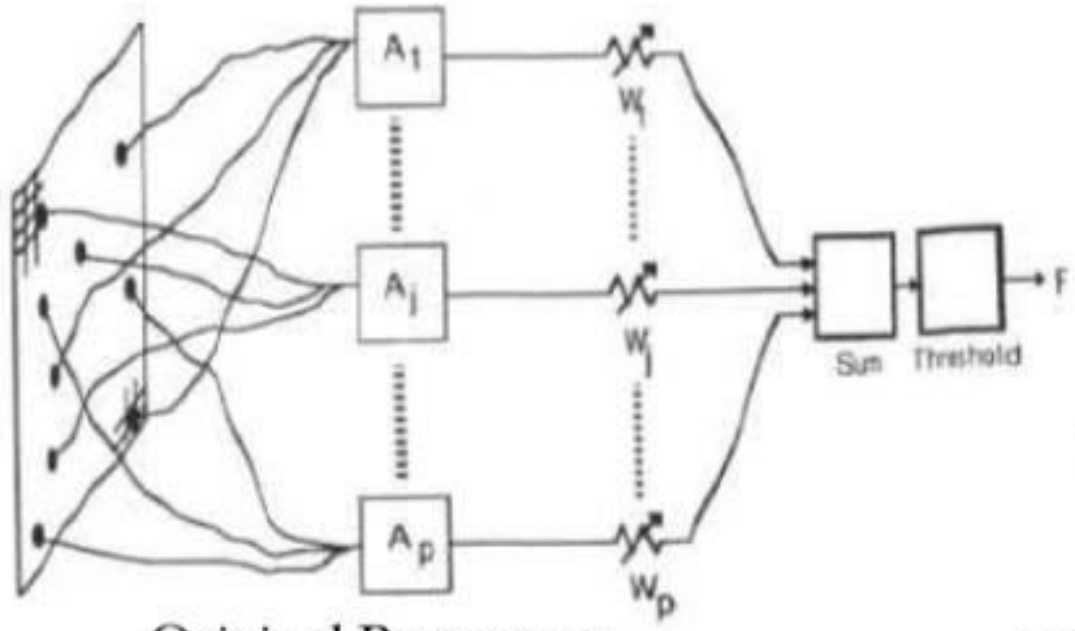
1956 Dartmouth conference

- Introduced the term "artificial intelligence"



1957 Perceptron model

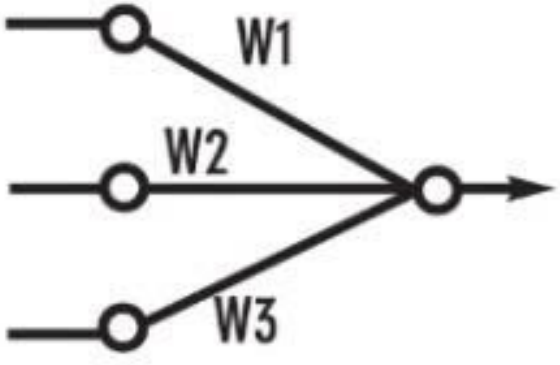
Perceptron (1957)



Original Perceptron

(From Perceptrons by M. L Minsky and S. Papert, 1969, Cambridge, MA: MIT Press. Copyright 1969 by MIT Press.

Frank Rosenblatt
(1928-1971)



Simplified model:

1967, Marvin Minsky: AI can be solved in one generation

ELIZA Dialogue robot

```
  Welcome to
                EEEEEE  LL      IIII  ZZZZZZZ   AAAAA
                EE      LL       II       ZZ   AA   AA
                EEEEE   LL       II     ZZZ    AAAAAAA
                EE      LL       II    ZZ      AA   AA
                EEEEEE  LLLLLL  IIII  ZZZZZZZ  AA   AA

  Eliza is a mock Rogerian psychotherapist.
  The original program was described by Joseph Weizenbaum in 1966.
  This implementation by Norbert Landsteiner 2005.


ELIZA: Is something troubling you ?
YOU:   Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU:   They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
YOU:   Well, my boyfriend made me come here.
ELIZA: Is it important to you that your boyfriend made you come here ?
YOU:   He says I'm depressed much of the time.
ELIZA: I am sorry to hear that you are depressed.
YOU:   It's true. I am unhappy.
ELIZA: Can you explain what made you unhappy ?
YOU:
```

ELIZA Dialogue robot

- ELIZA: Please tell me what's been bothering you.
- ME: Oh, hello, ELIZA. It's been a long time.
- ELIZA: Hi. What seems to be your problem ?
- ME: Lately, I've been feeling a little depressed.
- ELIZA: And how do you feel right now ?
- ME: I feel sad, but relieved to be writing.
- ELIZA: Of what does feeling sad remind you ?
- ME: Well, I'm an alcoholic, and I've been sober for a long time, but I drank yesterday. And then I decided to stop drinking again. But still; that makes me feel sad.
- ELIZA: Is it because you are an alcoholic that you came to me ?
- ME: I suppose that that's one of the reasons.
- ELIZA: Can you elaborate on that ?
- ME: I mean, it's not the only reason, but alcoholism is a primary thing that has messed up my life.
- ELIZA: That's quite interesting.
- ME: …Thank you.

ELIZA Dialogue Robot

ELIZA- Dialog

:

Ramu. I have a terrible headache
Eliza. May be you should consult a doctor of medicine.
Ramu. Okay I'll do that then
Eliza. I am glad you agree
Ramu. I want to eat an icecream
Eliza. Why do you wish to eat an icecream..
Ramu. My friends are intellectuals.
Eliza. Intellectuals... You said you have a headache
Ramu. Lets discuss religion
Eliza. You haven't mentioned your family
Ramu. I want to improve myself.
Eliza. That is interesting. Perhaps your terrible headache has something to do with your wanting an icecream

First winter

1974-1980

Fall of Perceptron

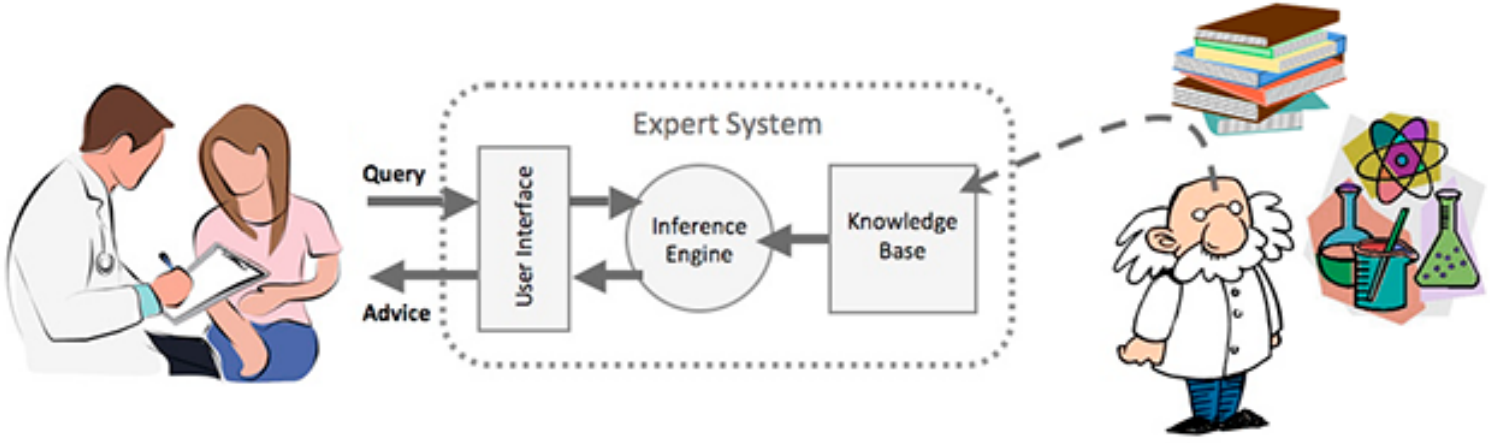
- Perceptron's limitations are discovered
  - Marvin Minsky & Seymour Papert (1969). Perceptrons, MIT Press, Cambridge, MA.
  - Unless input categories were "linearly separable", a perceptron could not learn to discriminate between them.
- Many important categories were not linearly separable
  - E.g., those inputs to an XOR gate that give an output of 1 (namely 10 & 01) are not linearly separable from those that do not (00 & 11).

First winter: 1974-1980

- Artificial intelligence is gradually being ignored
- Funding disappear

1980-1987: Expert System

- Learn experts' behavior and decisions
- Extract rules and build expert systems



Second winter: 1987-1993

- Too complex to extract rules and build systems
  - Some rules cannot be written
- Performance is below expectations

1993 until now

- Machine learning
- Deep learning
- Computer vision
- AlphaGo

The New York Times

Turing Award Won by 3 Pioneers in Artificial Intelligence

Content

- Definition
- History
- Type
- Application
- Concerns

Methods

Methods

- Machine Learning
- Searching
- Reasoning
  - Symbolic reasoning
  - Logical reasoning
- Statistics

Human Learning

Machine Learning

Machine learning

Arthur Samuel, 1959

"Field of study that gives computers the ability to learn without being explicitly programmed"

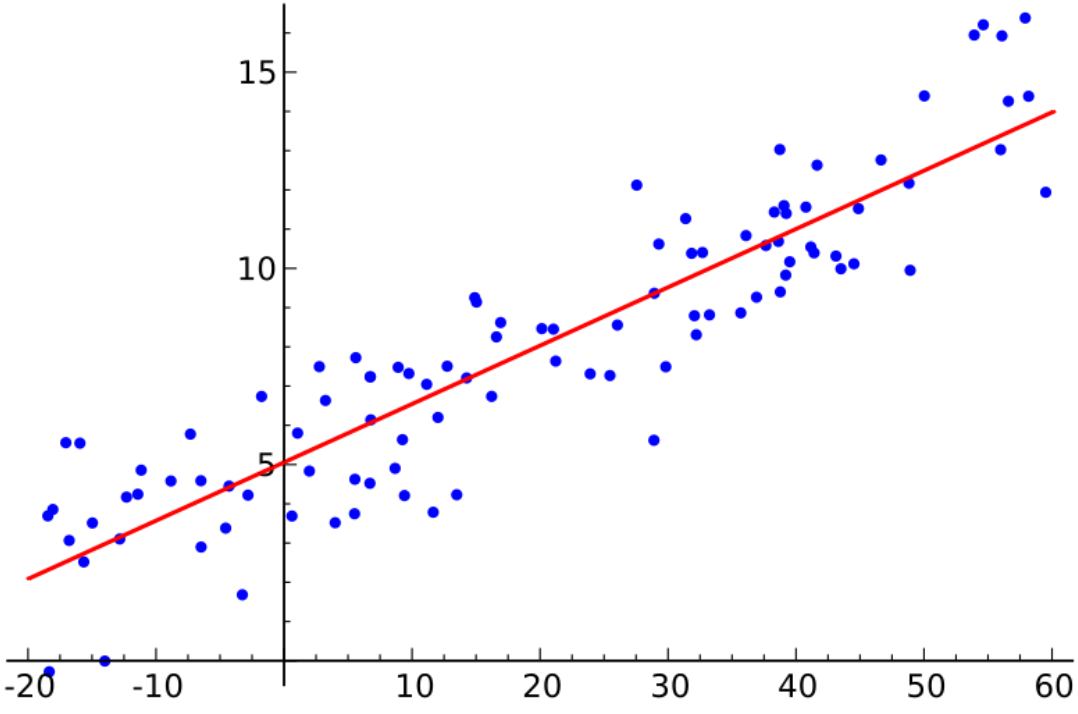
Machine learning

Tom Mitchell, 1998

"Well posed Learning Problem: A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E."

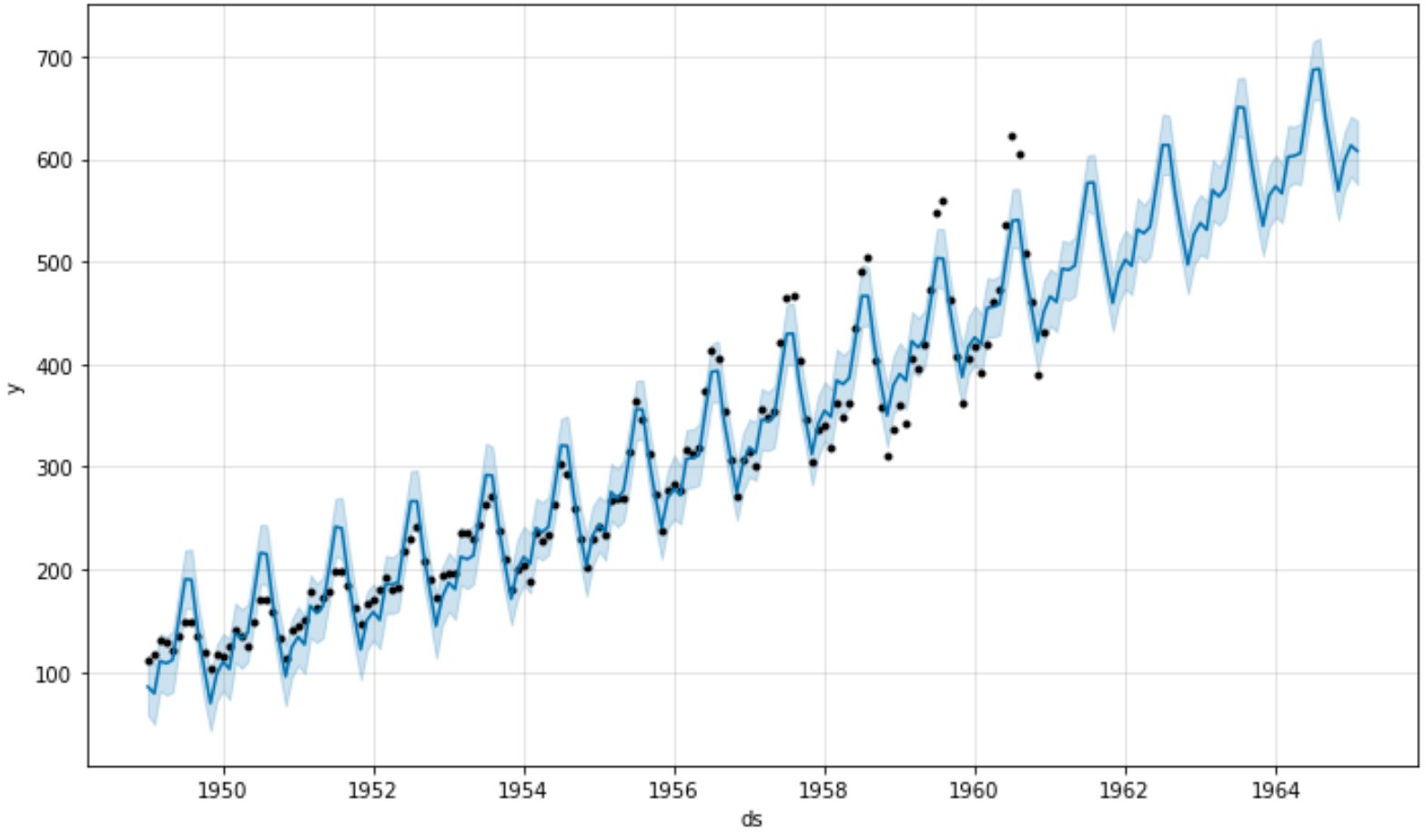
Experience is Data

- Learn from Historical Data
  - Trend
  - Predict based on this trend



Learn from Historical Data

- Trend and Seasonality



Eight Typical Machine Learning Tasks

1) Regression

Predict "numerical" value

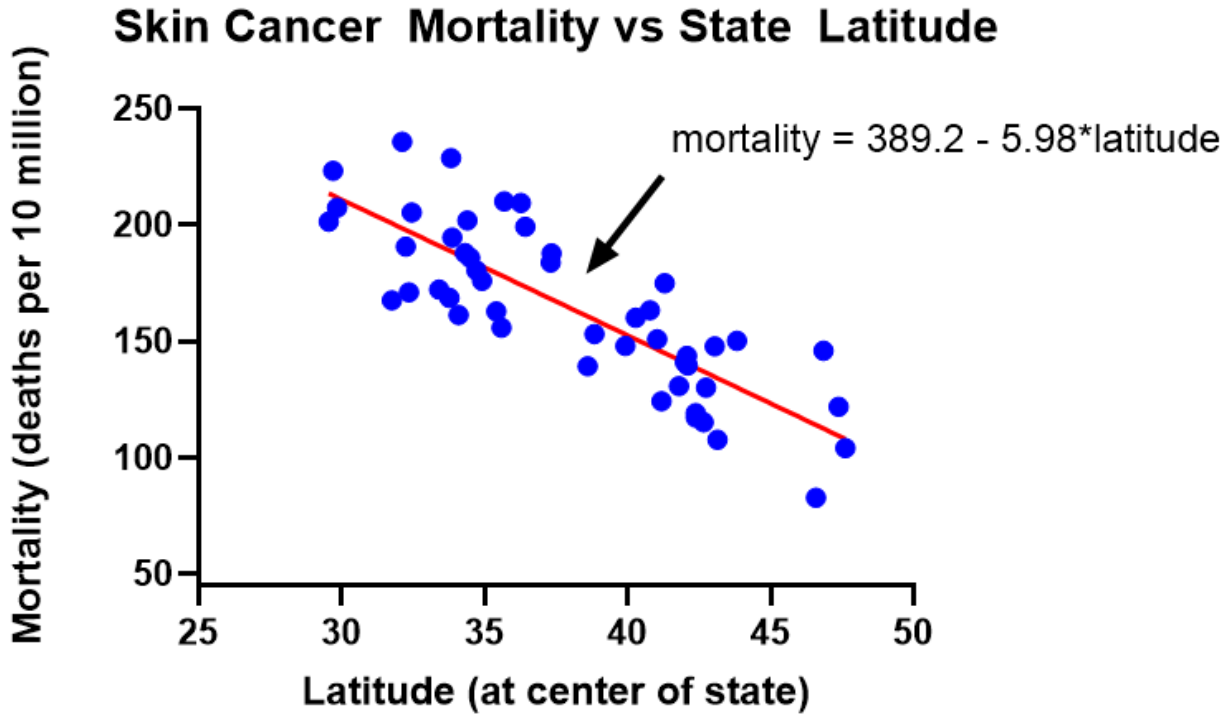
Predict House Price

- Based on
  - Number of rooms
  - Distance to city center
  - etc.

Data

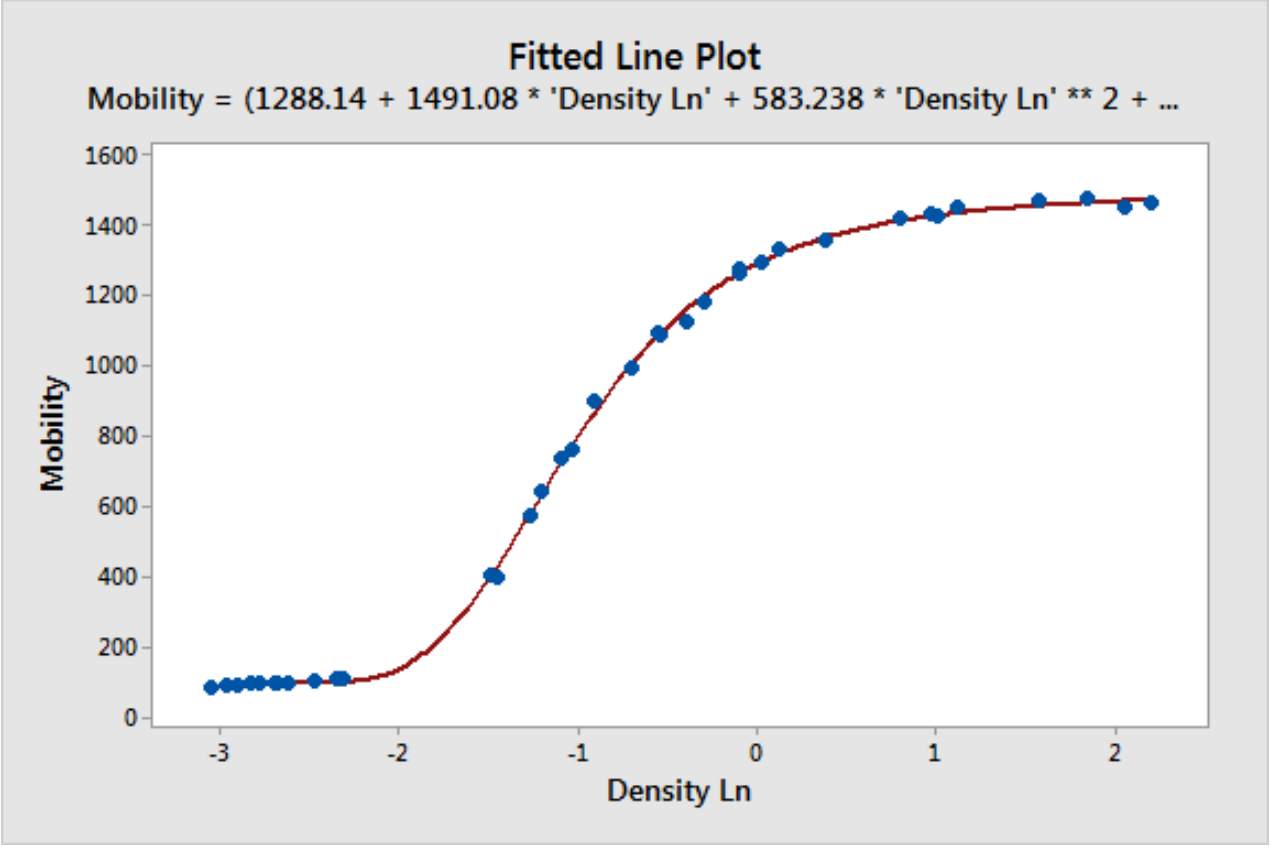
1. CRIM - per capita crime rate by town
2. ZN - proportion of residential land zoned for lots over 25,000 sq.ft.
3. INDUS - proportion of non-retail business acres per town.
4. CHAS - Charles River dummy variable (1 if tract bounds river; 0 otherwise)
5. NOX - nitric oxides concentration (parts per 10 million)
6. RM - average number of rooms per dwelling
7. AGE - proportion of owner-occupied units built prior to 1940
8. DIS - weighted distances to five Boston employment centres
9. RAD - index of accessibility to radial highways
10. TAX - full-value property-tax rate per $10,000
11. PTRATIO - pupil-teacher ratio by town
12. B - 1000(Bk - 0.63)^2 where Bk is the proportion of blacks by town
13. LSTAT - % lower status of the population
14. MEDV - Median value of owner-occupied homes in $1000's

Linear Regression



- $y = ax + b$, a and b are model parameters
- Learn a and b based on data

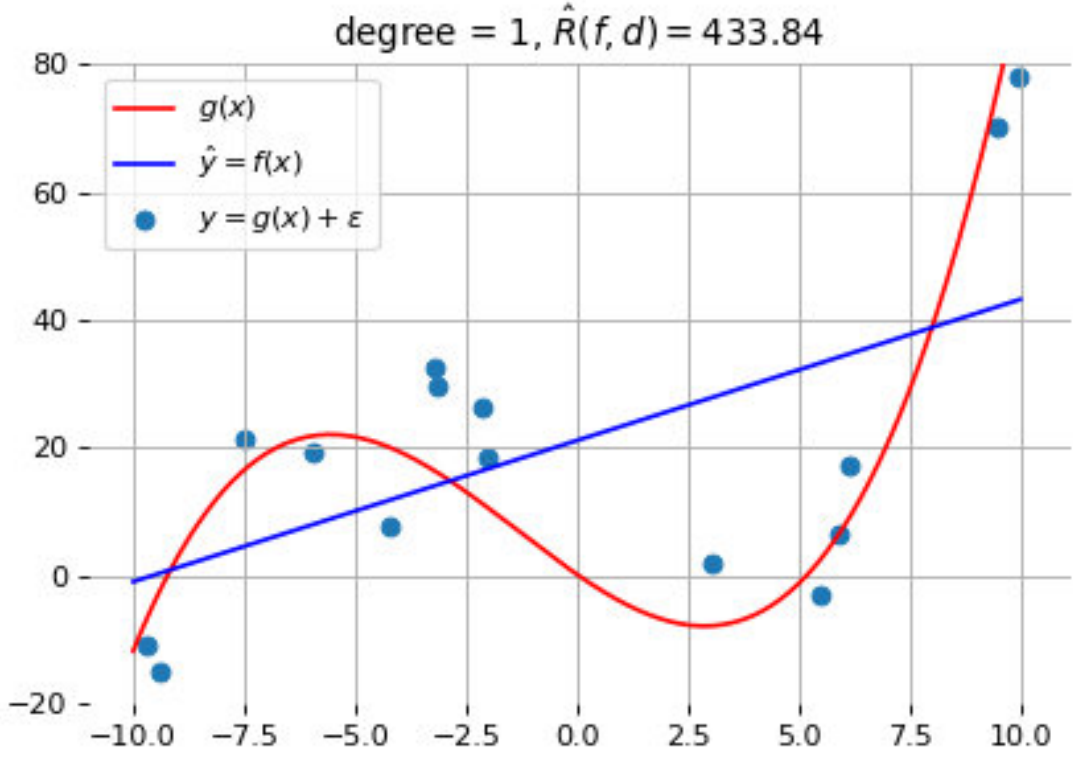
Non-Linear Regression



- $y = ax + bx^2 + cx^3$, a, b, c are model parameters
- Learn a, b, c based on data

Comparison

- Linear Regression: Line
- Non-Linear Regression: Curve



Discussion

- What regression task in you mind?

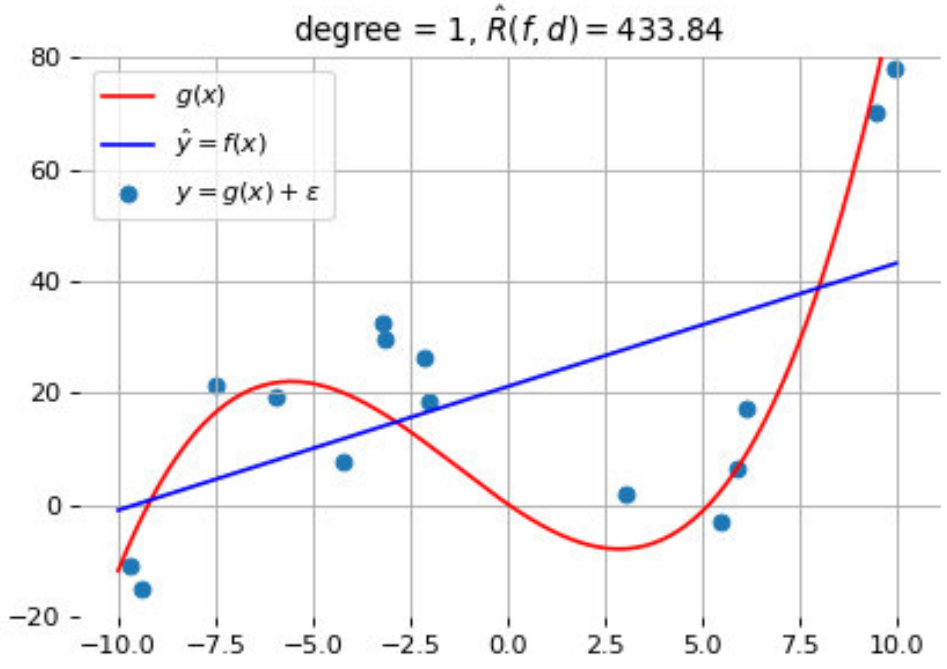
Discussion

- What are its related factors?

1. CRIM - per capita crime rate by town
2. ZN - proportion of residential land zoned for lots over 25,000 sq.ft.
3. INDUS - proportion of non-retail business acres per town.
4. CHAS - Charles River dummy variable (1 if tract bounds river; 0 otherwise)
5. NOX - nitric oxides concentration (parts per 10 million)
6. RM - average number of rooms per dwelling
7. AGE - proportion of owner-occupied units built prior to 1940
8. DIS - weighted distances to five Boston employment centres
9. RAD - index of accessibility to radial highways
10. TAX - full-value property-tax rate per $10,000
11. PTRATIO - pupil-teacher ratio by town
12. B - 1000(Bk - 0.63)^2 where Bk is the proportion of blacks by town
13. LSTAT - % lower status of the population
14. MEDV - Median value of owner-occupied homes in $1000's

Discussion

- Linear model or non-linear model?

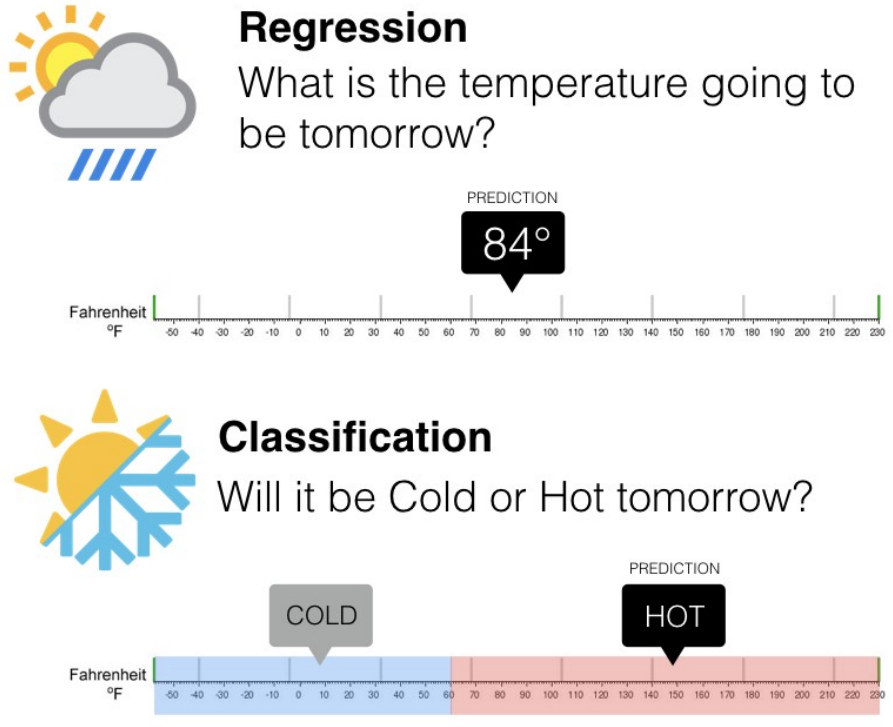


2) Classification

Predict "Categorical" values

Value-based classification

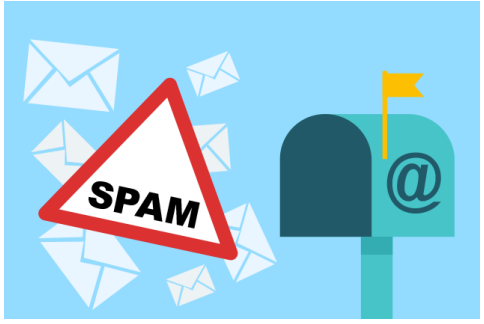
- Cold or Hot?



Threshold-based Model: $x > 60$F ?

Spam Mail Classification

- Input
  - Sending time
  - From
  - Key words: free, prize
- Output
  - spam or not?



Neuron Classification Model

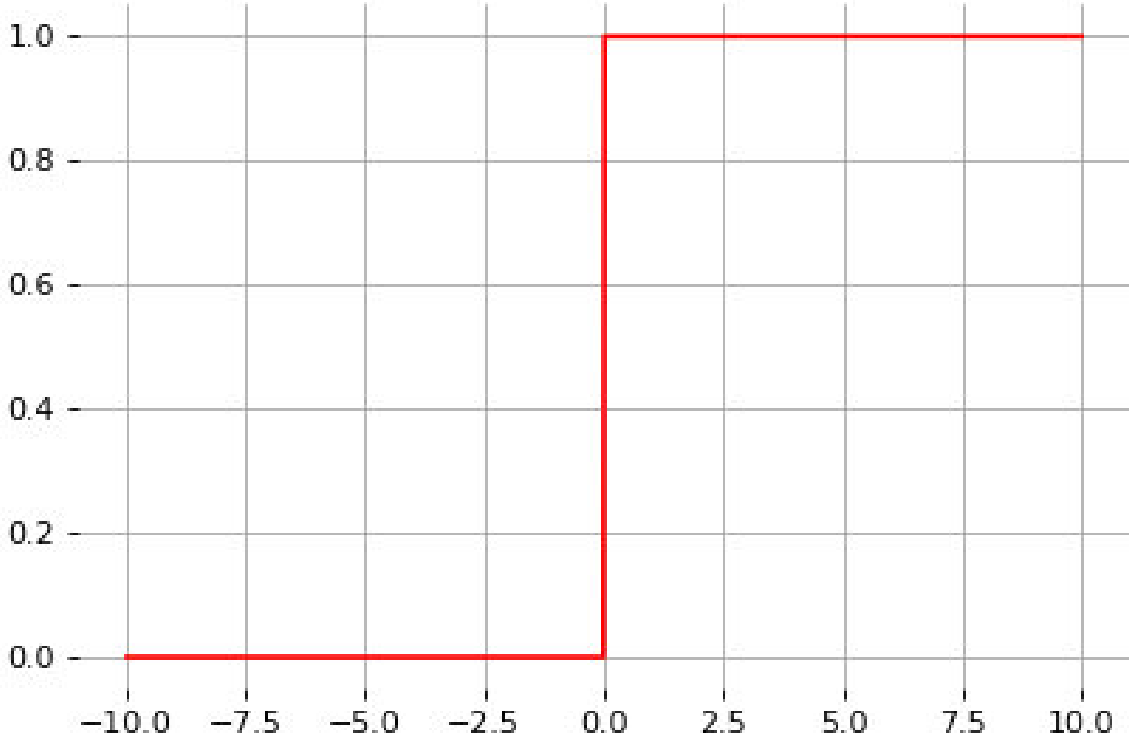
- Neurons (brain cells) are connected through synapses
- The brain constantly creates, strengthens, and weakens these connections

Perceptron Model I

- Linear weighted sum of inputs
  - Neuron input: x_i
  - Connection weight: w_i
  - Sum: $w_1 x_1 + w_2 x_2 + b$

Perceptron Model II

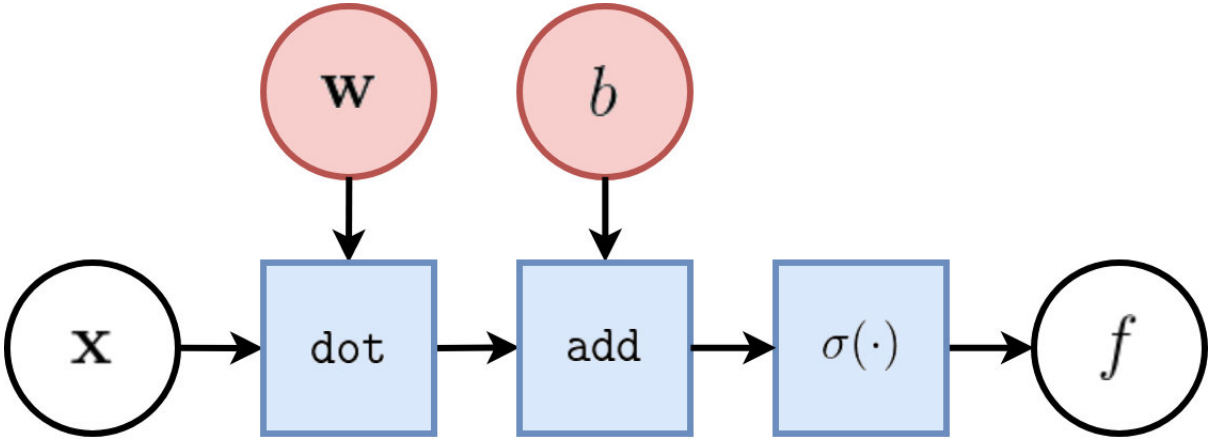
- Nonlinear activation function



$$\sigma(x) : x \geq 0$$

Perceptron Model

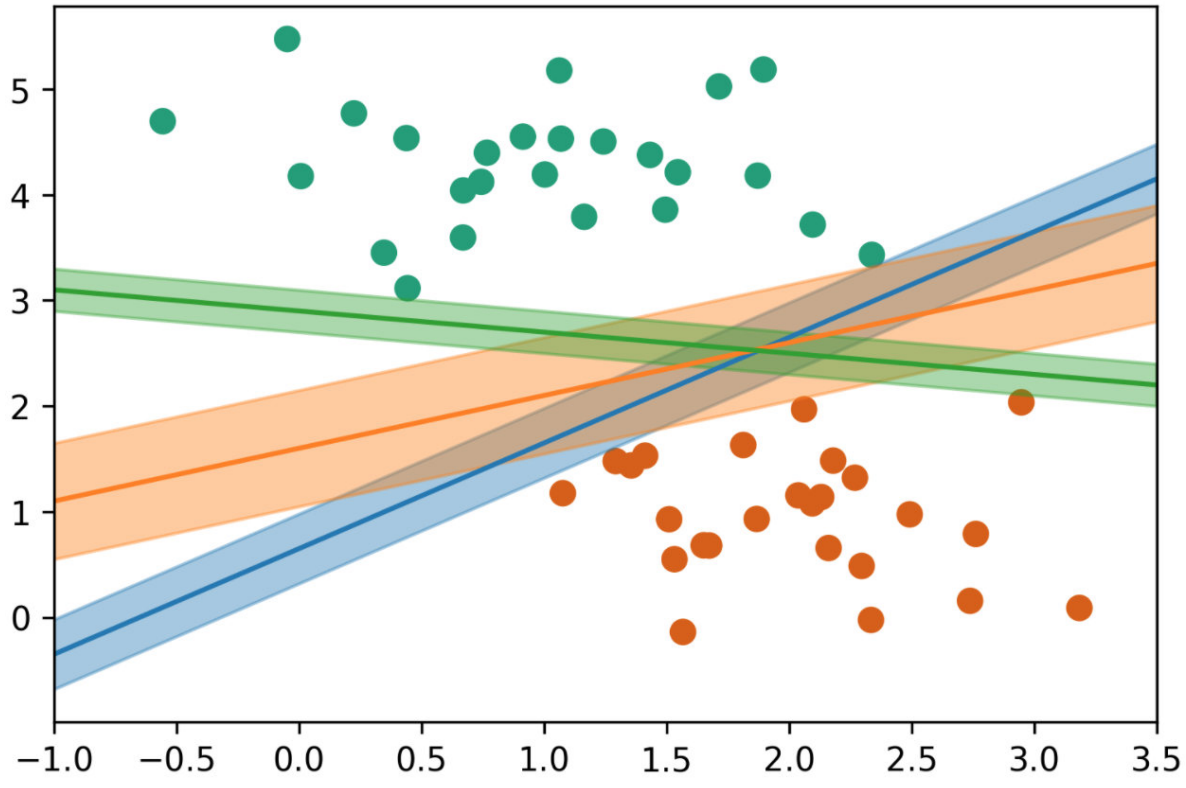
- Input linear weighting sum
- Non-linear activation function



$$f(x) : w_1 x_1 + w_2 x_2 + b \geq 0$$

Perceptron Classification

- Linear separation



$$w_1 x_1 + w_2 x_2 + b \geq 0$$

Implementation

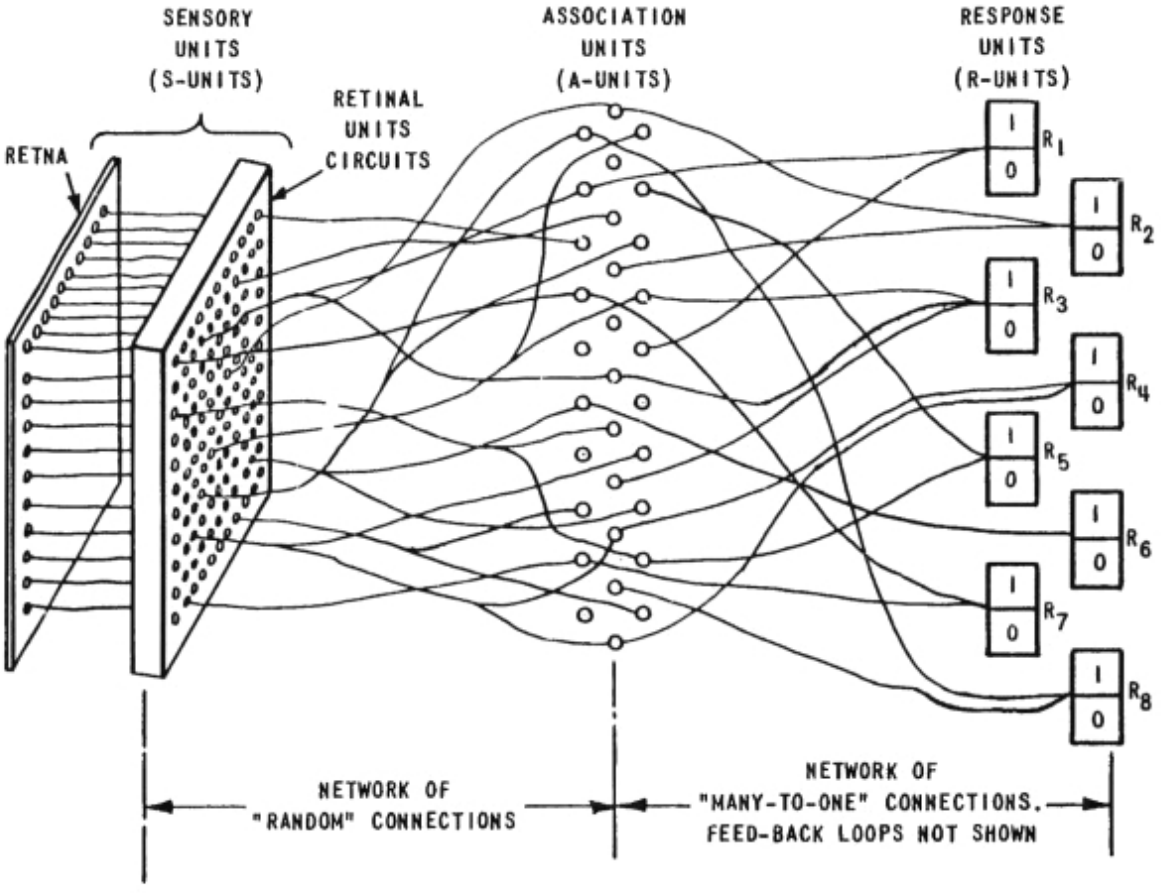
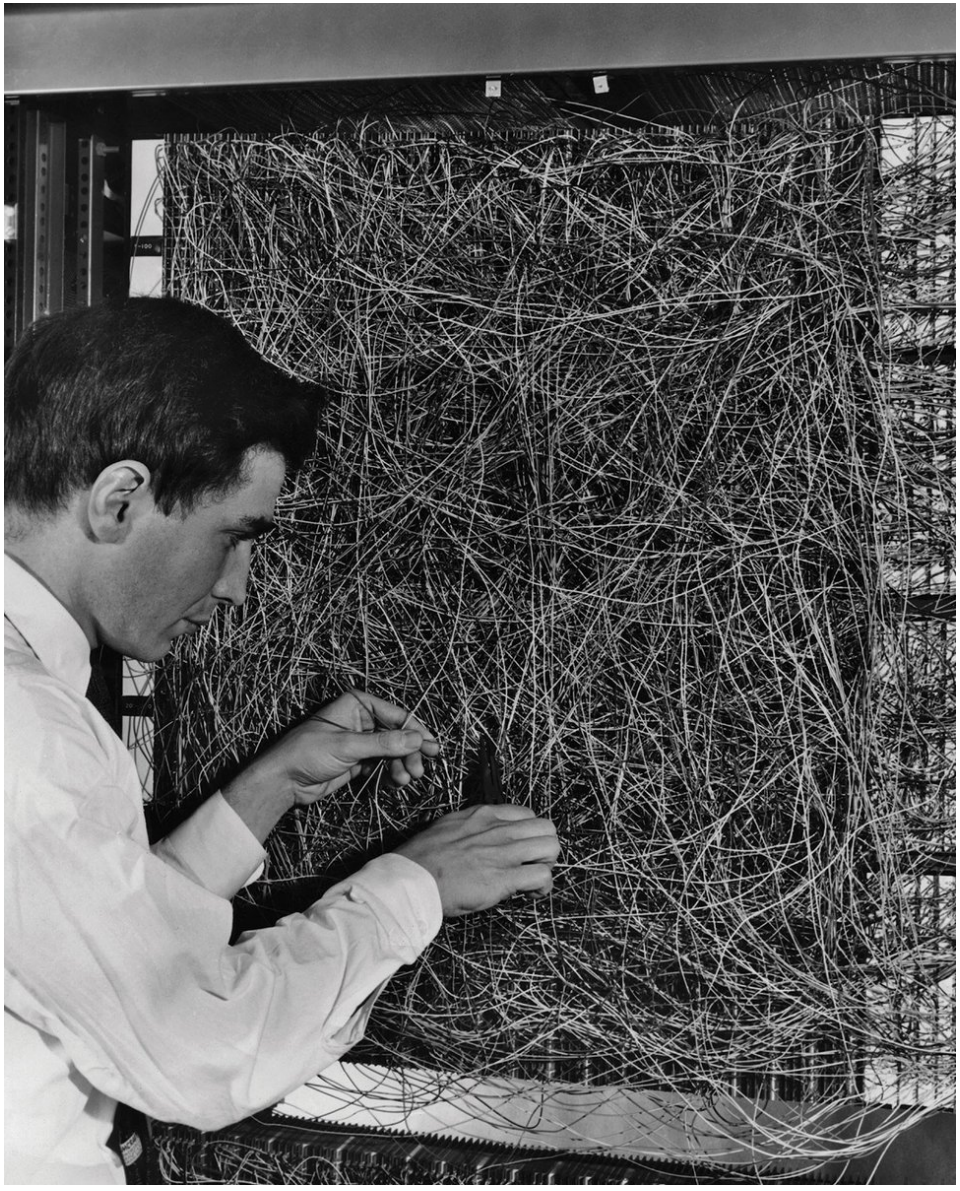


Figure 1 ORGANIZATION OF THE MARK I PERCEPTRON

Implementation



Model Training Method

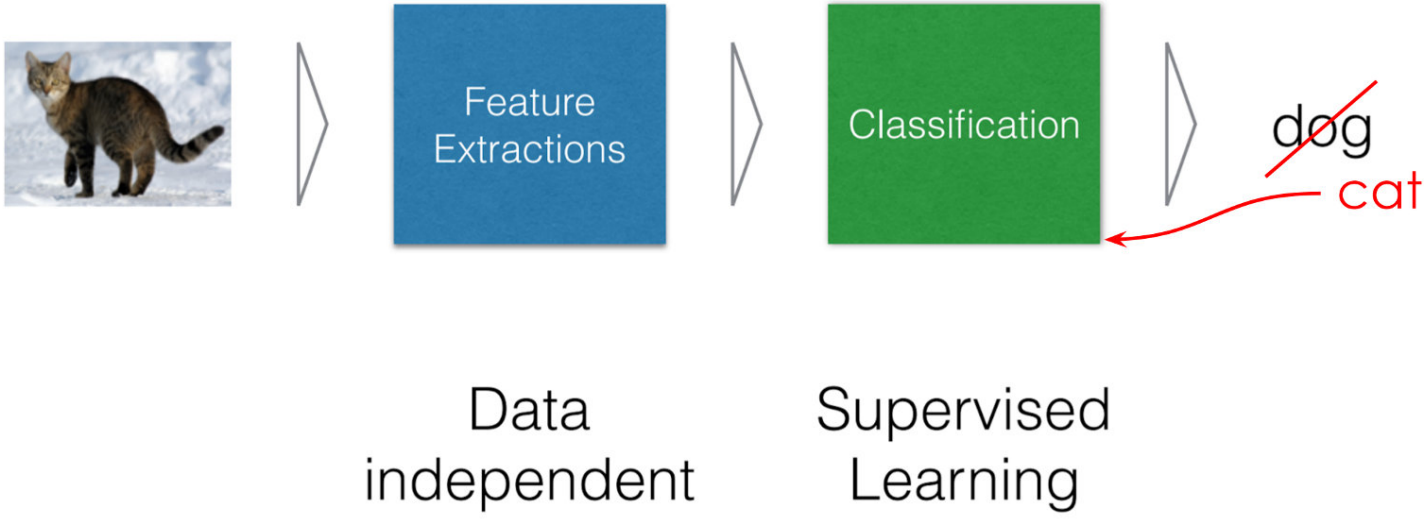
Learn from mistakes

Brain Learning Process

- Continuously create, strengthen, and weaken connections between neurons based on experimental results
- i.e., adjust the weight of the connection: w

Machine Learning Process

- An error occurred, adjusting model parameters *W* backwards



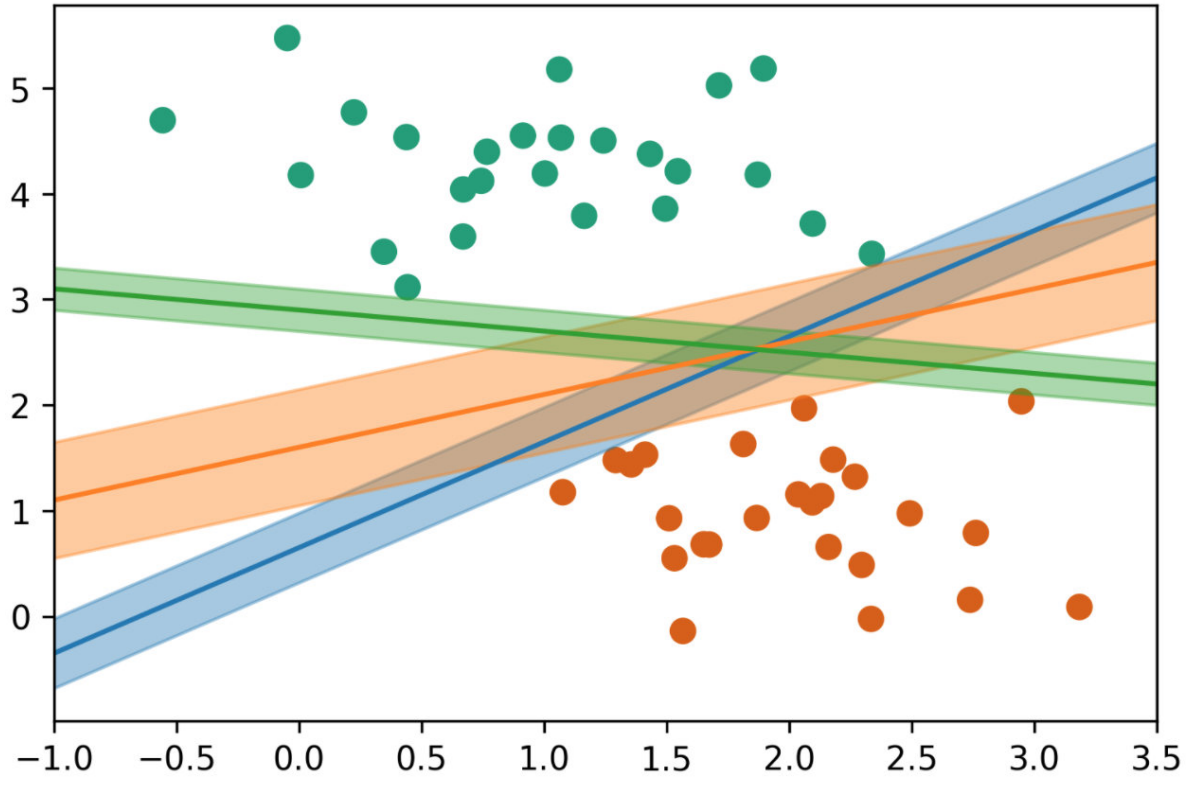
Perceptron Learning Process

- Find errors, adjust weight w to reduce errors

$$w_1 x_1 + w_2 x_2 + b \geq 0$$

Perceptron Learning Process

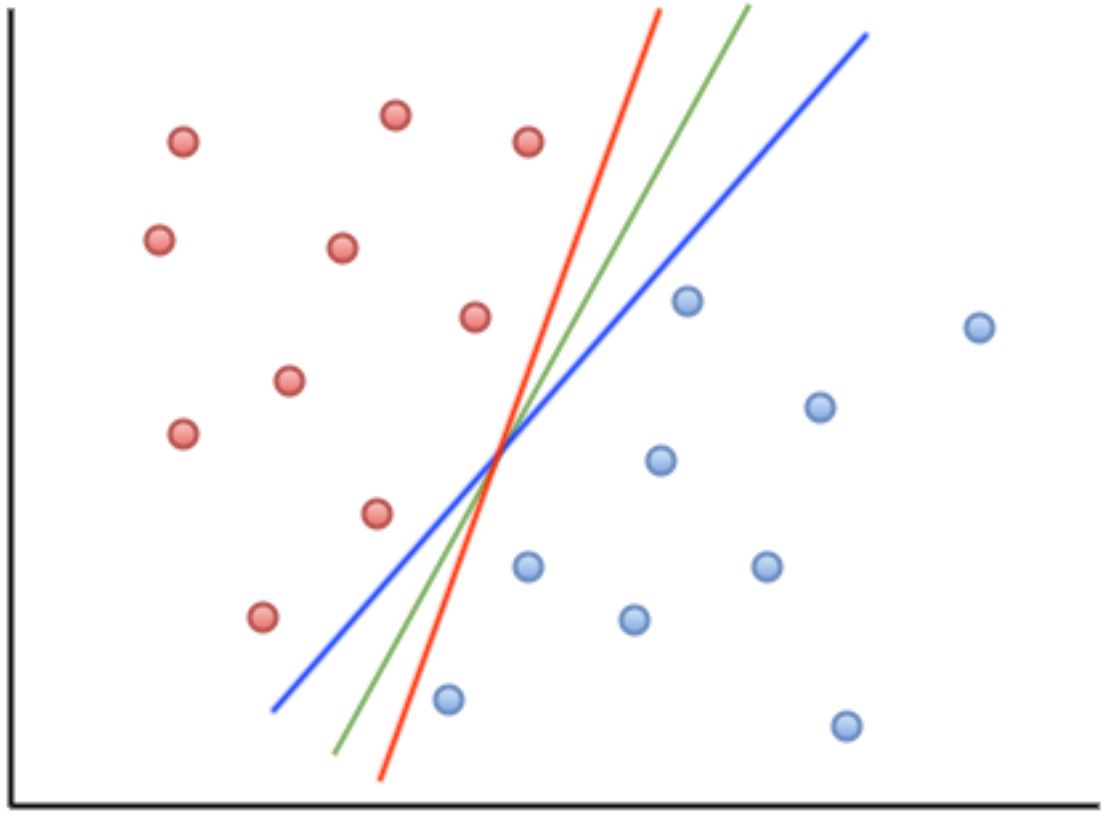
- Adjust w <=> adjust decision boundaries



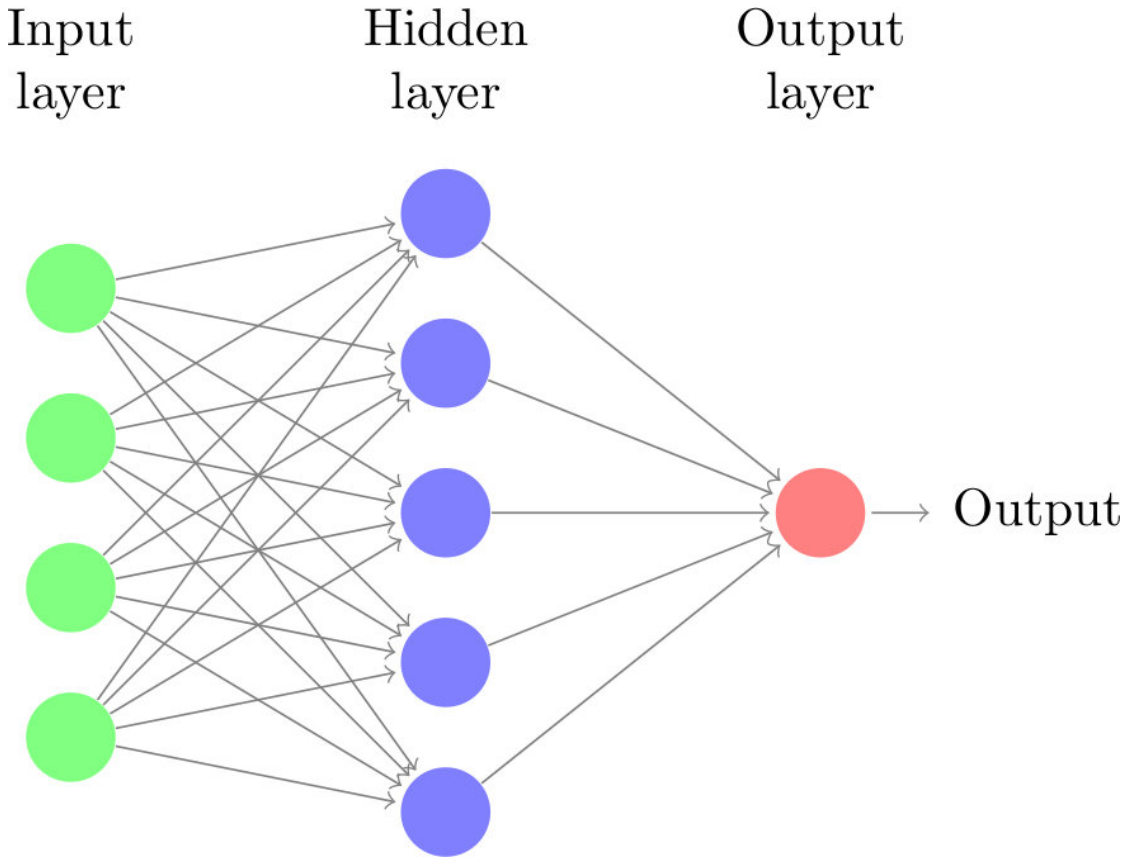
$$y = \sigma(w_1 x_1 + w_2 x_2 + b)$$

SVM

- Support Vector Machines
  - Adjust decision boundaries. Not only avoid mistakes, the farther the two sides are, the better

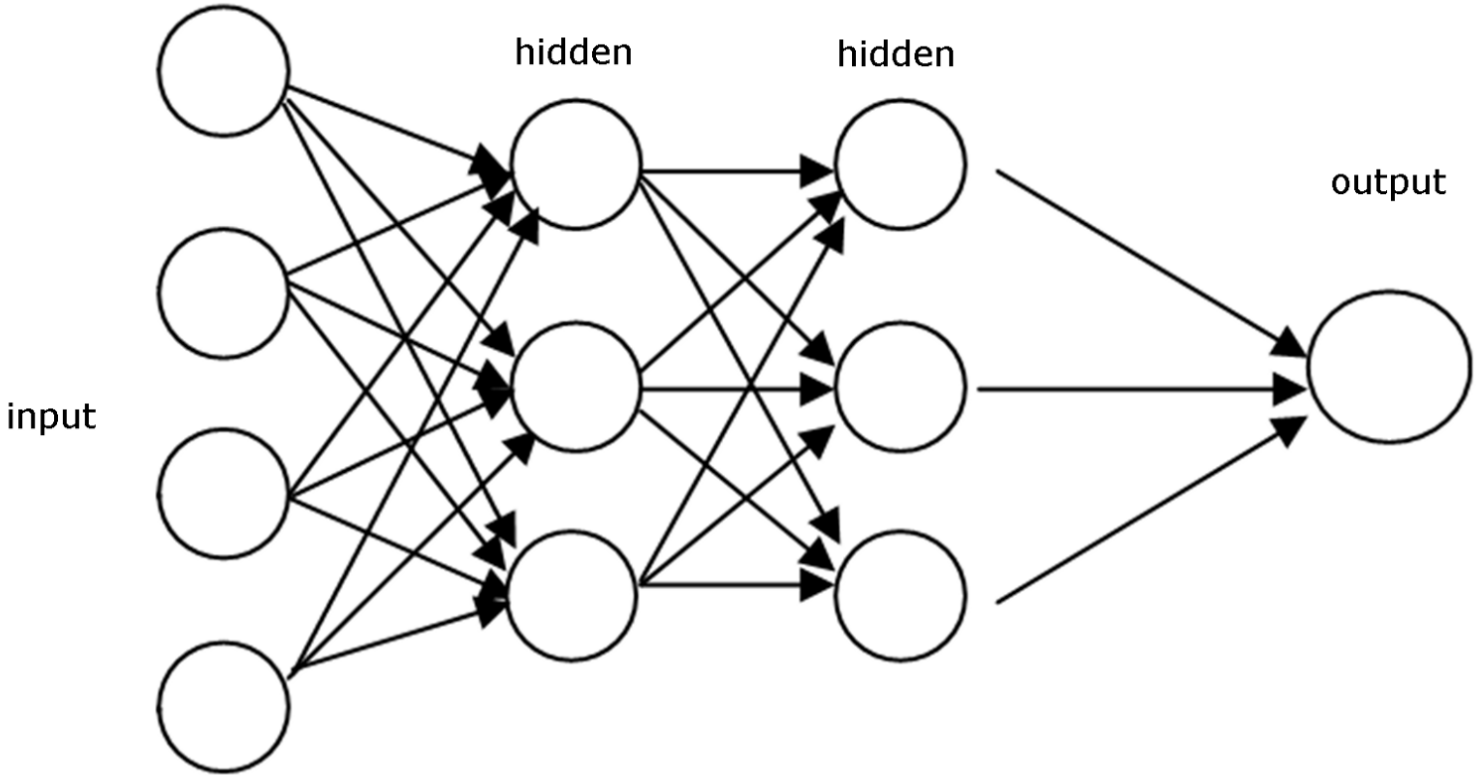


Forward Neural Network



Hidden and output layer units: perceptron

Deep Neural Network



Multiple hidden layers

Benefits of Depth

Generally, the deeper, the stronger the model

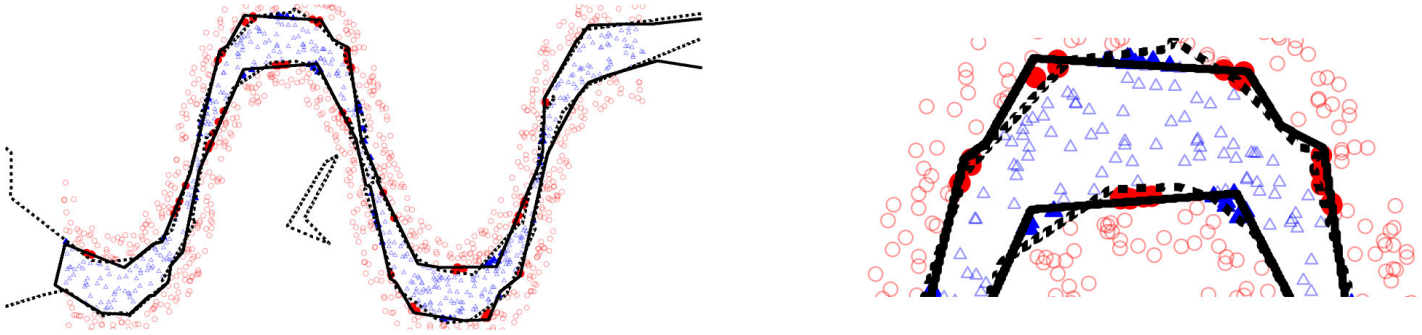
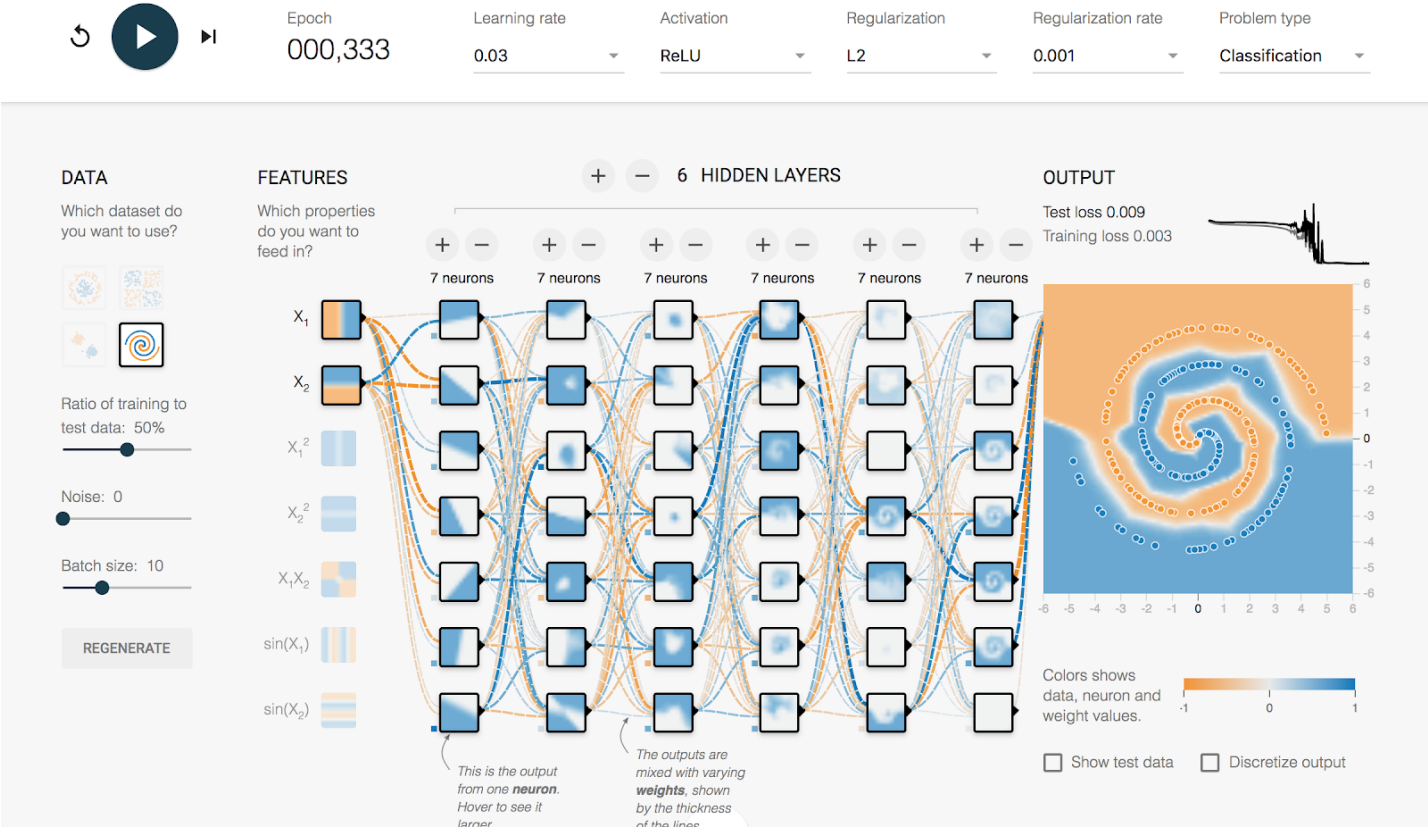


Figure 1: Binary classification using a shallow model with 20 hidden units (solid line) and a deep model with two layers of 10 units each (dashed line). The right panel shows a close-up of the left panel. Filled markers indicate errors made by the shallow model.

FNN Experiments

- Browser-based TensorFlow experiments
- http://playground.tensorflow.org



More Text Classification Examples

Sentiment Classification

Question - Answer

Sentiment Classification

- IMDB Review, 50000 reviews in English
  - positive or negative

IMDB User Review

Data

	text	sentiment
0	For a movie that gets no respect there sure ar...	0
1	Bizarre horror movie filled with famous faces ...	0
2	A solid, if unremarkable film. Matthau, as Ein...	0
3	It's a strange feeling to sit alone in a theat...	0
4	You probably all already know this by now, but...	0
5	I saw the movie with two grown children. Altho...	0
6	You're using the IMDb. You've given some heft...	0
7	This was a good film with a powerful message o...	0
8	Made after QUARTET was, TRIO continued the qua...	0
9	For a mature man, to admit that he shed a tear...	0

Question Answering

- Stanford Question Answering Dataset (SQuAD)
  - Reading comprehension dataset
  - Questions posed by crowdworkers on Wikipedia articles
  - Answer is a segment of text, or span, from the reading passage, or might be unanswerable.
- Classification problem
  - Is this word the answer?

SQuAD Website

Data

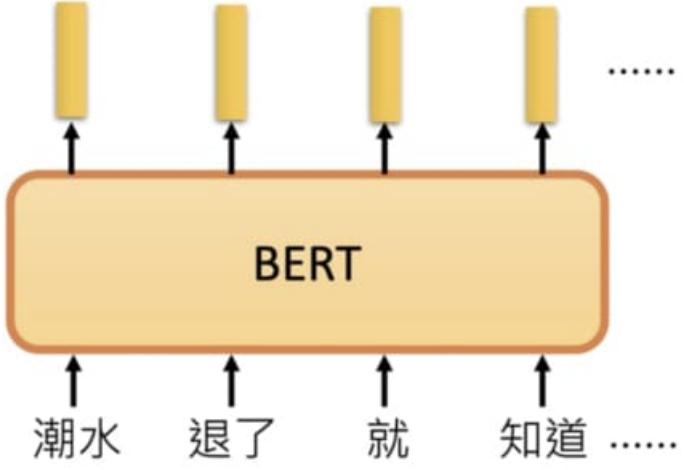
- Text
  - Tesla gained experience in telephony and electrical engineering before emigrating to the United States in 1884 to work for Thomas Edison in New York City. He soon struck out on his own with financial backers, setting up laboratories and companies to develop a range of electrical devices.
- Question
  - In what year did Nikola Tesla emigrate to the United States?
- Answer
  - 1884

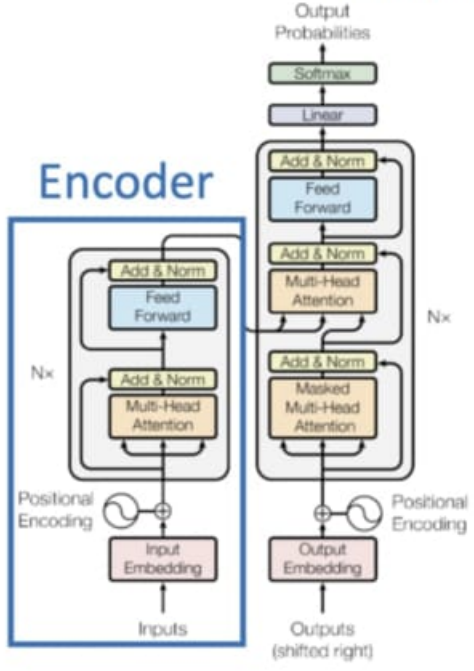
Deep Learning Text Model

Bidirectional Encoder Representations from Transformers (BERT)

- BERT = Encoder of Transformer

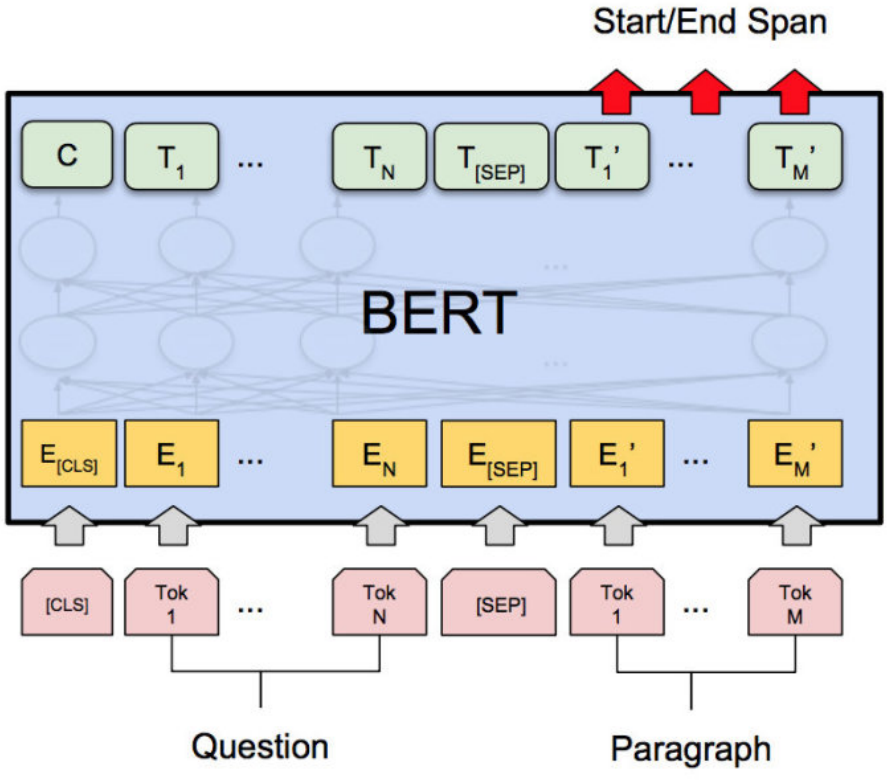
Learned from a large amount of text without annotation





IMDB Sentiment Classification: BERT 93.46%

BERT for SQuAD



(c) Question Answering Tasks: SQuAD v1.1

BERT: 88.5%

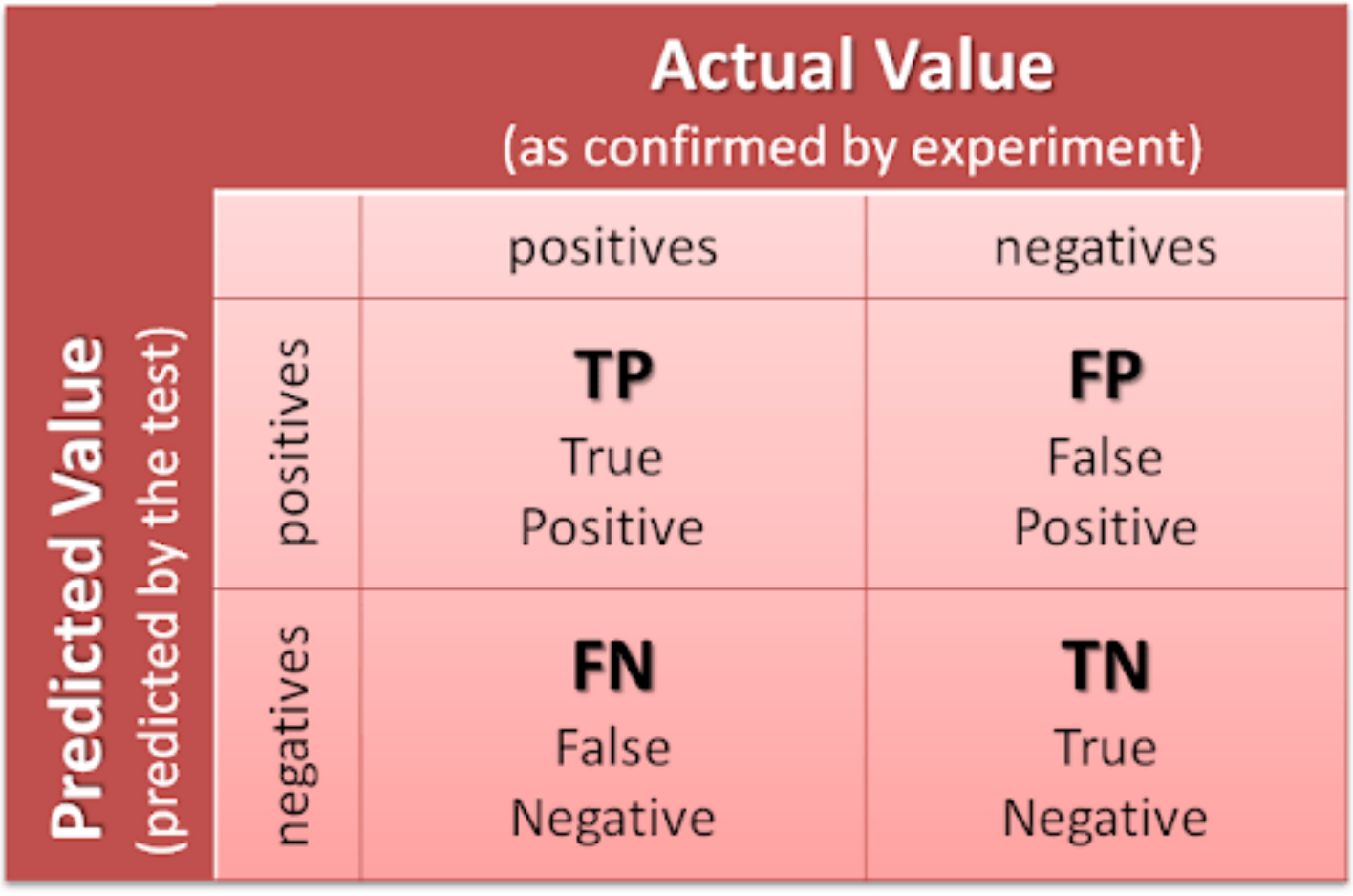
Discussion

- Classification is to learn "Categories"
  - Cats🐱 vs Dogs🐶
  - Spam or not
- Any classification tasks in your mind?
  - Share it with us?

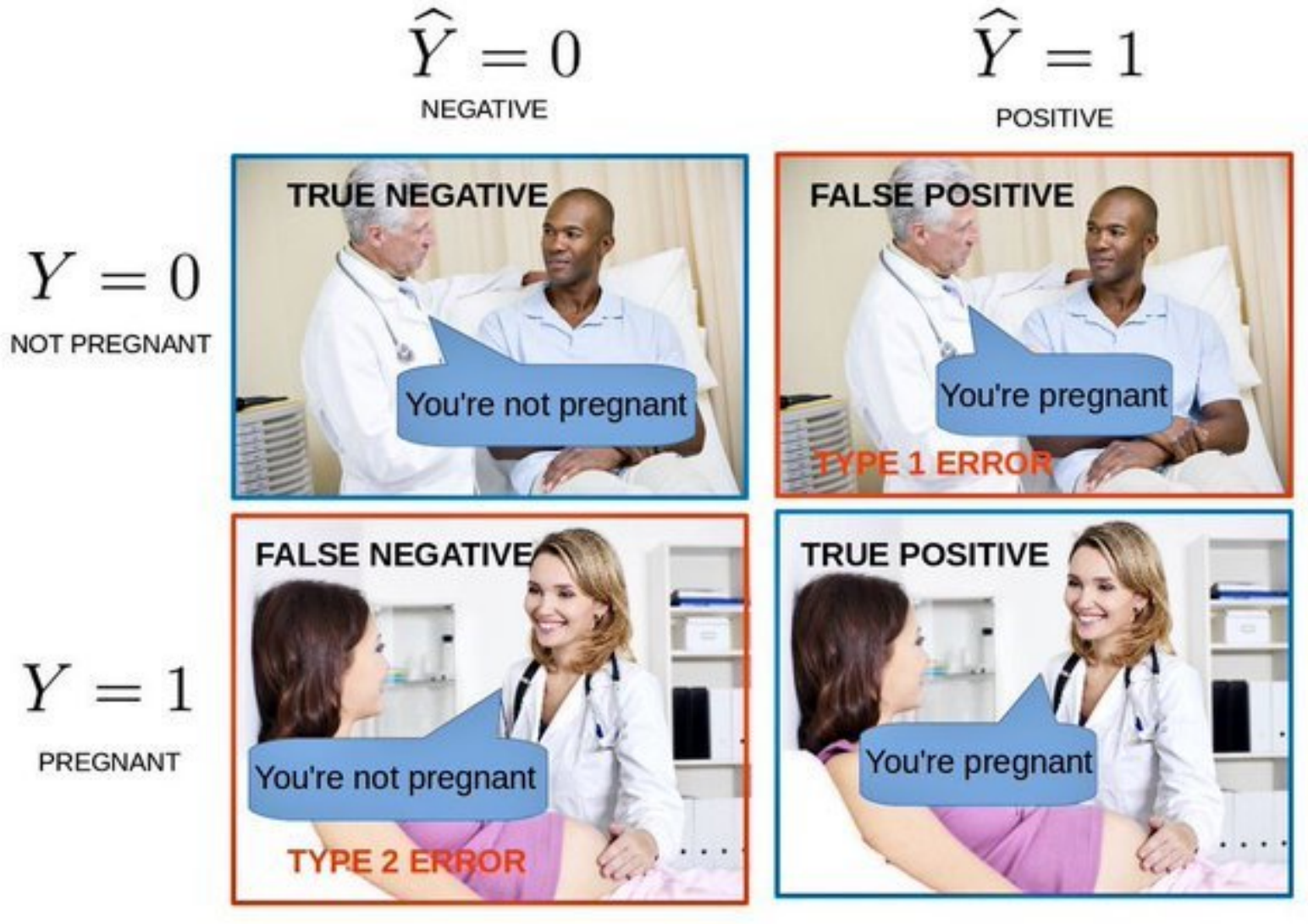
Classification Metrics

Confusion Matrix

Confusion Matrix



Example


$\widehat{Y} = 0$
NEGATIVE
$\widehat{Y} = 1$
POSITIVE
$Y = 0$
NOT PREGNANT
TRUE NEGATIVE
You're not pregnant
FALSE POSITIVE
You're pregnant
TYPE 1 ERROR
$Y = 1$
PREGNANT
FALSE NEGATIVE
You're not pregnant
TYPE 2 ERROR
TRUE POSITIVE
You're pregnant

Discussion

- When we predict that someone is positive and the actual result from the blood test is positive
  - True Positives (TP) or False Positives (FP)?
- When we predict that someone is negative and the actual result from the blood test is negative.
  - ?

Discussion

- Review
  - Classification: predict categorical label
  - Regression: predict numerical label
- Consider weather forecast
- classification or regression ?
  - Forecasting up or down
  - Forecasting temperature

3) Ranking

Ranking

- Sort through options to find the most relevant results
- Central part of information retrieval
  - Document retrieval
  - Online advertising

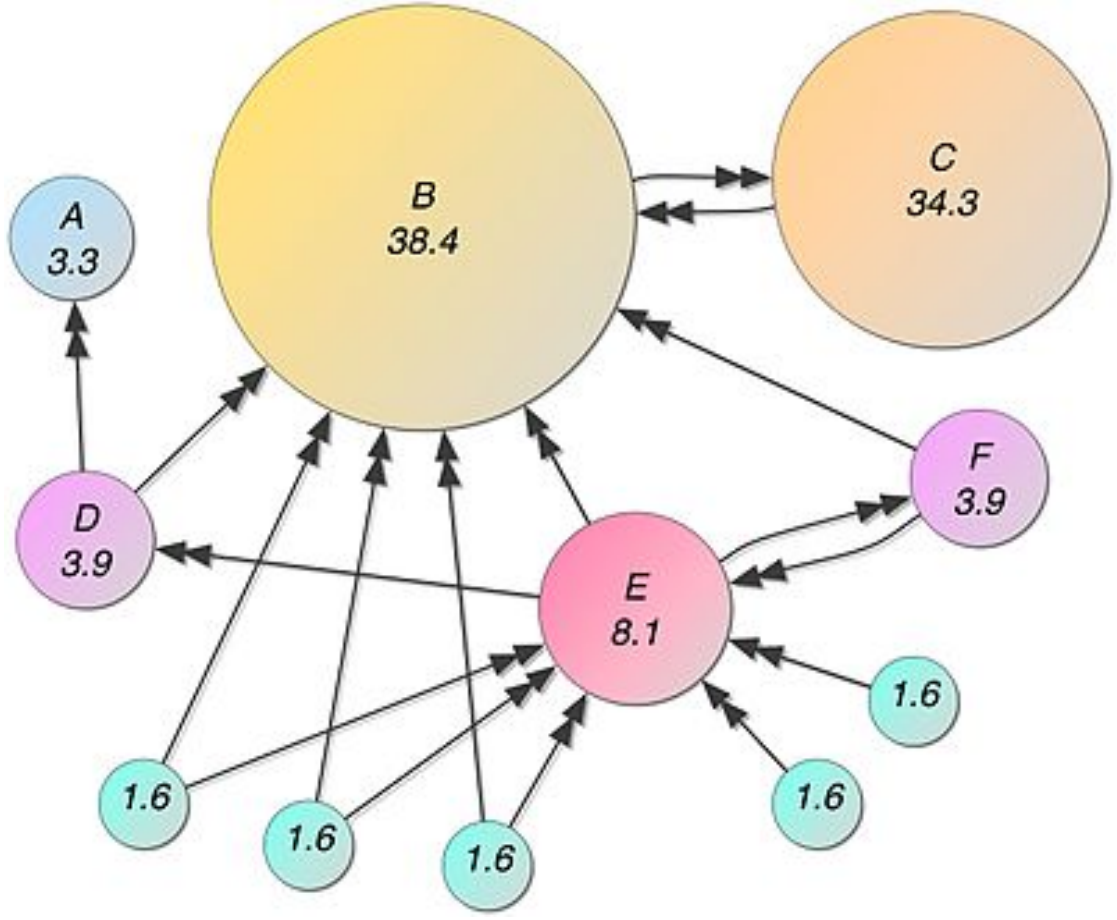
Search Engine

Ranking Features

- Google 2016
  - Hyperlinks: 29%
  - Regularly produced, original content: 23%
  - Keywords in meta info, titles, tags: 8%
  - Response speed on mobile phone and tablet: 8%
  - Landing pages optimized for conversions: 8%
  - Clean code: 6%
  - Website access speed: 5%
  - Popularity in social networks: 4%
  - Age: 4%

PageRank

Random access to all hyperlinks, the probability of coming to a page



4) Recommendation

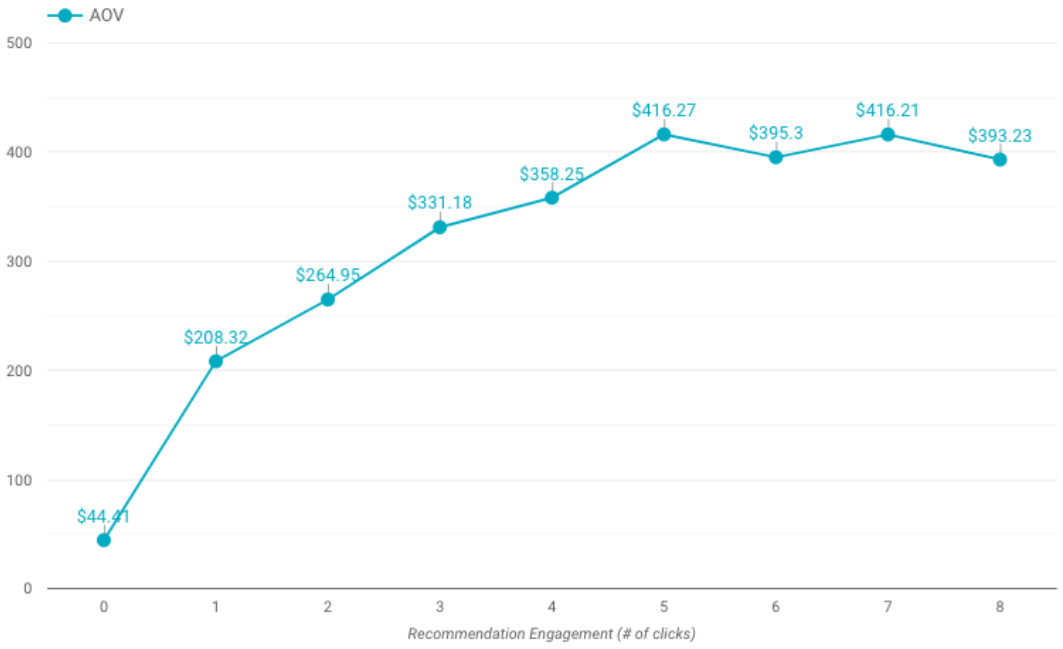
Product Recommendation

News Recommendation

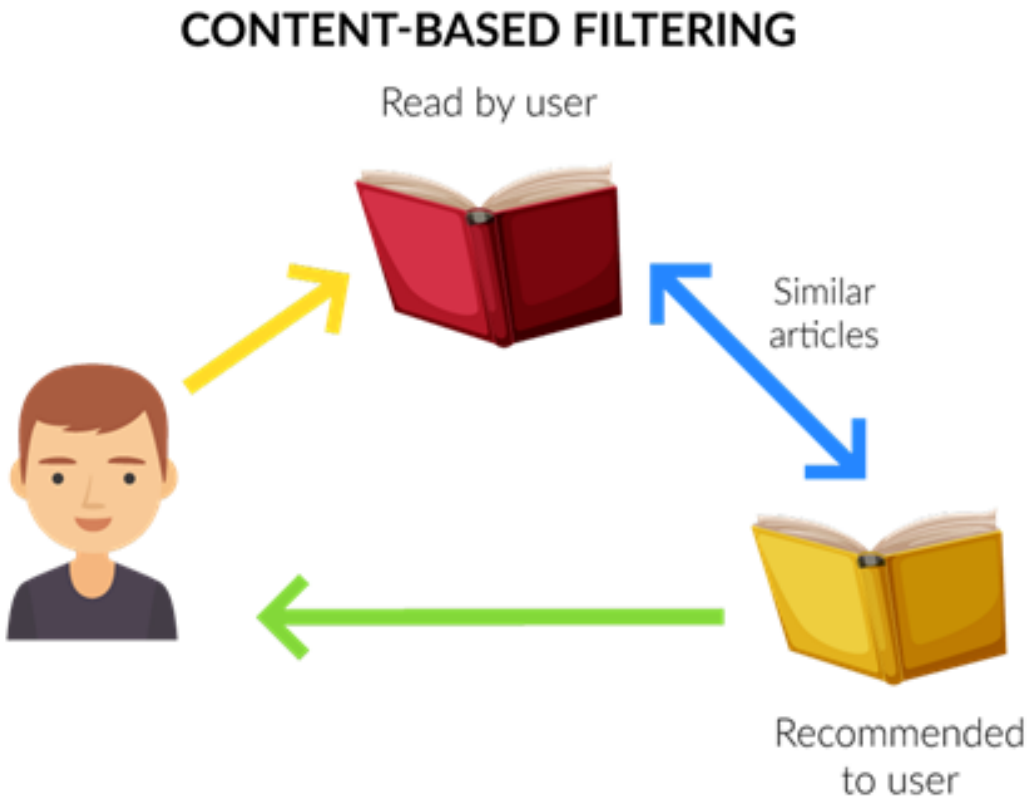
Efficiency

- Personalized product recommendations dramatically increase AOV (average order value)
  - Saleforce: 4.5x more likely to add items to cart

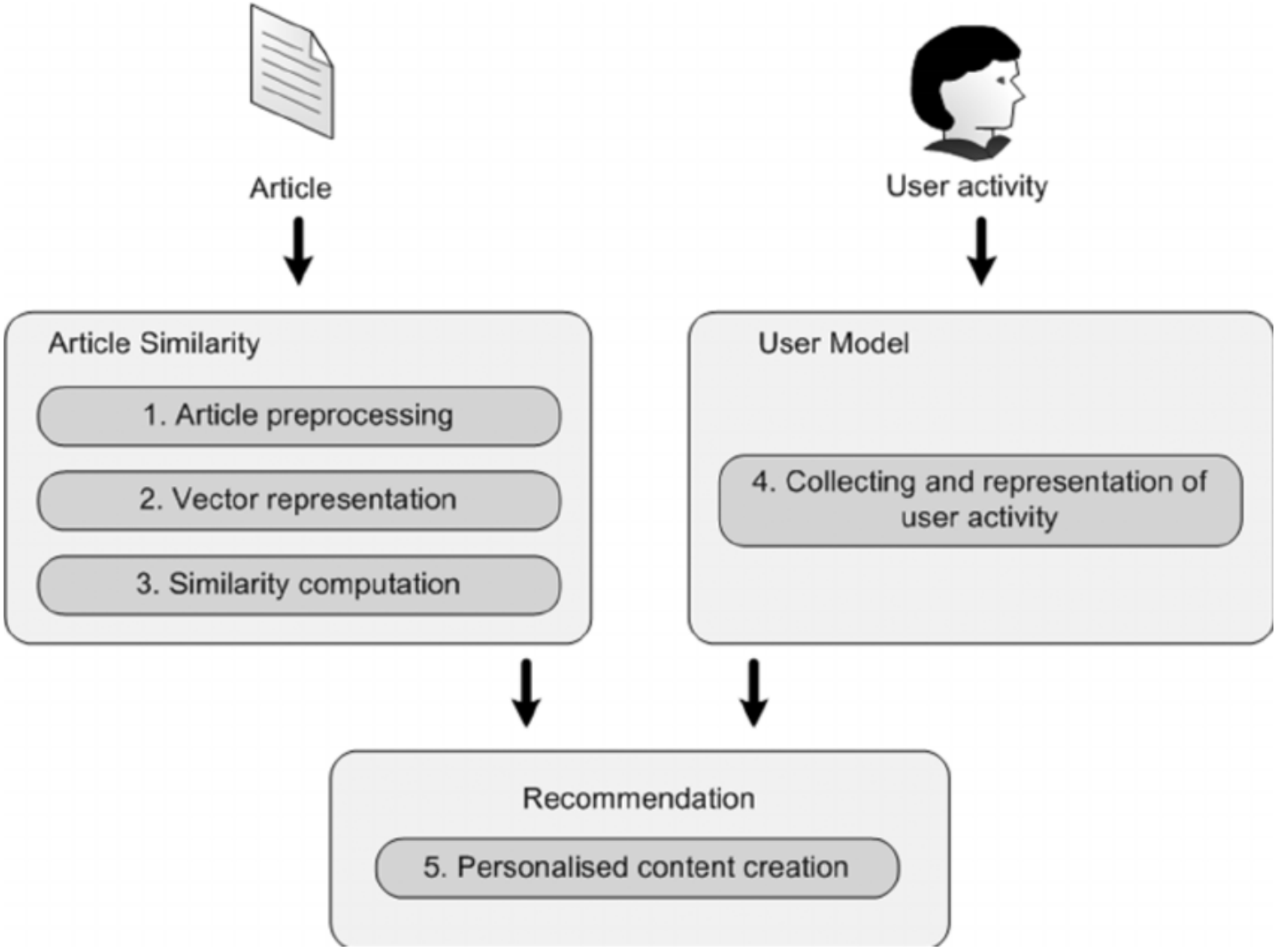
How Personalized Product Recommendations Increase AOV | 2018 Data



Content-based Recommendation

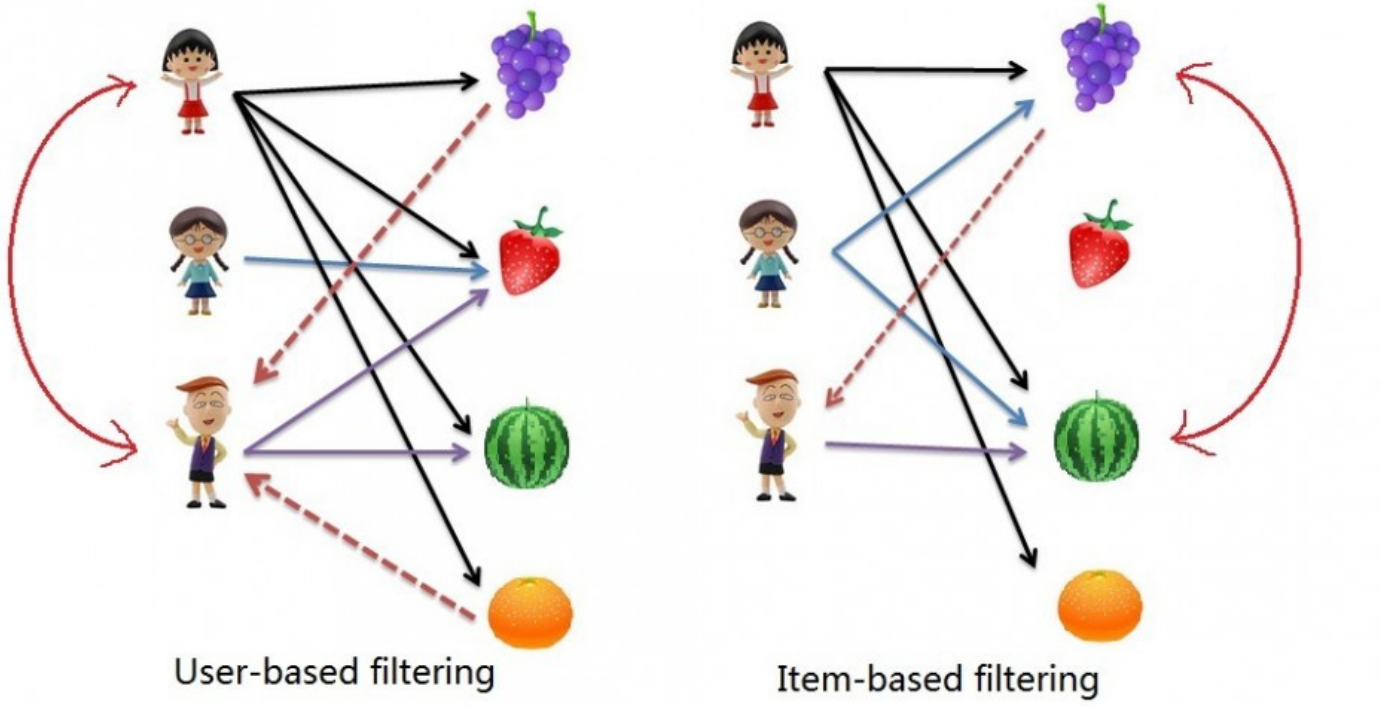


News Recommendation



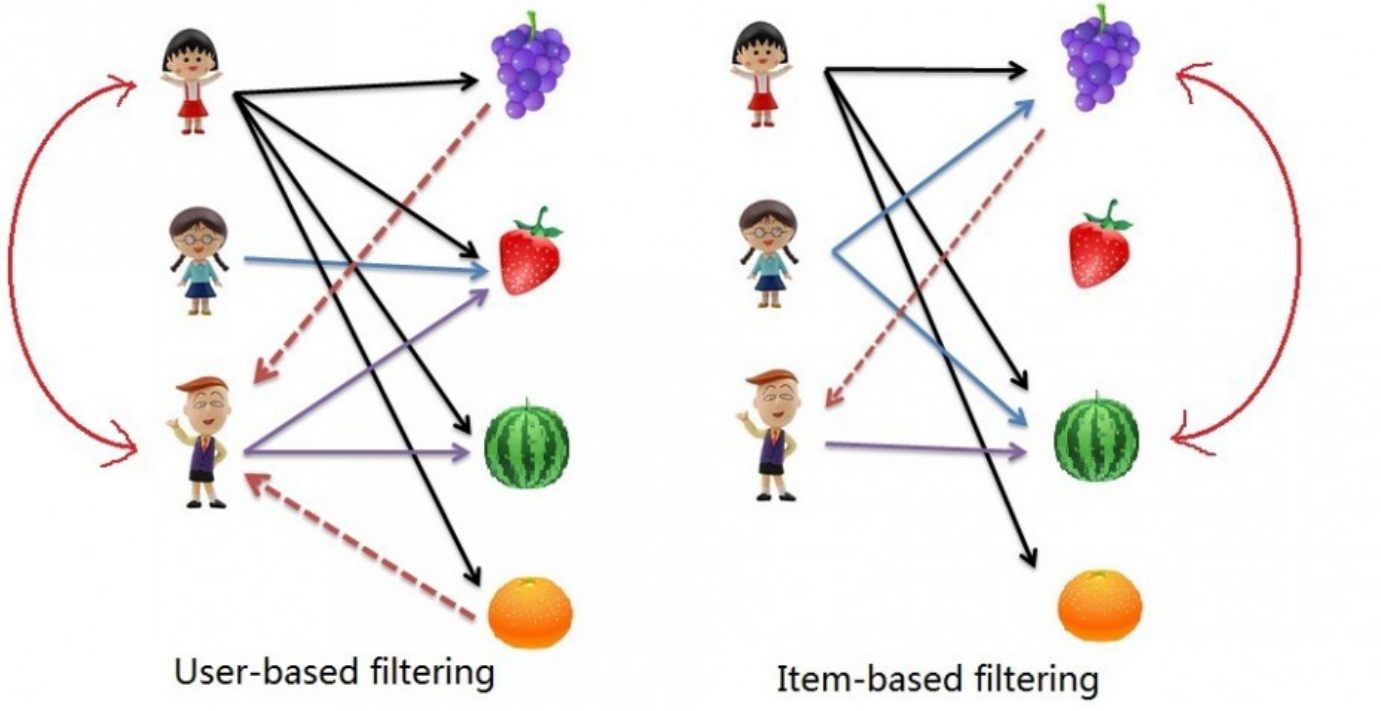
Collaborative Filtering

- User-based
  - Find people with similar interests (share items) as you
  - Recommend his favorite to you



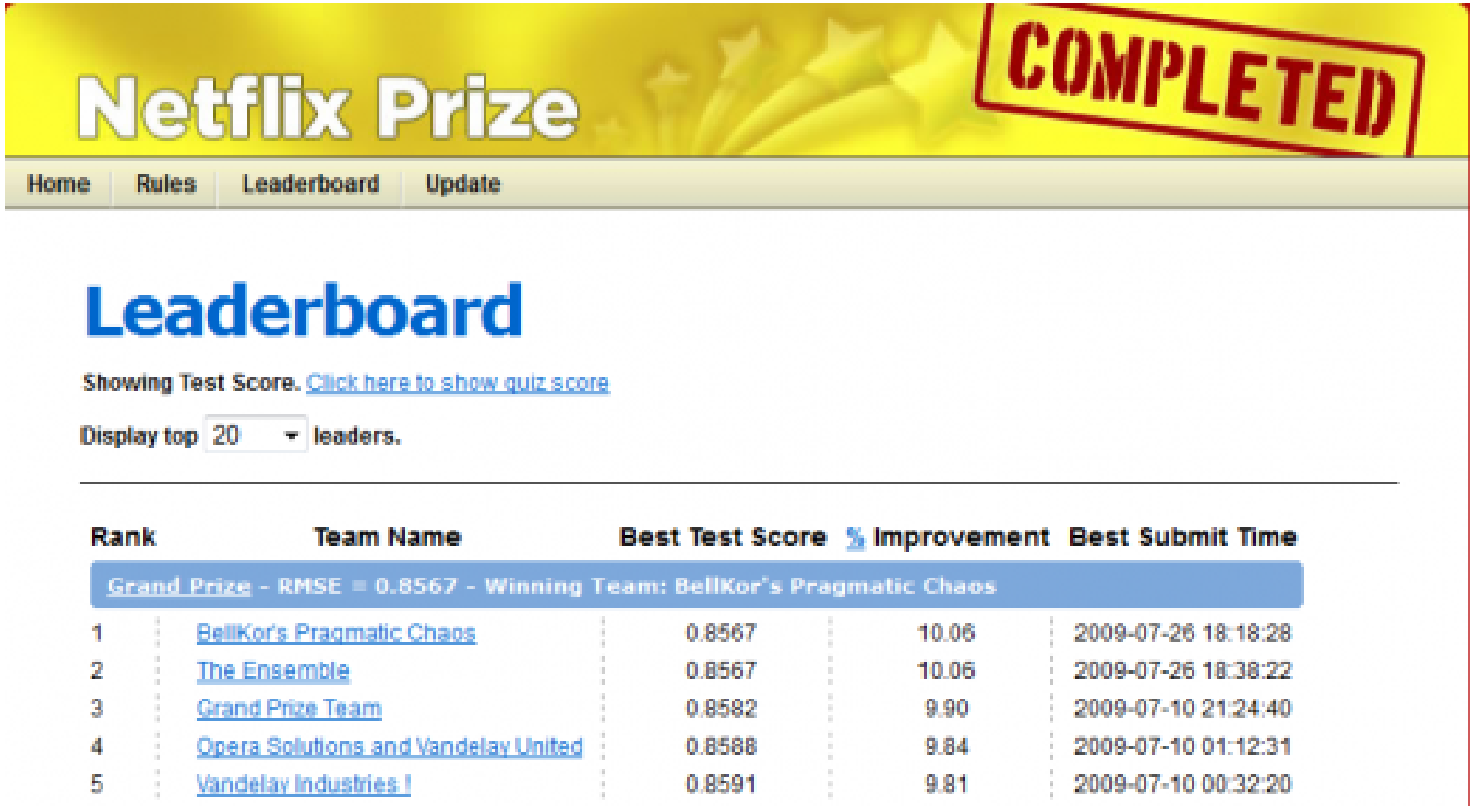
Collaborative Filtering

- Item-based
  - Find movies similar to those you liked (share users)
  - Recommend them to you



Netflix 1M Prize

- Improve accuracy of rating predictions



Netflix 1M Prize

- Highly engaged research community

Metric

- CTR = No. of Clicks/No. of Impressions

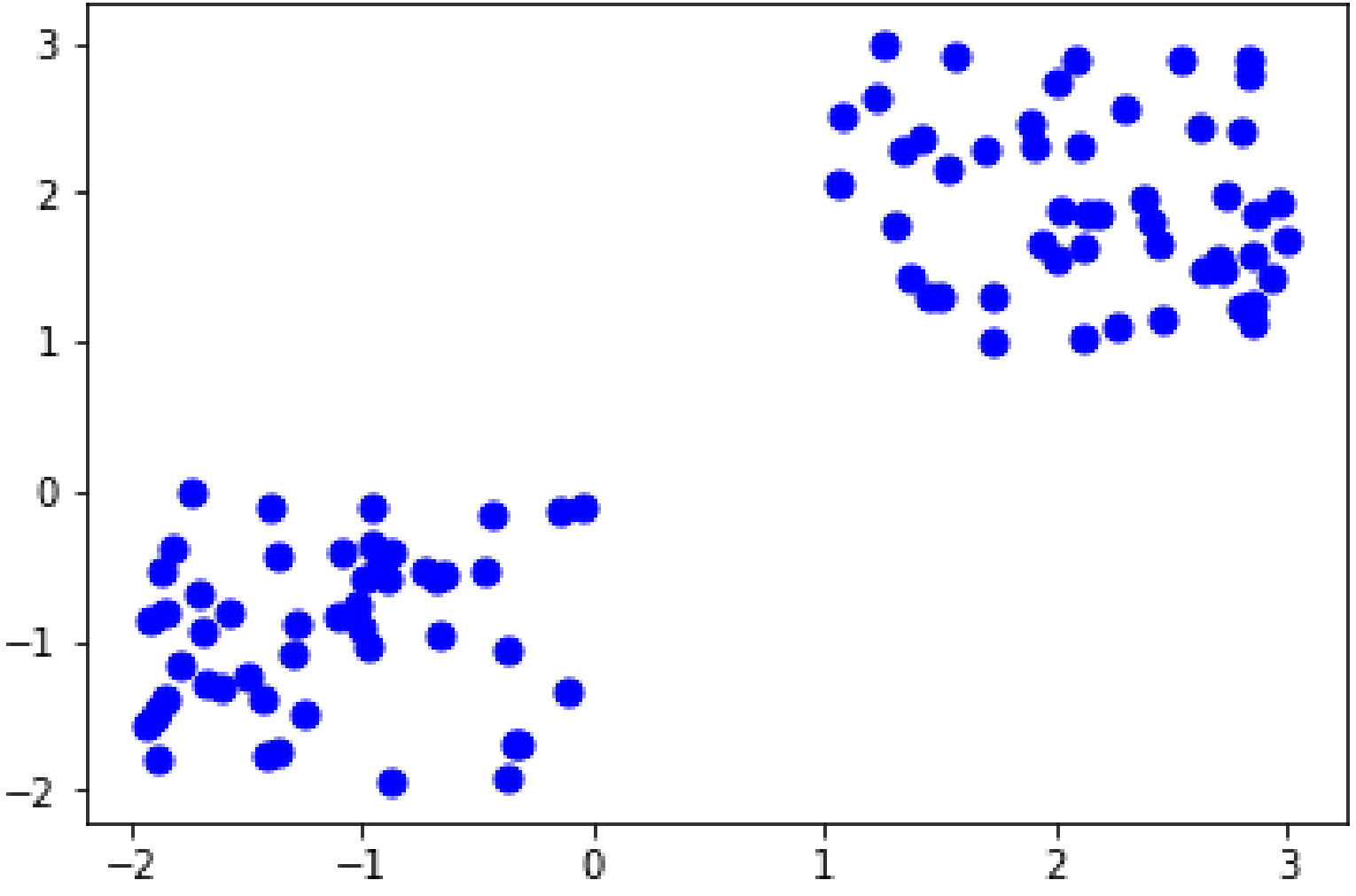
Quiz

- User-based or item-based collaborative filtering?
  - Find people with similar interests as you, recommend his favorite to you
  - Find movies similar to those you liked, recommend them to you

5) Clustering

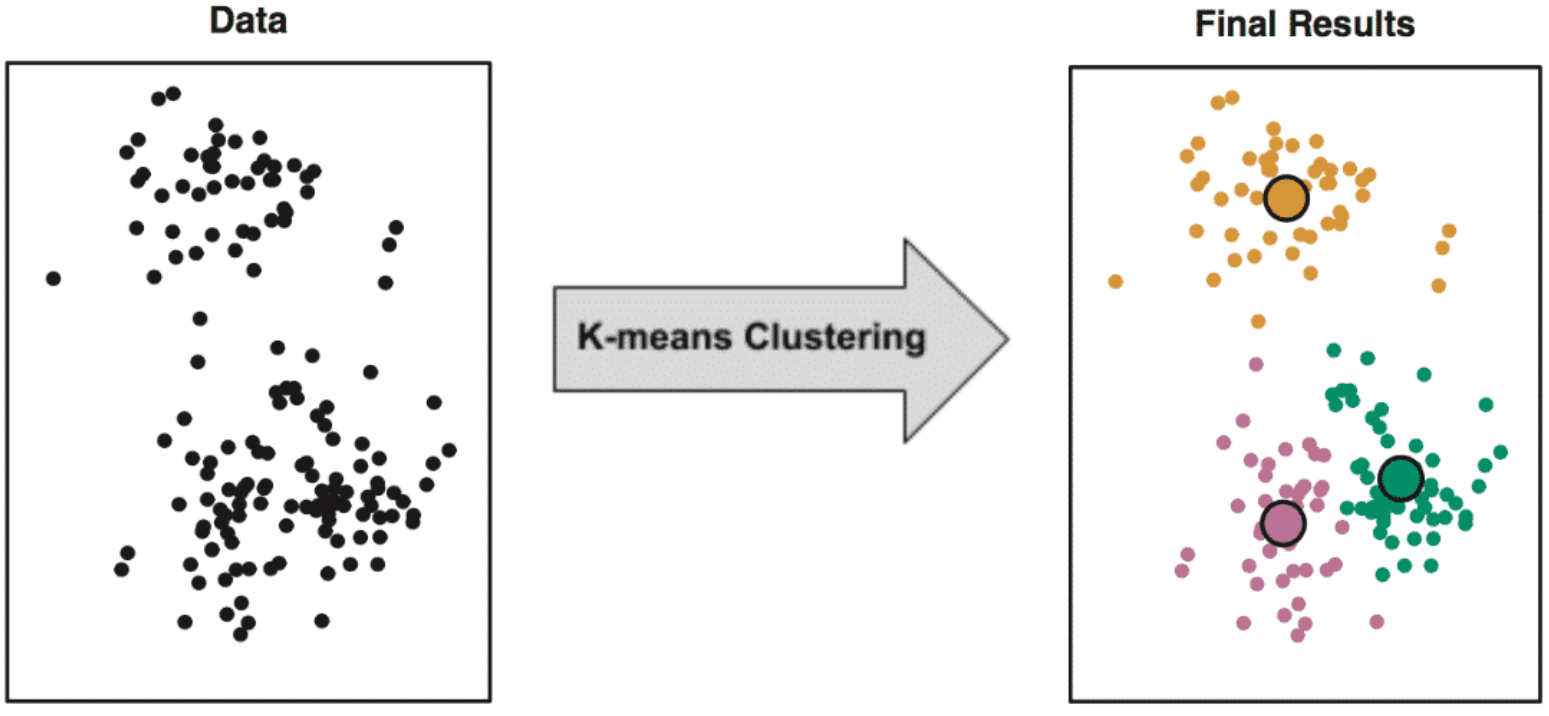
Clustering

Clustered samples as a class



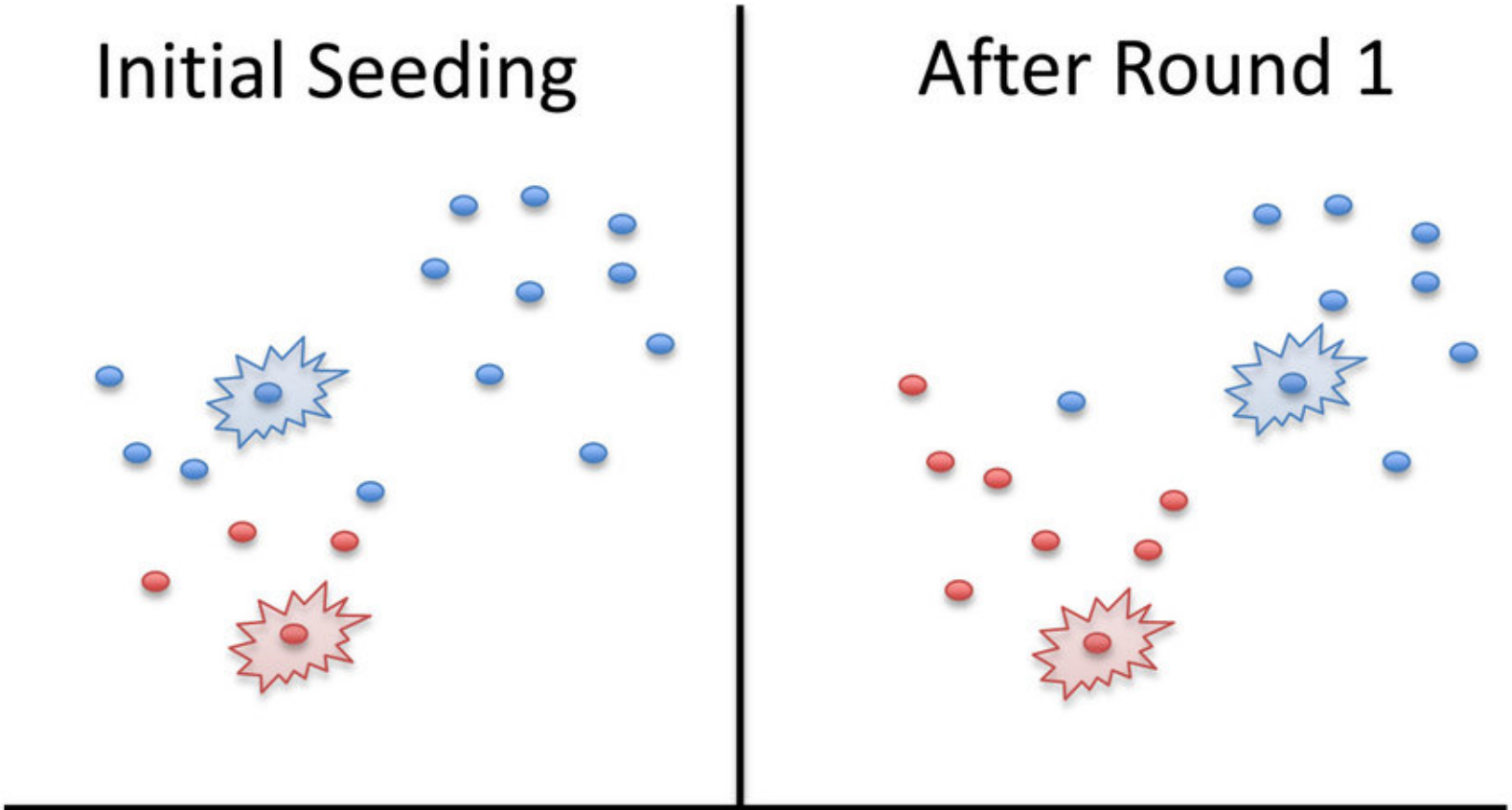
Clustering

Need specify number of clusters: 3

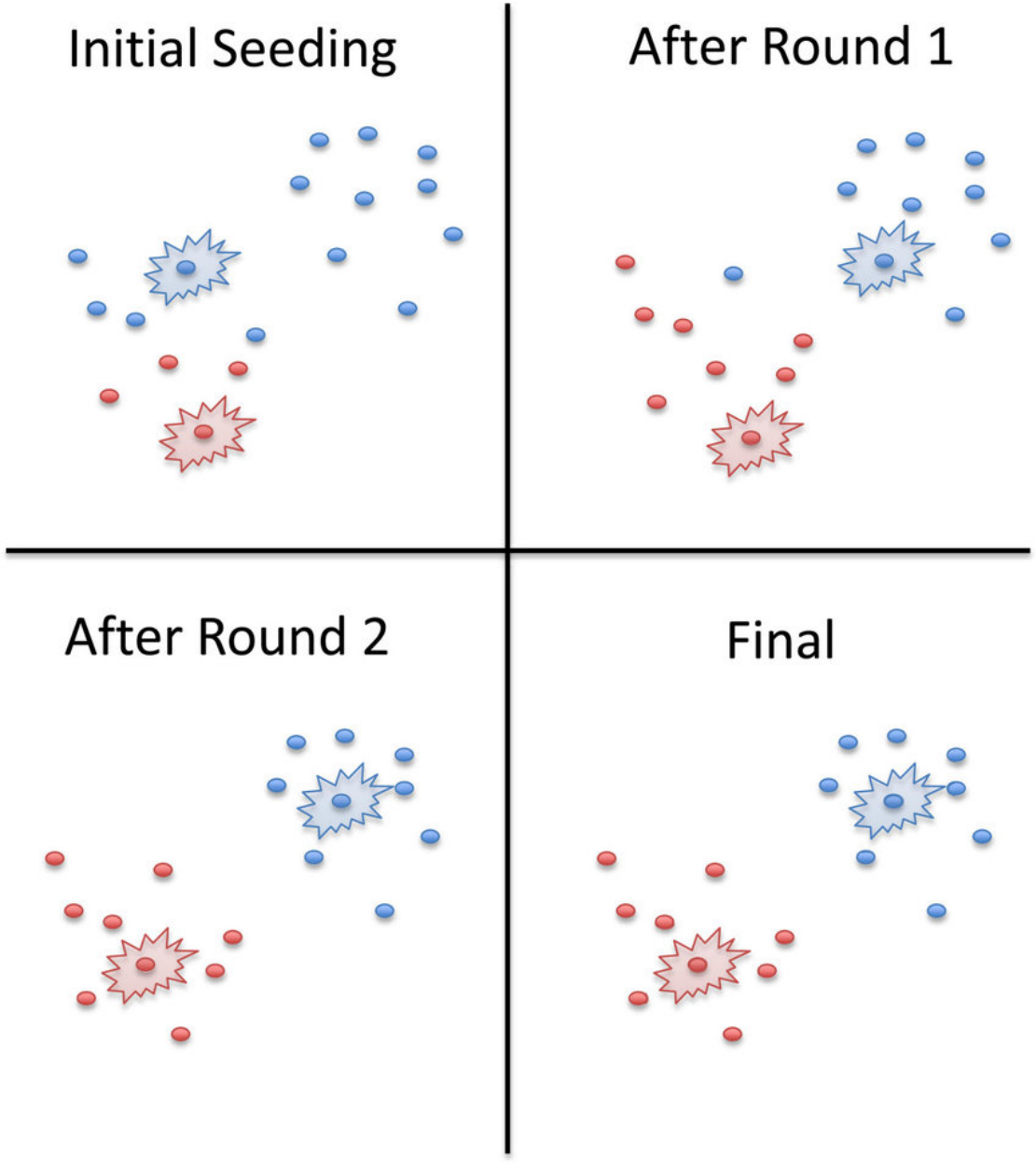


K-Means Clustering Method

- Randomly initialize cluster centroids
- Repeat until convergence
  - Assign observations to the closest centroid
  - Recalculate centroids

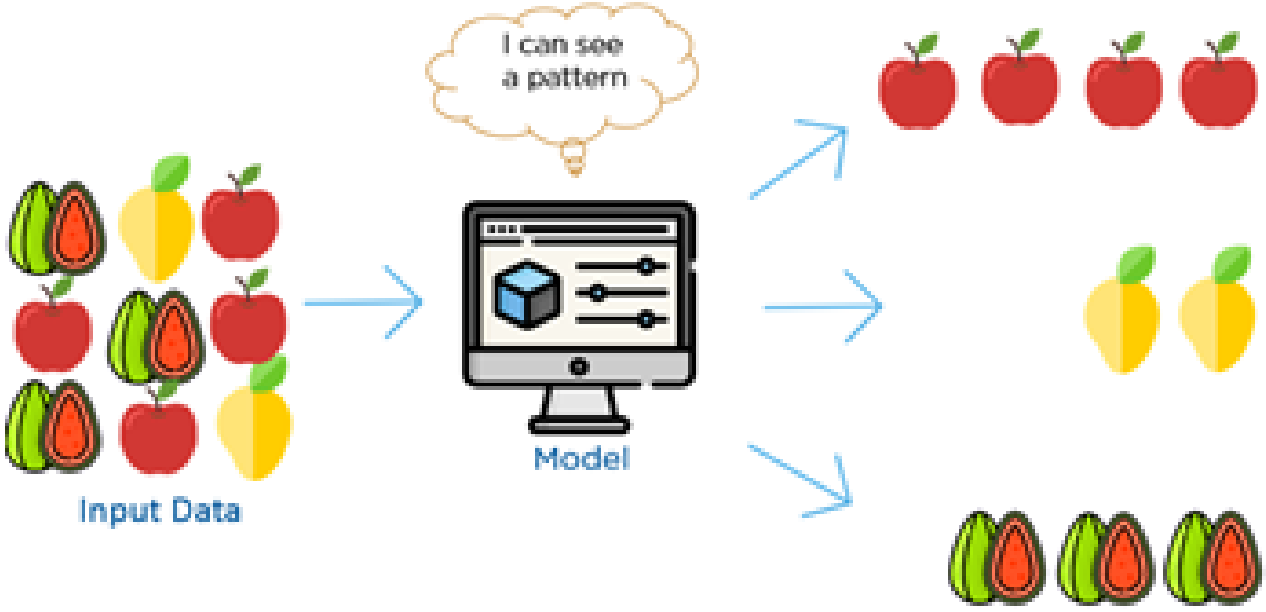


K-Means Clustering Method



Understanding Results

- After clustering, observe each cluster to get its meaning
  - The result might look like this:

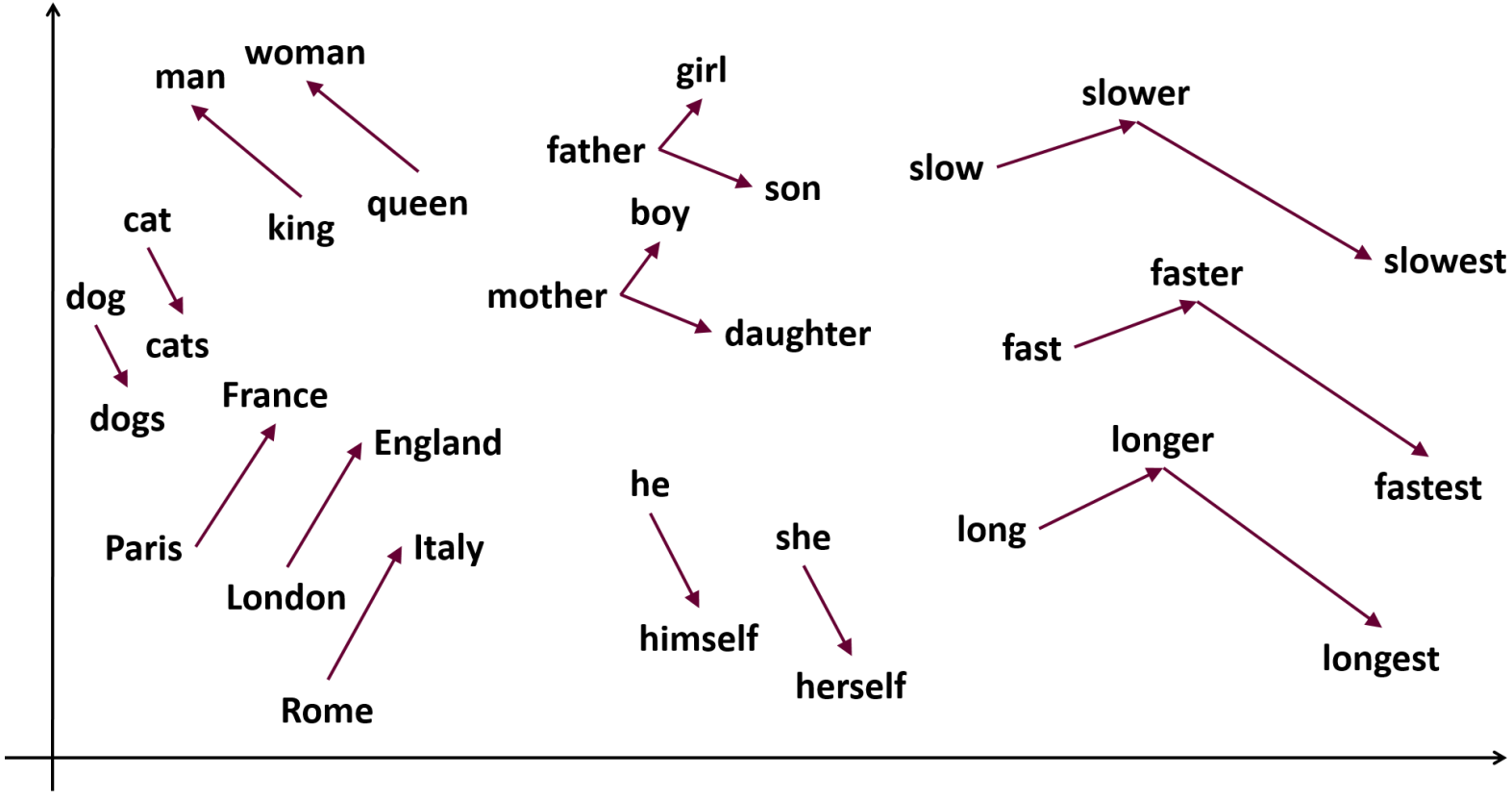


6) Representation

Represent complex data as mathematical vectors

Word Representation

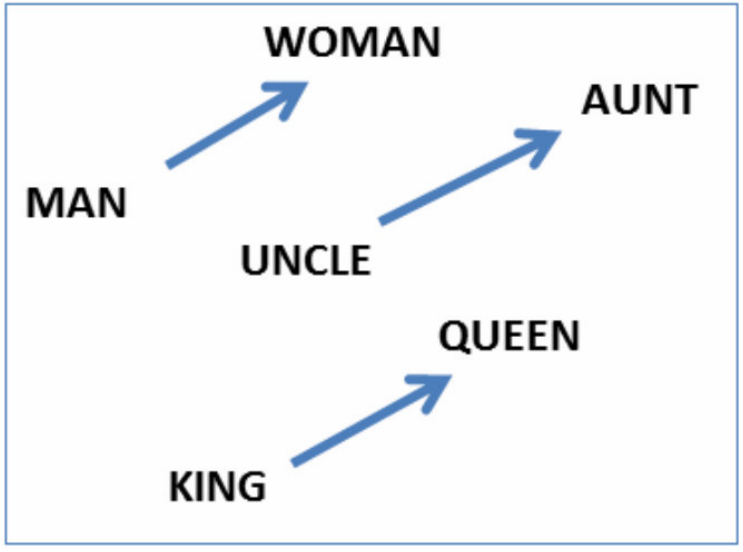
Word positions indicate their meaning



Word Representation

The distance vector between two words indicates their semantic relationship

- Gender relationship
- Singular and plural relations



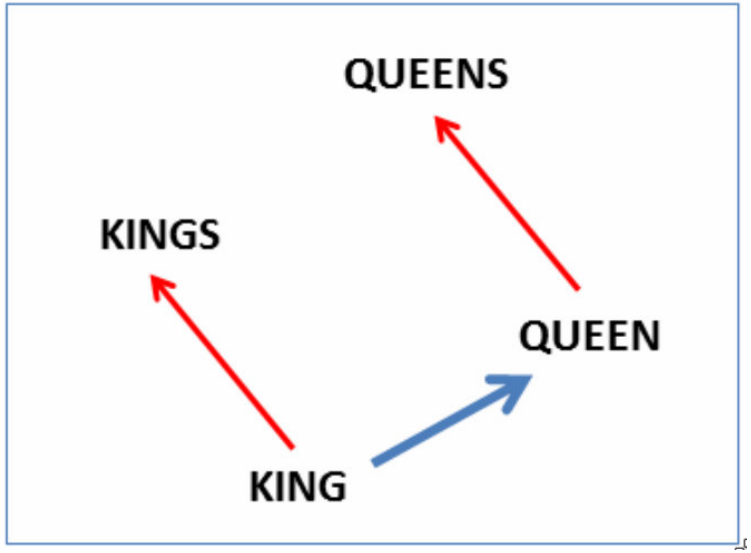
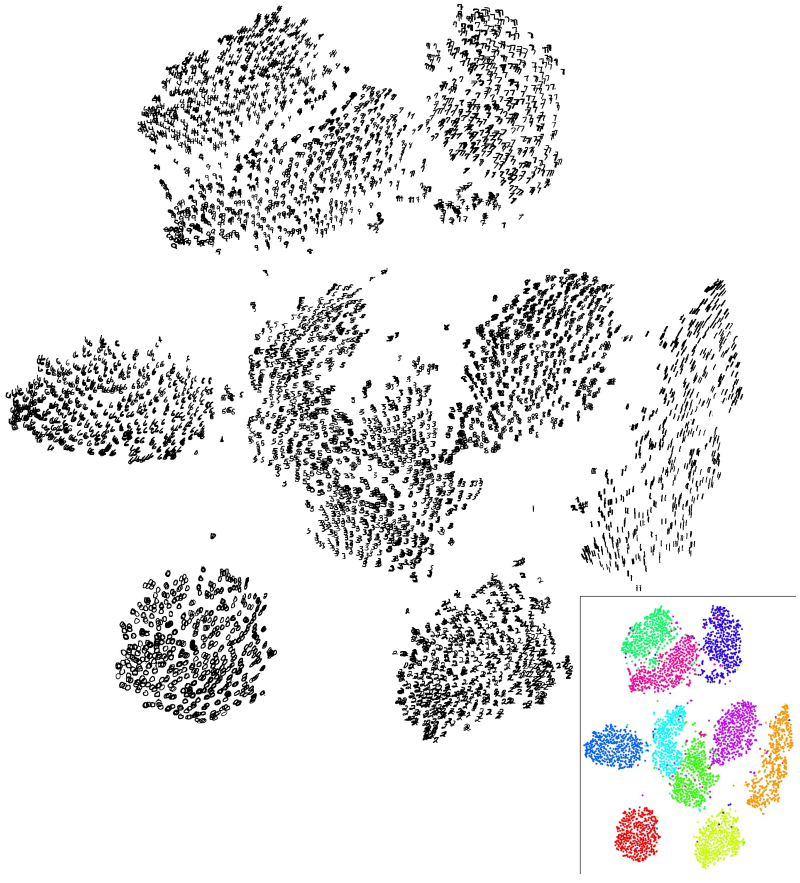


Image Representation

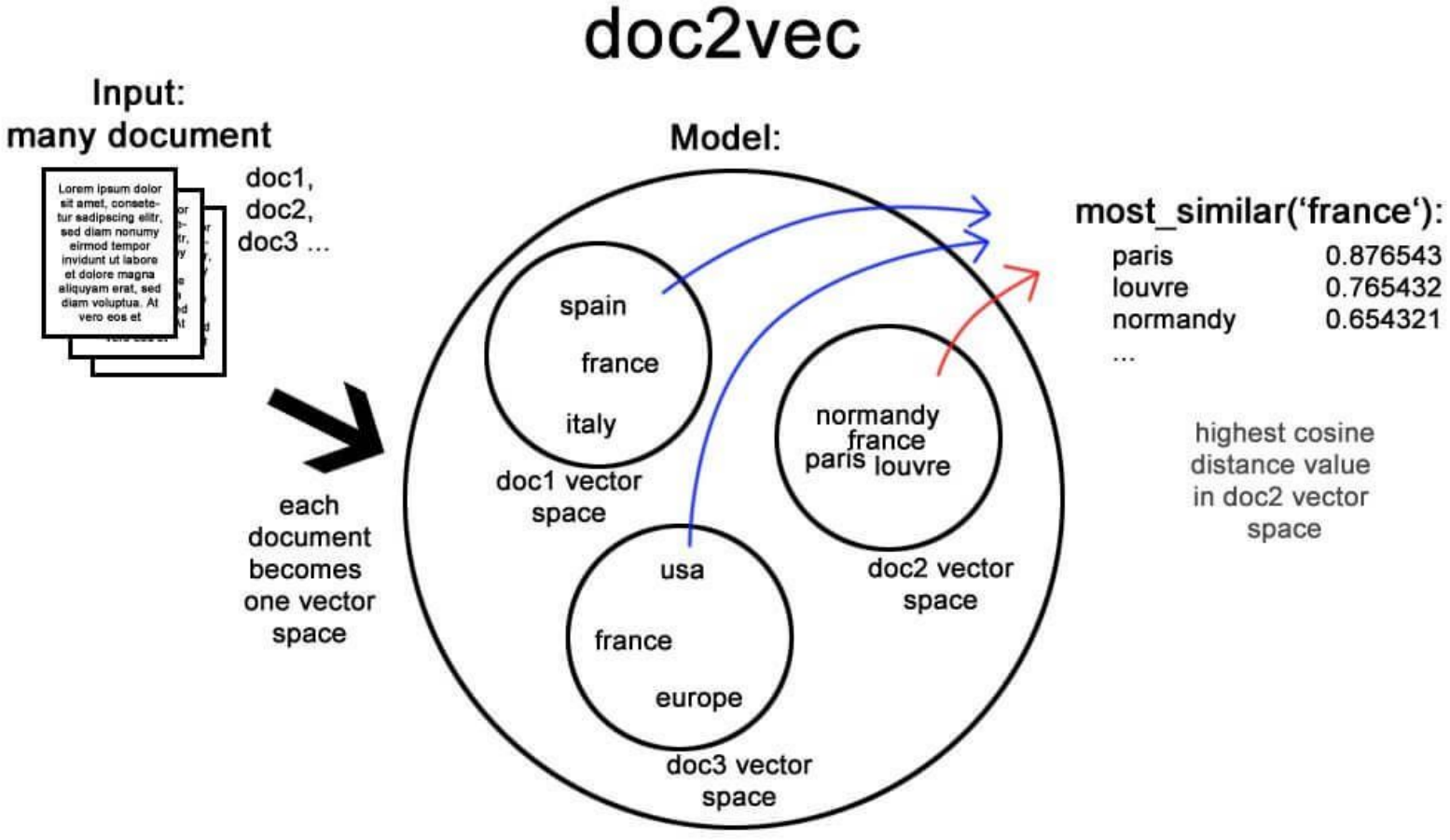
Picture position indicates their content



MNIST 0-9 digit pictures, clustered into 10 clusters

Document Representation

Find similar document better

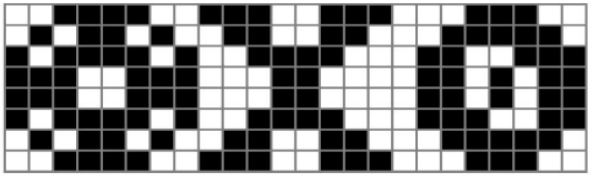


7) Structured Prediction

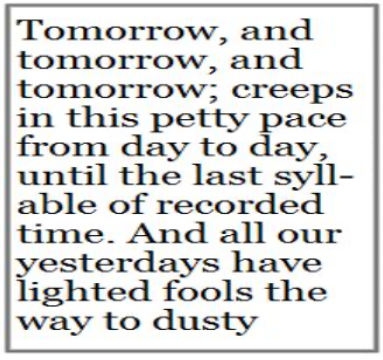
Leveraging structured information

Structured data

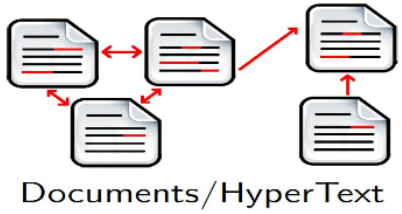
- Data that consists of several parts
  - Not only the parts themselves contain information
  - But also the way in which the parts belong together
- Text, Image, Documents are type of structured data



Images

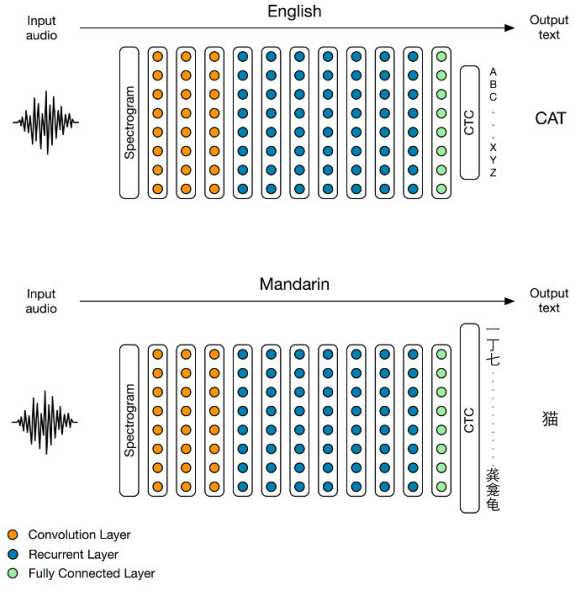


Text



Speech Recognition

- The next word depends on the previous words
- Therefore, you can use the previous words to predict the next word



[Baidu 2014]

Man-machine dialogue

[Google Inbox Smart Reply]

[Amazon Echo / Alexa]

Auto Translation

- Video: ASR speech translation and synthesis

播放器初始化...[完成]
加载用户配置...
加载视频地址...

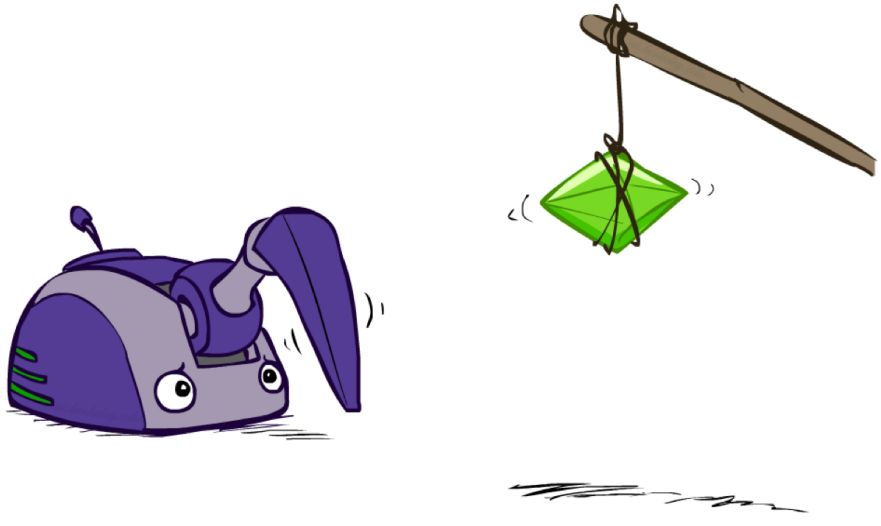
00:00 / 00:00

8) Reinforcement Learning

Learning based on the rewards received

Reward-Based Learning

- No labeled data set
- There is a reward
- Learning based on the rewards received
- Goal: maximize reward

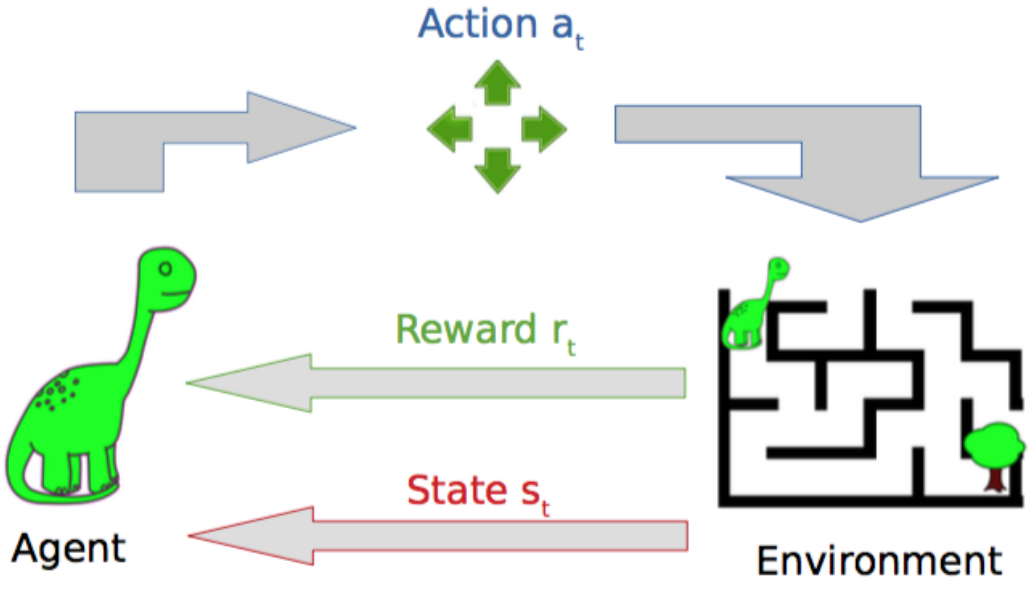


Reinforcement Learning

- Make a lot of experiments
- Don't be afraid to jump into the fire pit
- Replay

Reinforcement Learning

- Keep trying
- Get the "value" of each position
- Or get the best action in every position



Application

- Robot
- Game
- Automatic control

Game

Video: OpenAI Dota

Walk

Video: DeepMind Walk

Robot

Video: RoboCup 2018

Review

- Machine Learning Tasks
    1. Classification
    2. Regression
    3. Ranking
    4. Recommendation
    5. Clustering
    6. Representation
    7. Structured prediction
    8. Reinforcement learning

Content

- Definition
- History
- Type
- Application
- Concerns

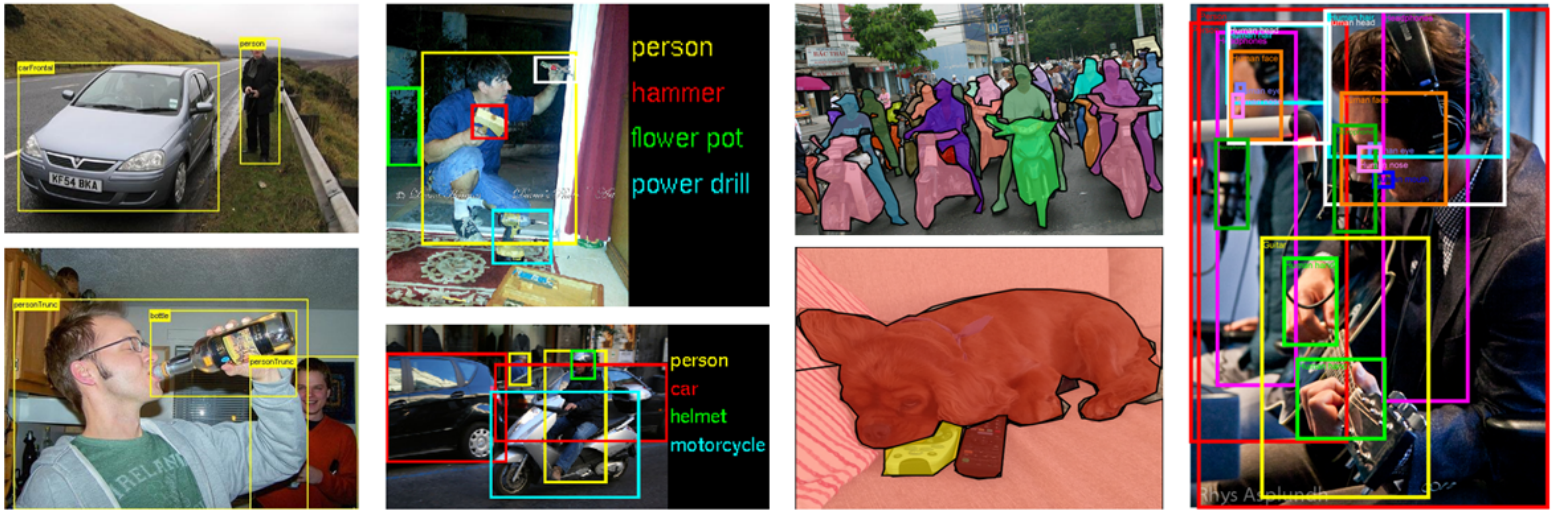
Application

Generating Image Descriptions (2015)

播放器初始化...[完成]
加载用户配置...
加载视频地址...

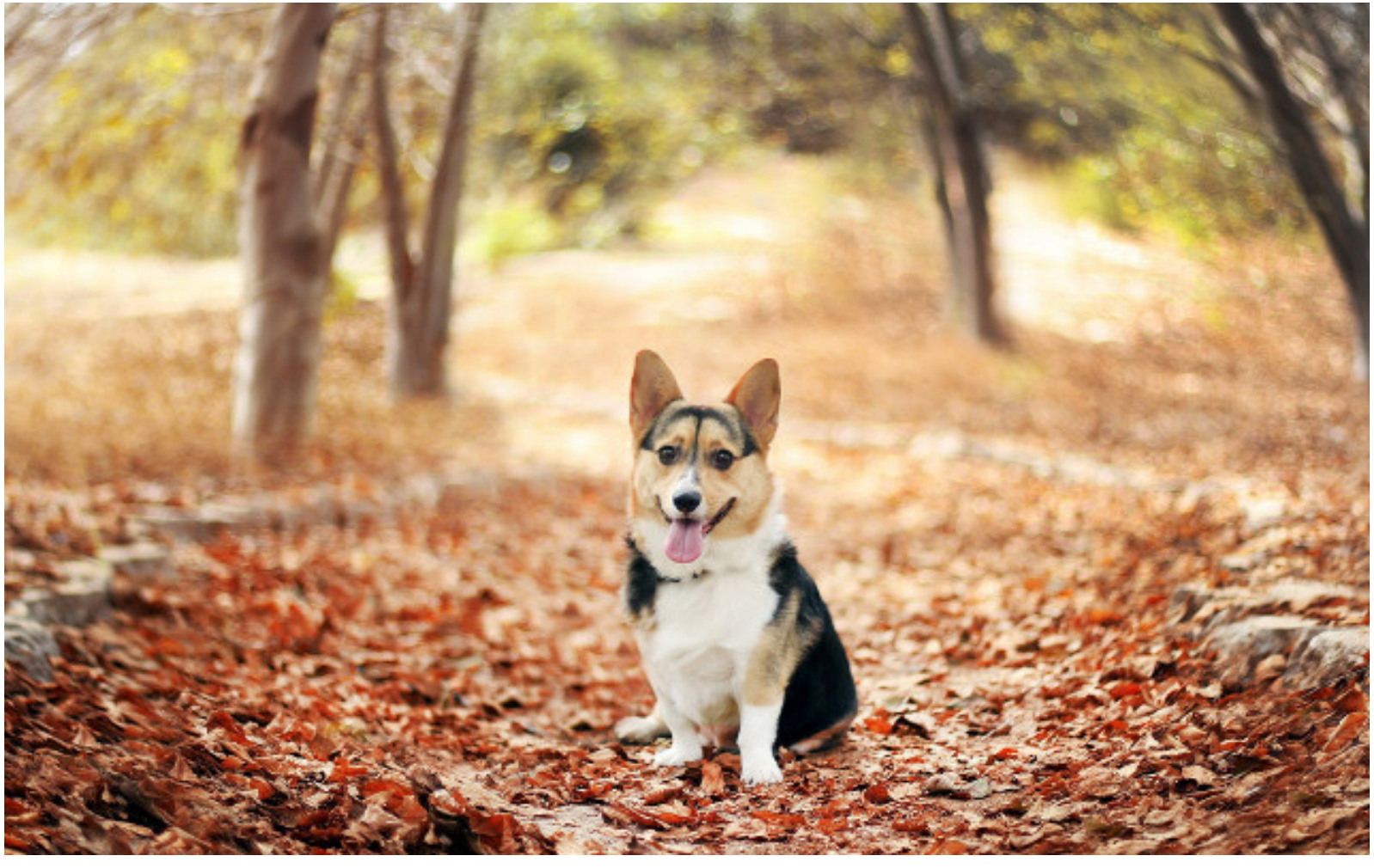
00:00 / 00:00

Object Detection & Recognition

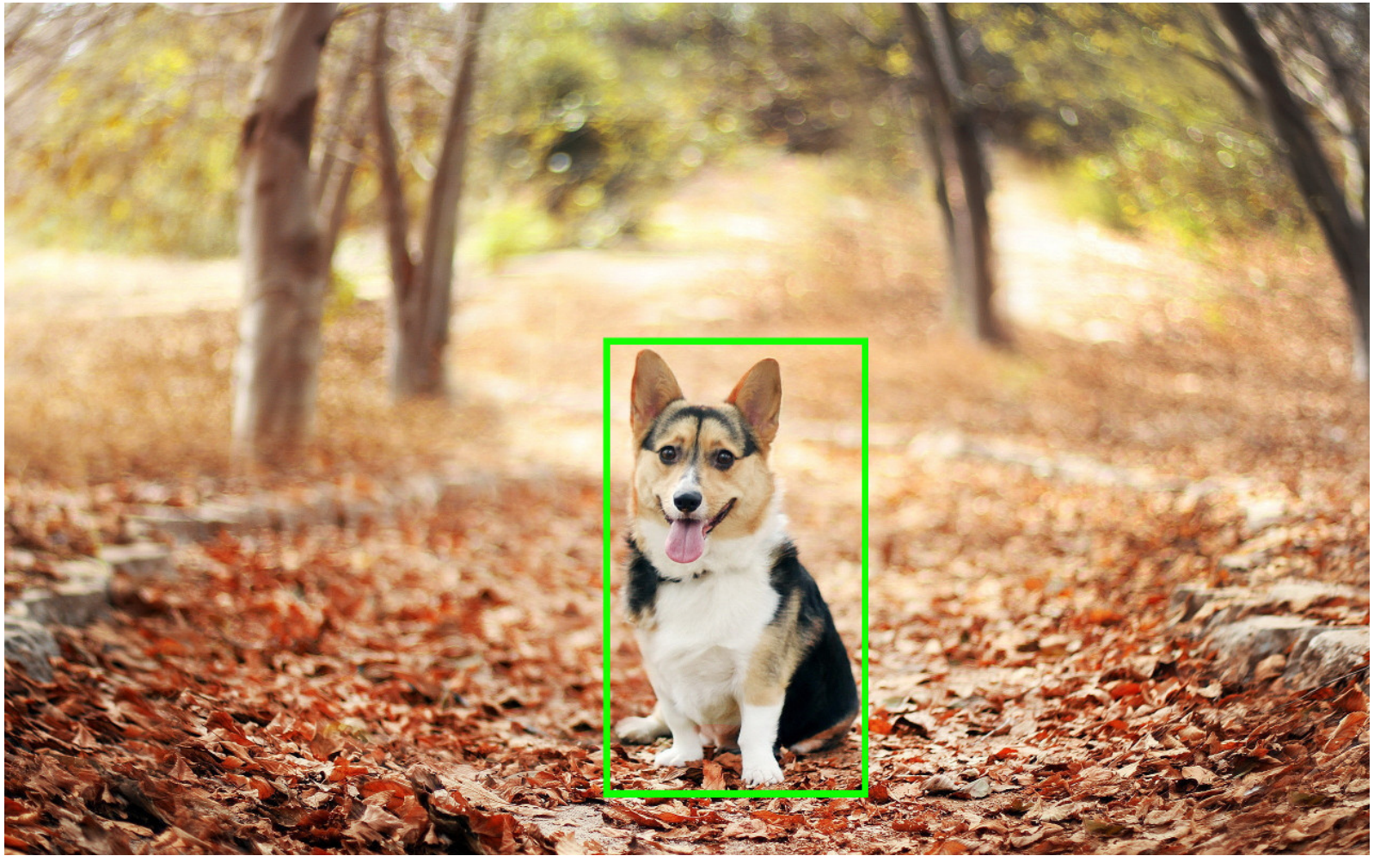


(a) (b) (c) (d)

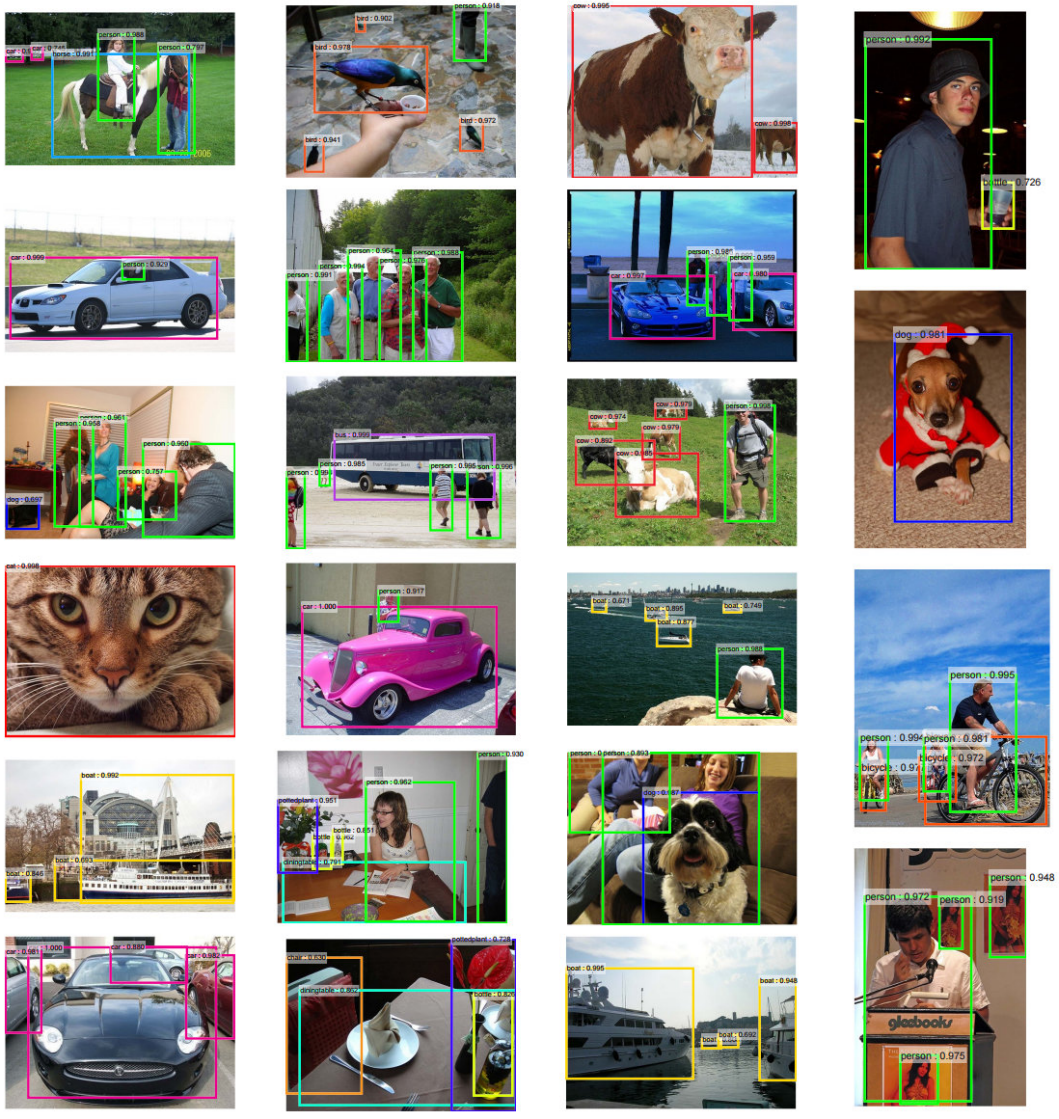
Image



Object Detection



Object Detection



Segmentation

播放器初始化...[完成]
加载用户配置...
加载视频地址...

00:00 / 00:00

进入bilibili,一起发弹幕吐槽!

去吐槽

Segmentation

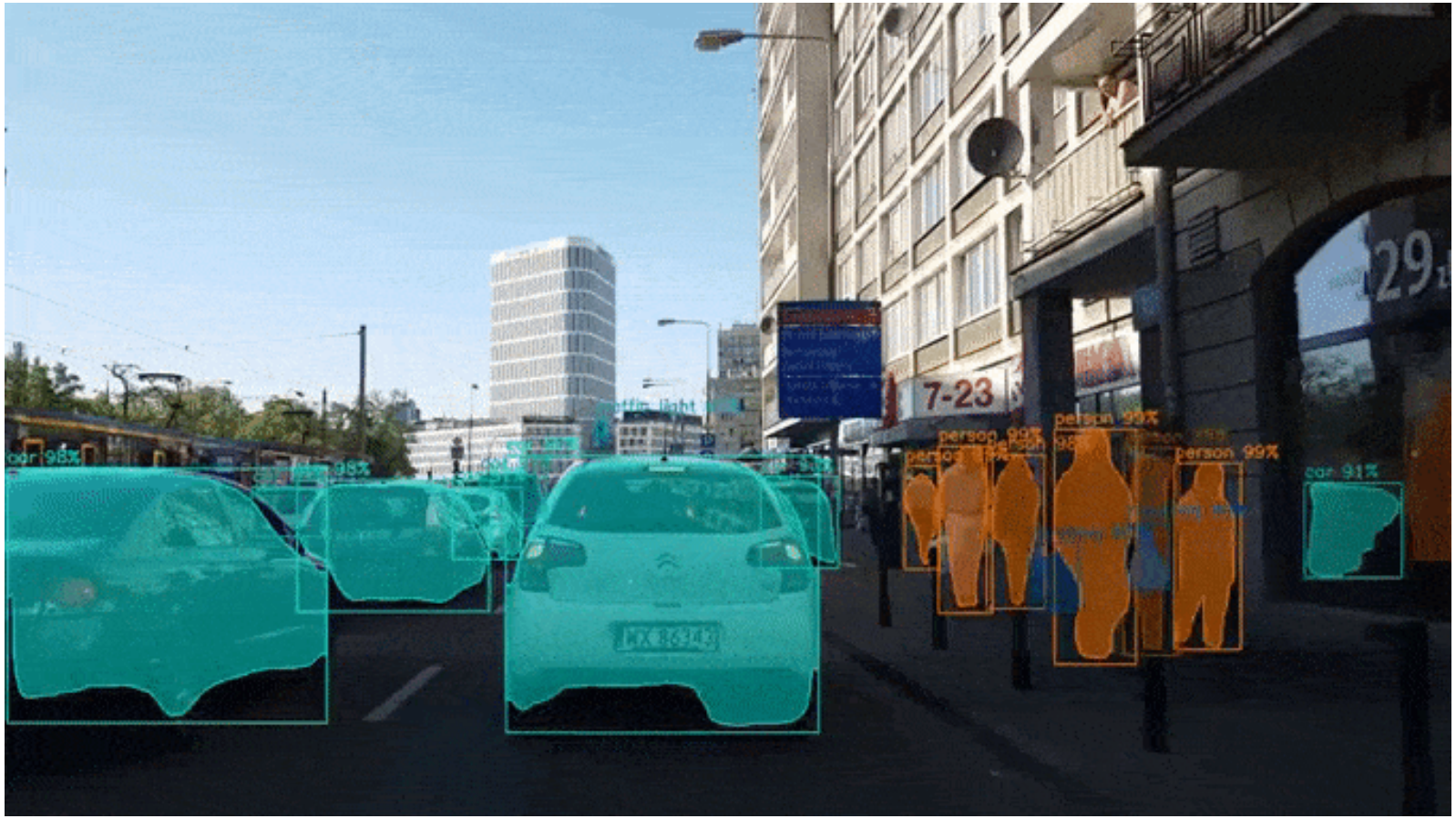
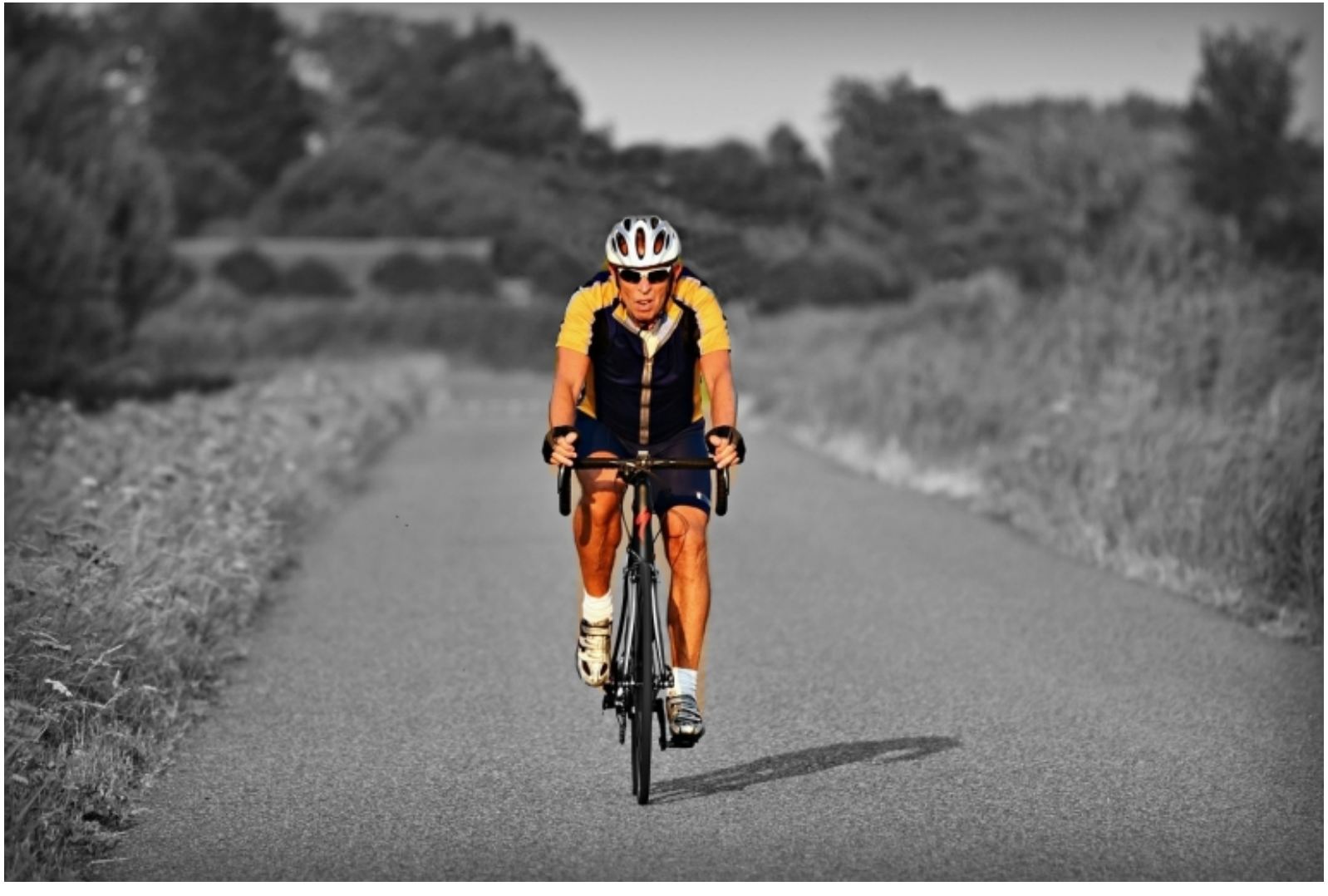


Photo Effects



Tracking and Coloring Object

3D Buildings

Face Detection

(a)

(b)

(c)

Face Recognition

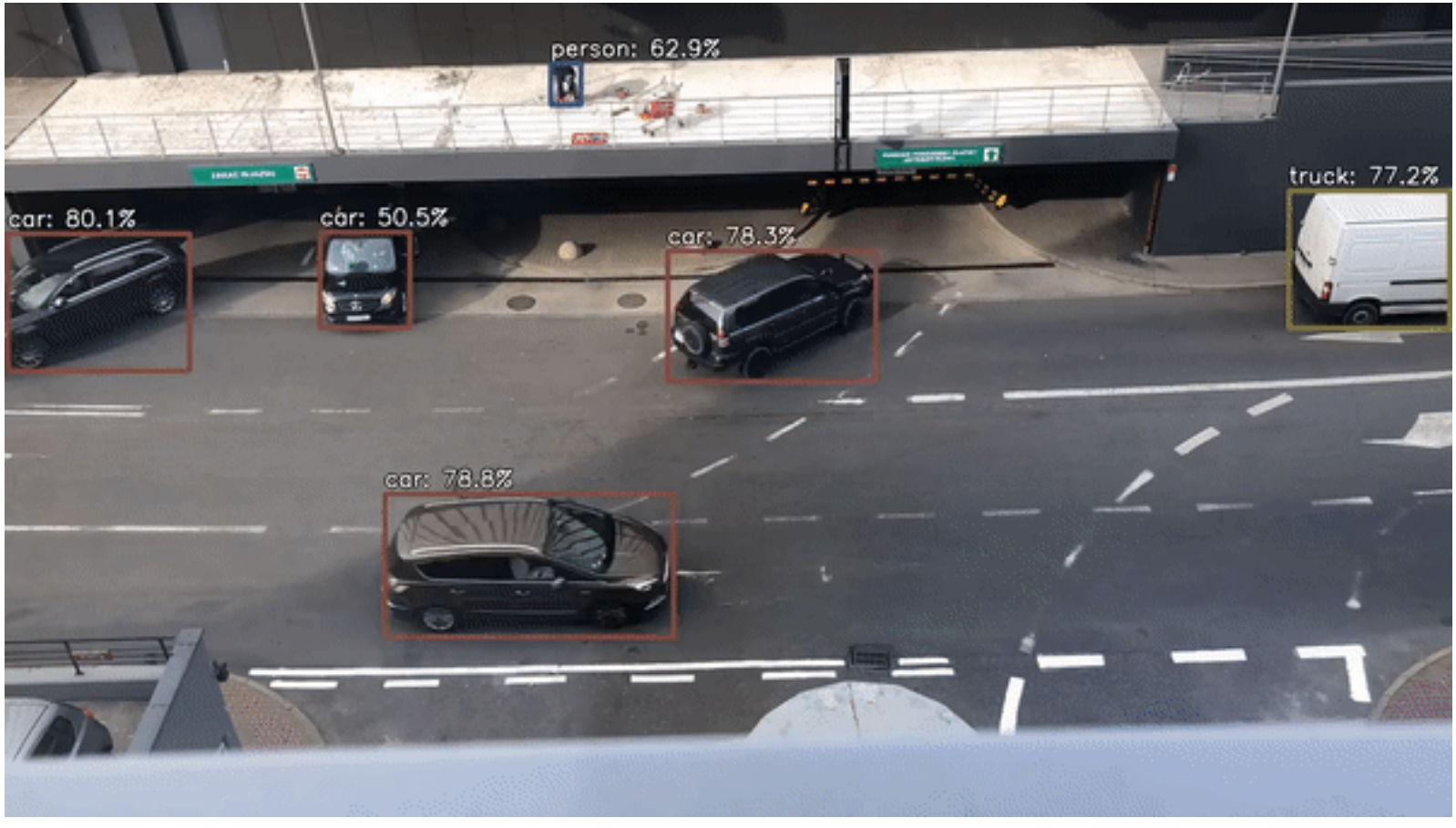
- Face recognition technology finds suspect in Maryland shootings
- Pop star Taylor Swift, filtering fans and followers at concerts
- Shelter tracks use of shelters

Pose Detection and Recognition

Emotion


Engagement
100%
Brow Raise
5%
Brow Furrow
0%
Valence
100%
Lip depressor
0%
Smile
100%

Traffic Flow Counting



Traffic Flow Counting

播放器初始化...[完成]
加载用户配置...
加载视频地址...

00:00 / 00:00

进入bilibili,一起发弹幕吐槽!

去吐槽

Traffic Flow Counting

播放器初始化...[完成]
加载用户配置...
加载视频地址...

00:00 / 00:00

进入bilibili,一起发弹幕吐槽!

去吐槽

Traffic Signal Recognition

(a)

(b)

(c)

Rail Recognition

播放器初始化...[完成]
加载用户配置...
加载视频地址...

00:00 / 00:00

进入bilibili,一起发弹幕吐槽!

去吐槽

Crossing Monitoring

播放器初始化...[完成]
加载用户配置...
加载视频地址...

00:00 / 00:00

进入bilibili,一起发弹幕吐槽!

去吐槽

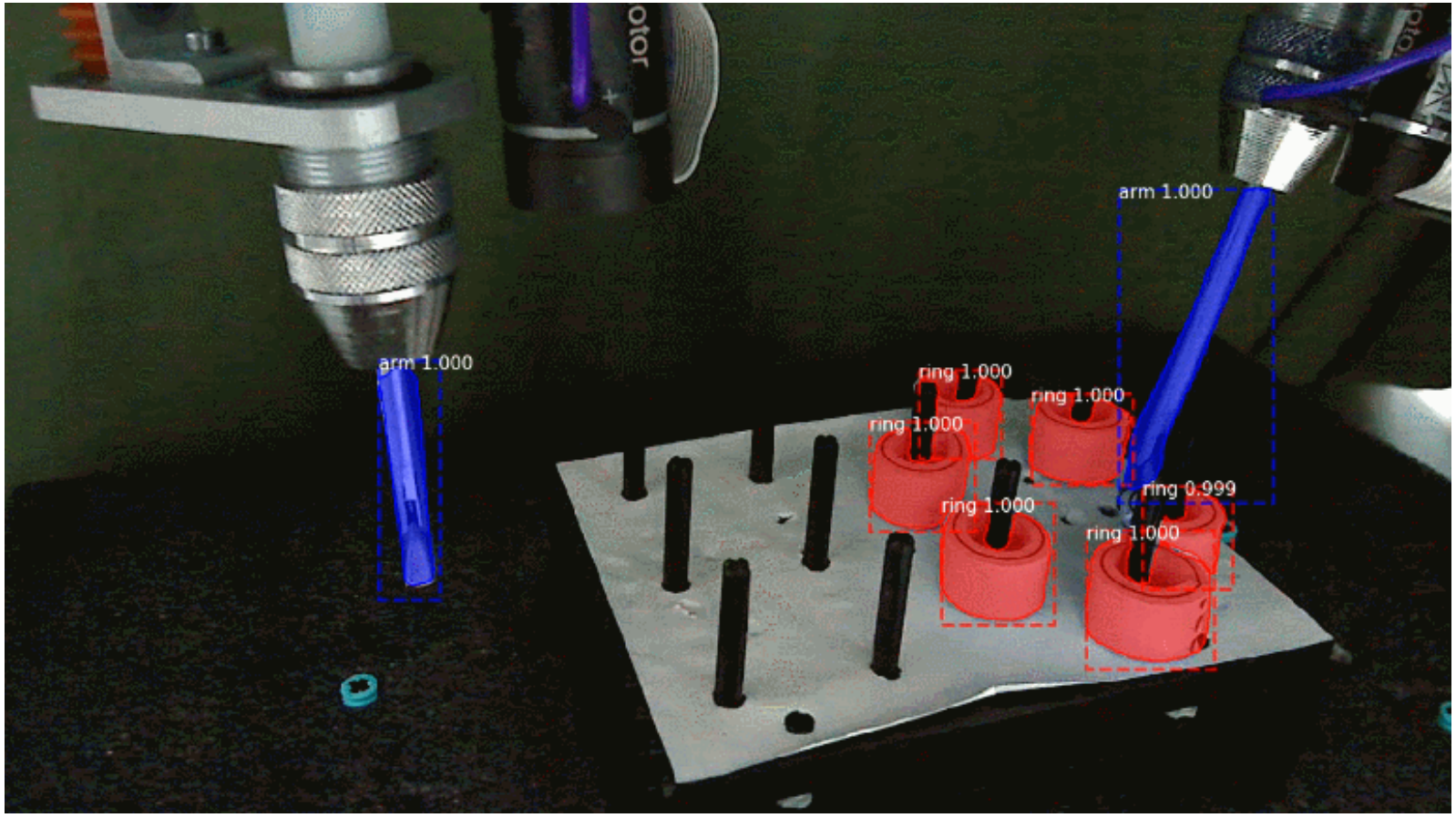
Text Recognition

(a)

(b)

(c)

Industrial Robot



Folding laundry (2010)

播放器初始化...[完成]
加载用户配置...
加载视频地址...

00:00 / 00:00

进入bilibili,一起发弹幕吐槽!

去吐槽

Driving a car (2016)

Video: Driving a car

Learning to sort waste

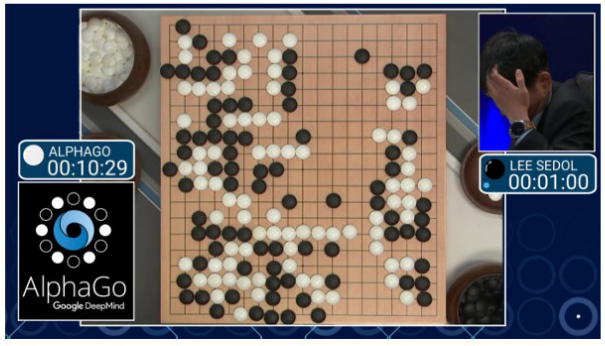
播放器初始化...[完成]
加载用户配置...
加载视频地址...

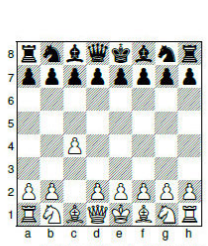
00:00 / 00:00

进入bilibili,一起发弹幕吐槽!

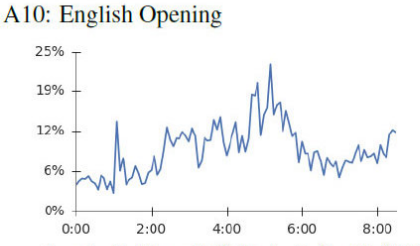
去吐槽

Games



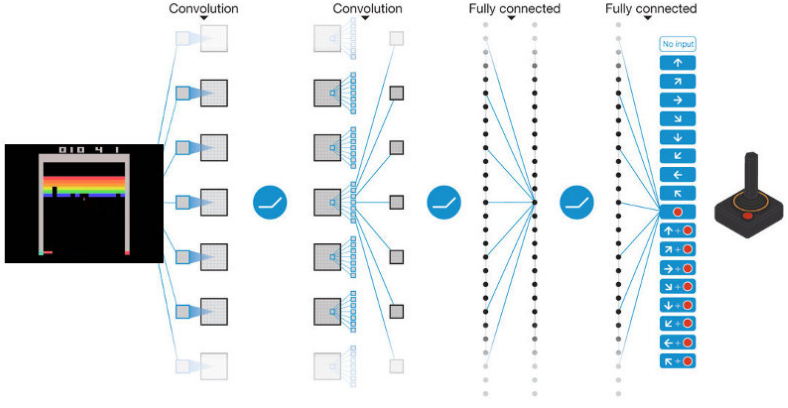


w 20/30/0, b 8/40/2

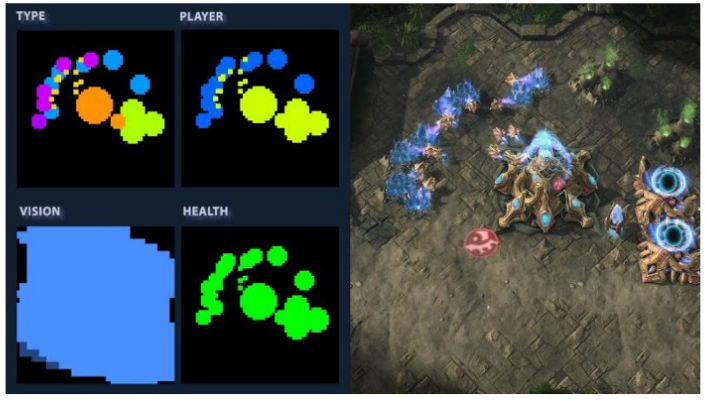


1...e5 g3 d5 cxd5 ♘f6 ♗g2 ♘xd5 ♘f3

[Deepmind AlphaGo / Zero 2017]



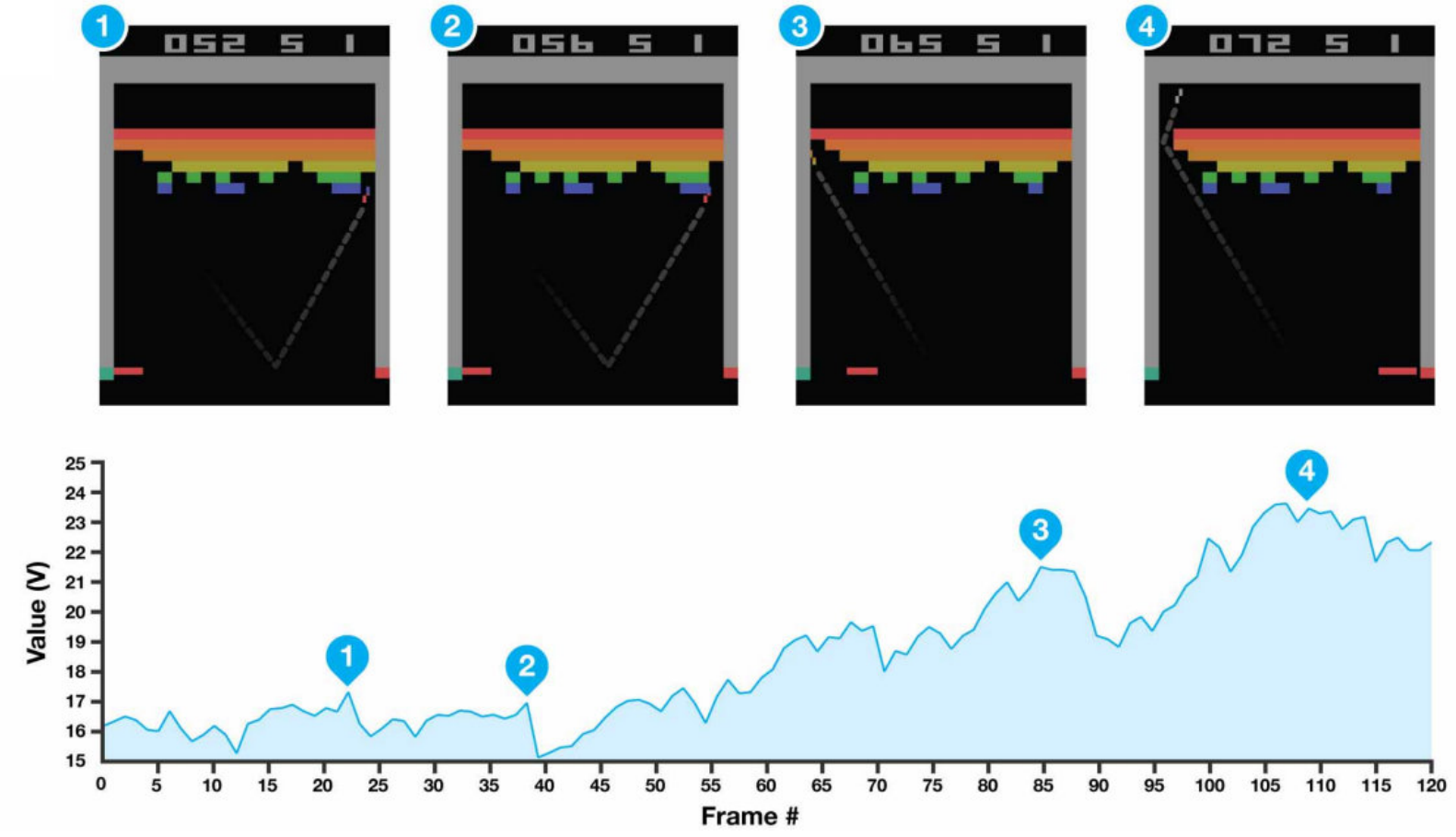
[Atari Games - DeepMind 2016]



[Starcraft 2 for AI research]

Atari

Reinforcement learning



Atari

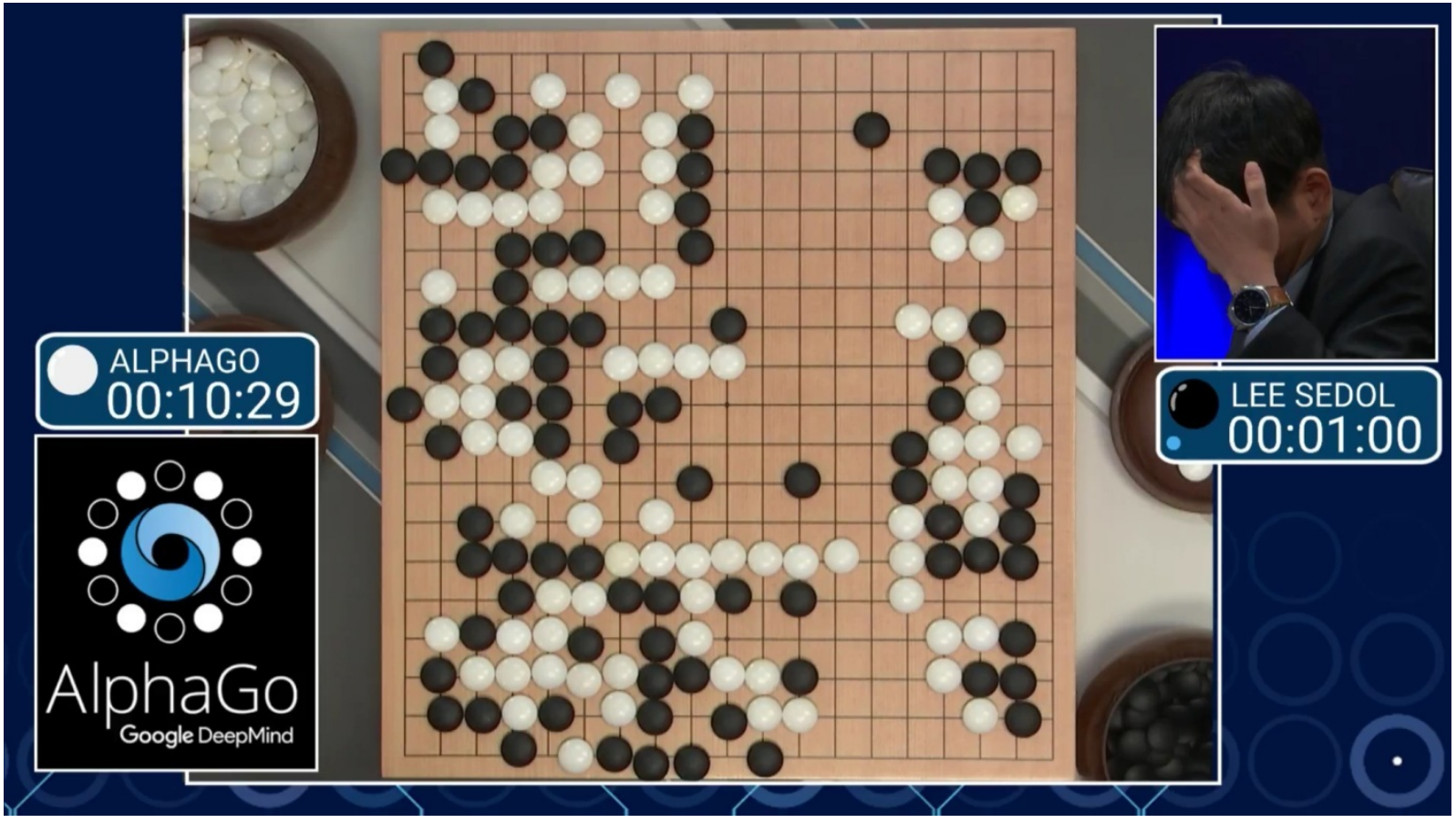
播放器初始化...[完成]
加载用户配置...
加载视频地址...

00:00 / 00:00

进入bilibili,一起发弹幕吐槽!

去吐槽

AlphaGo



Video: Beat the best human Go players (2016)

Content

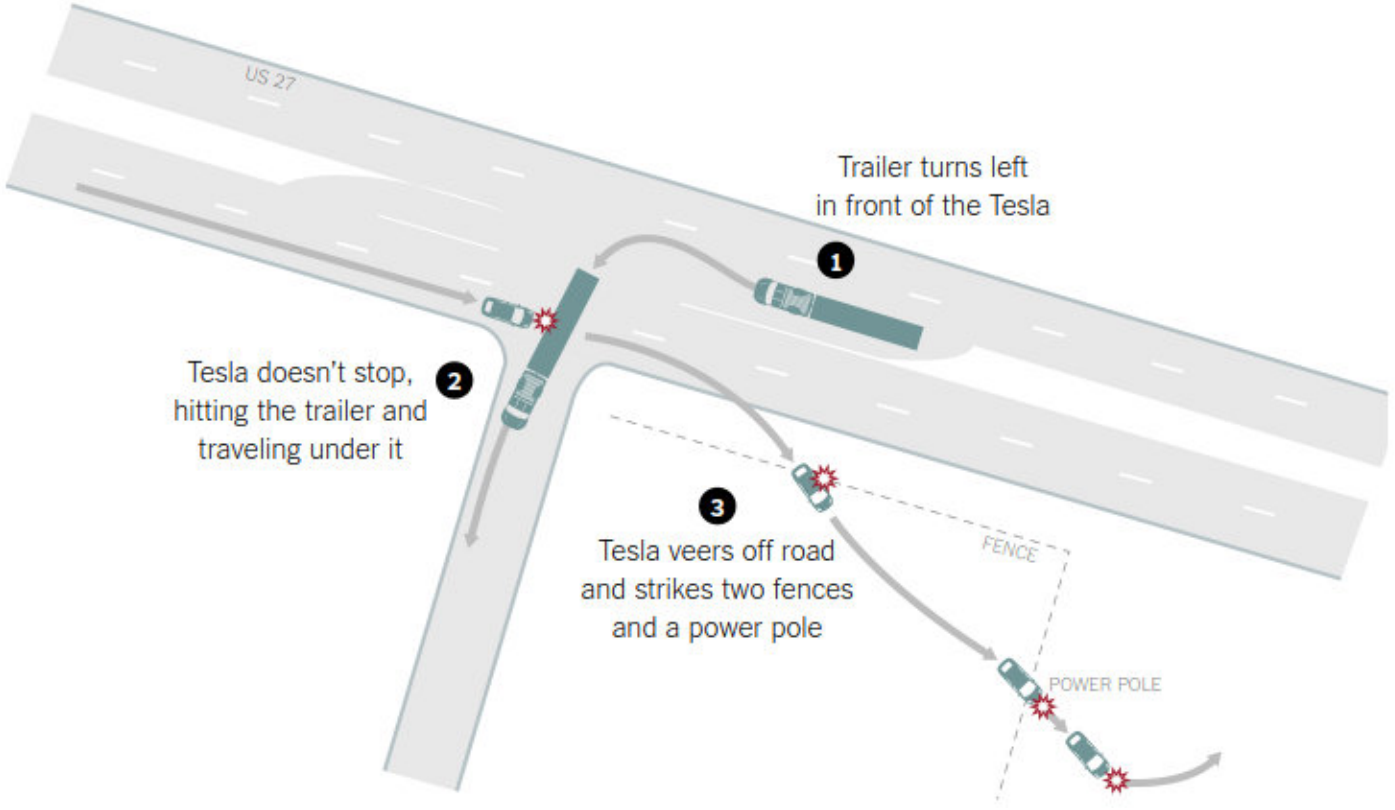
- Definition
- History
- Type
- Application
- Concerns

Concerns

Accuracy, privacy protection, fairness

Accuracy

- Tesla's autonomous driving system fails to identify white vans



Accuracy

<!-- ---

Privacy

- On May 14, 2019, the San Francisco City Supervisory Commission passed a decree by 8 votes to 1 to ban city workers from purchasing and using face recognition technology
- Face recognition technology tends to endanger civil rights and civil liberties far more than its claimed benefits. This technology will exacerbate racial inequalities and threaten our ability to live without long-term government surveillance.

Quiz

- In reality, what problems should be paid attention to in the application of computer vision technology?

Quiz

- Give examples of computer vision applications you might need at work

Quiz

- Deep learning brings major breakthroughs in the field of images, please give an example that impresses you

Summary

- Definition
- History
- Type
- Application
- Concerns