

Informatics & Learning Systems

Week 1:
Introduction

Tommy W S Chow
EE4146
January 2025



J Hopfield



A Turing

Source Materials

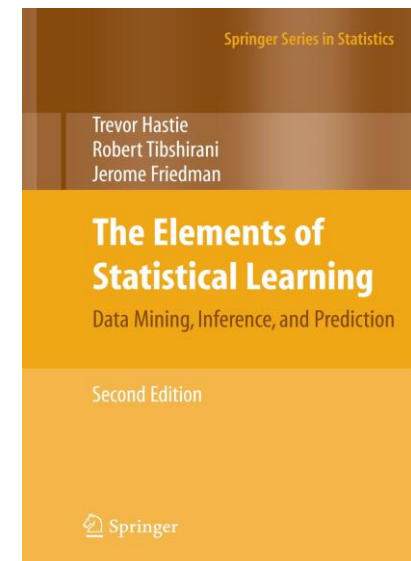
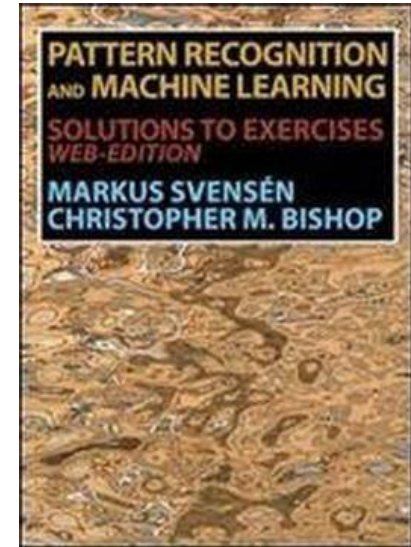
❑ Pattern Recognition and Machine Learning.
Christopher M. Bishop. Springer, 2006

<https://www.microsoft.com/en-us/research/people/cmbishop/prml-book/>

❑ The Elements of Statistical Learning. Trevor Hastie, Robert Tibshirani, Jerome Friedman. Springer, 2009

<https://web.stanford.edu/~hastie/ElemStatLearn/>

***Many topics of this course cannot be found in these two books due to the quick development of machine learning.**



Theme

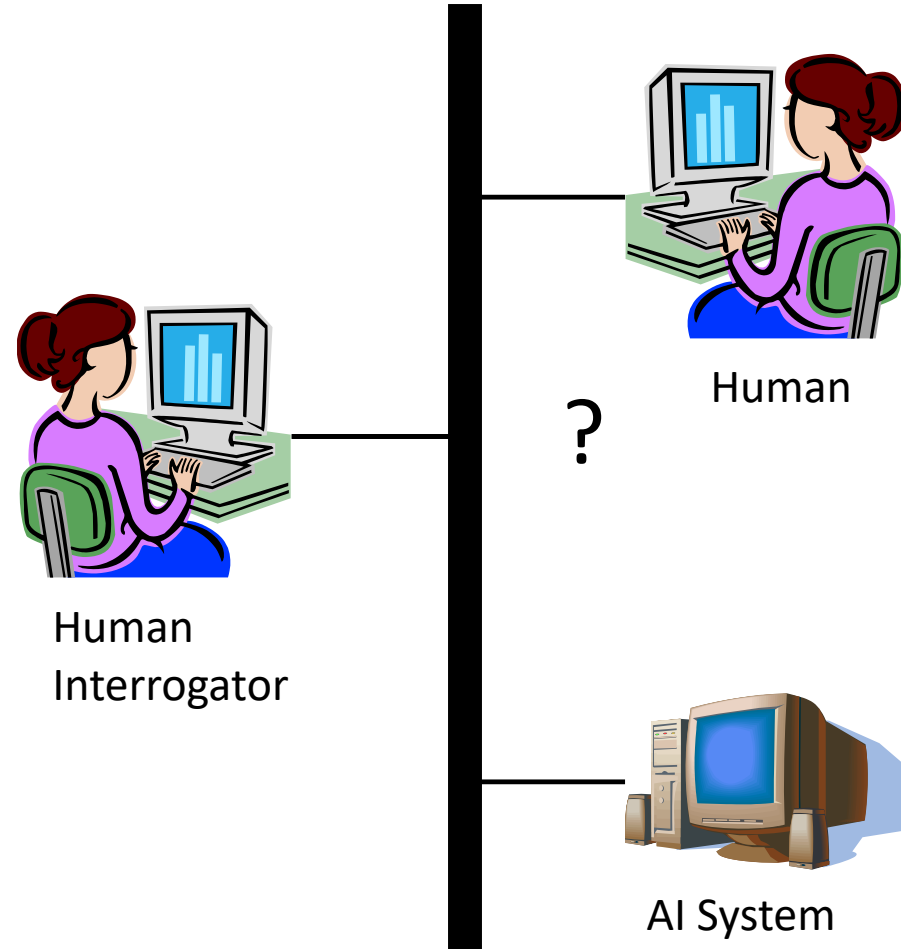
- The course Informatics & Learning Systems teaches basic machine learning techniques with applications to engineering and non-engineering problems.
- Here we learn and appreciate some of the most fundamental machine learning techniques
- This is not a programming course so we do not teach coding and do not require student to code
- Machine learning means how we teach computer machines to learn.
- We will only focus on algorithms and methods

What do we do

- We deal with using computer to handle complicated tasks:
 - extracting key information from millions of data (big data)
 - relating different data, or different classes of data so that we can data mine important information
 - making classification, data retrieval, prediction etc. For example train the model with a lot of cats, dogs pictures and use the trained model to classify dogs from cats.
 - based on given info, machine learning makes decision (use mathematical/computational ways)
 - via: supervised, unsupervised, or semi-supervised methods.

Acting Humanly

- The very famous Turing Test (1950)
 - Can machines think?
 - Can machines behave intelligently?
- Operational test for intelligent behavior
 - The Imitation Game



Acting Humanly

- Turing Test (cont)



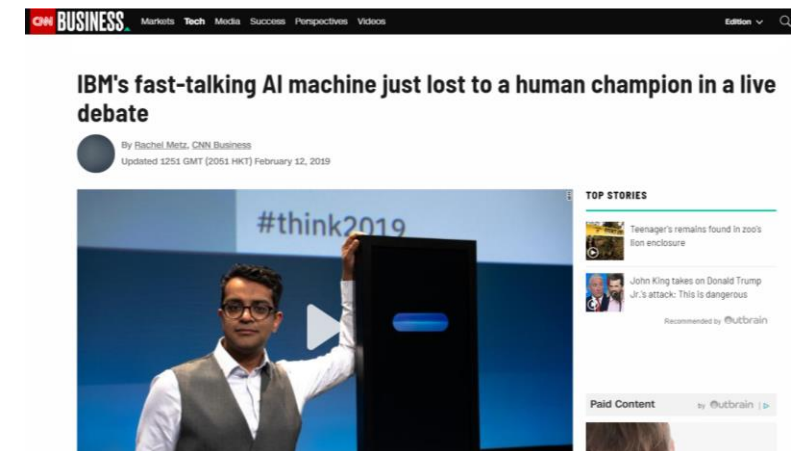
1912-1954

- The halting problem and the Turing machine
- Predicted that by 2000, a machine might have a 30% chance of fooling a lay person for 5 minutes
- Anticipated all major arguments against AI in the following 50 years
- Suggested major components of AI: knowledge, reasoning, language understanding, learning

Natural Language Processing, NLP

https://www.youtube.com/watch?v=BXJn_aETY5Y

<https://www.youtube.com/watch?v=naQujxmg9gg>



Birth of AI (1943-1956)



Marvin Minsky MIT, Turing Award (died 2016)

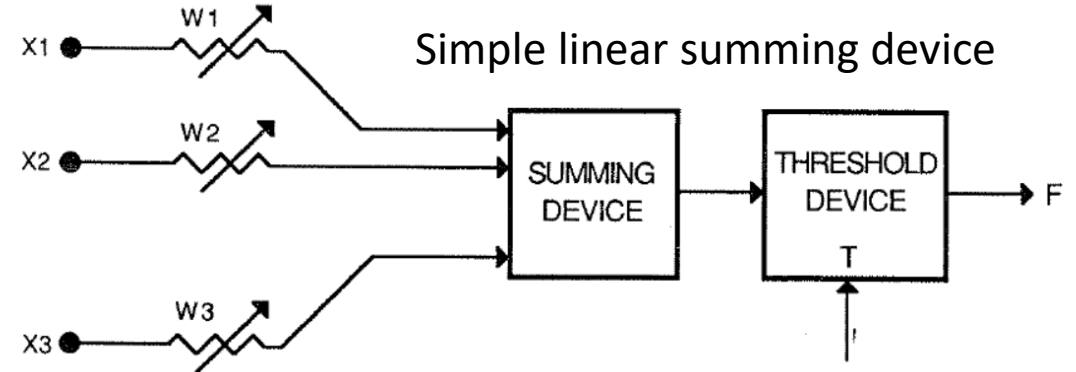
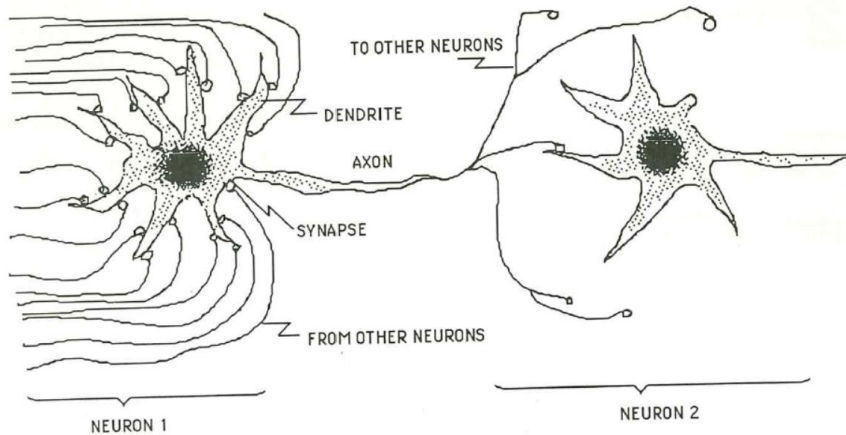
- 1943 – The McCulloch and Pitts Model (simplest Neural Network model)
- 1950 – Turing machine and the Turing test
- 1955-56 Newell & Simon (Carnegie Mellon U, CMU) Logic Theorist
- 1956 – Dartmouth 2-month conference (10 attendees), organized by John McCarthy, Minsky, Rochester, Claude Shannon, Herbert Simon, Allen Newell
- 1956 – Newell & Simon stole the show, showing they had developed program that is capable of thinking, NON_NUMERICALLY
- Next 30 years: this research area had been dominated by these people (and their students) at MIT, CMU, Stanford and IBM



Herbert Simon; CMU, Turing Award, Nobel Price Economic, (died 2001)



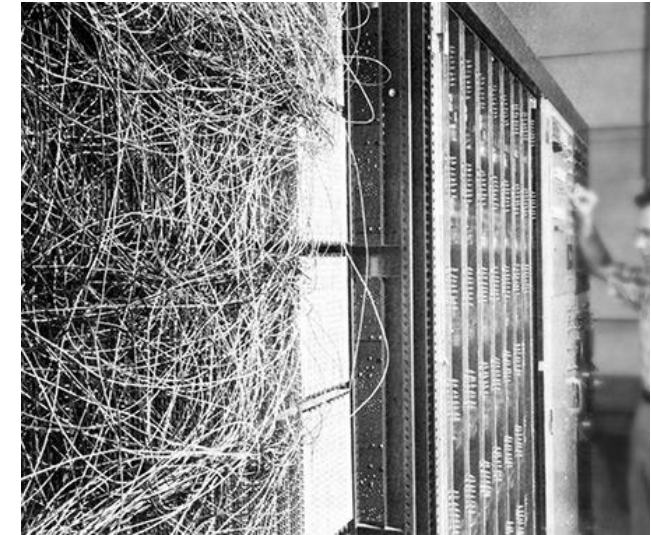
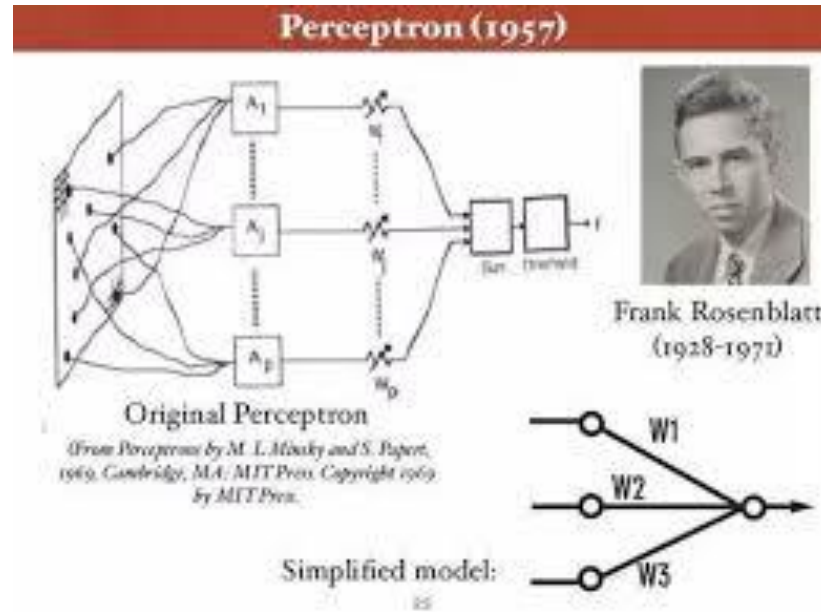
The McCulloch and Pitts Model



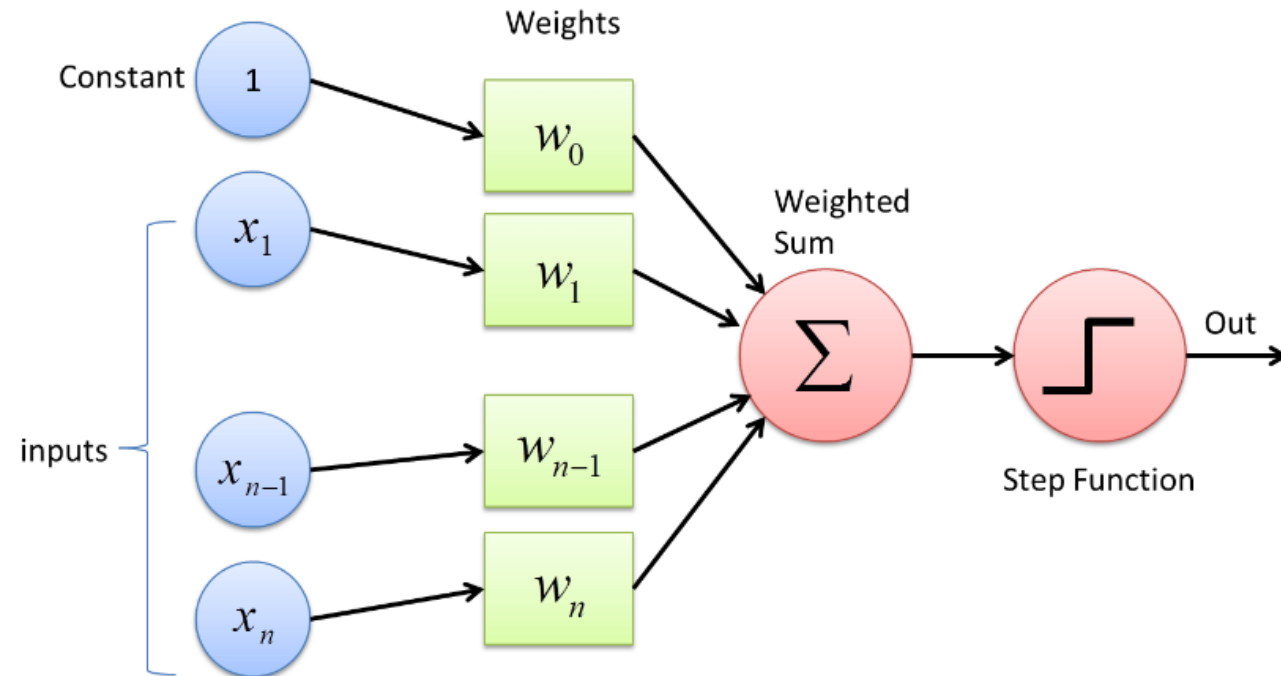
- There are about 10^{11} neurons in the average human brain.
- Some most recent research says people above 60 start to have 5% neurons decayed in every year if not doing good protective procedures.
- In the M&P model, the threshold device is a voltage comparator to control the output being 1 or 0.
- Have a kind a training procedures to adjust the weights and threshold to model a simple logic table. [Birth of AI \(1943-1956\)](#)



Frank Rosenblatt (1962)



Simple Perceptron Model

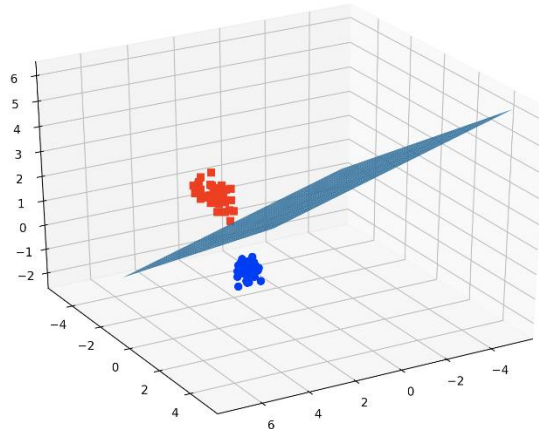


Single layer (no hidden layer) simple linear model

Timeline of Neural Networks Research (1)

- 1943: McCullough and Pitts - Modeling the Neuron
- 1958: Rosenblatt – Perceptron model attracted huge attention
- 1969: Minsky and Papert discredited Perceptron being unable to solve XOR problem
- 1970's and 1980's: DARPA cut funding on NN/AI
- 1973: Lighthill report (NN no practical usage) ended AI funding in the UK
- 1982 John Hopfield developed Hopfield model to model human brain's content addressable Memory
- 1983: Hinton developed a kind of extended Hopfield model named Boltzmann machine
- 1986: Rumelhart, Hinton developed Backpropagation training algorithm
- 1989 Hornik proved Multilayer FNNs are universal approximators
- 1990-2000: FNN, Recurrent NN, lots of applications

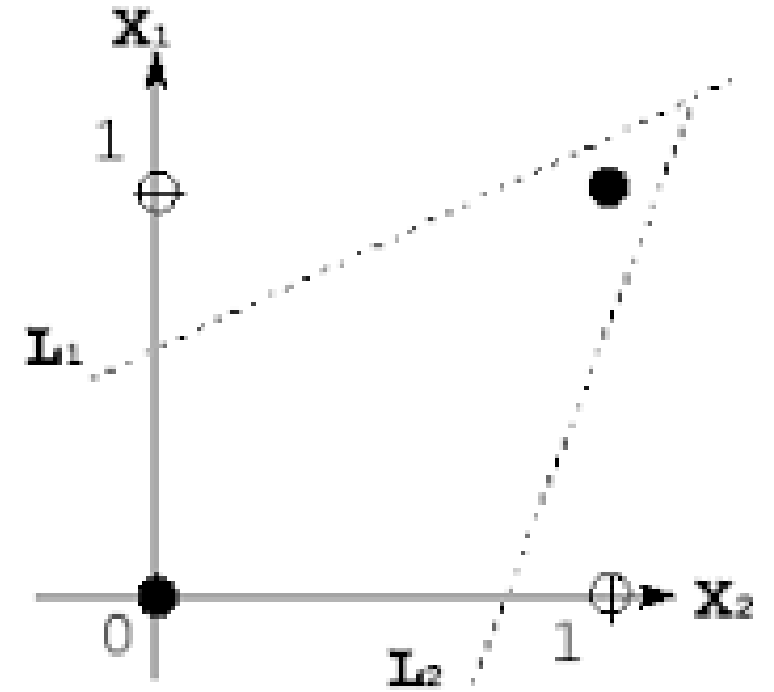
Separable vs Inseparable Problems



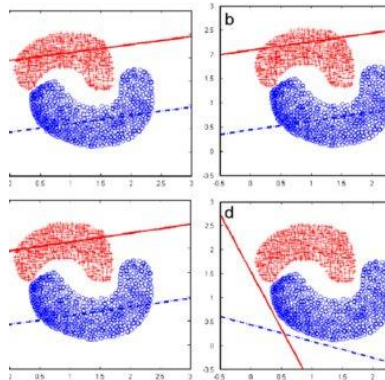
Separable problem

X_1	X_2	Y
0	0	0
0	1	1
1	0	1
1	1	0

$$Y = X_1 \oplus X_2$$



Linear Inseparable ex-or

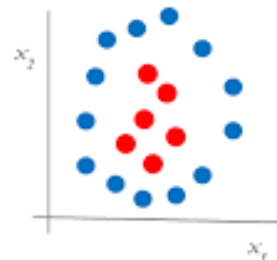


Linear Inseparable

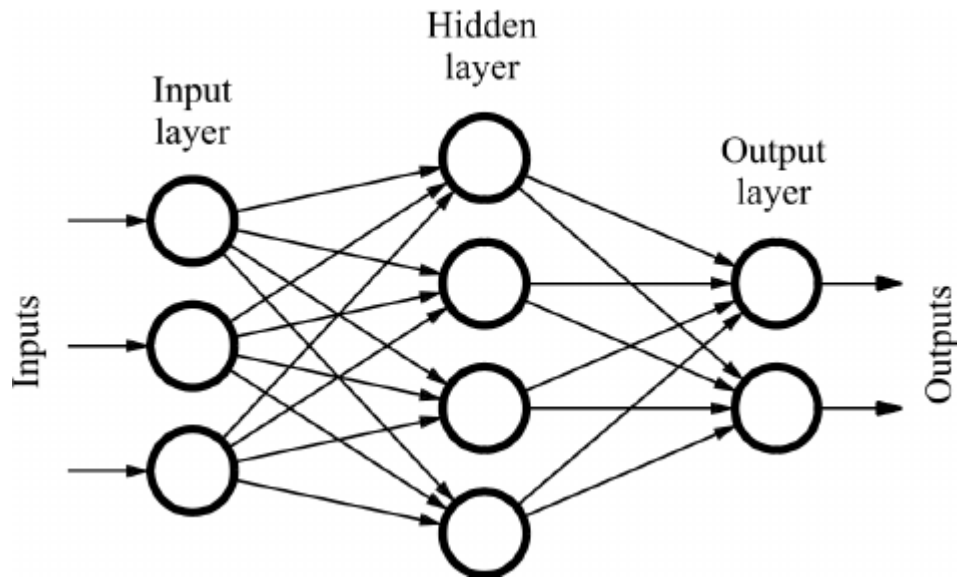
1-Dimensional Linearly Inseparable Classes



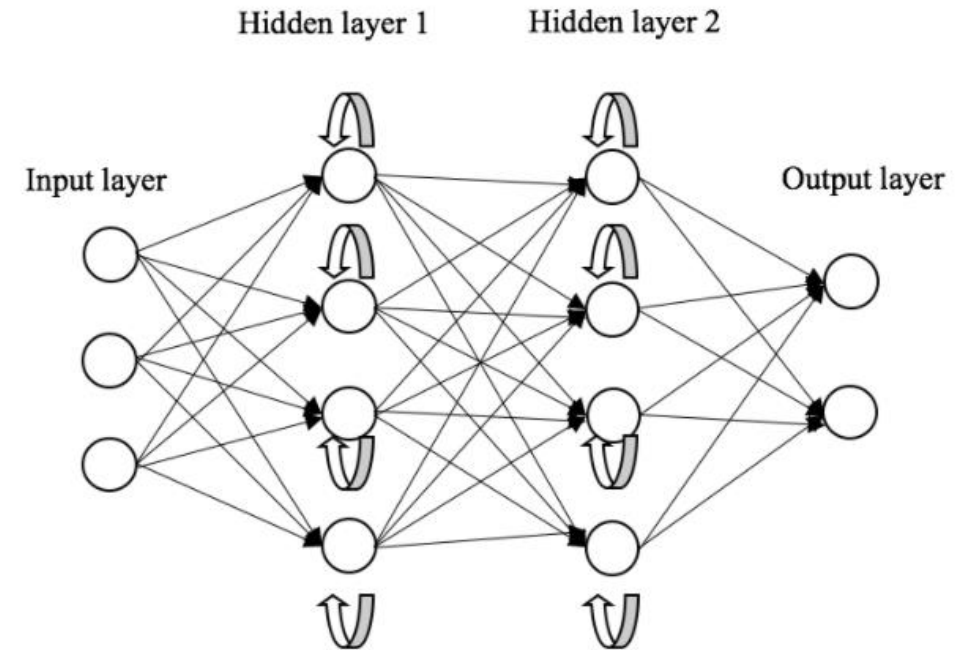
2-Dimensional Linearly Inseparable Classes



Multilayer Neural Network: Feedforward and Recurrent



Feedforward Neural Network (1 hidden layer), nonlinear because of a nonlinear activation function in every neuron



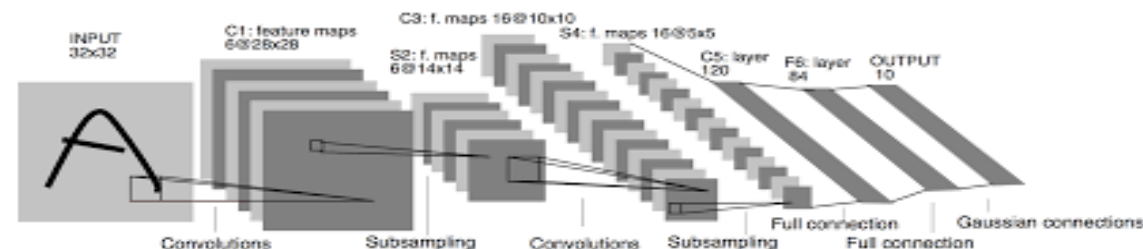
General Form of RNNs

Timeline of Neural Networks Research (2)

- Hinton continued the deep network research in UT after publishing BP algorithm with Rumelhart
- 1995+: Deep Learning, LeCun LeNet handwritten recognition (still restricted by computational issue/training speed)
- Around 1995 to 2000 we could not solve complex real-world problem because of hardware
- 2005: super long training time solved by Nvidia GPU
Computational power: H100 (80 billion transistors) > 1000 Tera Flops,
25 MHz i386 7 MIPS
- 2005 to 2010: more deep learning models or CNN were developed. We can now solve virtually most complicated real world problems with the powerful hardware



handwritten digits

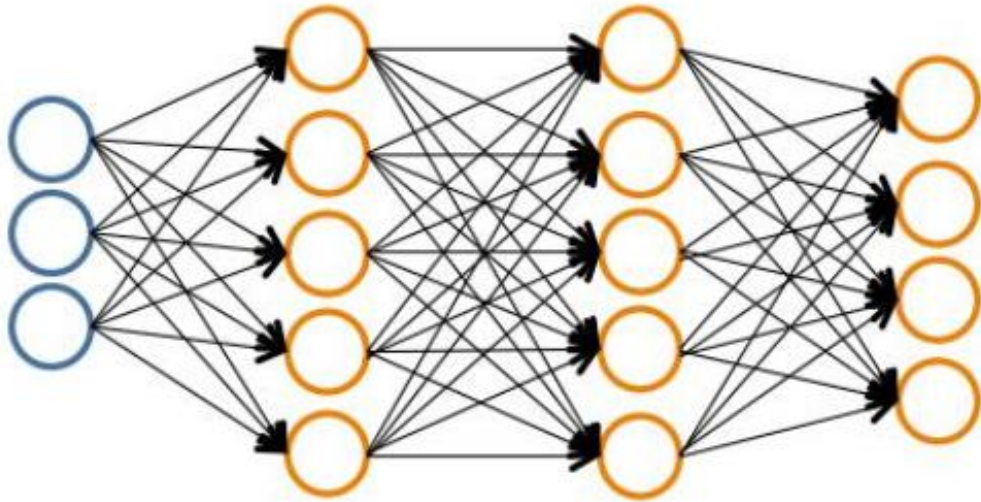


LeNet-5



Yann LeCun

Deep Learning: Number of Parameter



3 inputs 2 hidden layers with 5 hidden neurons
and 3 output feedforward NN

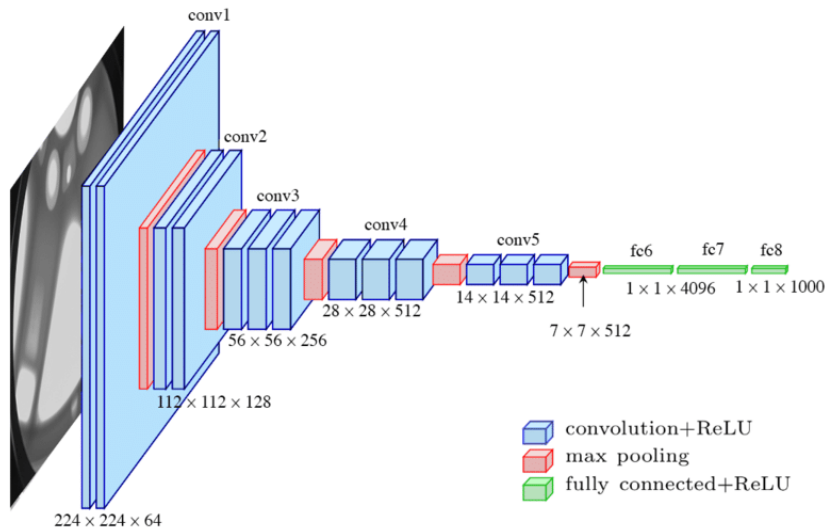
Forget the bias weight

Number of parameters equal

$$3 \times 5 + 5 \times 5 + 5 \times 4 = 15 + 25 + 20 \\ = 60 \text{ parameters}$$

Deep Neural Network may have over
100 millions parameters

2005+ CNN Variants: Aimed at Image Applications



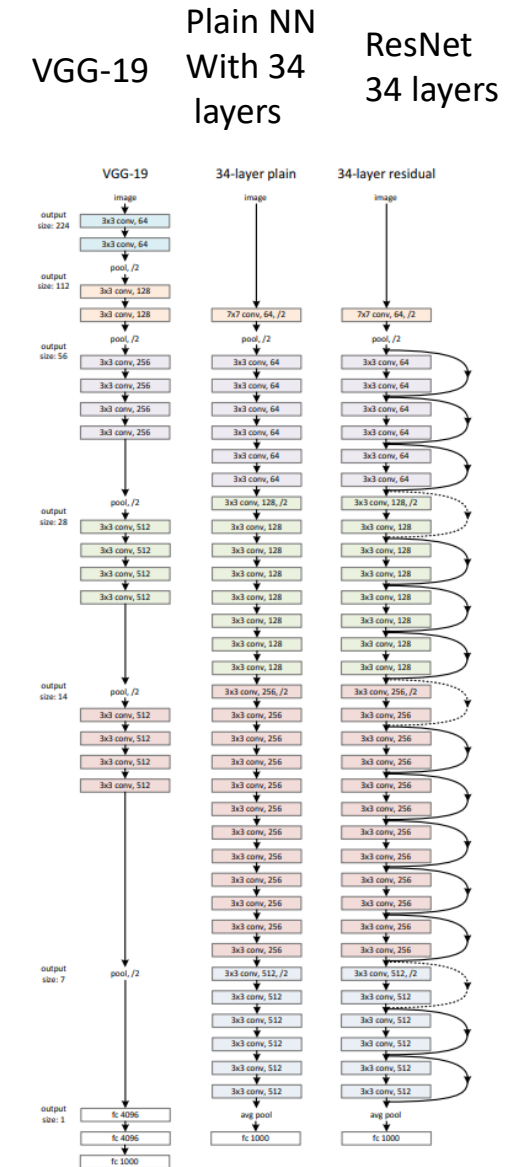
The architecture of VGG16*

- **Left:** The VGG-19 model (**19.6 billion FLOPs**).*
- **Middle:** a plain network with 34 parameter layers (**3.6 billion FLOPs**).
- **Right:** a residual network (ResNet) with 34 parameter layers (**3.6 billion FLOPs**).

VGG19 has over 144 million parameters

25 MHz i386 CPU: only 7 MIPS

Become a hardware dominate issue



*Model: K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," in *International Conference on Learning Representations*, 2015.

*Image: M. Ferguson, R. Ak, Y.-T. T. Lee, and K. H. Law, "Automatic localization of casting defects with convolutional neural networks," in *2017 IEEE International Conference on Big Data (Big Data)*, 2017, pp. 1726–1735. doi: 10.1109/BigData.2017.8258115.

^ K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90.

1990- 2010s: Focus on Applications



1997 – Deep Blue beats Gary Kasparov

2001-2010 – \$60 billion involved in combinatorial sourcing auctions



2005,2007 – Stanford and CMU respectively win DARPA grand challenge in autonomous driving

2019 IBM AI machine had debate with human

Winner Carnegie Mellon Uni (CMU)
2 million prize money (2007)
Then this technology turned into business that everyone knows now or in use now.

<https://www.youtube.com/watch?v=tXuC6gUO4mY>

Hardware & Deep Learning Changed the AI landscape

- Deep Learning and convolution NN (CNN) architectures:
- Nvidia: 1993 Nvidia's 3 founders including the current President, **Jensen Huang**, had coffee at a San Jose coffee. That started Nvidia and started to change the world.
- Nvidia originally aimed at video game GPU markets.
- “Titan X” 1995 cost US\$1.2k. After 2 decades, it still generated \$5 billion each year. Most recently, it still grew by 63%.
- This GPU technology changed the landscape of AI/machine learning completely.
- Current Tech giants like Google, Microsoft, Facebook, Amazon data centers all rely on Nvidia's technology.
- Now most world AI startups rely on Nvidia's platform, Nvidia A100 and H100, to power all AI real-time applications.
- With these powerful GPU, machine learning and AI based methods can do most real-life applications on real-time



NVIDIA Corp

NASDAQ: NVDA

Overview

Compare

Financials

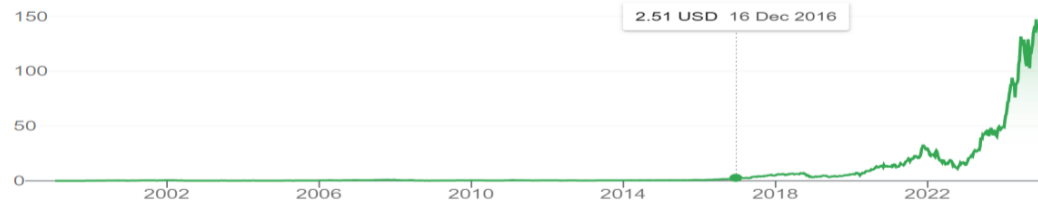
Market Summary > NVIDIA Corp

140.11 USD

+140.07 (350,175.00%) ↑ all time

Closed: 8 Jan, 8:00 pm GMT-5 • Disclaimer
After hours 138.54 -1.57 (1.12%)

1D 5D 1M 6M YTD 1Y 5Y Max



Open	142.58	Mkt cap	3.43T	52-wk high	153.13
High	143.95	P/E ratio	55.21	52-wk low	49.48
Low	137.56	Div yield	0.029%		

Share price increased from
\$0.4 to \$140 (350,000 times)
in the last 10 years

Market Summary > NVIDIA Corp

140.11 USD

+134.00 (2,193.13%) ↑ past 5 years

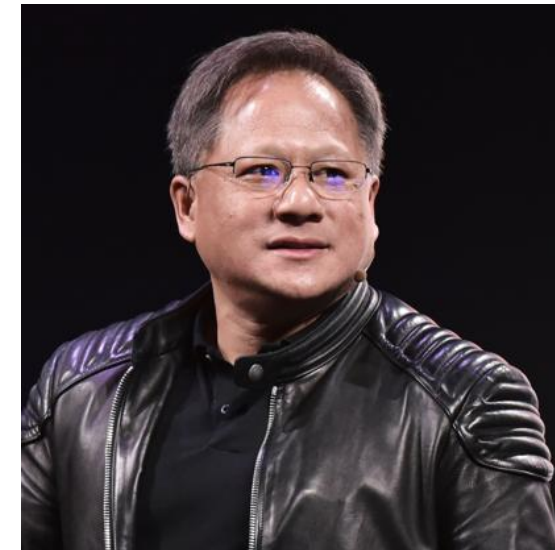
Closed: 8 Jan, 8:00 pm GMT-5 • Disclaimer
After hours 138.54 -1.57 (1.12%)

1D 5D 1M 6M YTD 1Y 5Y Max



Open	142.58	Mkt cap	3.43T	52-wk high	153.13
High	143.95	P/E ratio	55.21	52-wk low	49.48
Low	137.56	Div yield	0.029%		

Share price increased from
\$6.23 to \$140 (2,200 times)
in the last 5 years



- Jensen Huang 黃仁勳
- Nvidia CEO & president
- 2025 Worth US\$ 77B
- Born February 1963 Taiwan
- Moved to US Tacoma at age 9
- Nationality: Taiwan & USA
- Oregon State U (BS)
- Stanford (MS)

Nvidia

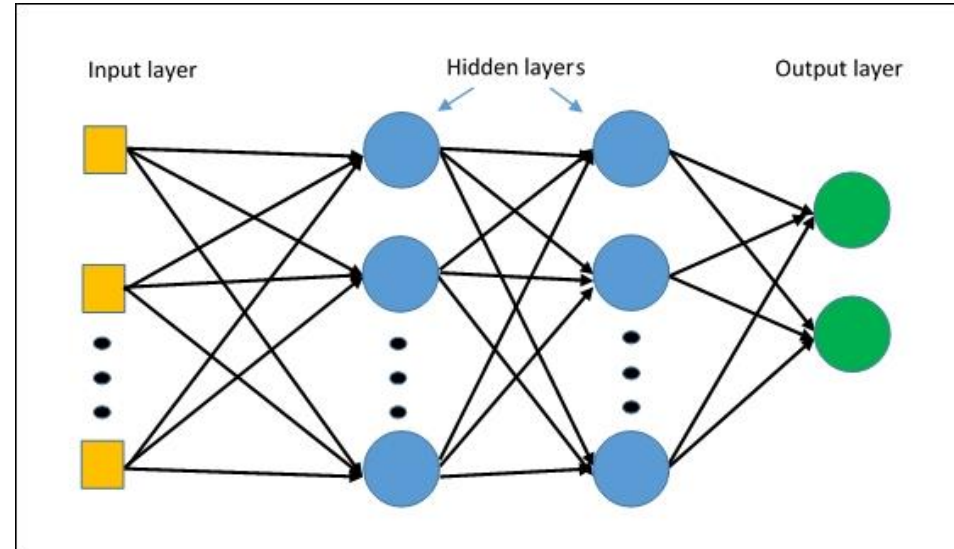
Market capitalization 3.3 trillion US

Second world largest Market Cap

HK 2023-24 government expenditure 613B HK or < 80 B US

HK2023 GDP 382 B US

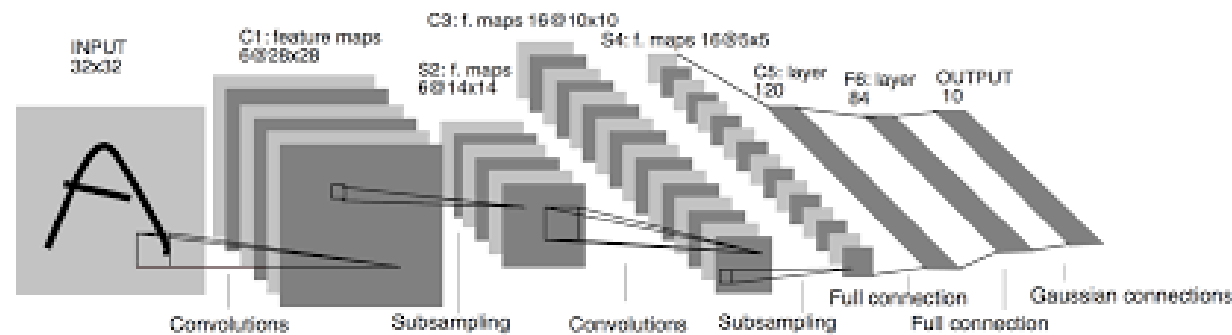
From Shallow NN to Deep NN



Conventional multi-layer Neural networks



Yann LeCun NYU



LeNet (1980)



2012 to 2022: Machine Learning (Deep learning NN models) shocked the world



- Google, Facebook, Microsoft, Amazon, Twitter all have large AI labs
- 2012 AlexNet (large layers of CNN) developed and won a major image classification contest. It achieved more than 10.8% lower error rate than the runner up. AlexNet was the first CNN powered by Nvidia GPU.
- 2015-2017 – superhuman speech understanding and speech synthesis using deep learning
- Speech and translation using RNN became sophisticated (there are still the training convergence problem)
- 2017 Google developed and published the Transformer model that pushed machine learning into another level; more NLP applications
- 2018 Google developed Bidirectional Encoder Representations from Transformers (BERT) language model
- 2020 OpenAI with 300+ engineers launched the GPT-3
- 2022 OpenAI launched the ChatGPT that shocked the world
- 2023 G Hinton expressed his concern about the abuse of AI; AI needs to be monitored and regulated



Real time speech recognition – real time recognized and typed in English
Real time speech translation – real time translated and typed in Chinese
Real time translation – real time translated and synthesized in his own voice in Chinese

Deep Learning can Do a lot of things and Fake a lot of Things



Source: K F Lee Talk on Deep Learning 李開復

https://www.google.com/search?q=KF+Lee+Deep+learning+youtube&rlz=1C1CHBF_enHK861HK862&oq=KF+Lee+Deep+learning+youtube&aqs=chrome..69i57j33i160.8872j0j15&sourceid=chrome&ie=UTF-8#fpstate=ive&vld=cid:6dc6e414,vid:ajGgd9Ld-Wc

2014 Generative Adversarial Network (GAN)



*I. Goodfellow et al., "Generative Adversarial Networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, Oct. 2020, doi: 10.1145/3422622.

^Y. Choi, Y. Uh, J. Yoo, and J.-W. Ha, "StarGAN v2: Diverse Image Synthesis for Multiple Domains," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 8185–8194. doi: 10.1109/CVPR42600.2020.00821.

2024 AI Tools Typical Image Change



Used stable-diffusion-xl-base 1.0 to change the left picture into the right picture

2025 Less expansive Market Nvidia AI Tools: Project DIGITS

Personal AI supercomputer that provides AI researchers, data scientists and students worldwide with access to the power of the NVIDIA Grace Blackwell platform

Can run 200Billion- parameters Deep learning models

Memory and up to 4TB of NVMe storage

Project DIGITS will be available in May 2025 from NVIDIA and top partners, starting at **US\$3,000**.

It is now the Hardware dominating era

<https://nvidianews.nvidia.com/news/nvidia-puts-grace-blackwell-on-every-desk-and-at-every-ai-developers-fingertips>

China Deepseek-V3 is making promising progress too

<https://www.deepseek.com/>

NVIDIA Puts Grace Blackwell on Every Desk and at Every AI Developer's Fingertips

NVIDIA Project DIGITS With New GB10 Superchip Debuts as World's Smallest AI Supercomputer Capable of Running 200B-Parameter Models

January 6, 2025



CES—NVIDIA today unveiled NVIDIA® Project DIGITS, a personal AI supercomputer that provides AI researchers, data scientists and students worldwide with access to the power of the NVIDIA Grace Blackwell platform.

Project DIGITS features the new NVIDIA GB10 Grace Blackwell Superchip, offering a petaflop of AI computing performance for prototyping, fine-tuning and running large AI models.

With Project DIGITS, users can develop and run inference on models using their own desktop system, then seamlessly deploy the models on accelerated cloud or data center infrastructure.

*AI will be mainstream in every application for every industry. With Project DIGITS, the Grace Blackwell Superchip

Back to what we could possibly do as
we are not Google, Microsoft, OpenAI

Typical applications

- **Recognizing patterns:**
 - Facial identities or facial expressions
 - Handwritten or spoken words
 - Medical images, DNA pattern, protein pair matching
- **Recognizing anomalies:**
 - Unusual sequences of credit card transactions
 - Unusual patterns of sensor readings in a nuclear power plant or unusual sound in your car engine.
 - Fault diagnosis and prognosis
- **Classification, Prediction:**
 - 2016, 2012 US Presidential election
 - Identify whales/hostile submarine based on audio signals

Our 2015 Keyword extraction results

Documents	JOHN2.1-2.11	JOHN2.13-2.22	TKDE.2009.46
Provided keywords	JESUS	JESUS	MINING
	WATER	CLEAR	FREQUENT
	WINE	TEMPLE	PATTERN
			INCREMENTAL
			INTERACTIVE
KE results (Only the top 10 are listed in descend order)	WINE	TEMPLE	TREE
	JESUS	JESUS	STRUCTURE
	FEAST	JEWS	PATTERN
	BRIDEGROOM	CHANGERS	MINING
	JEWS	DISCIPLES	TRANSACTION
	MIRACLES	MONEY	INCREMENTAL
	MOTHER	SHEEP	INTERACTIVE
	FIRKINS	ANSWERED	DATABASE
	STONE	OXEN	FREQUENCY
	PURIFY	JERUSALEM	CUSTOMER

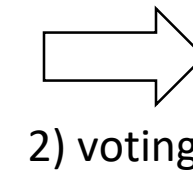
Original Doc: JOHN 2.1-2.11 (describe Jesus turns water into wine)

On the **third** day there was a **marriage** at **Cana** in **Galilee**, and the **mother** of **Jesus** was there; **Jesus** also was invited to the marriage, with his **disciples**. **When the wine failed**, the mother of **Jesus** said to him, “**They have no wine.**” And Jesus said to her, “O woman, what have you to do with me? **My hour has not yet come.**” His mother said to the **servants**, “Do whatever he tells you.” Now **six stone** jars were standing there, **for the Jewish rites of purification, each holding twenty or thirty gallons.** **Jesus** said to them, “Fill the jars with **water.**” And they filled them up to the brim.

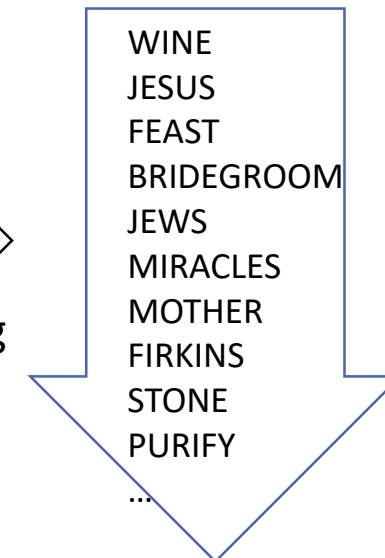
1) labelling



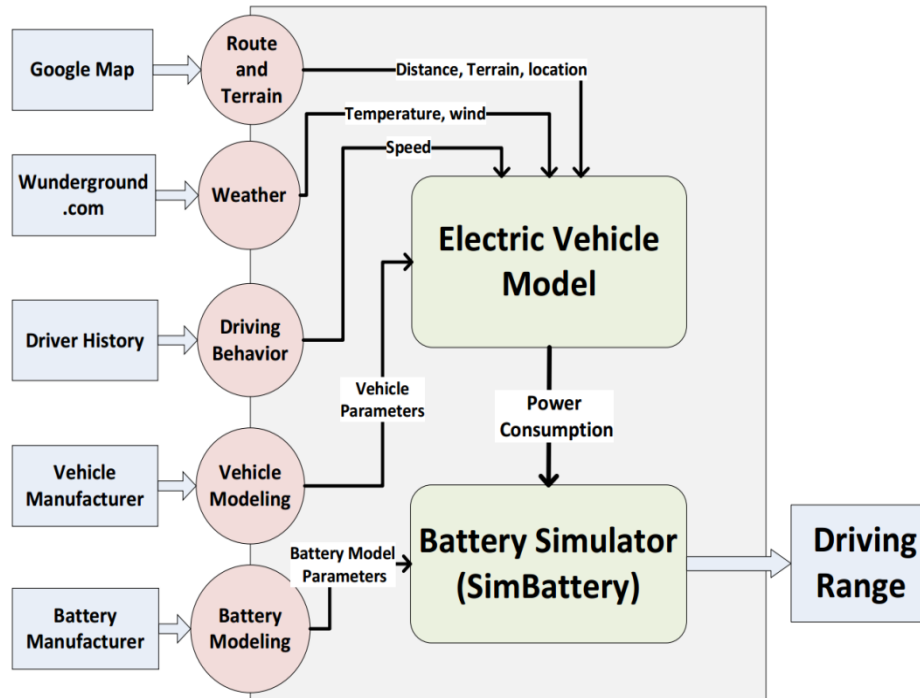
significance of words (1.0 to 0.1)



2) voting



Big data/Electrical Vehicle Range estimation (not just battery itself)



Range estimation framework block diagram.

Standard data: standard driving time from google

Historical data: Recent Miles performance

Real-time data: traffic, rain, weather etc

Electrical vehicle: How far can it go?

-There are algorithms estimating the remaining charge of the battery, but is it good enough?

-Drivers need to know how many miles left.

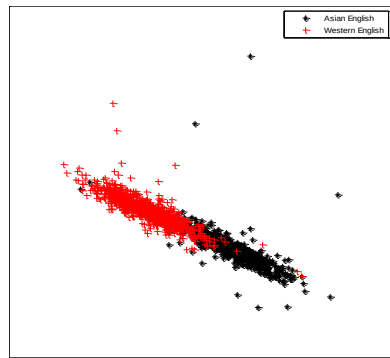
- But is that good enough?

- HK is small, 50 miles: from TST to Yuen Long that is very different from LV to Grand Canyon South regime, Arizona.

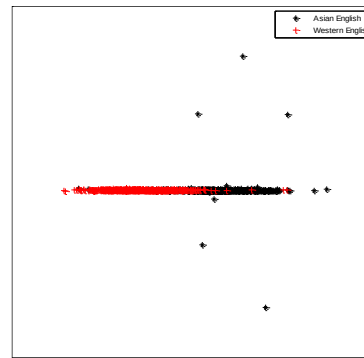
2015 results on Visualization of British, American & Asian English

Discrimination of Chinese, Japanese and Korean English from US, British and Australian English.

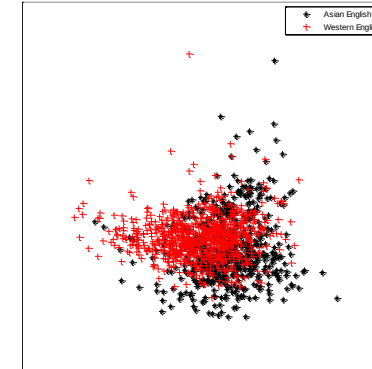
***** = Chinese, Japanese and Korean English **+** = the US, British and Australian



TR-LDA

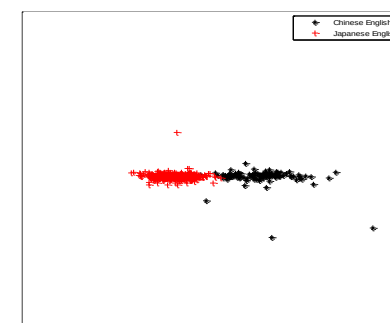
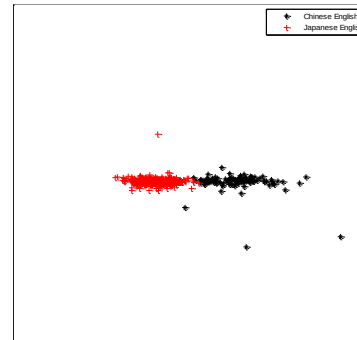
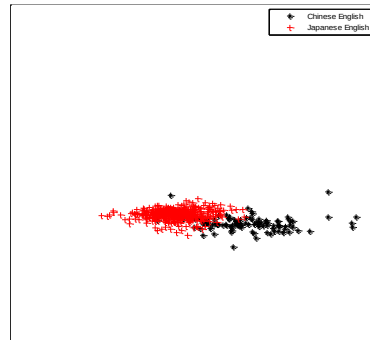


LDA



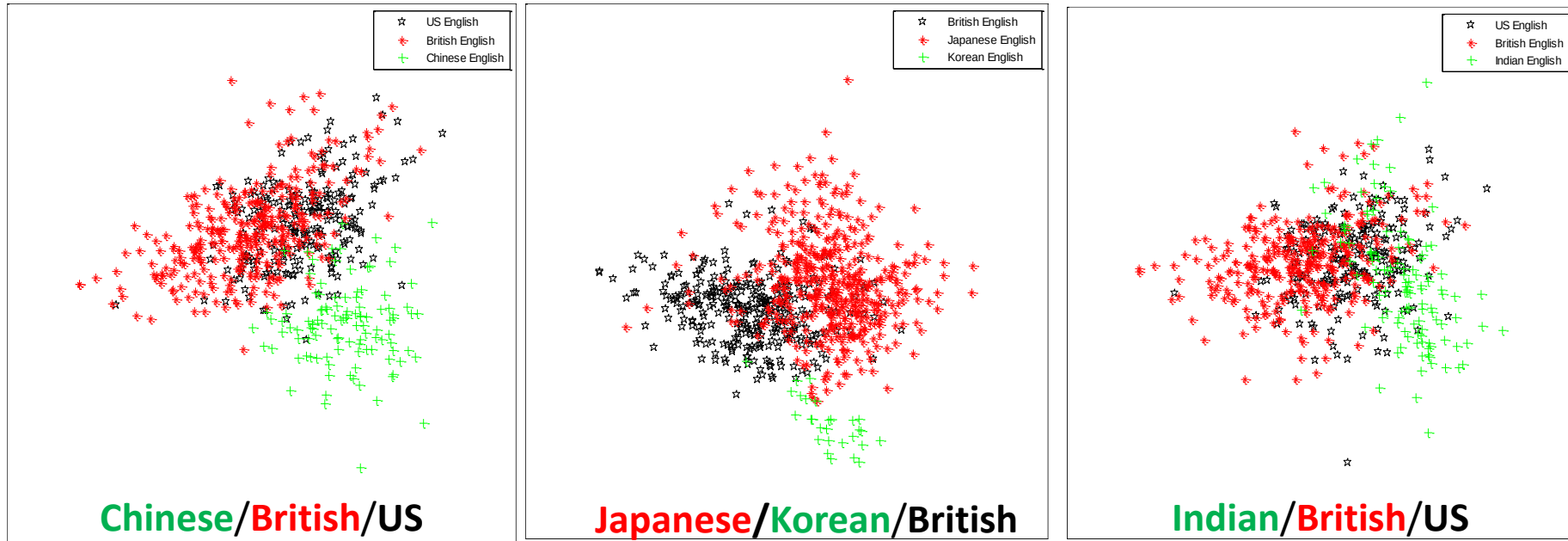
PCA

Discrimination of the **Japanese English** from **Chinese English**



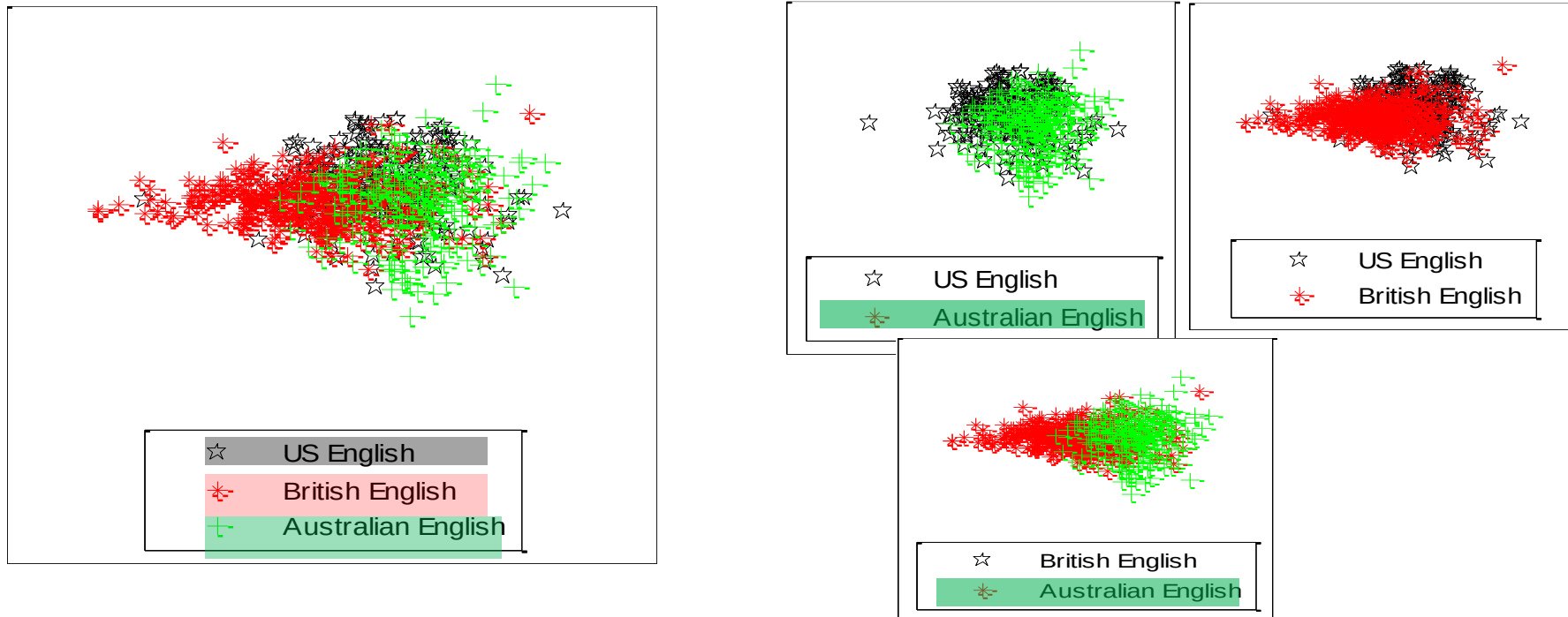
Asian English vs US and British English Visualization

Asian English (Chinese, Japanese, Korean and Indian English)



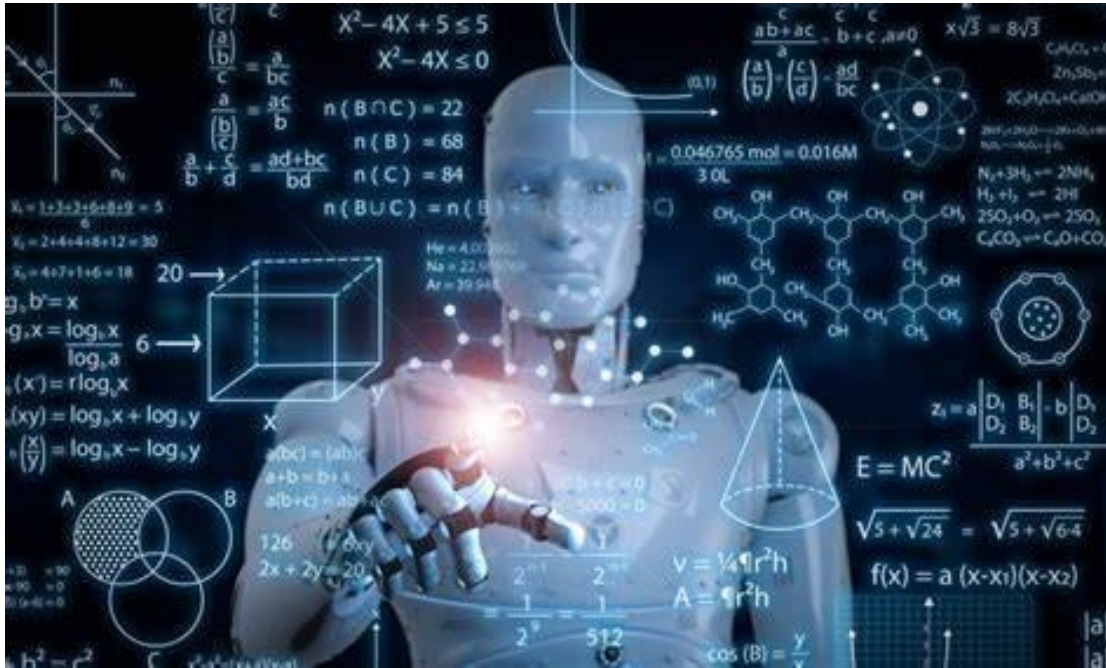
- Indian English are closer to the British English
- There is clean boundary between Japanese and Korean English, implying their written English are significantly different in styles
- There is a relatively clean boundary between Asian English and Native English; Asian English is easy to be discriminated from the native English

2015 Results: Native English Classification



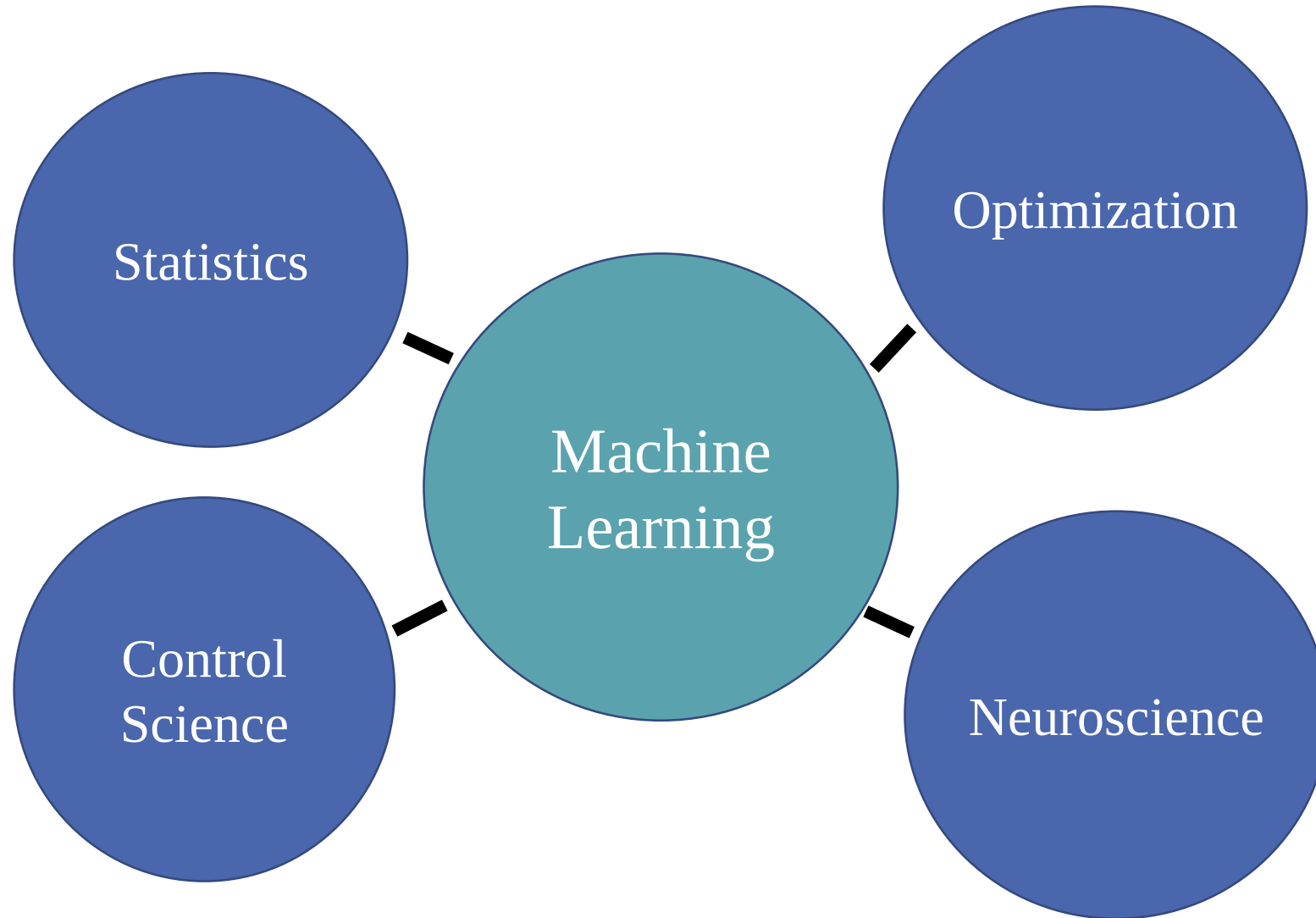
- We could not obtain a clean separation between US, British and Australian English
- The US and Australian English are close to each other
- The US and the British English exhibit significant overlapped, which means the difference from them is insignificant

What is machine learning?



- ❑ *Programming computers to optimize a performance criterion using example data or past experience*
- ❑ *Making machine to learn and make decision like human being*

Related areas



A 3D white figure is shown from the side, holding a large red and white target. Four red arrows are embedded in the bullseye of the target. The figure's right arm is raised in a celebratory gesture. The background is plain white.

-
- A digital cityscape with blue and white lines forming buildings and a perspective view of a street receding into the distance. The image is overlaid with various text elements, including code snippets and technical terms, suggesting a theme of data, technology, and digital infrastructure.



Application---Web Search/Mining



“cat” →



“dog” →



“cat and dog” →

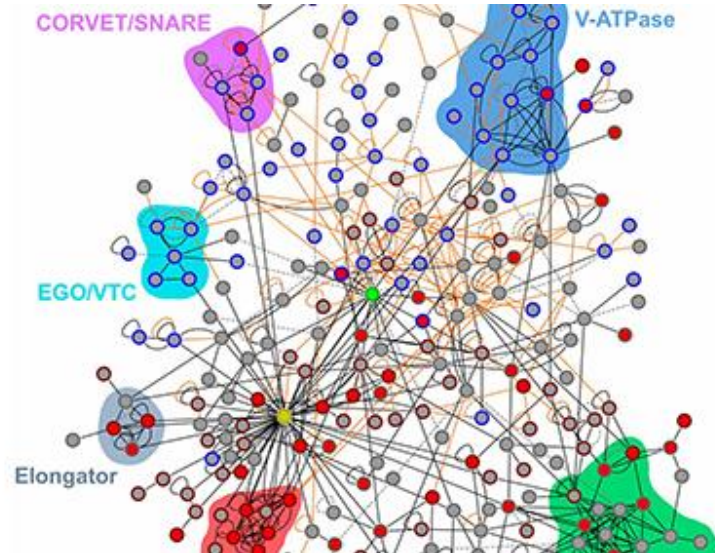


automatic image captioning

Misclassification!



Application--- Biomedical

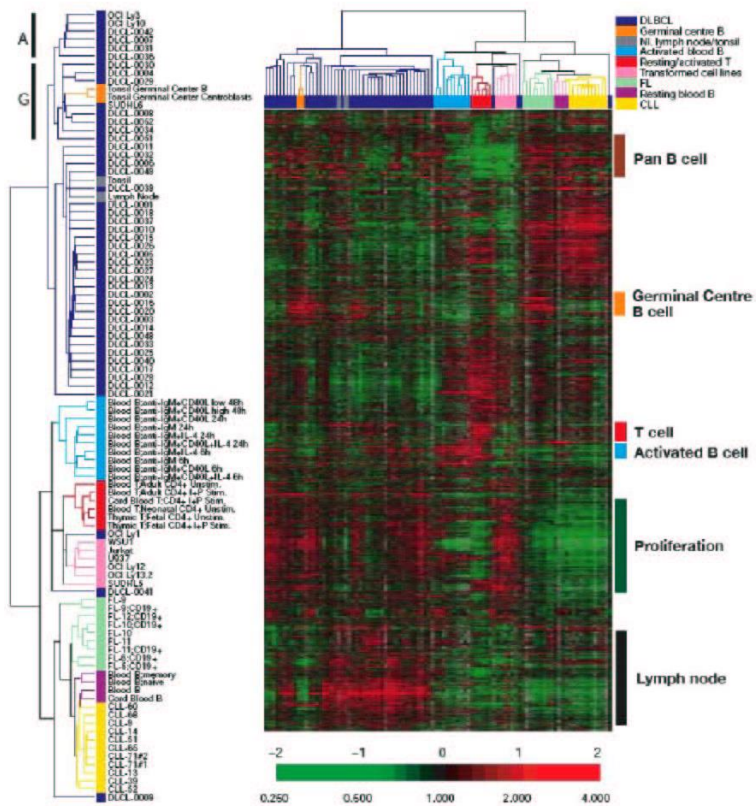
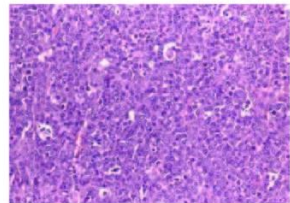


- Machine learning technology has significant contribution toward improved diagnosis and treatment
- Includes clinical diagnostics ie., (accurate and speedy CT scan, PET scan diagnosis), precision treatments and health monitoring
- Machine learning technology in drug discovery that help screening millions of compounds in hours that took year or years to do it in the past
- ML help identifying potential drug candidates that would have taken years to find using traditional methods

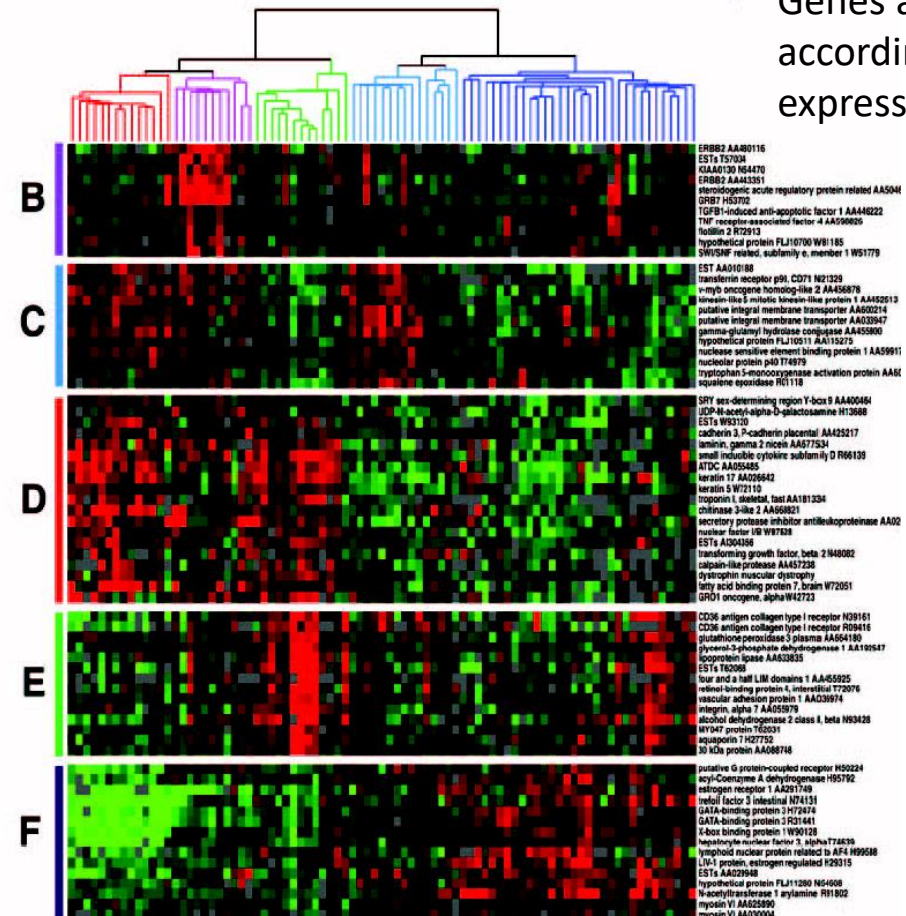
Machine Learning for Cancer Cells Identification

Subphenotyping Diffuse Large B-Cell Lymphoma

Alizadeh ... Staudt
Nature 2000



Gene expression patterns of 85 breast samples



Genes are clustered by rows according to similarity of expression

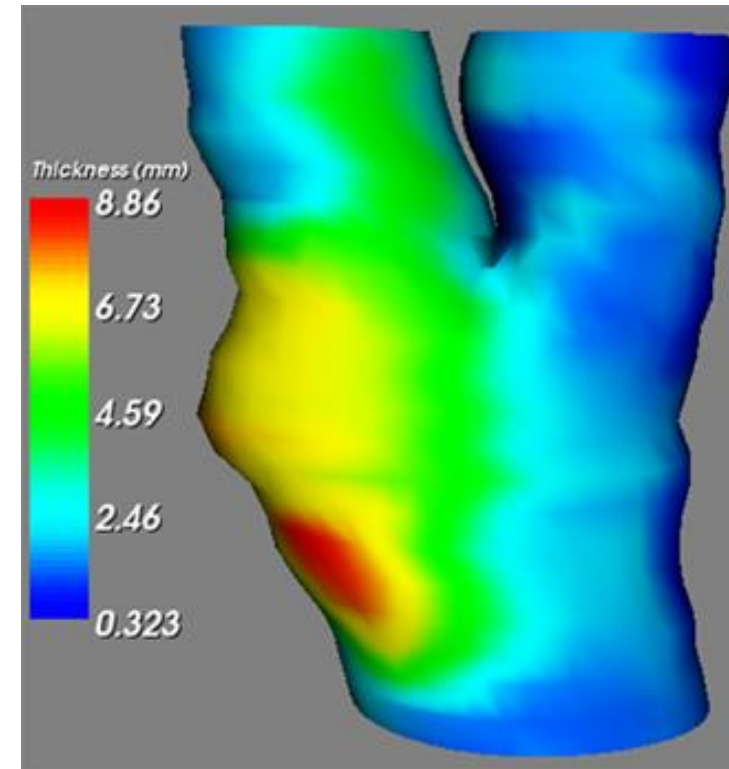
B, C, D, F are three different types of cancer gene

E is non-cancer type of gene

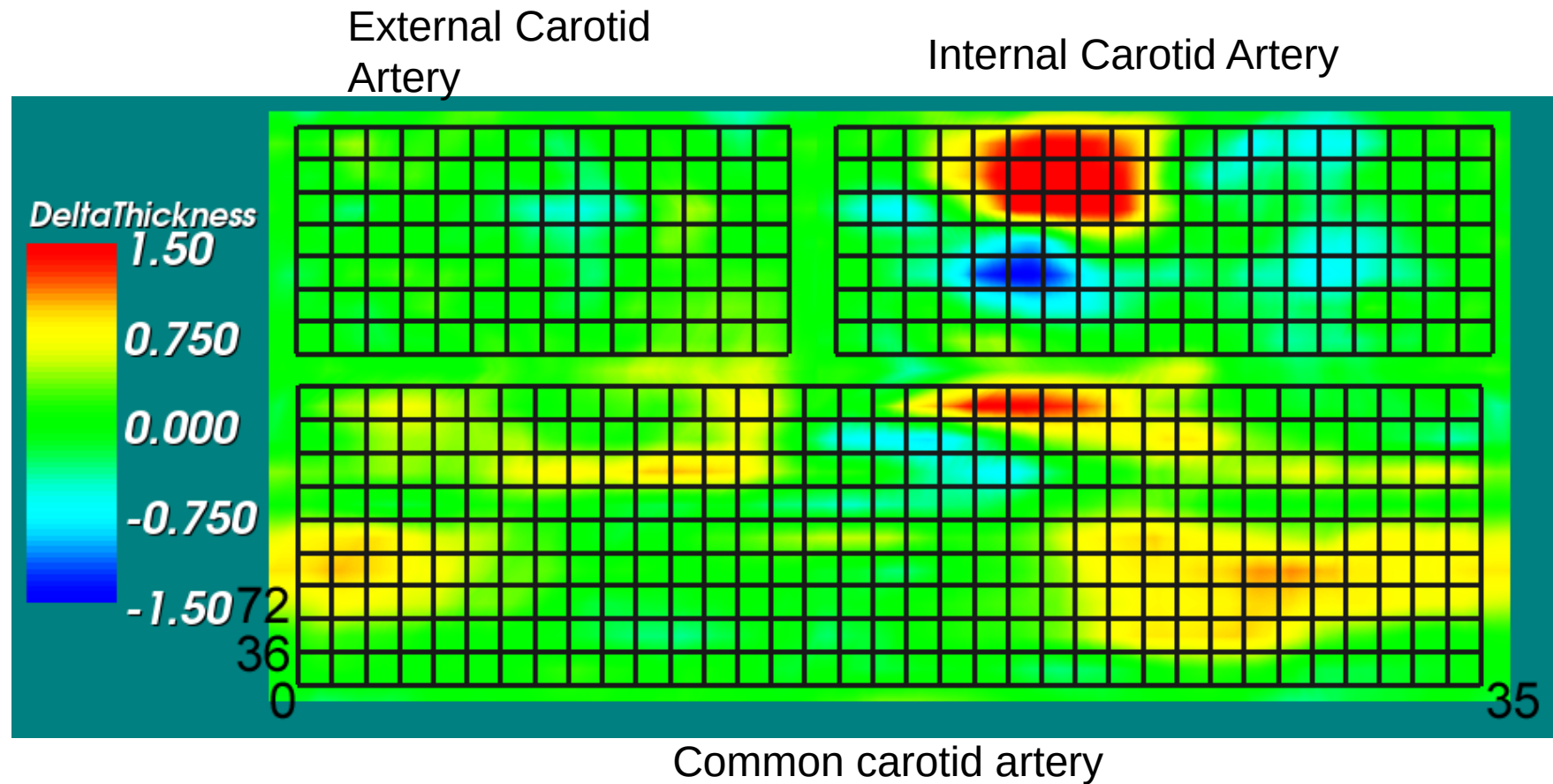
We need to perform clustering on this cDNA microarray

3D Vessel-Wall-Plus-Plaque Thickness (VWT)

- 3D IMT based on 3D ultrasound
- Only inputs are lumen and outer wall boundaries.
- 3D localized plaque thickness can be correlated with other local parameters:
 - Hemodynamic parameters: Shear stress, tensile stress
 - Location and thickness of plaque components (MR, histology)



Standardized 2D Map & Resampling



Application---E-Commerce



amazon



Netflix Research: Recommendations

<https://www.youtube.com/watch?v=f8OK1HBEgn0>

Problem: If Netflix asked me, I could not provide good answers as I had not watched many many movies

Netflix Price 2009: Recommender System

Collaborative Filtering and Matrix Completion

$$\begin{bmatrix} -4.70 & ? & 0.65 & -2.02 & ? \\ ? & 1.36 & ? & -0.86 & 1.97 \\ 4.08 & ? & -1.67 & ? & -2.70 \\ ? & 0.41 & ? & 5.23 & ? \\ 7.96 & ? & -0.33 & ? & 4.38 \end{bmatrix} \Rightarrow \begin{bmatrix} -4.70 & 0.47 & 0.65 & -2.02 & -1.10 \\ -3.77 & 1.36 & 1.39 & -0.86 & 1.97 \\ 4.08 & -1.67 & -1.67 & 0.78 & -2.70 \\ 9.69 & 0.41 & -0.11 & 5.23 & 6.30 \\ 7.96 & 0.07 & -0.33 & 4.09 & 4.38 \end{bmatrix}$$

Figure 1. An example of matrix completion ('?' indicates missing entry)

Netflix price:

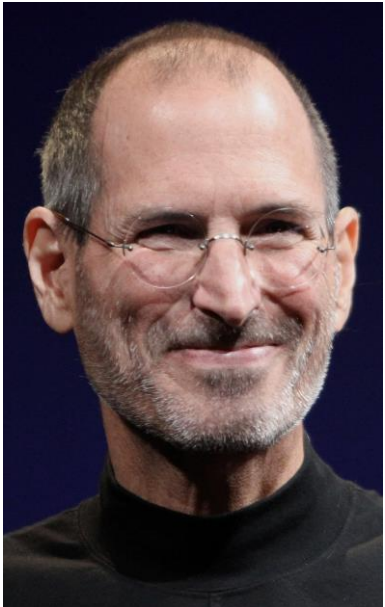
- Matrix of **480,189x17,770**
With 100,480,507 ratings
- It means **98.83%** are empty entries

Netflix Prize

- An open competition for the best algorithm to predict user ratings for films
- On 2009, the grand prize of **US\$1 million** was given to the BellKor's team which outperformed Netflix's own algorithm by 10.06%
- A *training* data set of 100,480,507 (100 millions) ratings that 480,189 users gave to 17,770 movies, $(480189 \times 17770 = 8,532,958,530)$ (8.5 billion) ratings)

<https://www.netflixprize.com/>

- **Two-thirds** of movies watched by **Netflix** customers are recommended movies
- 38% of click-through rates on **Google News** are recommended Links
- **35% of sales at Amazon** arise from recommended products



Steve Jobs: “A lot of times, people don’t know what they want until you show it to them.”

Recommender System is what business needs in many ways - brings up to **30%** of additional revenue to business.

Very Important Capabilities/Skills for Responding to Marketing Landscape Disruptions According to U.S. Marketers



Analytics capabilities to understand customer behavior



Developing content and experiences across the entire customer journey



Delivering relevant personalized customer experiences

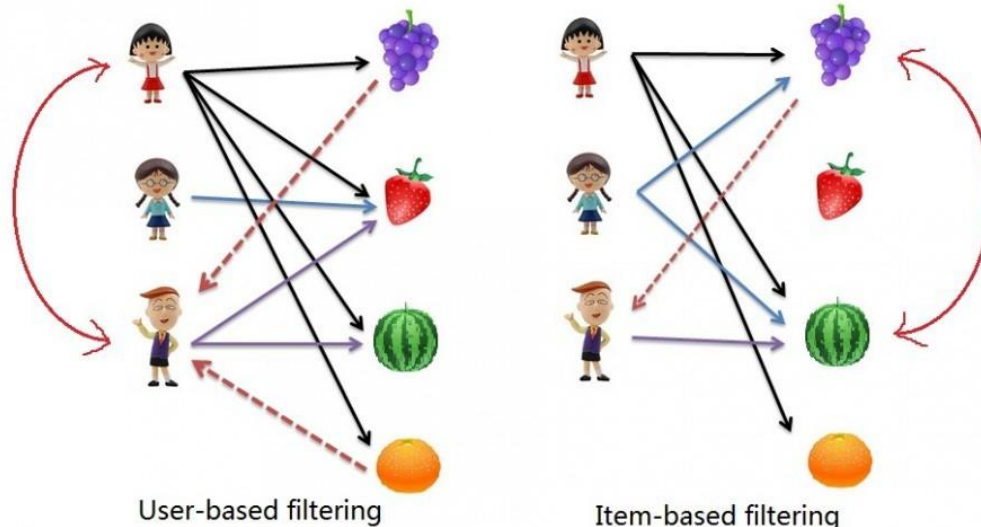


Types of Recommendation Systems

- **Content-Based Filtering**: Recommend items with similar content.
- **Collaborative Filtering**: Recommend items based on interests of a community of users.
- **Hybrid Content-based Collaborative Filtering**: Combination of the two above approaches to overcome the disadvantage of each approach.

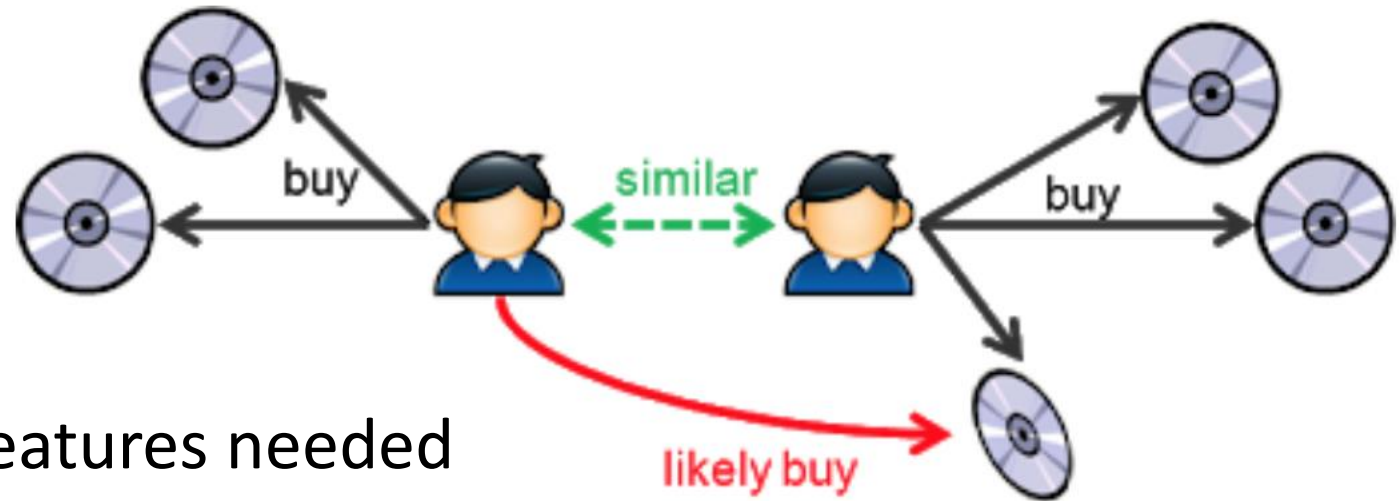
Collaborative Filtering

- It finds a subset of users who have similar tastes and preferences to the target user and use this subset for offering recommendations.
- Basic Assumptions:
 - Users with similar interests have common preferences.
 - Sufficiently large number of user preferences are available.
- Main Approaches:
 - User-Based
 - Item-Based



User-Based Collaborative Filtering

- Find a group of users that is 'similar' to user X
- Find all items liked by this group that has not been seen by user X
- Ranks these items and recommend to user X



Advantage:

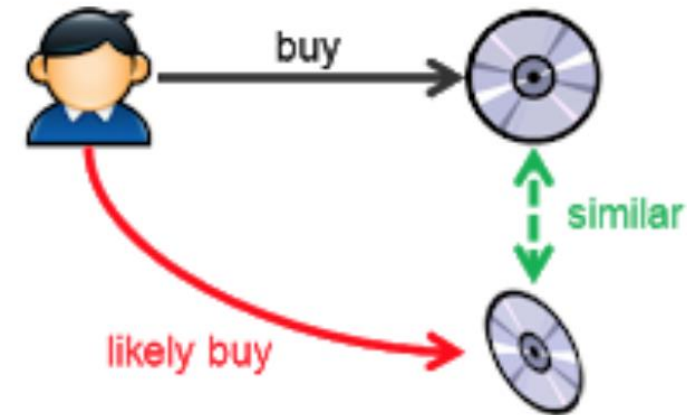
- No knowledge about item features needed

Problems:

- New user cold start problem
- New item cold start problem: items with few ratings cannot easily be recommended

Item-Based Collaborative Filtering

- Find the set of items that **user X likes** (from interaction data)
- Find a group of **items that is similar to these set of items** that we know user X likes
- Ranks these items and recommend to user X



Advantage:

- No knowledge about item features needed
- Better scalability, because correlations between limited number of items instead of very large number of users

Problems:

- New user cold start problem
- New item cold start problem

More you can learn from the below

How do recommendations on YouTube work?

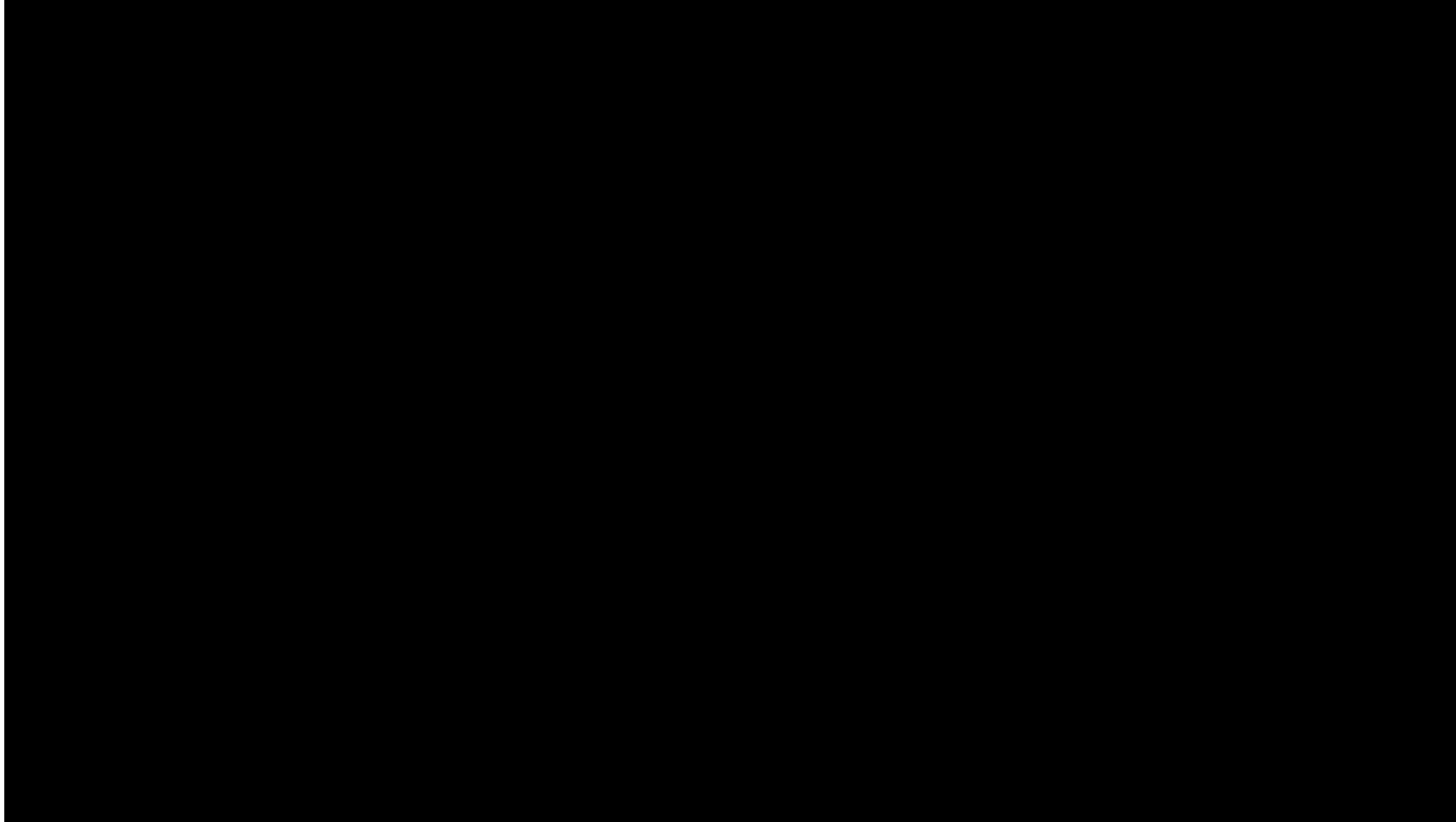
<https://www.youtube.com/watch?v=zChtNhBZmfs&t=314s>

Netflix Recommendation Systems

<https://www.youtube.com/watch?v=JwHzqYN0WTA>

An Engineer, any engineering application?

Dr. Zhiyao XIE, Asst Prof HKUST, PhD Duke USA 2022, BEng CityU EE 2017



Application---Games



IBM Deep Blue

Vs



World chess champion
Garry Kasparov in 1997

Deep Blue (chess computer)1997



AlphaGo (Weiqi) 2016

Application---Image Processing



Denoising



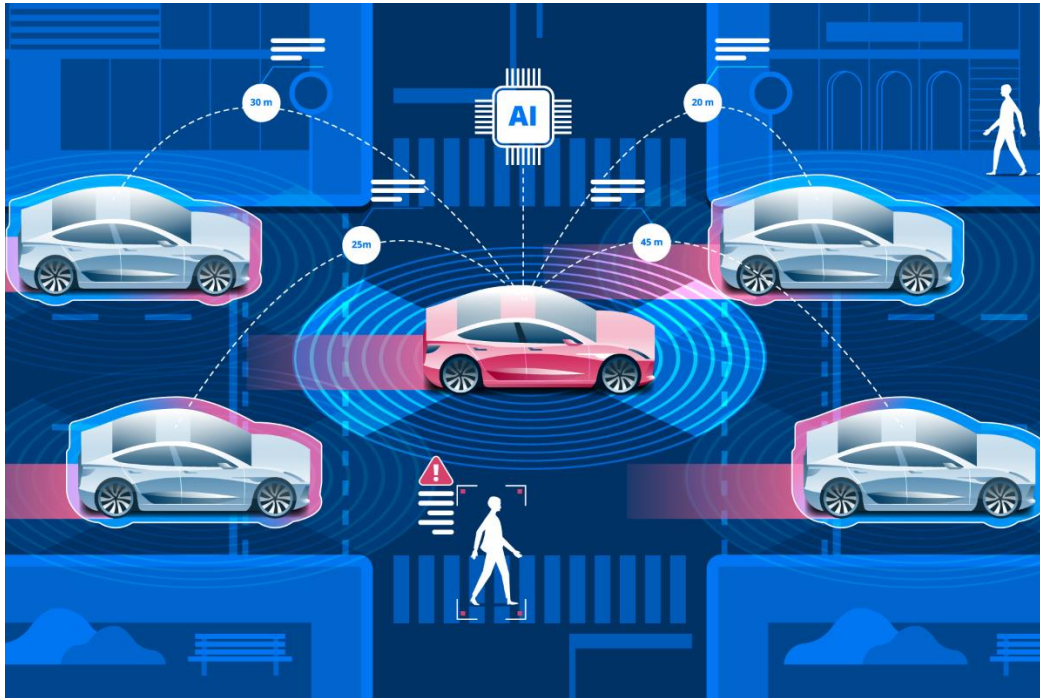
Super-resolution

PULSE: Self-Supervised Photo Upsampling via
Latent Space Exploration of Generative Models (CVPR2020)

Application-Video Processing

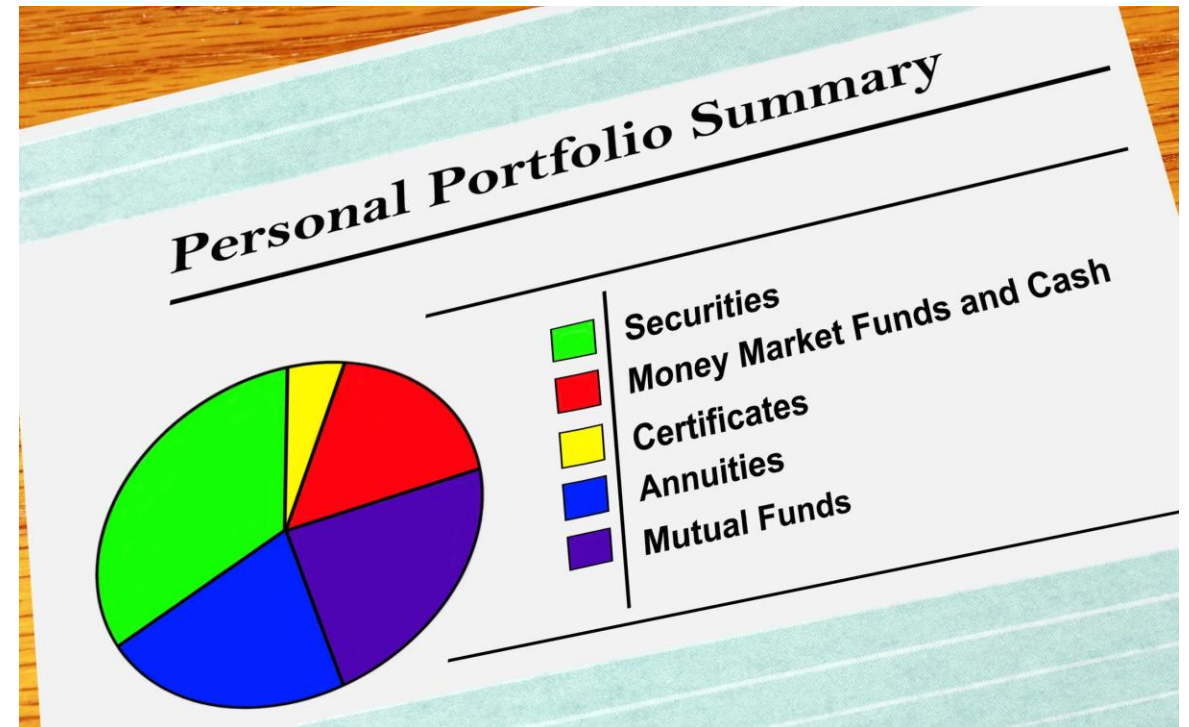


Application---Autonomous Vehicles



Actually the autonomous driving technology still have rooms to improve

Application---Finance



Application---Finance

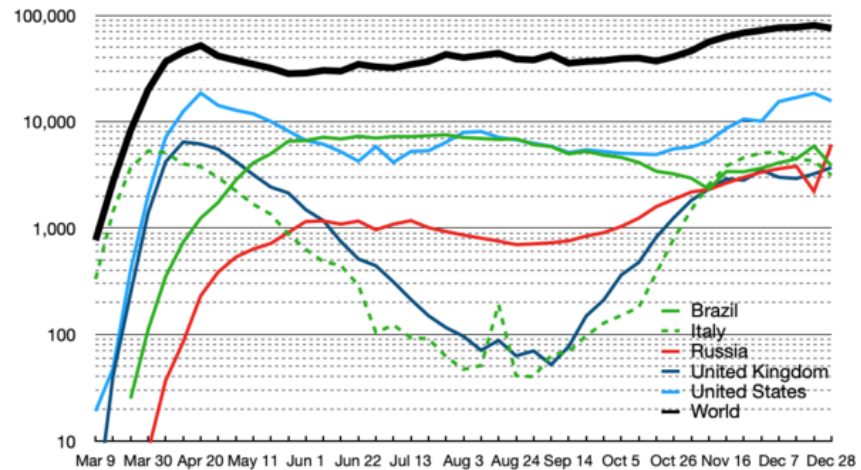
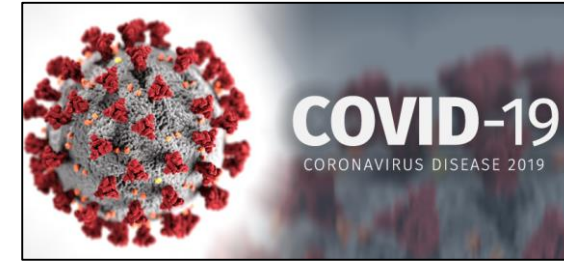
- Portfolio Management – Robo-Advisors
- Algorithmic Trading
- High-Frequency Trading (HFT)
- Fraud Detection
- Loan/ Insurance Underwriting
- Risk Management
- Chatbots
- Document Analysis
- Trade Settlements
- Money-Laundering Prevention

PCA, Clustering, different kinds of classifiers are widely used machine learning tools in finance applications

Application---Health Care



- Disease prediction/diagnosis
- Drug discovery
- Health monitoring



- Data visualization and prediction
- Treatment improvement

Machine Learning on Clinical Trials

- Improving and **speeding up** clinical trial immensely
- AI increase clinical trial success rates
- Use AI to find the “right” clinical trial
- Speed up new drug development by a 10-fold time reduction
 - ML reshaping the landscape of drug development
 - Able to analyze complex biological data. Predict molecular interactions.
 - Unintended consequences: ML decisions driven could lead to unintended consequences within biological system, such as optimizing certain drug parameters may result in new drug development that was not the intent at the first place.
 - **But there maybe disadvantages on long term safety and efficacy. ML driven model may not be able to capture human long-term safety which traditional drug development took years to monitor and pass the FDDA requirements.**

Machine Learning vs Statistical Modeling

- Probability models, i.e., Normal, Poisson, Gamma, binomial, Weibull distribution are used to model certain kinds of datasets with different behavior.
- Probability models have parameters such as mean, variance, machine learning model do not have these parameters. But some ML methods still rely on probabilistic theory; Naive Bayes classifier.
- Probability models require huge data point for accurate modeling; i.e, population of a city, probability distribution of CityU student behavior.
- For applications with a smaller sizes than the above, or with a specific behavior, we can use other modeling methods so that we can **classify new, unseen input data**; all sort of business and engineering applications.
- When the data becomes immensely large like the one at Google, they can do very accurate job without traditional statistical modeling.

Types of Data

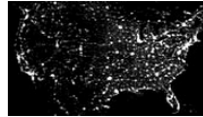
Machine Learning Repository

Center for Machine Learning and Intelligent Systems

Census Income Data Set

Download: [Data Folder](#), [Data Set Description](#)

Abstract: Predict whether income exceeds \$50K/yr based on census data. Also known as "Adult" dataset.



Data Set Characteristics:	Multivariate	Number of Instances:	48842	Area:	Social
Attribute Characteristics:	Categorical, Integer	Number of Attributes:	14	Date Donated	1996-05-01
Associated Tasks:	Classification	Missing Values?	Yes	Number of Web Hits:	93365

Source:

Donor:

Ronny Kohavi and Barry Becker
Data Mining and Visualization
Silicon Graphics
e-mail: ronnyk '@' sgi.com for questions.

Data Set Information:

Extraction was done by Barry Becker from the 1994 Census database. A set of reasonably clean records was extracted using the following conditions: ((AGE>16) && (AGI>100) && (AFNLWGT>1)&& (HRSWK>0))

Prediction task is to determine whether a person makes over 50K a year.

Attribute Information:

Listing of attributes:

>50K, <=50K.

age: continuous.
workclass: Private, Self-emp-not-inc, Self-emp-inc, Federal-gov, Local-gov, State-gov, Without-pay, Never-worked.
fnlwgt: continuous.
education: Bachelors, Some-college, 11th, HS-grad, Prof-school, Assoc-acdm, Assoc-voc, 9th, 7th-8th, 12th, Masters, 1st-4th, 10th, Doctorate, 5th-6th, Preschool.
education-num: continuous.
marital-status: Married-civ-spouse, Divorced, Never-married, Separated, Widowed, Married-spouse-absent, Married-AF-spouse.
occupation: Tech-support, Craft-repair, Other-service, Sales, Exec-managerial, Prof-specialty, Handlers-cleaners, Machine-op-inspct, Adm-clerical, Farming-fishing, Transport-moving, Priv-house-serv, Protective-serv, Armed-Forces.
relationship: Wife, Own-child, Husband, Not-in-family, Other-relative, Unmarried.
race: White, Asian-Pac-Islander, Amer-Indian-Eskimo, Other, Black.
sex: Female, Male.
capital-gain: continuous.
capital-loss: continuous.
hours-per-week: continuous.
native-country: United-States, Cambodia, England, Puerto-Rico, Canada, Germany, Outlying-US(Guam-USVI-etc), India, Japan, Greece, South, China, Cuba, Iran, Honduras, Philippines, Italy, Poland, Jamaica, Vietnam, Mexico, Portugal, Ireland, France, Taiwan, Haiti, Columbia, Hungary, Guatemala, Nicaragua, Scotland, Thailand, Yugoslavia, El-Salvador, Trinidad&Tobago, Peru, Hong, Holand-Netherlands.

- Numerical,
- Non-numerical: categorical, nominal
- Number of attributes (features)
- With label or without label
- Typical applications including supermarket dataset and ecommerce datasets

Problem of real dataset:

Could be sparse (empty space)

Hybrid: combine numerical and non-numerical

Types of learning

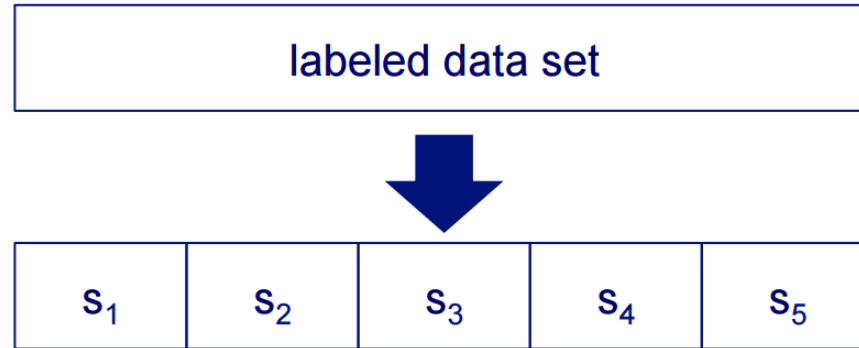
- ❑ Supervised learning
 - Regression
 - Classification
- ❑ Unsupervised learning
 - Clustering
 - Dimensionality reduction
 - Probability distribution estimation
- ❑ Semi-supervised learning
- ❑ Reinforcement learning

Supervised vs Unsupervised

- Probabilistic statistical approach rely on distribution assumption, ie., Gaussian, or uniform distribution, and huge data points to obtain mean, variance for modeling.
- Machine Learning approaches require a learning dataset to model the interested behavior, no assumption on data distribution. The larger the dataset, the more accurate the performance, **Bigdata era**.
- The dataset, i.e., CityU EE Dept student performance behavior, **consists of a label for teaching and learning (Supervised learning)**.
- Super market dataset only consists of each transactions. Amazon database consists of customer basic info; name, & address etc. **No label for teaching and leaning the model (Unsupervised approach)**.
- Labeling huge datasets is a costly procedure.

K-Fold Cross Validation Test

partition data
into n subsamples



In this case it
is 5-fold

iteratively leave one
subsample out for
the test set, train on
the rest

iteration	train on	test on
1	s_2 s_3 s_4 s_5	s_1
2	s_1 s_3 s_4 s_5	s_2
3	s_1 s_2 s_4 s_5	s_3
4	s_1 s_2 s_3 s_5	s_4
5	s_1 s_2 s_3 s_4	s_5

K can be 3, 5 and 10.
But widely used is 10-fold test

Testing Accuracy & Model comparisons

Suppose we have a small dataset of 100 instances/data point (5-fold cross validation test)

iteration	train on	test on	correct
1	s ₂ s ₃ s ₄ s ₅	s ₁	11 / 20
2	s ₁ s ₃ s ₄ s ₅	s ₂	17 / 20
3	s ₁ s ₂ s ₄ s ₅	s ₃	16 / 20
4	s ₁ s ₂ s ₃ s ₅	s ₄	13 / 20
5	s ₁ s ₂ s ₃ s ₄	s ₅	16 / 20

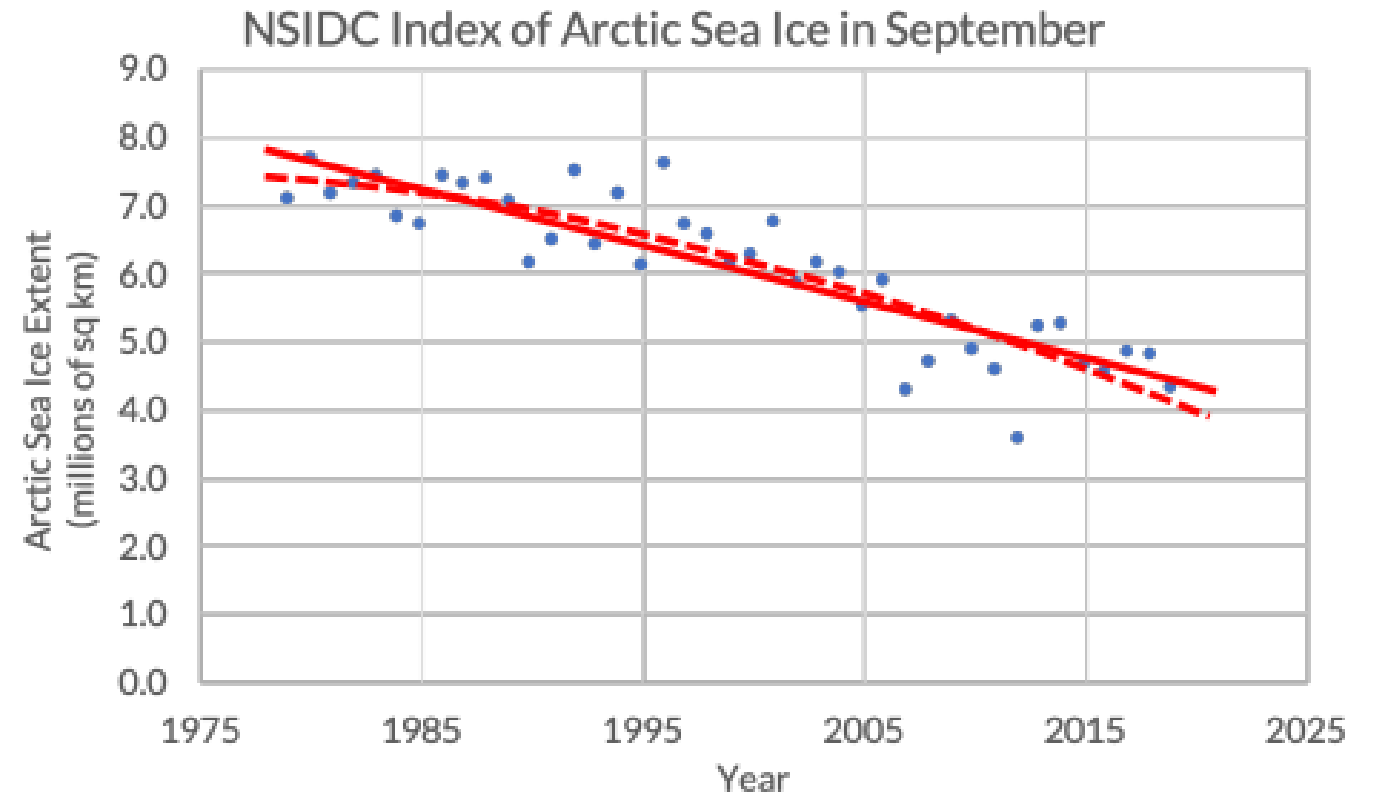
accuracy = 73/100 = 73%

Usually we use 10 fold test. K fold test is used for performance or models evaluation.

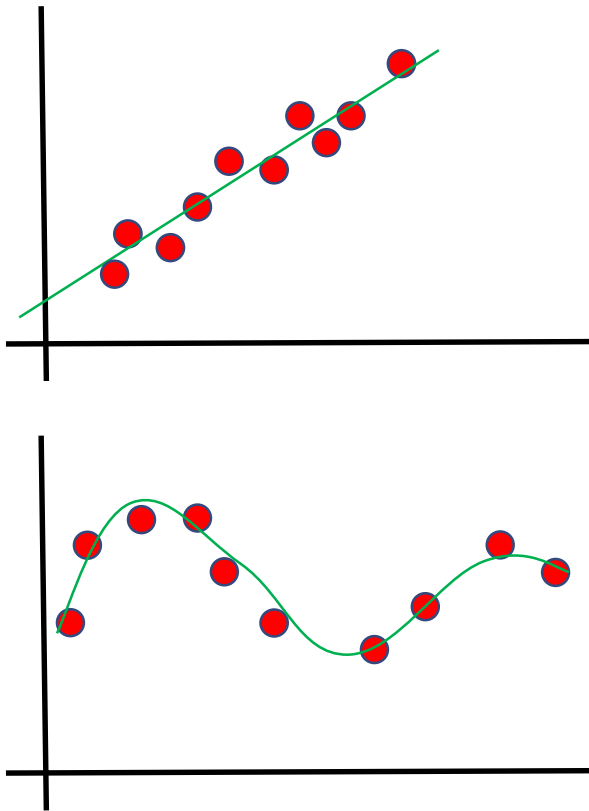
Classifier and Classification

- After comparing with different classifier models, SVM, Decision tree, deep learning, feedforward NN, LDA, KNN etc using 10-fold CV test, we form the classifier using all the given data
- We use the model to predicate, estimate, or group new unseen data.
- We call the process classification and the model used for classifying new unseen data points **CLASSIFIER**.
- Widely used ML classifiers include Feedforward Neural Network, Support vector machine (SVM), Softmax, Logistic regression, K-nearest neighbor (KNN), decision tree, LDA etc.

Supervised Learning---Regression



Supervised Learning---Regression



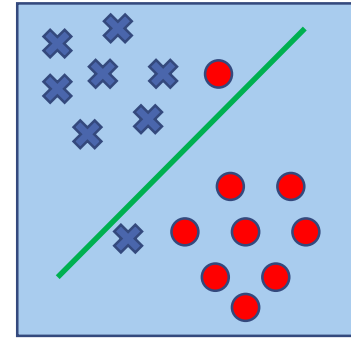
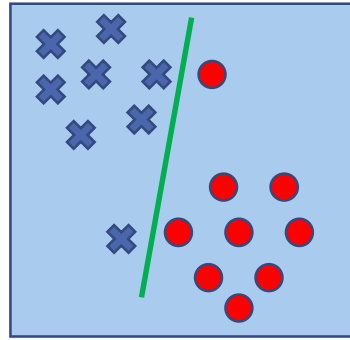
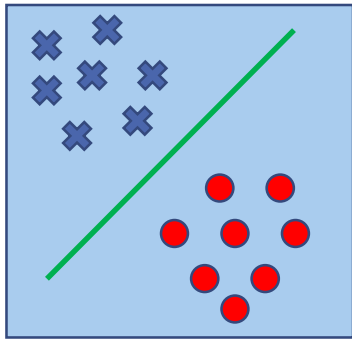
2-D and 3-D data are easy to handle and perceivable.

How about high-dimensional space, ie., 100 dimensions?

Techniques:

- least square regression
- polynomial regression
- neural networks

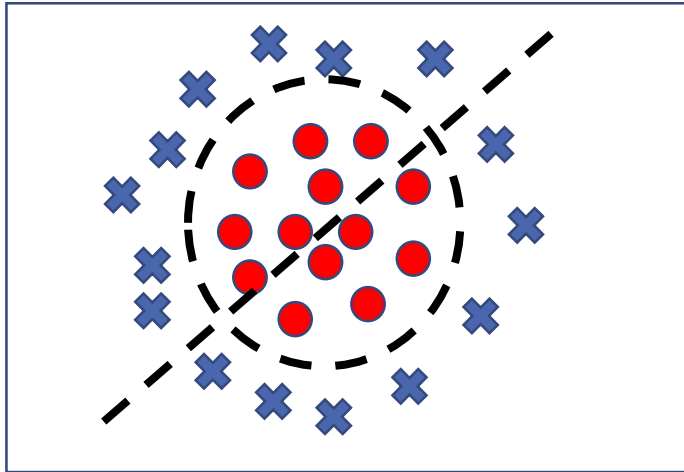
Supervised Learning---Linear Classification



Techniques

- K-nearest neighbors (kNN)
- FNN
- Logistic regression
- Support vector machine (SVM)

Supervised Learning---Non-Linear Classification



Techniques:

- K-nearest neighbors (kNN)
- Support vector machine (SVM)
- Feedforward Neural Networks (FNN)
- Deep learning

Supervised Learning

Applications---handwritten digits recognition

Training examples



MNIST (image size=28x28)

<http://yann.lecun.com/exdb/mnist/>

Testing images



Supervised Learning

Applications----facial recognition (ORL database of faces)

Training images



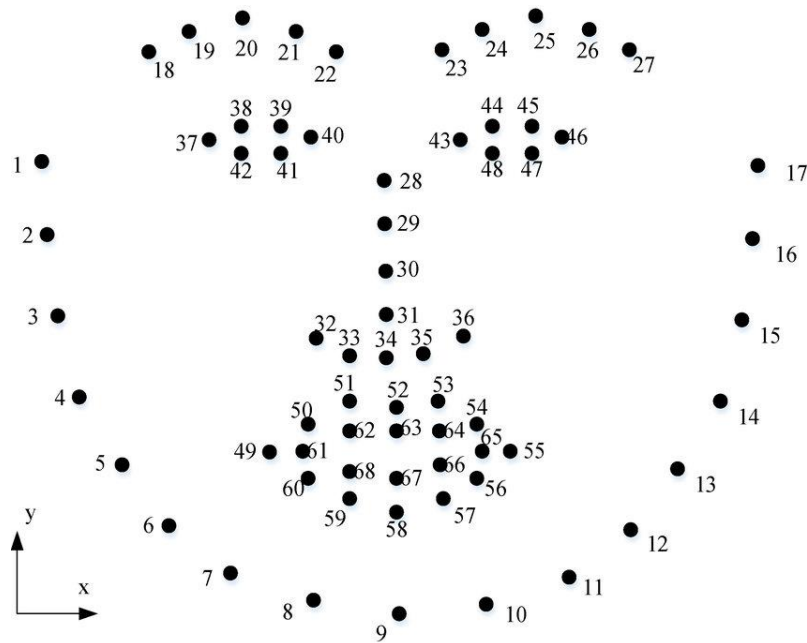
Testing images



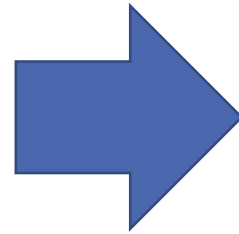
Testing images



Facial Recognition and Face Landmarks



Early work relied on 68 (or more) facial landmark points to form a 68 dimensional feature vector for facial recognition

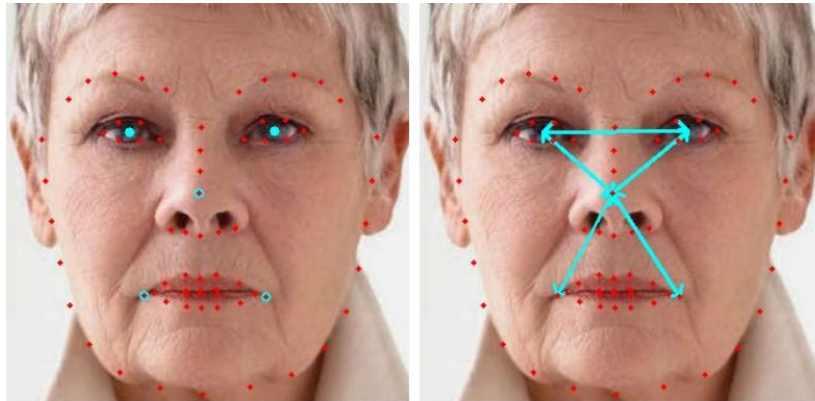


To improve classification results



468 facial landmark points

Facial Landmarks

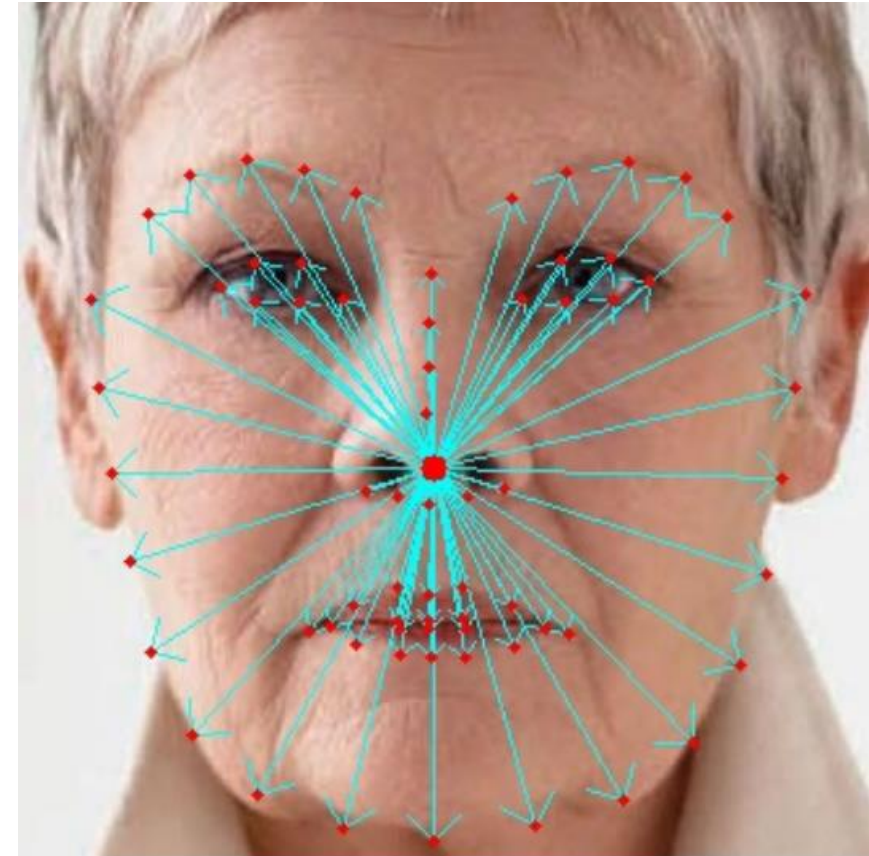


(a) Selected 5 nodal points.

(b) 5 nodal points distances.

Feature	Top 5 accuracy
5 points feature	47%
68-points feature	76%

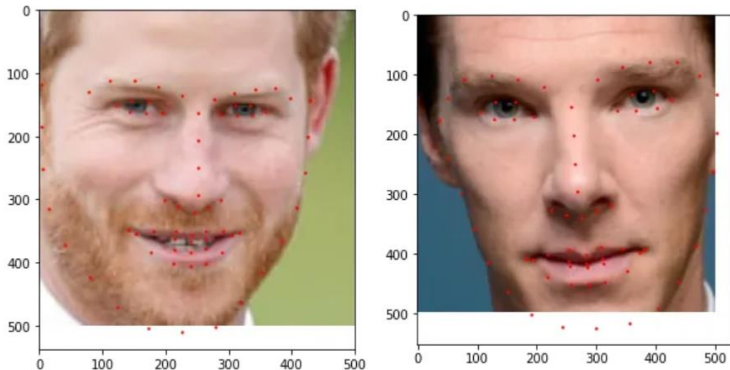
Still low compared to CNN deep technology
which reached 99%



68 facial landmark points are obtained
from centroid of the face measurement

Facial Landmark points

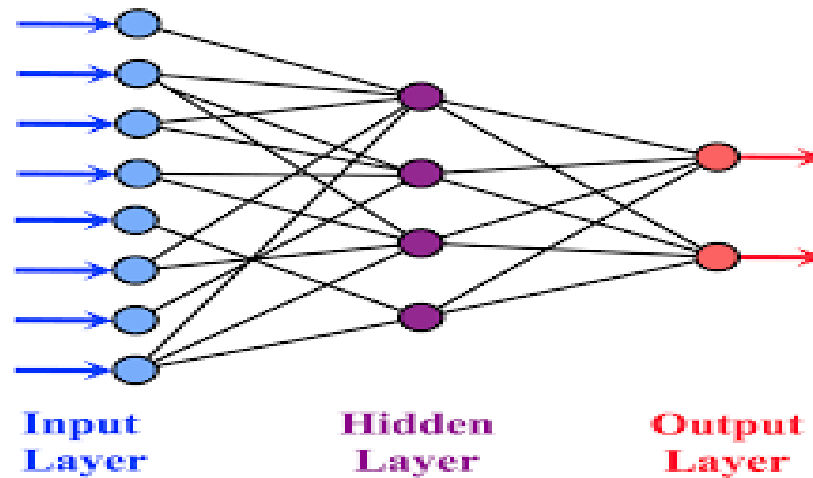
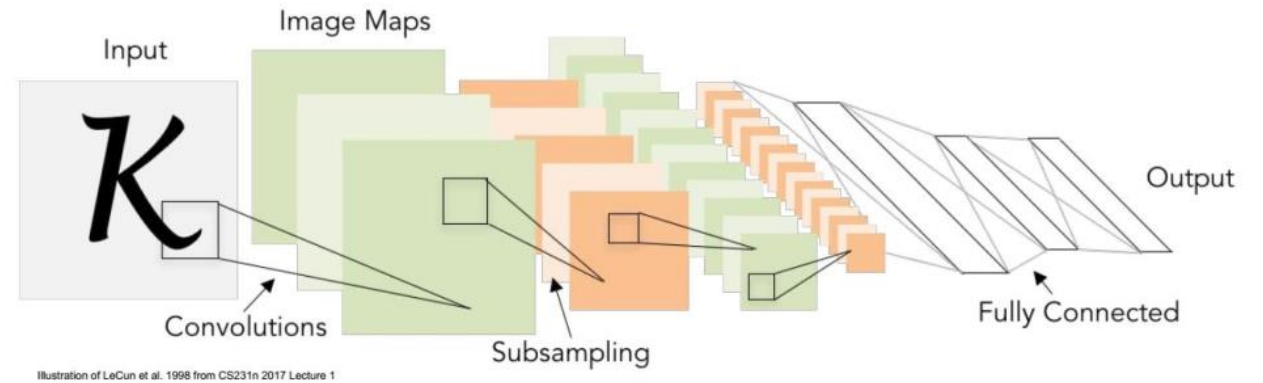
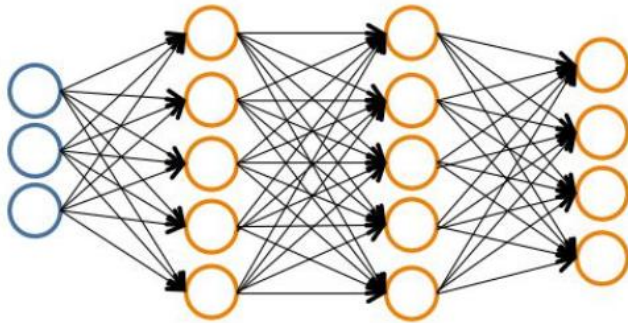
- Since Deep learning and CNN have widely been used and have provided excellent classification accuracy of up to 99% for frontal facial recognition, facial landmark points shift to other applications
- Used for facial gesture ie., angry , depressed, nervous, clam, happy, ecstasy classification
- Used for distinguishing and compare one face from another face. To determine how 2 given faces are different ie., eye width, big/small nose
- Also useful in forensics cases to be discussed in court as they provide objective measures



- We still need to use computational method to determine the facial landmarks from given images especially when there are huge number of images
- Still an interesting research topic. In 2021 there were researchers proposed to use deep recurrent neural network for determining facial landmark points

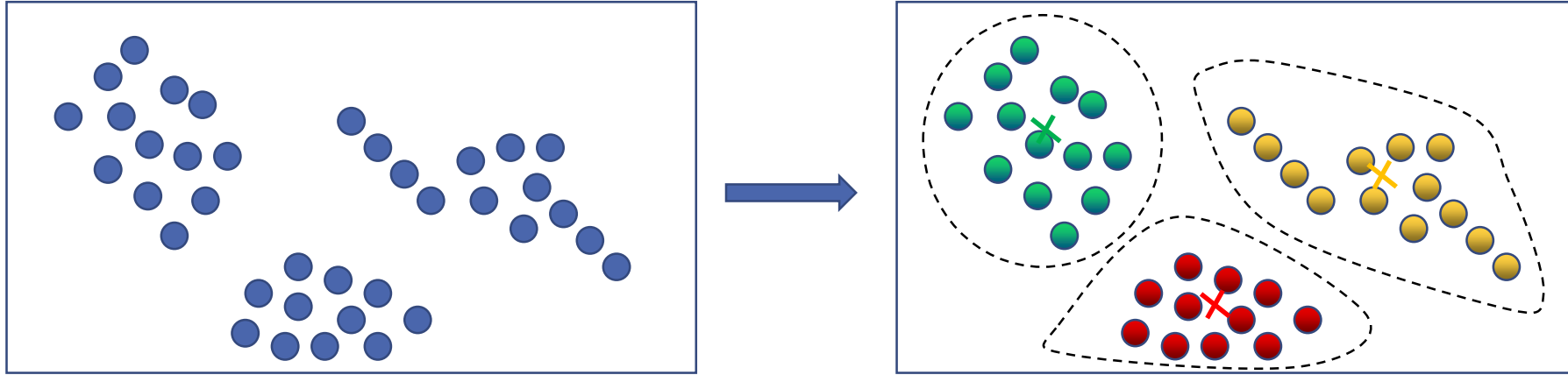
Supervised Learning

Deep learning



Multi-layer or shallow
neural network

Unsupervised Learning-Clustering



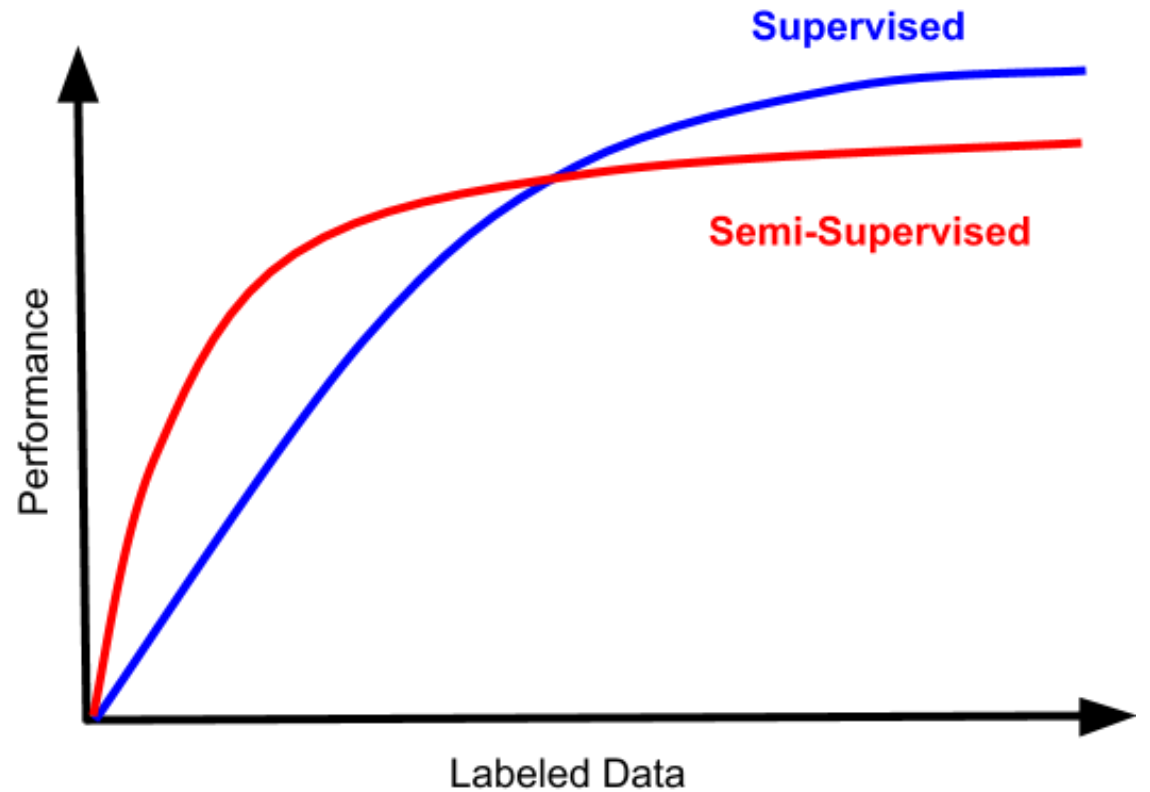
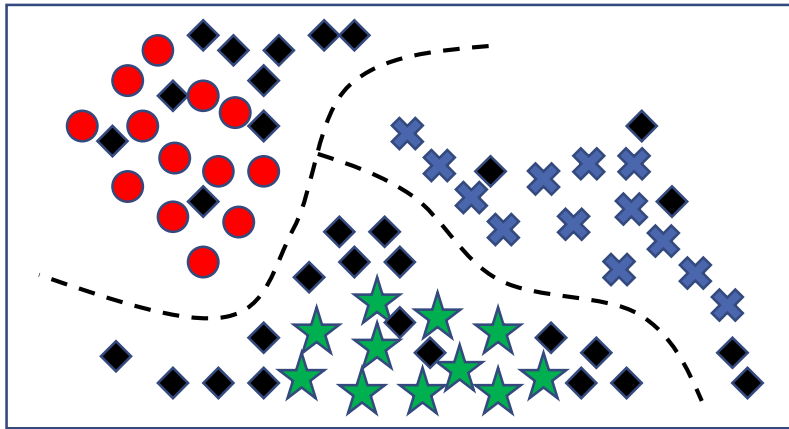
Methods:

- hierarchical clustering
- **k-means clustering**
- Gaussian mixture models
- spectral clustering
- deep learning clustering

Why unsupervised?

No label dataset, we need to find the information based on the given dataset

Semi-Supervised Learning (SSL)



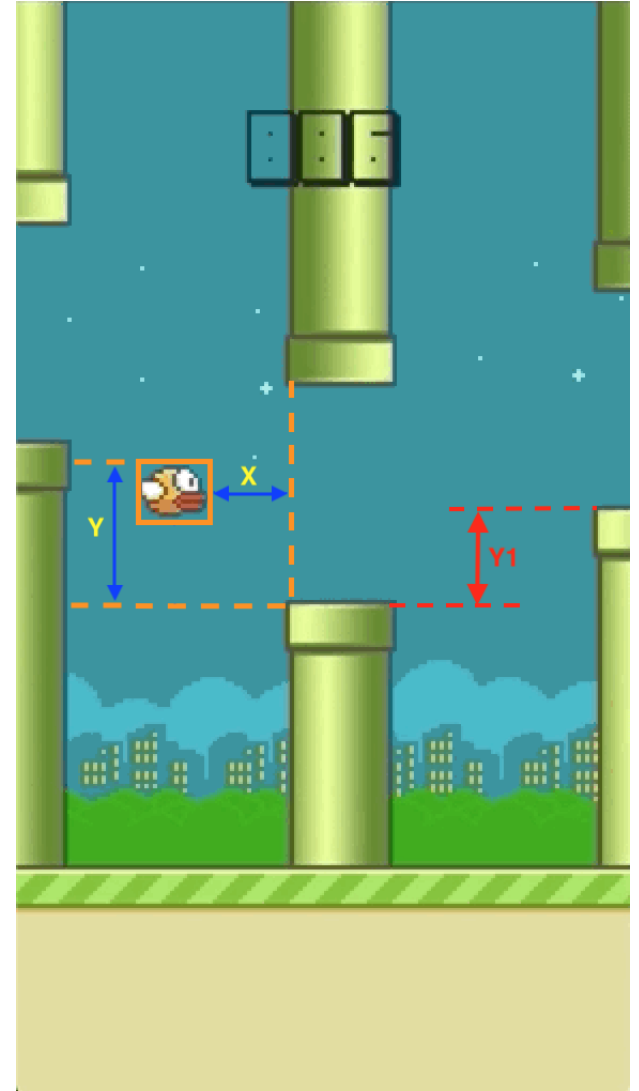
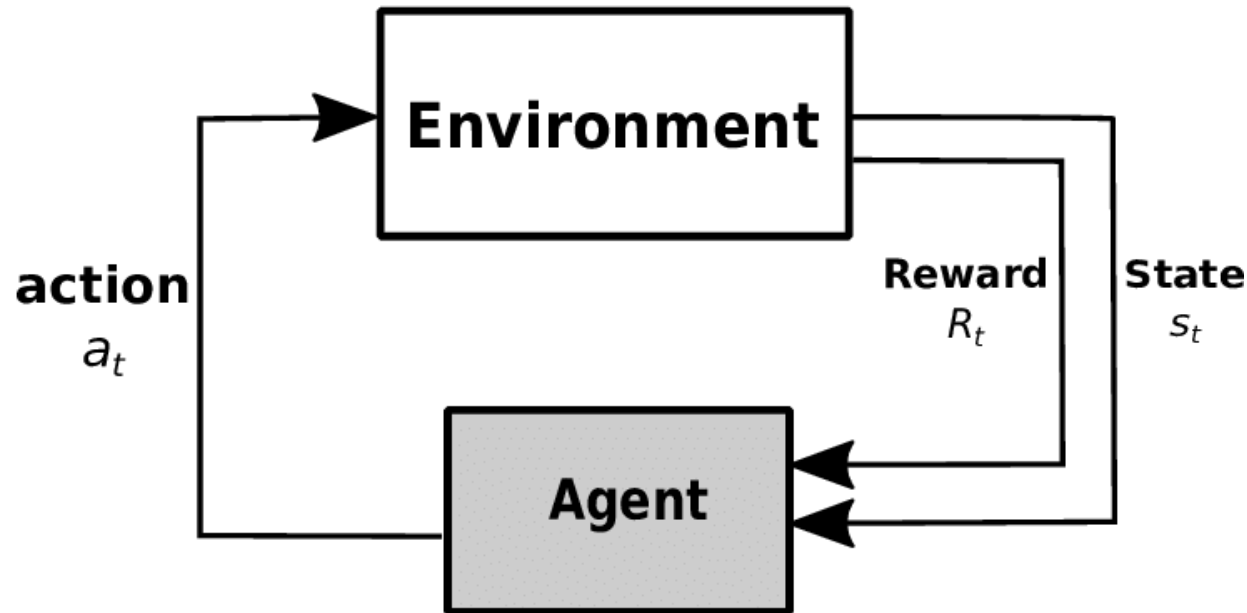
Why SSL?

- Labeled data are “expensive”
- Large number of unlabeled data are available (millions of data points)
- Unlabeled data help learning

Reinforcement Learning

Features of RL

- Learning a policy: a sequence of outputs
- No supervised output but delayed reward



Reinforcement Learning

Real applications

- Bidding and advertising
- Game playing
- Robot



AI Safety Discussions

Cyber attacks

- Attackers can use generative AI and large language models (LLM) to scale attacks at an unseen level of speed and complexity. They may use generative AI to find fresh ways to undermine networks and take advantage of geopolitical tensions for advanced attacks.

Physical safety

- As more systems such as autonomous vehicles, manufacturing and construction equipment, and medical systems use AI, risks of artificial intelligence to physical safety can increase.
- For example, an AI-based true self-driving car that suffers a cyber security breach could result in risks to the physical safety of its passengers. Similarly, the dataset for maintenance tools at a construction site could be manipulated by an attacker into creating hazardous conditions
- Can we imagine that cyber attack that hijacking hundreds of self driving cars to act as whatever purposes.

AI Privacy risks

- A hacker who breaches an AI system could access different kinds of sensitive information.
- An AI system designed for marketing, advertising, profiling, or surveillance could also threaten privacy

2014 Generative Adversarial Network (GAN)



*I. Goodfellow et al., "Generative Adversarial Networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, Oct. 2020, doi: 10.1145/3422622.

^Y. Choi, Y. Uh, J. Yoo, and J.-W. Ha, "StarGAN v2: Diverse Image Synthesis for Multiple Domains," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 8185–8194. doi: 10.1109/CVPR42600.2020.00821.

Impersonation (Deep Fake): Lost of TRUST

- Deep fake, the real problems, use a form of artificial intelligence called deep learning to make images of fake events
- Deepfake can put words in a politician's mouth (in audio format), star in your favourite movie, or dance like a pro.
- Deepfake **generated** many pornographic under celebrities face image.
- The AI firm **Deeptrace** found 15,000 deepfake videos online in September 2019, a near doubling over nine months. A staggering 96% were pornographic and 99% of those mapped faces from female celebrities on to porn stars.
- Several companies will make fake image/movie for you and do all the processing in the cloud. There's even a mobile phone app, Zao, that lets users add their faces to a list of TV and movie characters on which the system has trained.
- Voice –cloning
- **DETECTION RESEARCH**

Deep Learning can Do a lot of things and Fake a lot of Things



Source: K F Lee Talk on Deep Learning

李開復

https://www.google.com/search?q=KF+Lee+Deep+learning+youtube&rlz=1C1CHBF_enHK861HK862&oq=KF+Lee+Deep+learning+youtube&aqs=chrome..69i57j33i160.8872j0j15&sourceid=chrome&ie=UTF-8#fpstate=ive&vld=cid:6dc6e414,vid:ajGgd9Ld-Wc

Can You Sue Someone for a Deepfake?



AI-generated videos or images falsely depict someone saying or doing something they never did can have severe consequences, ranging from reputational damage to emotional distress.

Can you sue someone for a deepfake?

The answer is complex and varies based on several factors, including jurisdiction, the nature of the deepfake, and the specific damages incurred.

Deepfake Tech Is Being Used to Create Fictitious Faces, Cats, Airbnbs - Business Insider

Thank You
END