Machine Learning
機器學習理論

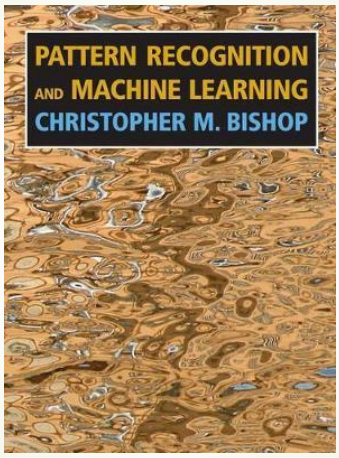
1

林嘉文 (Chia-Wen Lin)
清華大學電機系
cwlin@ee.nthu.edu.tw

2/13/2022

2

Textbook



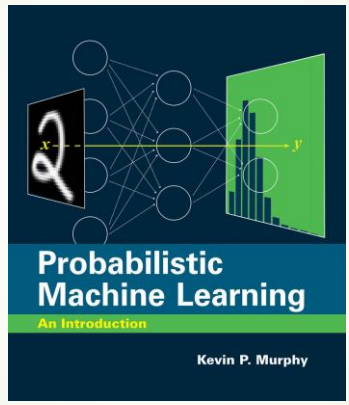
Christopher Bishop
Laboratory Director,
Microsoft Research
Cambridge

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

2/13/2022

3

Reference



Kevin P. Murphy
Research Scientist.
Google
Co-EiC, JMLR

Kevin P. Murphy, Probabilistic Machine Learning: An Introduction, The MIT Press, 2022

2/13/2022

Course webpage: https://eeclass.nthu.edu.tw/

4

Course Content

- Introduction
- Probability Distributions
- Linear Regression Models
- Linear Classification Models
- Neural Networks & Deep CNNs
- Kernel Methods & SVM
- Graphical Models
- Mixture Models & EM
- Beyond Supervised Learning: Clustering, Transfer Learning, Few-Shot Learning, …

2/13/2022

5

Grading Policy

- Homework (60%)
  - Written exercises (selected problems) (20%)
  - Computer assignments (40%)
- No midterm & final exams
- Course project for 3~4-person team (40%)

6

Course Project

- 3~4 persons/team
- 工研院巨資中心 Aidea 平台
- 2~3 topics determined by lucky draw
- Grading
  - Midterm proposal presentation (25%)
  - Rank in the topic (40%)
  - Final report (35%)

7

Aidea Projects

議題	資料提供單位
AOI瑕疵分類	工研院光電所
載客熱點預測	台灣大車隊
地層透水性參數分析	國立中央大學
排煙脫硫警示預測	中國石油化學
長期安置機構類別預測	臺中市家防中心
尋找病媒蚊孳生源—積水容器影像物件辨識	疾管署
馬拉松運動博覽會訪動線類別預測	展盟展覽
台灣牧場乳量預測	中華民國乳業協會
臺灣海洋廢棄物預測	荒野保護協會
語音資料辨識分析	AI語音數據資料集
圖書館資源對學習成效之影響預測	亞洲大學圖書館
員工離職預測	某企業
腦腫瘤分割	臺大醫神

8

Academic Integrity

- Can discuss HW with peers, but cannot copy and/or share code
- Carefully document any sources within HW hand-in
- Do not use your published work as your final project
- Plagiarism. Zero tolerance.

9

What is Machine Learning

A popular definition of machine learning or ML, due to Tom Mitchell [CMU], is as follows:

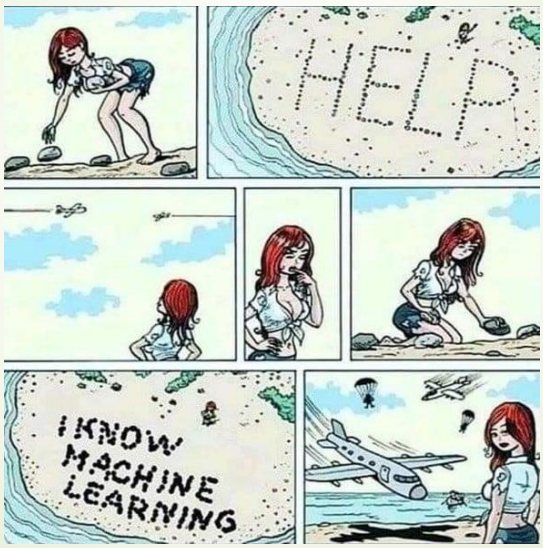
A computer program is said to learn from experience E with respect to some class of tasks T, and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.

- The task T we wish the system to learn
- The nature of the performance measure P we use to evaluate the system
- The nature of the training signal or experience E we give it

2/13/2022

10

Why to Learn Machine Learning?



2/13/2022

11

ABCD for AI

Steering Wheel (Business)

Fuel (Data)

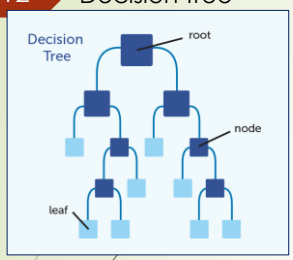
Engine (Algorithm)

Wheels (Computing)

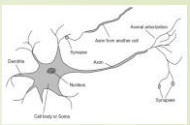
12

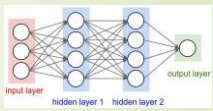
Schools of Machine Learning **Algorithms**

Decision Tree

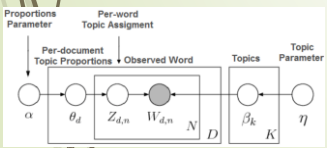


Neural Net

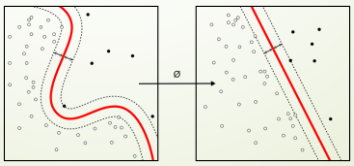




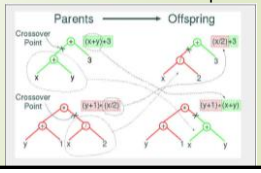
Probabilistic Model



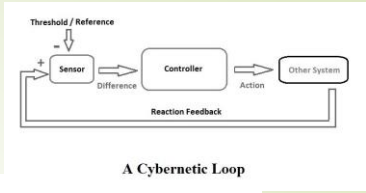
SVM



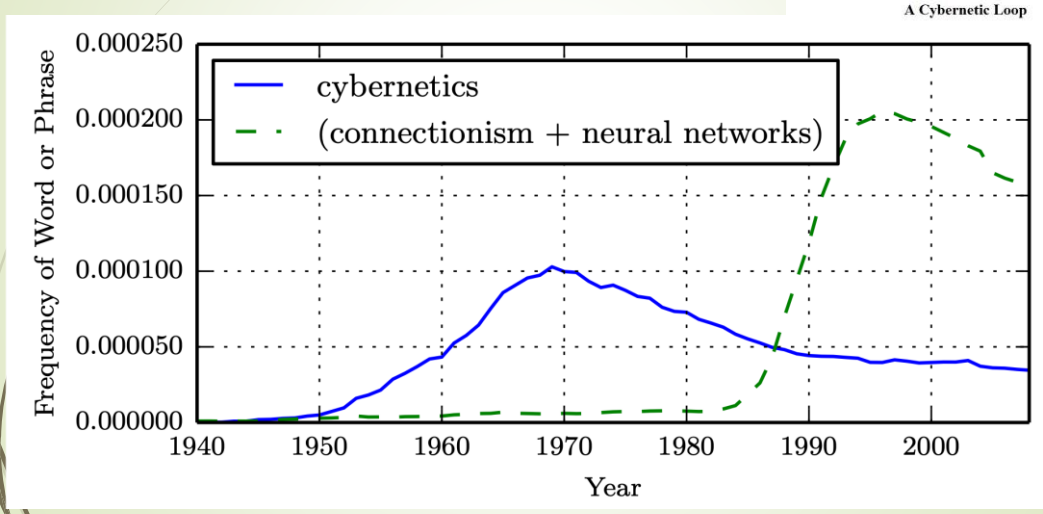
Evolutionary Computing



13 Historical Waves



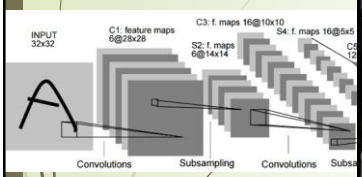
A Cybernetic Loop



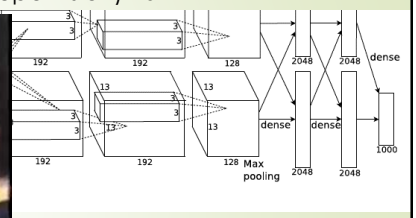
/13/2022

14 Connectionists: CNNs

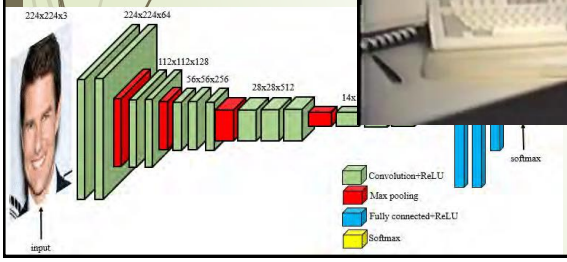
LeNet 1998



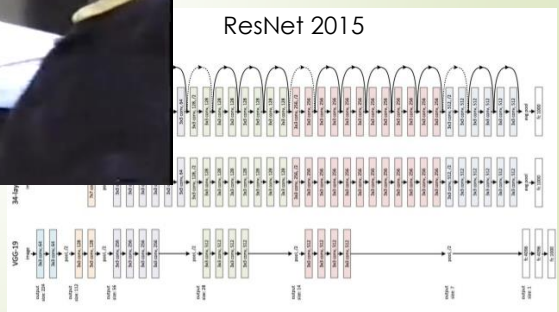
AlexNet (Supervision) 2012



VGG 2014

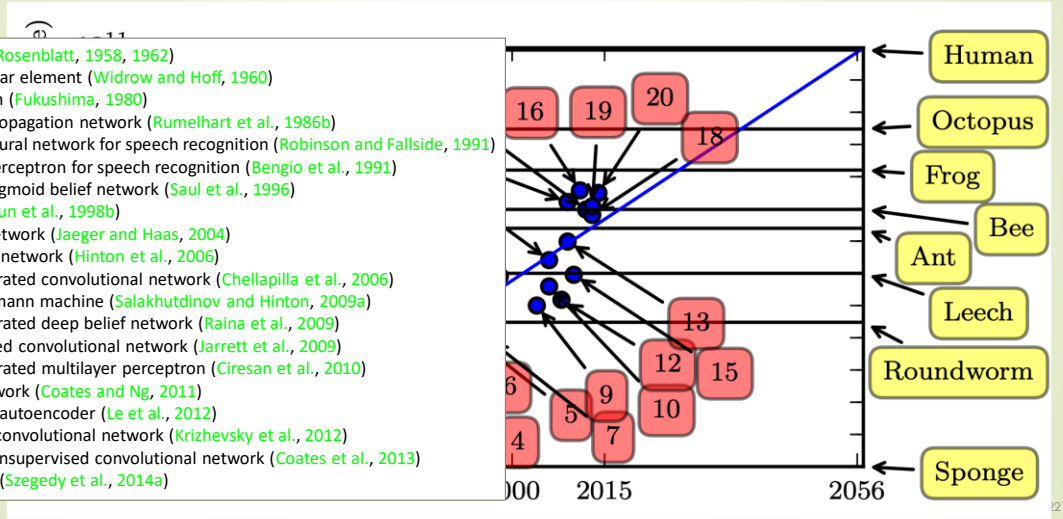


ResNet 2015



15 Connectionists: CNNs

1. Perceptron (Rosenblatt, 1958, 1962)
2. Adaptive linear element (Widrow and Hoff, 1960)
3. Neocognitron (Fukushima, 1980)
4. Early back-propagation network (Rumelhart et al., 1986b)
5. Recurrent neural network for speech recognition (Robinson and Fallside, 1991)
6. Multilayer perceptron for speech recognition (Bengio et al., 1991)
7. Mean field sigmoid belief network (Saul et al., 1996)
8. LeNet-5 (LeCun et al., 1998b)
9. Echo state network (Jaeger and Haas, 2004)
10. Deep belief network (Hinton et al., 2006)
11. GPU-accelerated convolutional network (Chellapilla et al., 2006)
12. Deep Boltzmann machine (Salakhutdinov and Hinton, 2009a)
13. GPU-accelerated deep belief network (Raina et al., 2009)
14. Unsupervised convolutional network (Jarrett et al., 2009)
15. GPU-accelerated multilayer perceptron (Ciresan et al., 2010)
16. OMP-1 network (Coates and Ng, 2011)
17. Distributed autoencoder (Le et al., 2012)
18. Multi-GPU convolutional network (Krizhevsky et al., 2012)
19. COTS HPC unsupervised convolutional network (Coates et al., 2013)
20. GoogLeNet (Szegedy et al., 2014a)



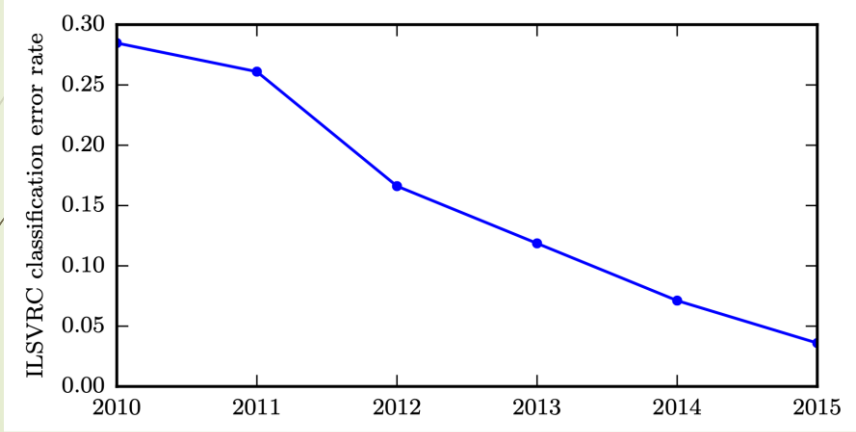
16 Connectionists: CNNs

		place	top 5 error
2011	Compressed Fisher kernel + SVM	1st	25.8%
2012	SIFT + GIST + LBP + PA classifier	2nd	26.1%
2012	Supervision	1st	16.4%
2013	Clarifai	1st	11.5%
2014	VGG	2nd	7.3%
2014	GoogLeNet / Inception	1st	6.6%
2014	*Andrej Karpathy* 	n/a	5.1%
2015	Batch Normalization Inception	n/a	4.8%
2015	Inception v3	2nd	3.6%
2015	ResNet	1st	3.6%
2016	Inception-ResNet	n/a	3.1%

2/13/2022

17

Connectionists: CNNs



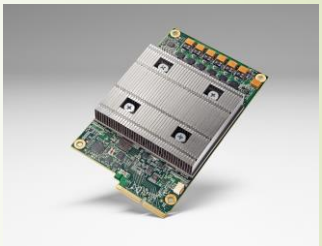
2/13/2022

18

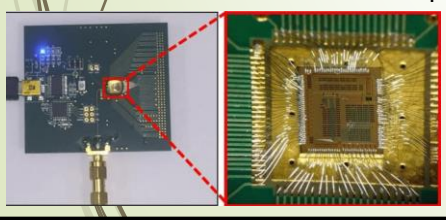
Computing for AI

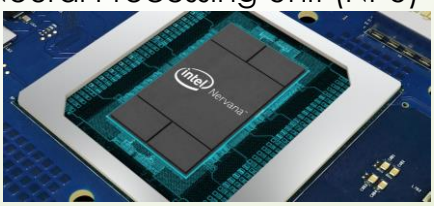
Graphics Processing Unit (GPU)

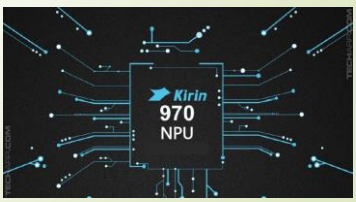
Tensor Processing Unit (TPU)



Neural Processing Unit (NPU)

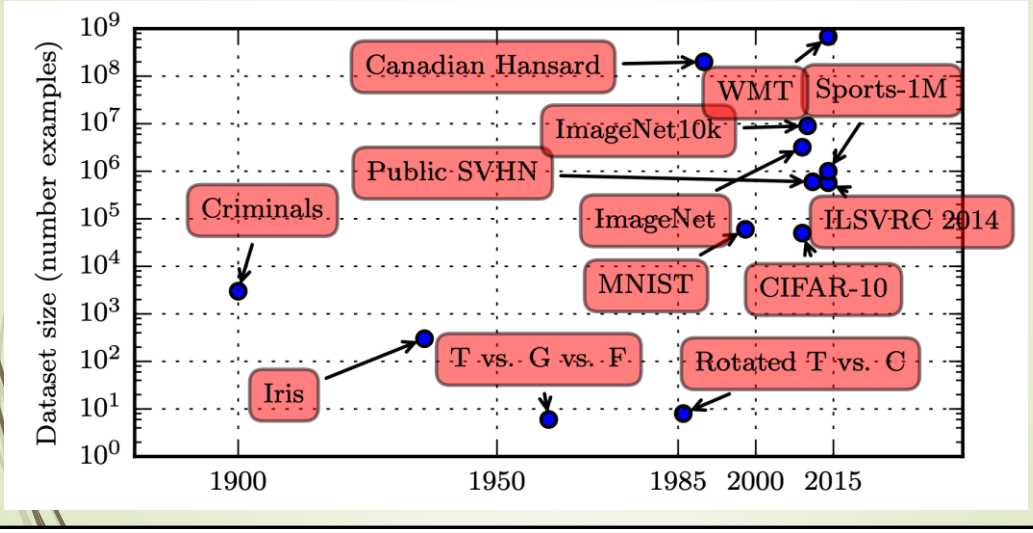






19

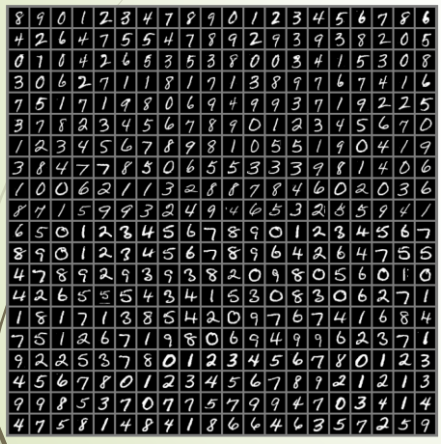
Data for AI



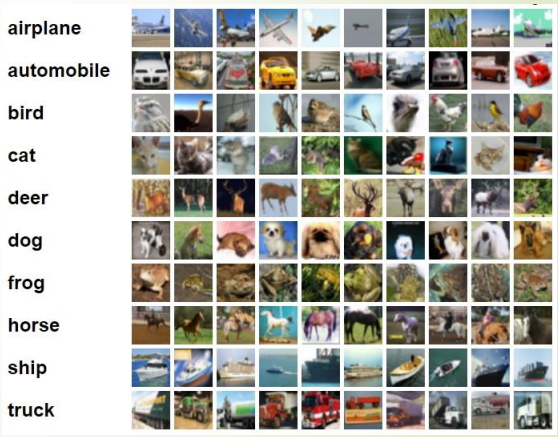
20

Data for AI

MNIST

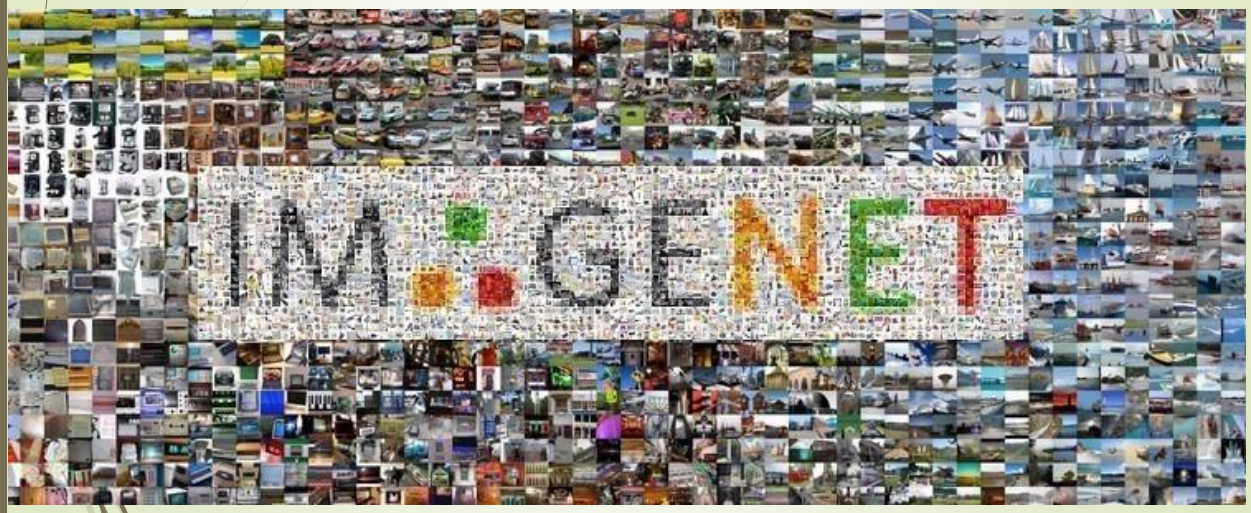


CIFAR-10



21

Data for AI



22

Data for AI

Era of Data Exploration

- 90% ----- Currently, 90% data were generated in recent two years
- 1.7 MB ----- In 2020, each person will generate 1.7 MB per second on average
- 50 B ----- In 2020, there will be ≥ 50 B Internet devices

Status of Data Usage

- 0.5% ----- The current percentage of data analyzed/used
- 10% vs. 65M ----- We can obtain 65M revenue for every increased 10% data usability

2/13/2022

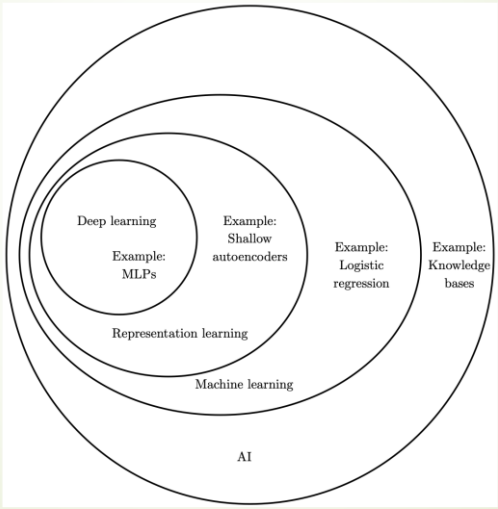
23

Why "Learn" ?

- Machine learning is programming computers to optimize a performance criterion using example data or past experience
- There is no need to "learn" to calculate payroll
- Learning is used when:
  - Human expertise does not exist (navigating on Mars),
  - Humans are unable to explain their expertise (speech/object recognition)
  - Solution changes in time (routing on a computer network)
  - Solution needs to be adapted to particular cases (user biometrics)

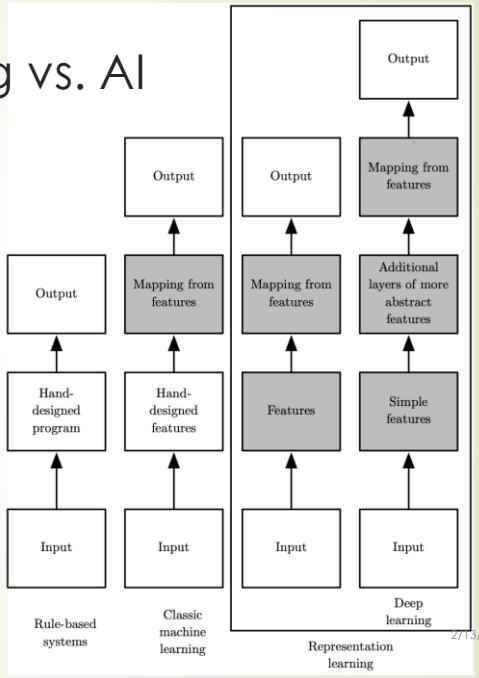
24

Machine Learning vs. AI



25

Machine Learning vs. AI



26

What We Talk About When We Talk About "Learning"

- Learning general models from data of particular examples
- Data is cheap and abundant (data warehouses, data marts); knowledge is expensive and scarce.
- Example in retail: Customer transactions to consumer behavior: *People who bought "Da Vinci Code" also bought "The Five People You Meet in Heaven" (www.amazon.com)*
- Build a model that is a good and useful approximation to the data.

27

Role of CS & Statistics in Machine Learning

- Optimize a performance criterion using example data or past experience.
- Role of Statistics: Inference from a sample
- Role of Computer science:
  - Efficient algorithms to Solve the optimization problem
  - Representing and evaluating the model for inference

28

Issues in Machine Learning?

- What algorithms can approximate functions well (and when)?
- How does number of training examples influence accuracy?
- How does complexity of hypothesis representation impact it?
- How does noisy data influence accuracy?
- What are the theoretical limits of learnability?
- How can prior knowledge of learner help?
- What clues can we get from biological learning systems?
- How can systems alter their own representations?

29

Machine Learning Problems

	Supervised Learning	Unsupervised Learning
Discrete	Classification	Clustering
Continuous	Regression	Dimensionality reduction

2/13/2022

30

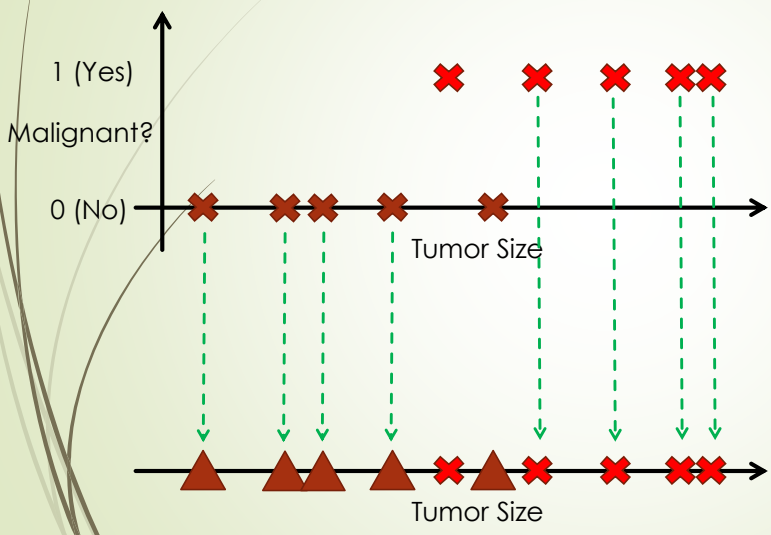
Machine Learning Problems

	Supervised Learning	Unsupervised Learning
Discrete	Classification	Clustering
Continuous	Regression	Dimensionality reduction

2/13/2022

31

Breast Cancer (Malignant, Benign)



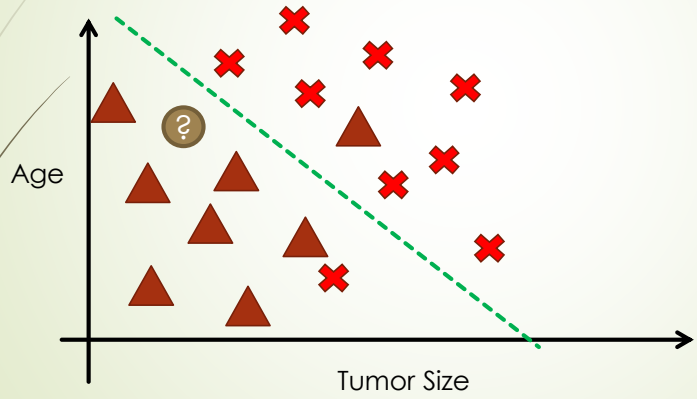
Classification problem
Discrete valued output
e.g., 0 or 1

Multi-class classification
e.g., 0 or 1 or 2 or 3

2/13/2022

32

Breast Cancer (Malignant, Benign)



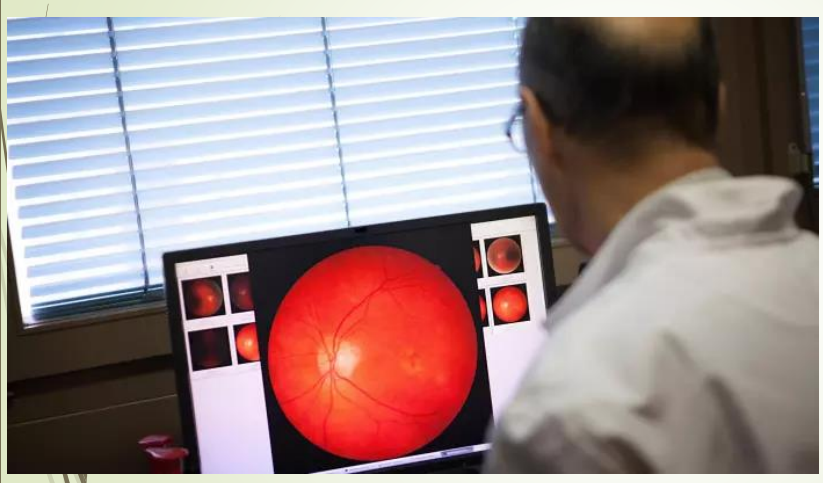
Multiple features

- Clump thickness
- Uniformity of cell size
- Uniformity of cell shape
- …

2/13/2022

33

Retinal Image Analysis for Spotting Eye Disease



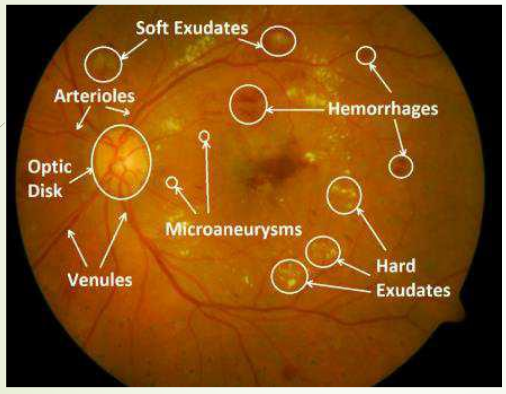
- Recognize 50 sight-threatening eye diseases
- As accurately as world-leading expert doctors

https://www.youtube.com/watch?v=MCI0xEGvHx8

Clinically applicable deep learning for diagnosis and referral in retinal disease, Nature Medicine, 2018

34

Retinal Image Analysis for Spotting Eye Disease



Syndromes

2/13/2022

35

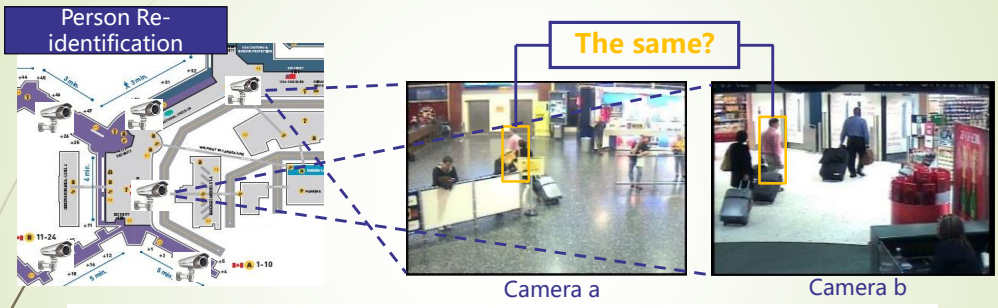
Face Recognition

Facebook auto-tagging

2/13/2022

36

Person Reidentification

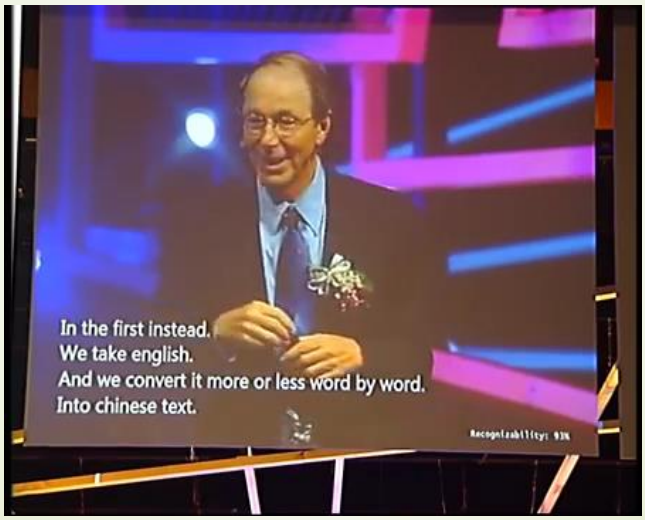


Low-resolution | shade | Pose variation | illumination | Blurred vision

2/13/2022

37

Speech Recognition



38

Machine Translation

39

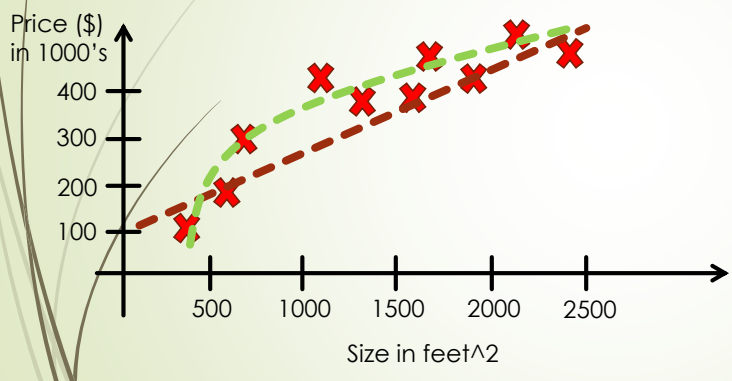
Machine Learning Problems

	Supervised Learning	Unsupervised Learning
Discrete	Classification	Clustering
Continuous	Regression	Dimensionality reduction

2/13/2022

40

Housing Price Prediction

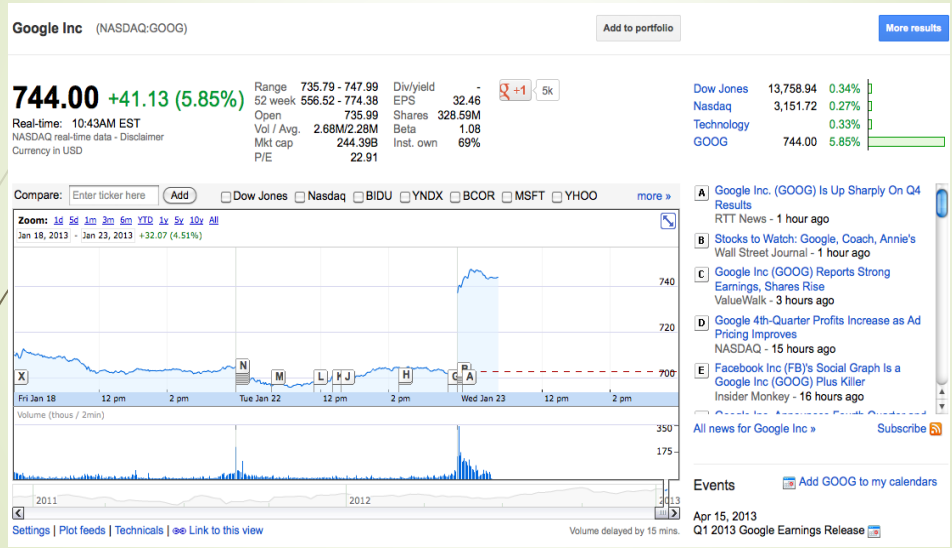


Regression problem
Continuous valued output (price)

2/13/2022

41

Stock Market



2/13/2022

42

Weather Prediction

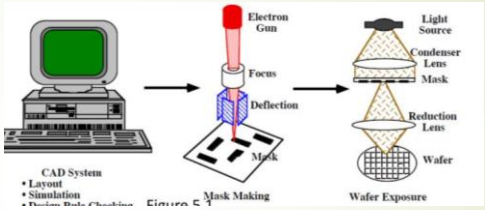
Temperature

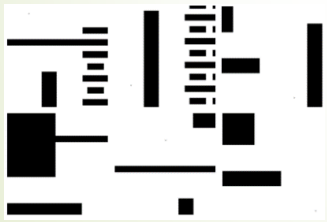
27°C

2/13/2022

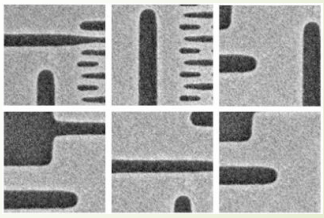
43

IC Fabrication Prediction





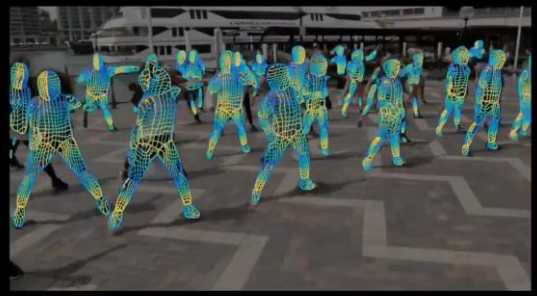
IC Layout



SEM of Fabricated

44

Human Pose Estimation



We introduce a system that can associate every image pixel with human body surface coordinates.

45

Facial Landmark Alignment

2/13/2022

Snapchat filter

https://www.youtube.com/watch?v=Pc2aJxnmzh0

46

Machine Learning Problems

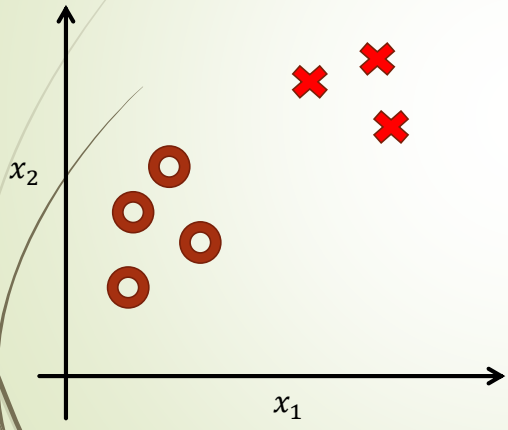
	Supervised Learning	**Unsupervised Learning**
Discrete	Classification	Clustering
Continuous	Regression	Dimensionality reduction

2/13/2022

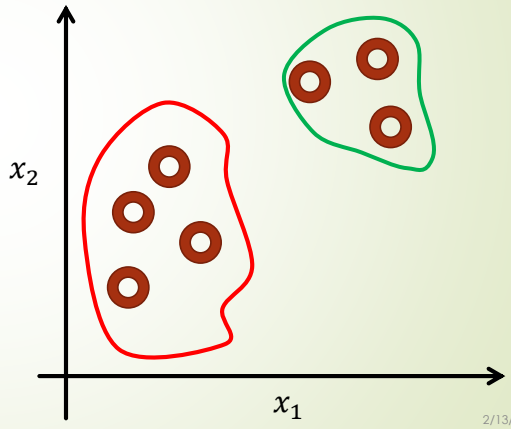
47

Supervised vs. Unsupervised Learning

Supervised Learning



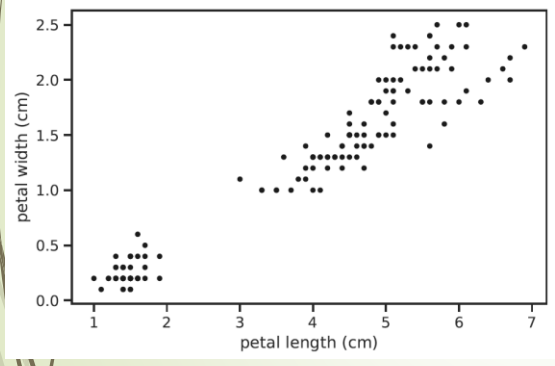
Unsupervised Learning

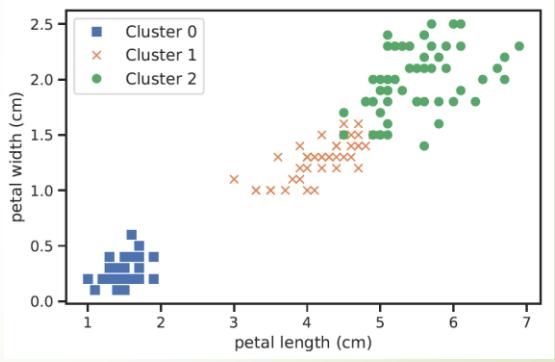


2/13/2022

48

Clustering

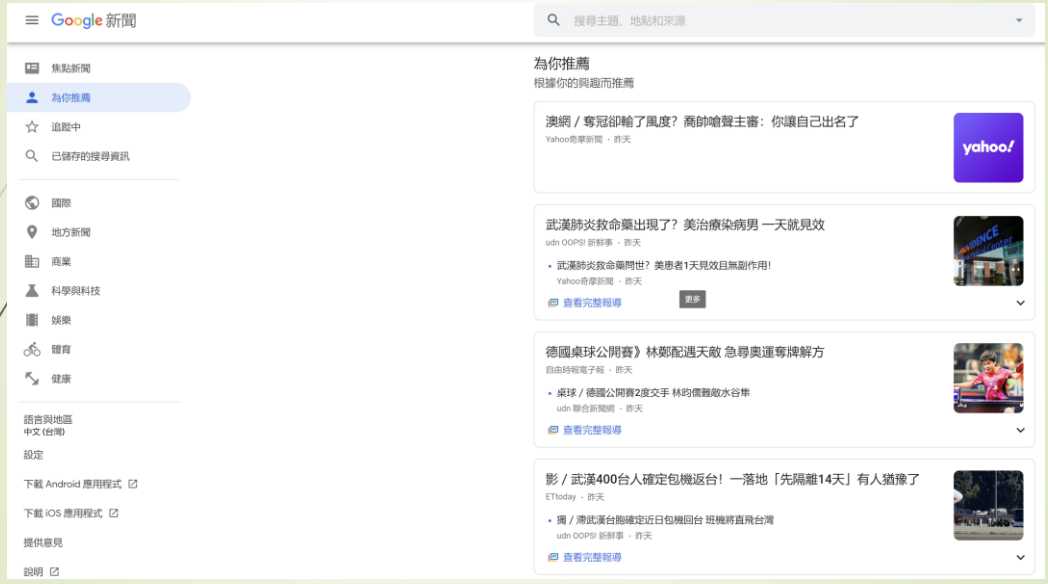




2/13/2022

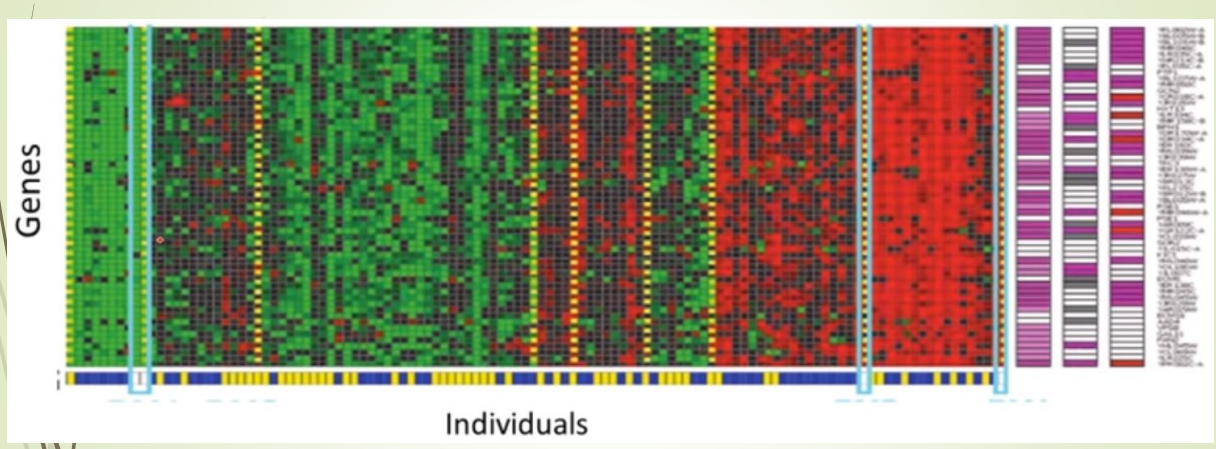
49

Google News (Feb. 3, 2020)



50

Clustering DNA Microarray Data



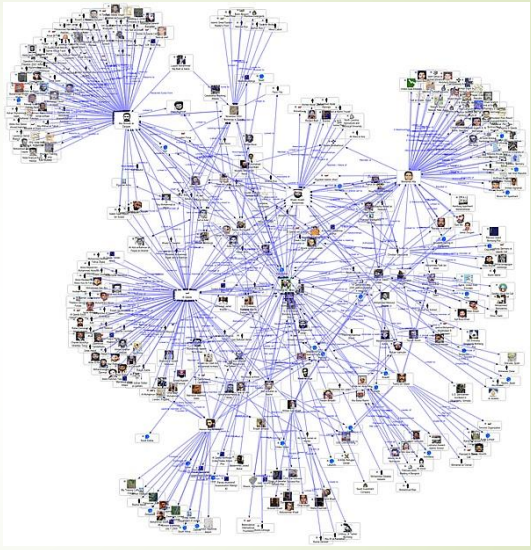
build groups of genes with related expression patterns (also known as coexpressed genes)

2/13/2022

51

Clustering

Market Segmentation



Social Network

2/13/2022

52

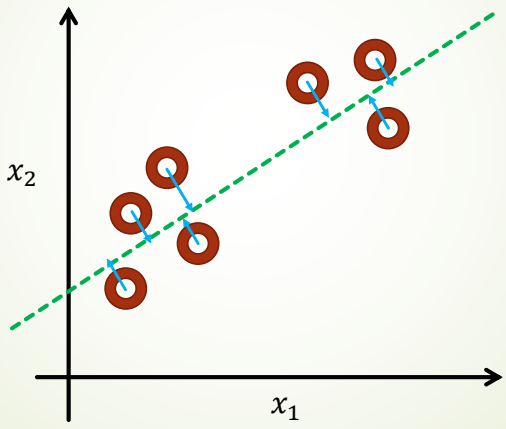
Machine Learning Problems

	Supervised Learning	**Unsupervised Learning**
Discrete	Classification	Clustering
Continuous	Regression	Dimensionality reduction

2/13/2022

53

Dimensionality Reduction



2/13/2022

54

3D face Modeling

2/13/2022

55

Shape Modeling