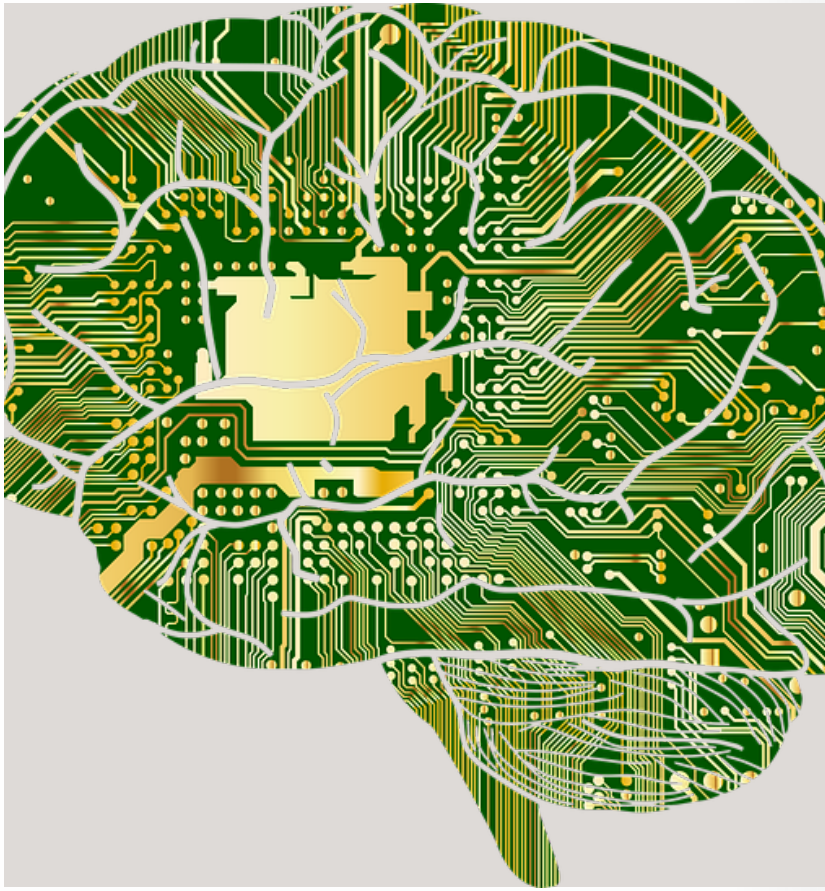
[RE50800]

INTRODUCTION TO DEEP LEARNING (DL)

Chih-Chung Hsu (許志仲)
Institute of Data Science
National Cheng Kung University
https://cchsu.info



Grading

- 平時成績: 70%
  - 程式作業
  - 報告 (in Word / PDF)
  - 論文報告投影片
- 期中考: 10%
- 期末專題: 20% (N人一組)
  - 期末投影片報告、論文 (請以正式submitted paper的格式)、GitHub
- 課堂加分: 10%
  - 互動、作業多做、回答問題等

TA group & Supporting Line Group

Support Line Group:
http://tinyurl.com/nfe88teh

廖涴婷（TA head）
woting1210@gmail.com

黃韋皓
robert20000831@gmail.com

李家銘
zuw408421476@gmail.com

王藝帆
re6113018@gs.ncku.edu.tw

Event Type	Description
Lecture 0	Course Introduction
Lecture 1	Computer vision overview Historical context Course logistics
Lecture 2	Recap: Image Classification The data-driven approach Python / numpy / Google Cloud
Assignment-1 (pre-test)	Image classification
Lecture 3	Loss Functions and Optimization Linear classification, Higher-level representations, image features Optimization, stochastic gradient descent
Lecture 4	Introduction to Neural Networks, Backpropagation Multi-layer Perceptron, Guidelines for Picking a Project
Lecture 5	Convolutional Neural Networks & RNN Convolution and pooling & RNN, LSTM, GRU Example: LeNet
Lecture 6	Deep Learning Hardware and Software Intro to Pytorch and Tensorflow under Ubuntu System
Assignment-2	Hand-crafted LeNet、GRU for image classification and ionosphere prediction using Pytorch
Lecture 7	Training Tricks Update rules, ensembles, data augmentation, transfer learning
Proposal due	Final Project Proposal due
Lecture 8	Traditional CNN Architectures: AlexNet, VGG, Inception SOTA CNN Architectures: GoogLeNet, ResNet, DenseNet, EfficientNet, ViT, Swin, etc

Event Type	Description
Assignment-3	Comparison of SOTA CNN and ViT for image recognition
Lecture 9 **Midterm**	In-class midterm
Discussion Section	Midterm Discussion
Lecture 10	Object Detection & Semantic Segmentation Det: M2Det, OCR, EfficientDet, YOLOv7, EVA, InternImage Seg: FCN to DeepLabv3-plus, HANet, HRNet, SAM,…
Assignment-4	Low-Power multitasking learning: joint segmentation & object detection
Milestone	Project Milestone due
Lecture 11	Unsupervised Learning: Generative Models GAN, WGAN, CycleGAN, Diffusion models, ChatGPT
Lecture 12	Security in DL: Adversarial attack, Deepfake detection, anti-spoofing
Assignment-5	Forgery detection
Lecture 13	Restoration in DL: Toward image and high-dimensional data super-resolution
Lecture 14	LLM/Foundation Models and its applications
Guest Lecture	TBA
Final Project Due	Project Report due,

期末專題說明

- 大約 Tensorflow / Pytorch 之後課程開始準備
  - 以你閱讀論文的角度，去剖析方法的特點，以及定義問題
  - Tips (要仔細看這個)
    - 題目、作者、其實驗室簡介、被引用次數
    - 論文導論簡介
    - 文獻探討: 相似的過往研究，其缺點優點
    - 論文方法介紹: 必須介紹其方法的細節，越詳細越好，含數學
    - 實驗結果：介紹其實驗並分析，不能只是帶過
    - 結論：總結論文的優缺點 (以你自己的角度)
  - 實作與原理剖析
- 找出目前大家沒有考慮的議題、或是方法上有缺陷你可以改進，忌諱只有套用現有的方法論

DEEP LEARNING? DO WE NEED IT?

請描述

請描述一

1. 專業知
   域或一
2. 獨立研
   教授，
3. 提高創
   在日後
4. 能力全
   動，如
5. 增加就
   業市場

學術領

勺指導

重能力

究活

在就

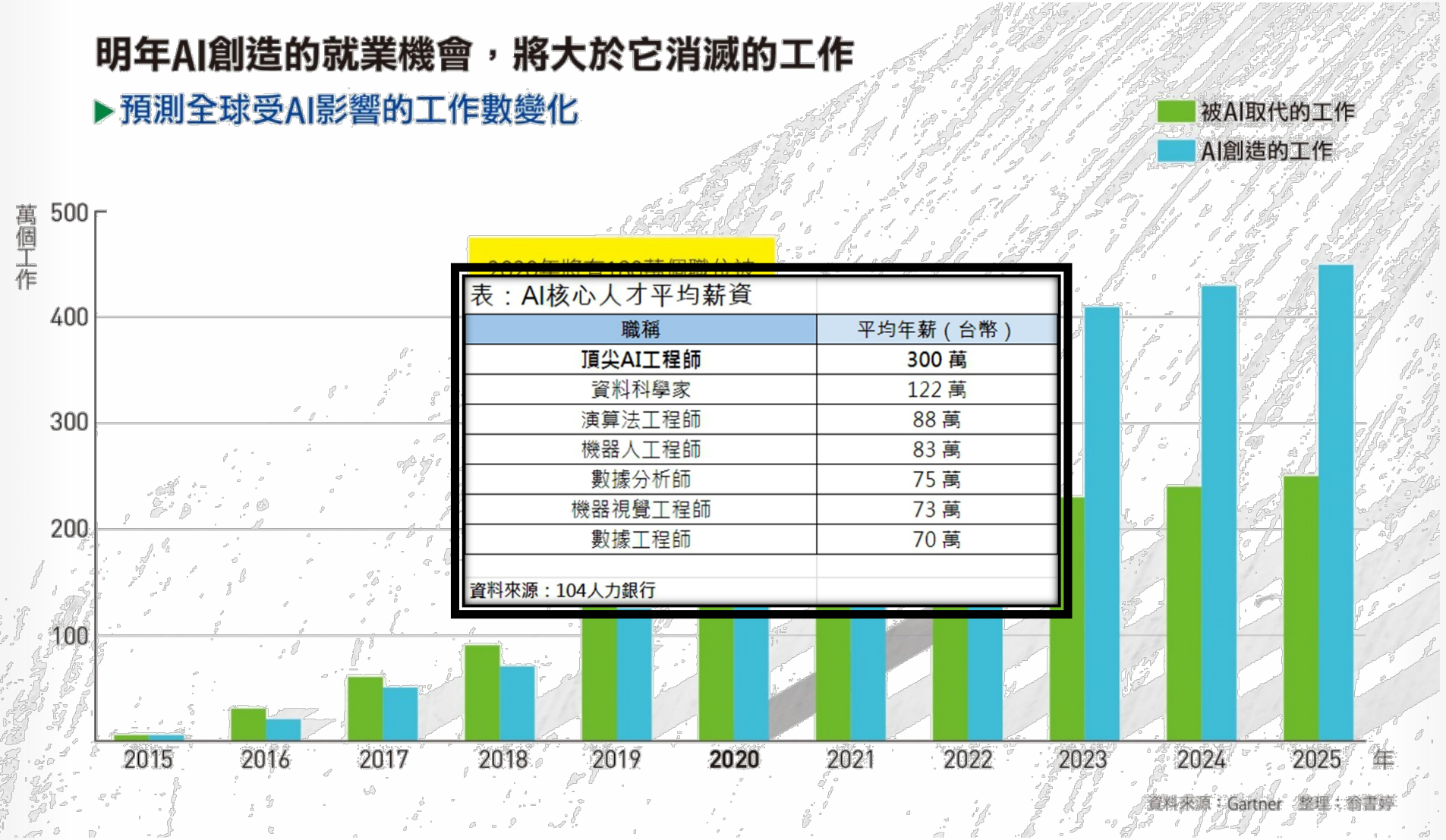
活

Capmus view of National Cheng Kung Universiry in Taiwan

Enter a negative prompt

Generate image

Source: https://www.bnext.com.tw/article/51971/ai-talent-cultivate-playbook
Source: 遠見雜誌



表：AI核心人才平均薪資

職稱	平均年薪（台幣）
頂尖AI工程師	**300 萬**
資料科學家	122 萬
演算法工程師	88 萬
機器人工程師	83 萬
數據分析師	75 萬
機器視覺工程師	73 萬
數據工程師	70 萬

資料來源：104人力銀行

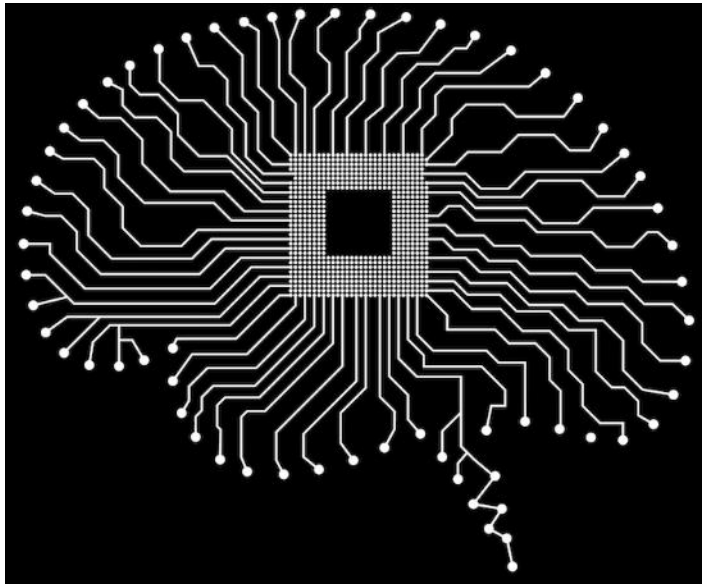
OK，學深度學習看起來很賺錢

- 沒有錯，確實很賺錢，前提是你/妳學得”精通”
  - 對學界來說，做做研究、產學合作什麼的，可能套套件就OK了
  - 對於業界來說，更重要的是你能不能implement it from scratch.
    - Oh, 更別說還要考慮Python 以外的程式語言了

- 重點：請別以為會深度學習就能賺大錢或成大事
  - 你會大家也會
  - 找出你自己的賣點
    - 學術頂尖：CVPR/ICCV/ECCV/ICLR/ACMMM/ICASSP/ICIP/ICME等頂尖研討會
      - 期刊也可以但通常碩班沒辦法那麼快發期刊
    - 實作頂尖：你附上你用 C/C++/C#/Java/JS implement的 LeNet 展現硬實力
    - 累積作品：沒有上述但很多side project
      - 可能在GitHub, 也可能累積多篇國際論文、或是競賽 (只看國際的)、或是實際產學合作的 “產品” 成果等

Deep Learning in AI

- A good one, or the best one
  - 成熟的硬體環境 (Hardware)
  - 快速平行化的演算法 (Software)
  - 巨量資料時代 (Data)
  - 幾個天才(Genius)
- Now we may have a AI
  - Based on
    - Neural Network
      - Much deeper
  - GPU-based optimization
  - Big data to learn



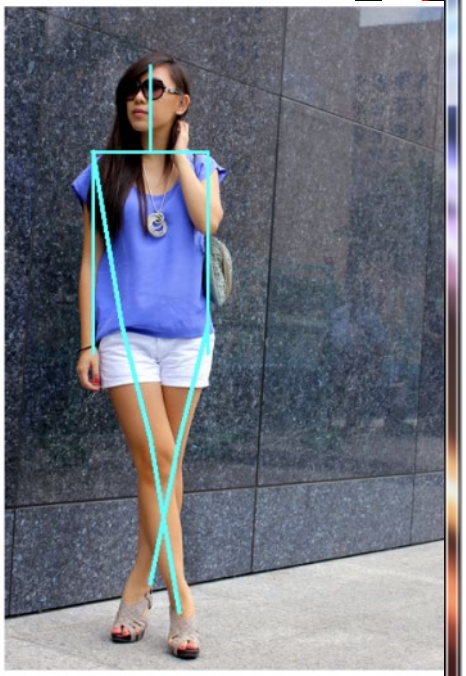
AI 什麼時候會成功?

- 你有足夠多的資料
  - 如果以分類任務 (分人車貓狗)，每個類別至少要有上百張照片供學習
  - 資料的數量往往決定效能
  - 標註品質

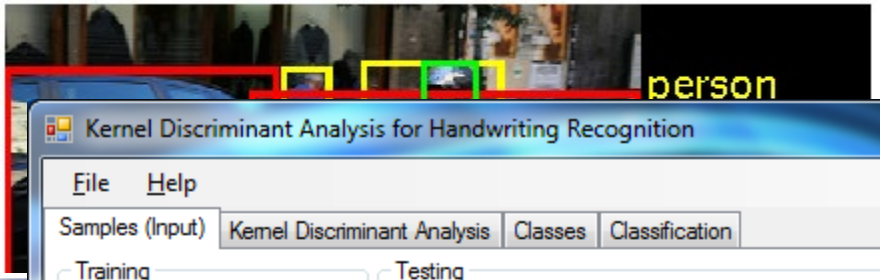
  - 重要的一點
    - 目前AI的成功，
      是基於上一波熱潮 【大數據】

- 少量樣本學習技術?
  - 還在實驗室研發階段，
    尚未有知名商用技術

AI 應用於電腦視覺之範例



(b) Pose estimation



憑什麼要選深度學習 (Deep Learning)?

Task: video activity recognition

Method	Accuracy
Hessian + ESURF [Williems et al 2008]	38%
Harris3D + HOG/HOF [Laptev et al 2003, 2004]	45%
Cuboids + HOG/HOF [Dollar et al 2005, Laptev 2004]	46%
Hessian + HOG/HOF [Laptev 2004, Williems et al 2008]	46%
Dense + HOG / HOF [Laptev 2004]	47%
Cuboids + HOG3D [Klaser 2008, Dollar et al 2005]	46%
Unsupervised feature learning (Deep learning)	**52%** ←

[Le, Zhou & Ng, 2011]

超大型影像資料庫分類

14M images, 22k categories

0.005%	9.5%	?
Random guess	State-of-the-art (Weston, Bengio ‘11)	Feature learning From raw pixels

Le, et al., *Building high-level features using large-scale unsupervised learning*. ICML 2012

超大型影像資料庫分類

14M images, 22k categories

0.005%	9.5%	21.3%
Random guess	State-of-the-art (Weston, Bengio ‘11)	Feature learning From raw pixels

Le, et al., *Building high-level features using large-scale unsupervised learning*. ICML 2012

那麼，人工智慧已經到來?

- 注意，人工智慧 ≠ 深度學習
  - 由人製造出來的機器所表現出來的智慧
  - 理解環境、推理原因
- AI的幾個階段
  - 機器學習：使用一系列演算法從經驗中進行學習。(傳統方法)
  - 機器智能：使用一系列演算法從資料中自動學習出何種經驗有效，例如深度神經網絡。 (現在的方法)
  - 機器認知：不需要外部數據就能從經驗中自學習。
    - 普遍認為 Reinforcement learning 接下來有機會做到
  - 人工智慧：自動發掘某特定任務的解法、收集資料等
    - 通過圖靈測試

Deep Learning怎麼做到這些的?

- Actually, deep learning 還是 Machine learning的一種
  - 在Deep learning 出現以前…
  - Computer vision 中最強大的方法是…
    - Multiple features fusion (ILSVRC 2011 as an example)
- Yes, 特點就在 “多種特徵” 上
  - Deep learning
    - Learning a lot of features from data
    - Called data-driven approach
- 要了解原理，就要先懂機器學習 (Machine learning)基本概念

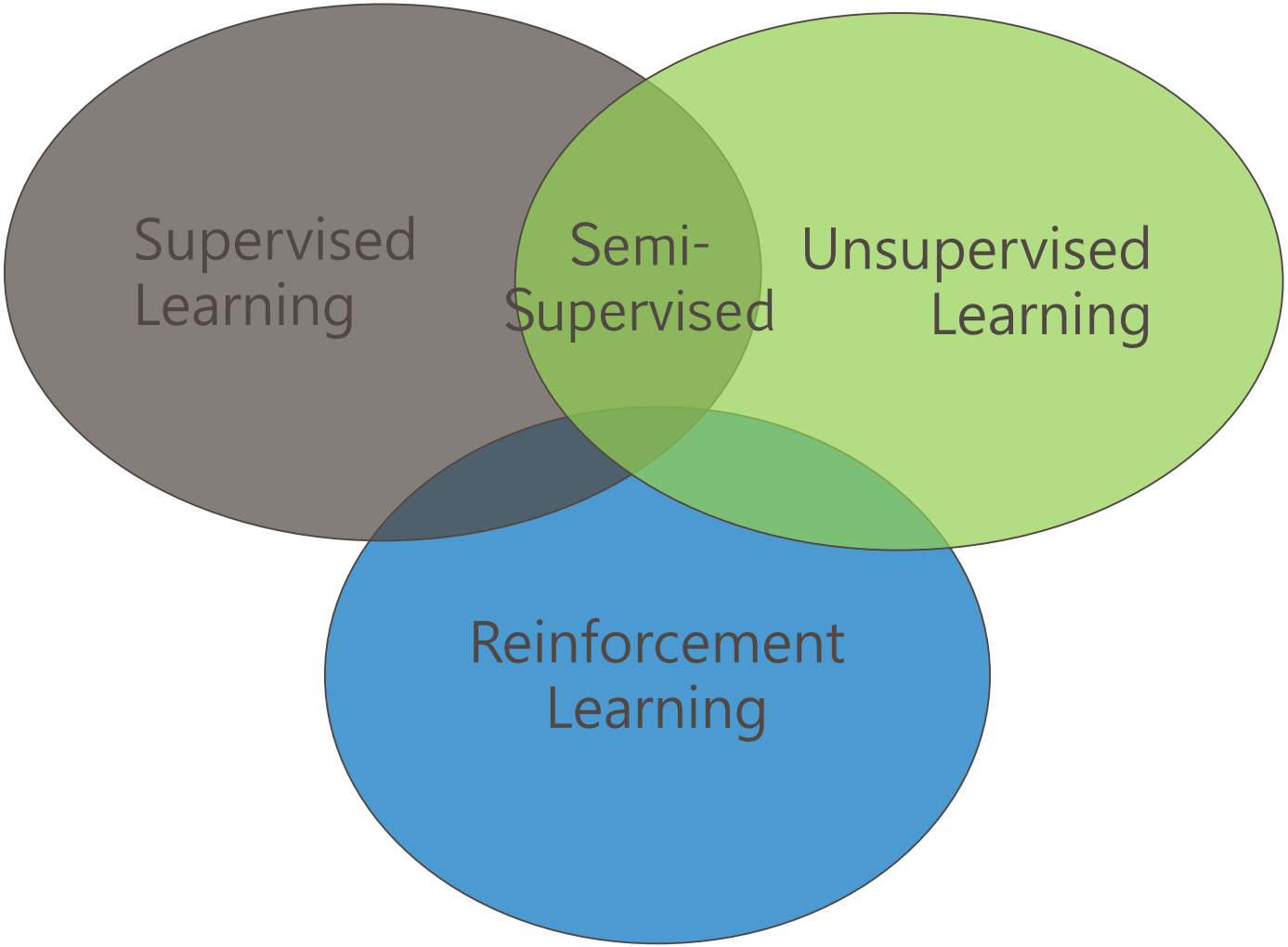
什麼是特徵表達

- 樂高積木組好了：資料

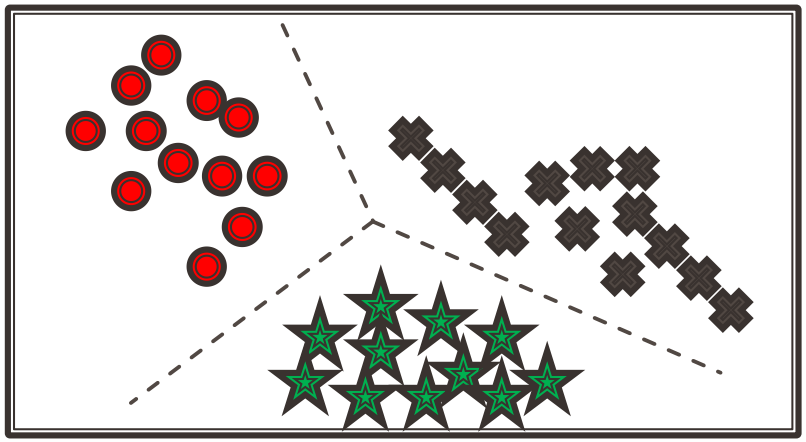
- 什麼是特徵表達? 每一塊積木就是特徵
  - 特徵有大有小，有不同形狀、顏色亮度
  - 不同特徵組合，便可以找出一個組合來“表達”資料

MACHINE LEARNING概念

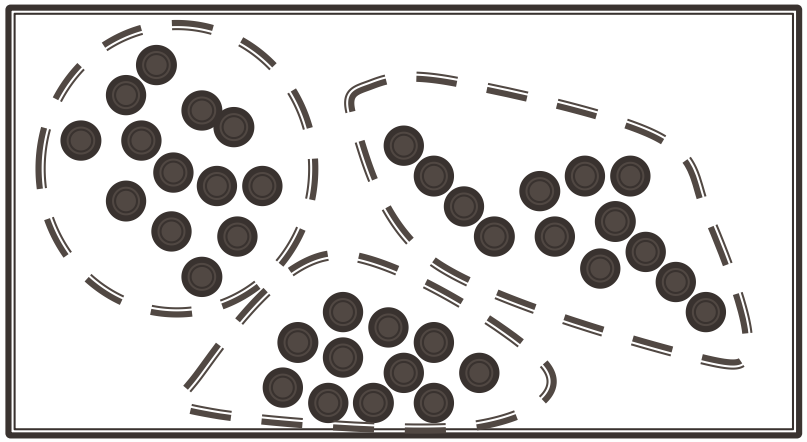
機器學習概觀 Taxonomy of ML



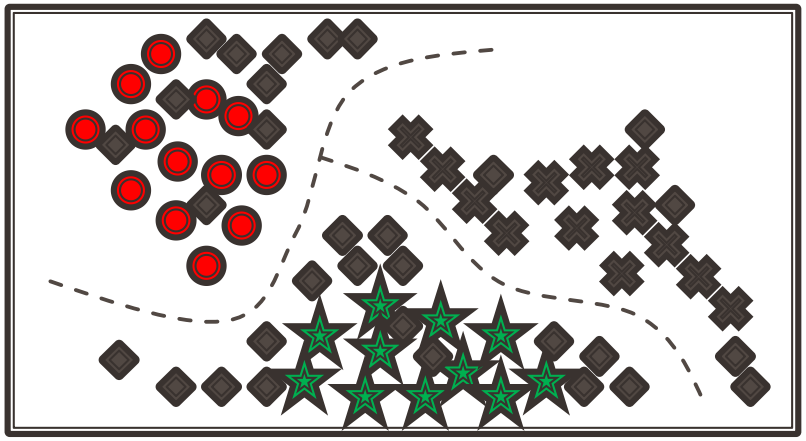
Example to Machine Learning



Supervised learning

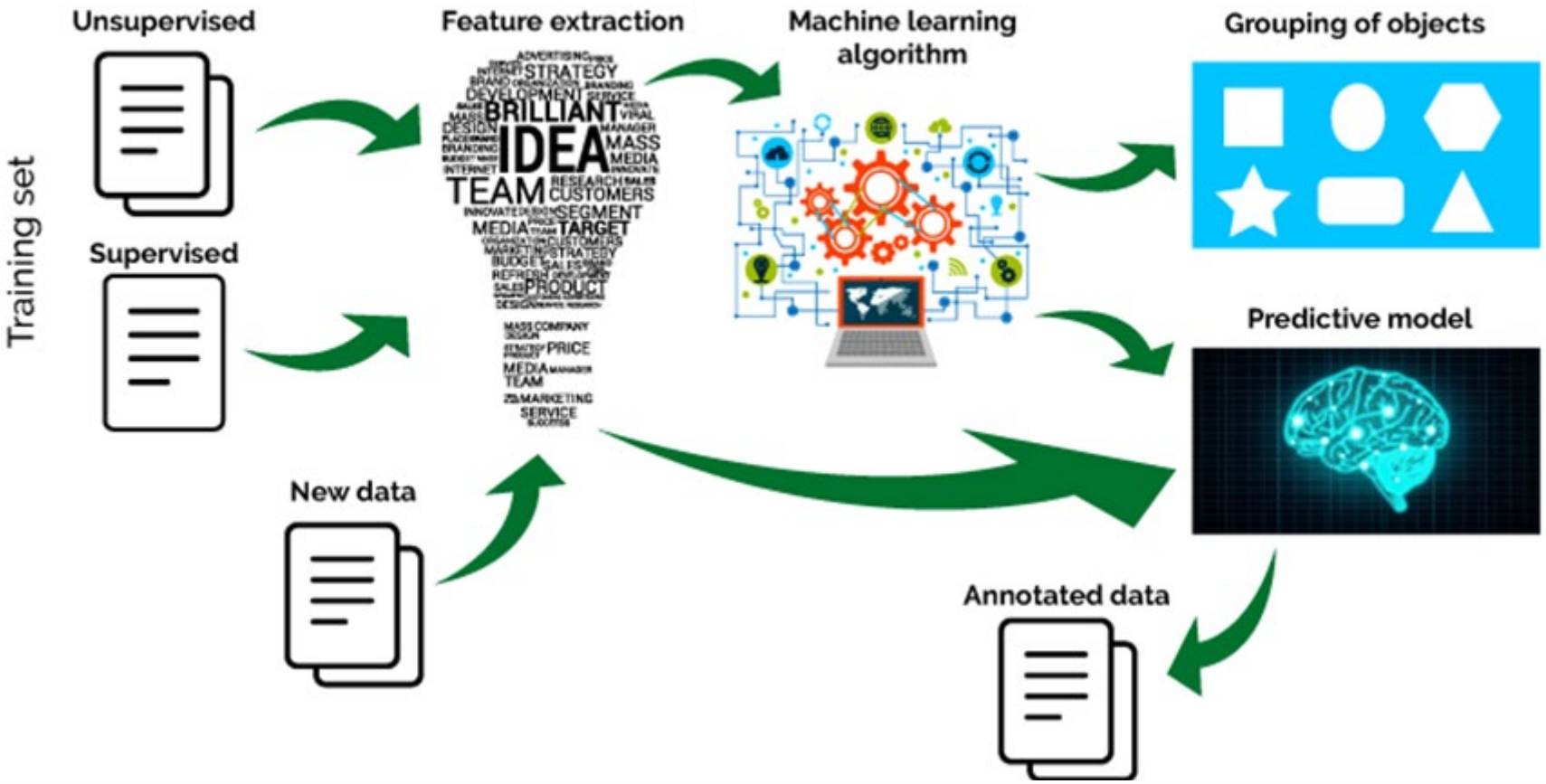


Unsupervised learning

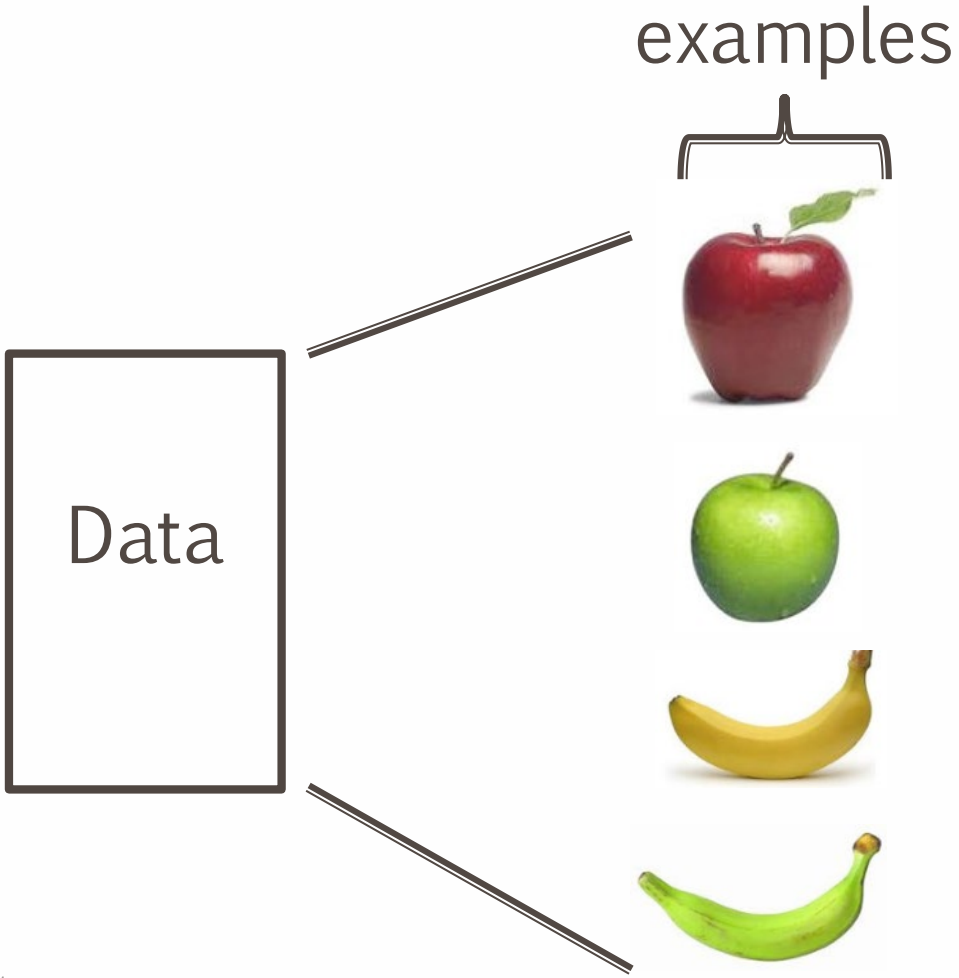


Semi-supervised learning

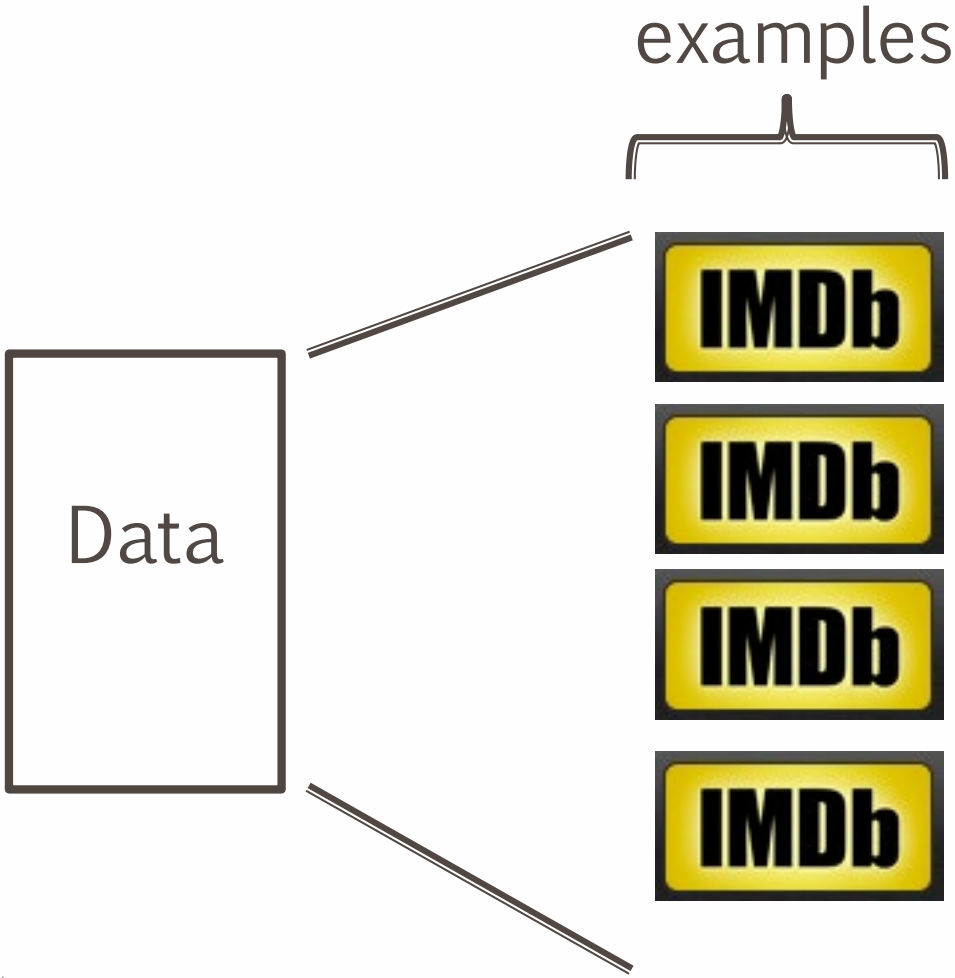
Basic Concept of Machine Learning



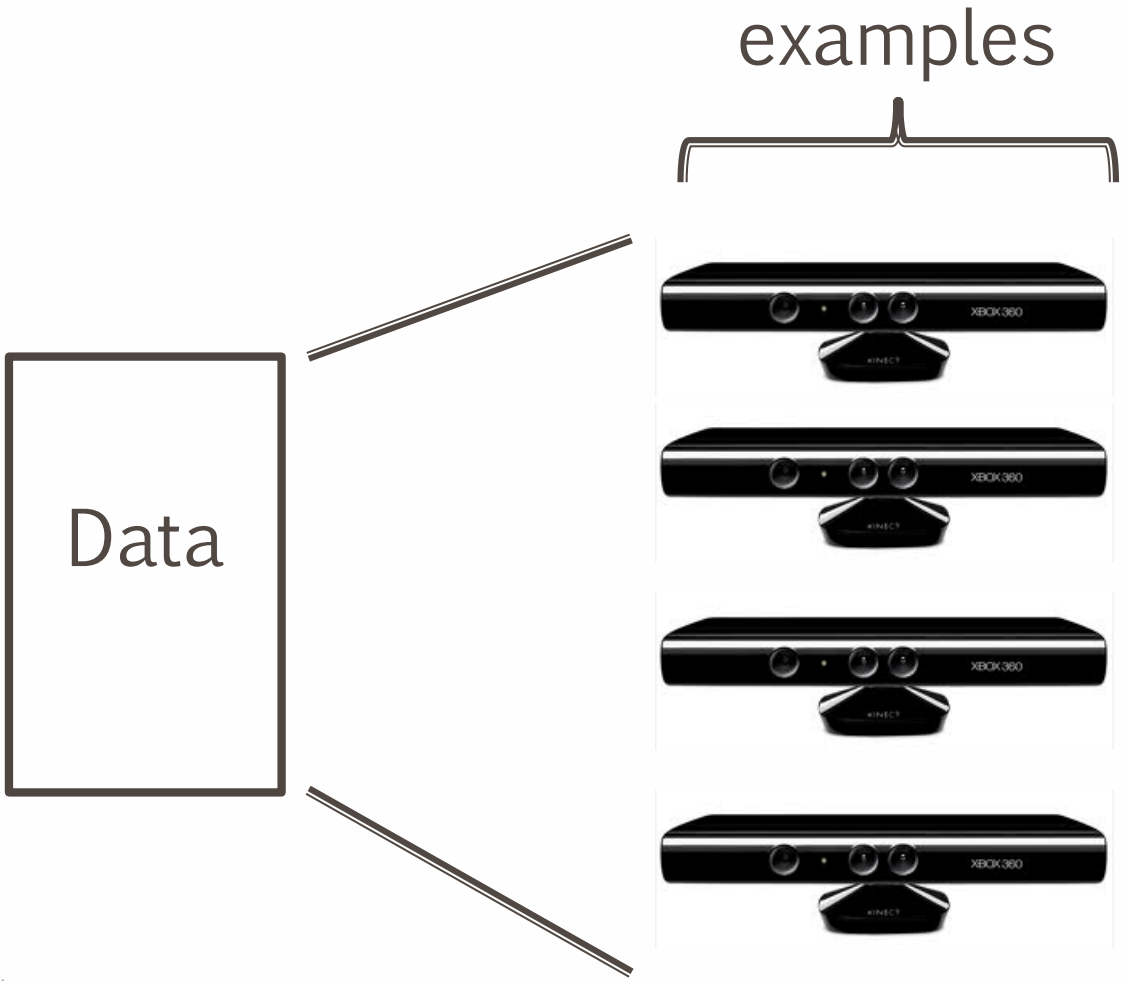
Training Set ⇔ Data



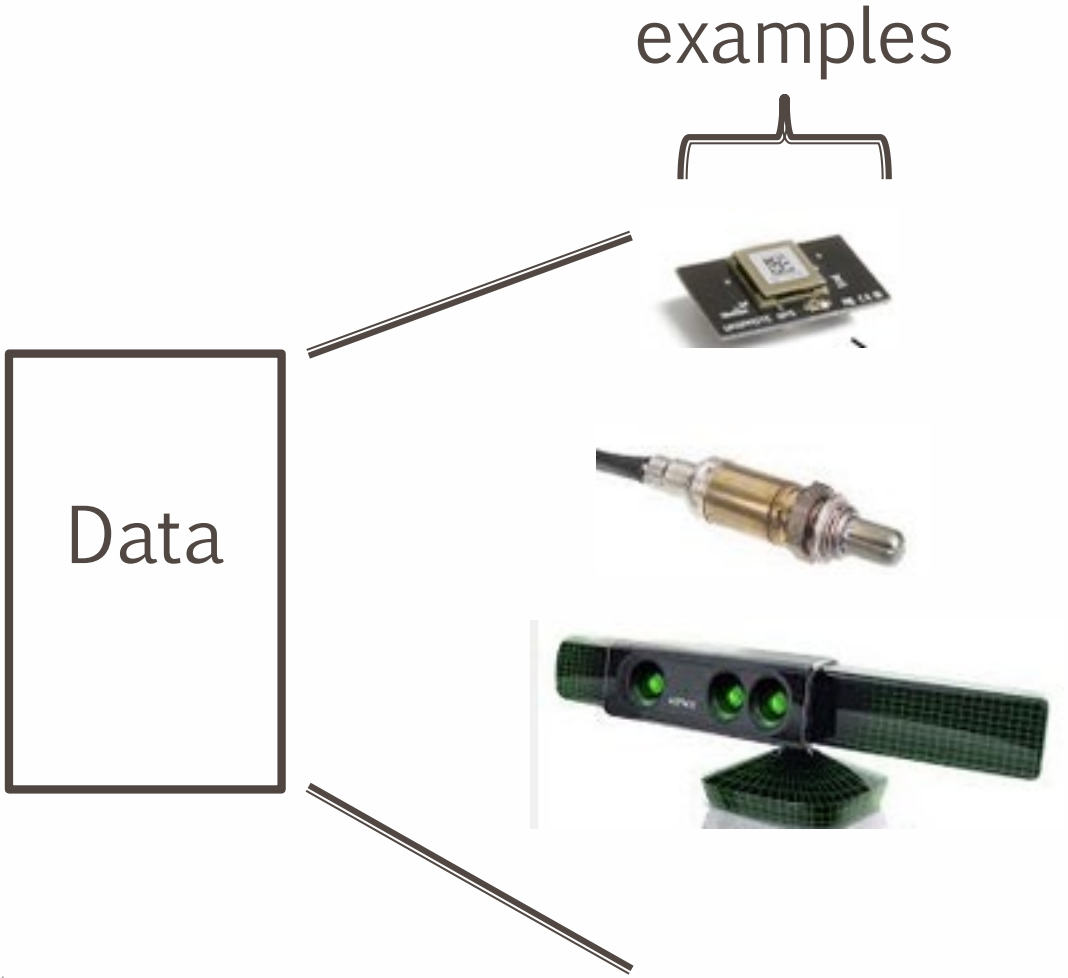
Data



Data



Data



Supervised learning

examples label

Apple: 1

Greed apple: 1

Banana: 2

G-Banana: 2

labeled examples

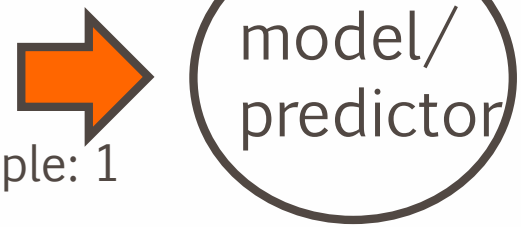
Supervised learning: given labeled examples

Supervised learning

label

Apple: 1

Greed apple: 1

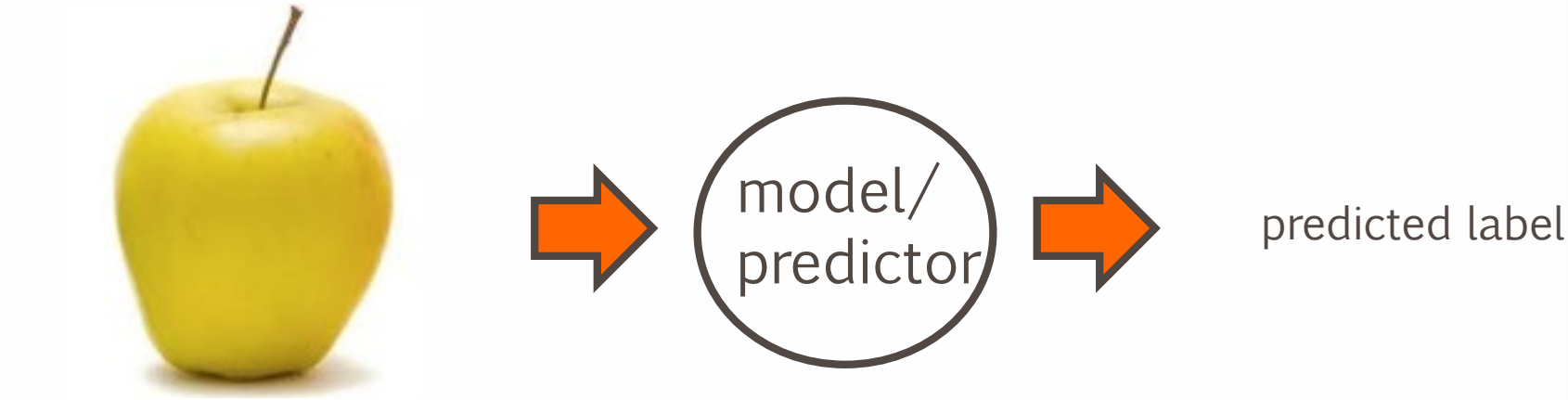


Banana: 2

G-Banana: 2

Supervised learning: given labeled examples

Supervised learning



Supervised learning: given labeled examples

Supervised learning: classification

label

Apple: 1

Greed apple: 1

Banana: 2

G-Banana: 2

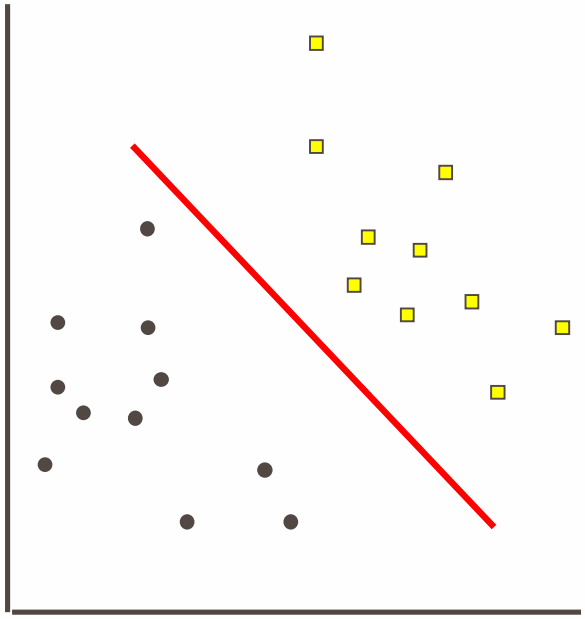
Classification: a finite set of labels

Supervised learning: given labeled examples

常見的Supervised Learning方法 (分類器)

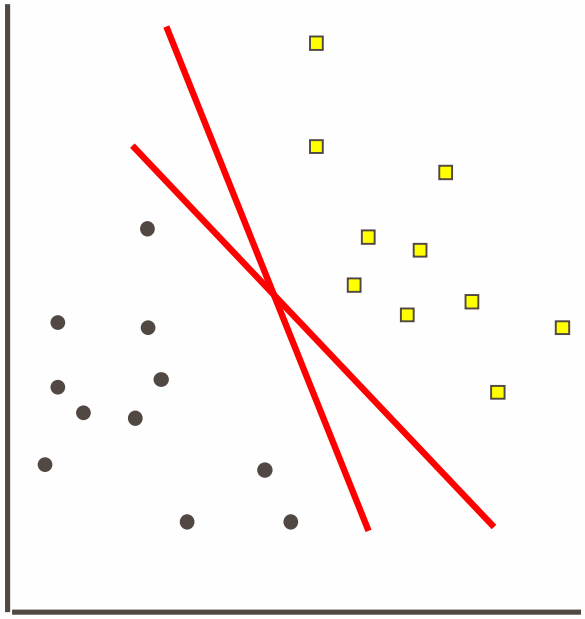
- Decision trees
- Neural networks
- K-nearest neighbors
- Naïve Bayes classifier
- Support vector machines (SVMs)
- Boosted decision stumps
- …

Perceptron (neural net with no hidden layers)

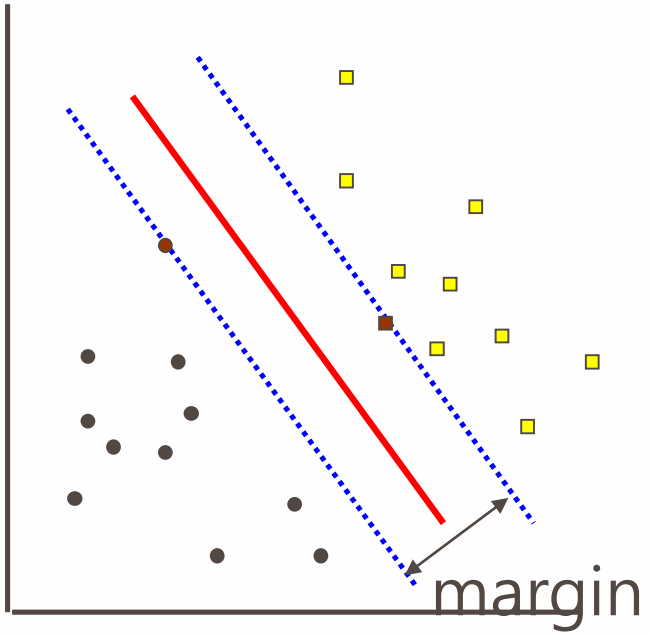


Linearly separable data

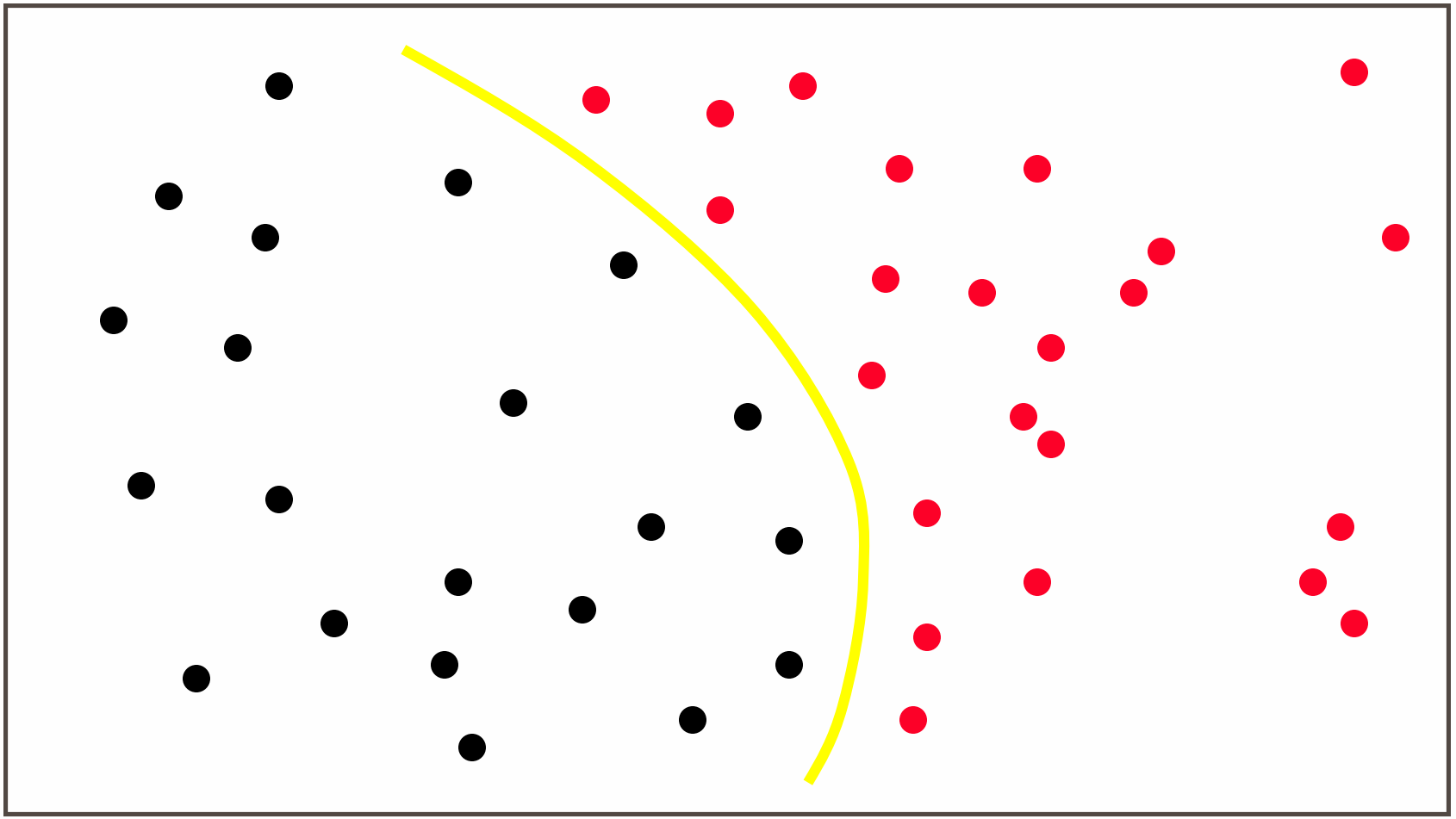
分類線的選取



Maximum Marginal (SVM概念)



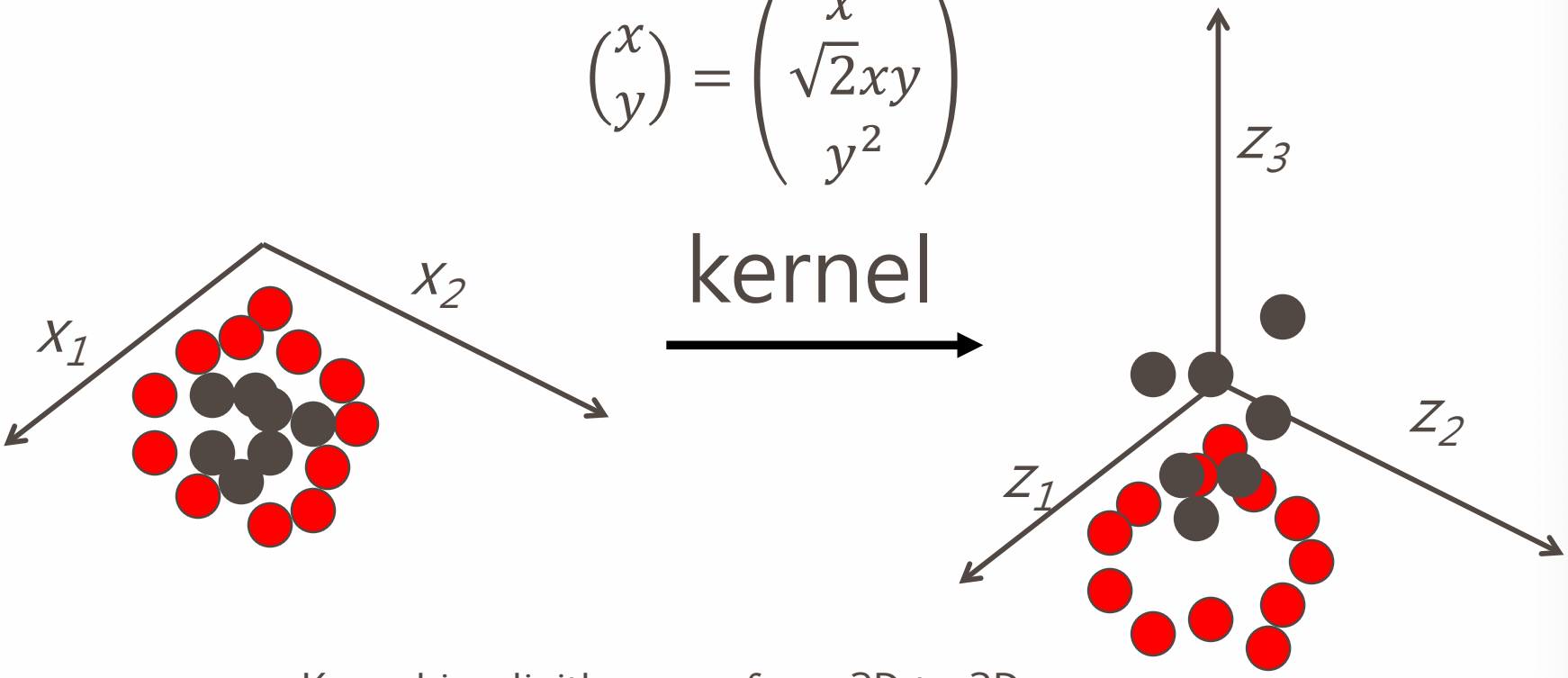
Data Points 若是非線性?



常見的招式：Kernel Mapping

$$K(\vec{x_i}, \vec{x_j}) = (\vec{x_i} \cdot \vec{x_j})^2 = \Phi(\vec{x_i}) \cdot \phi(\vec{x_j})$$

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x^2 \\ \sqrt{2}xy \\ y^2 \end{pmatrix}$$

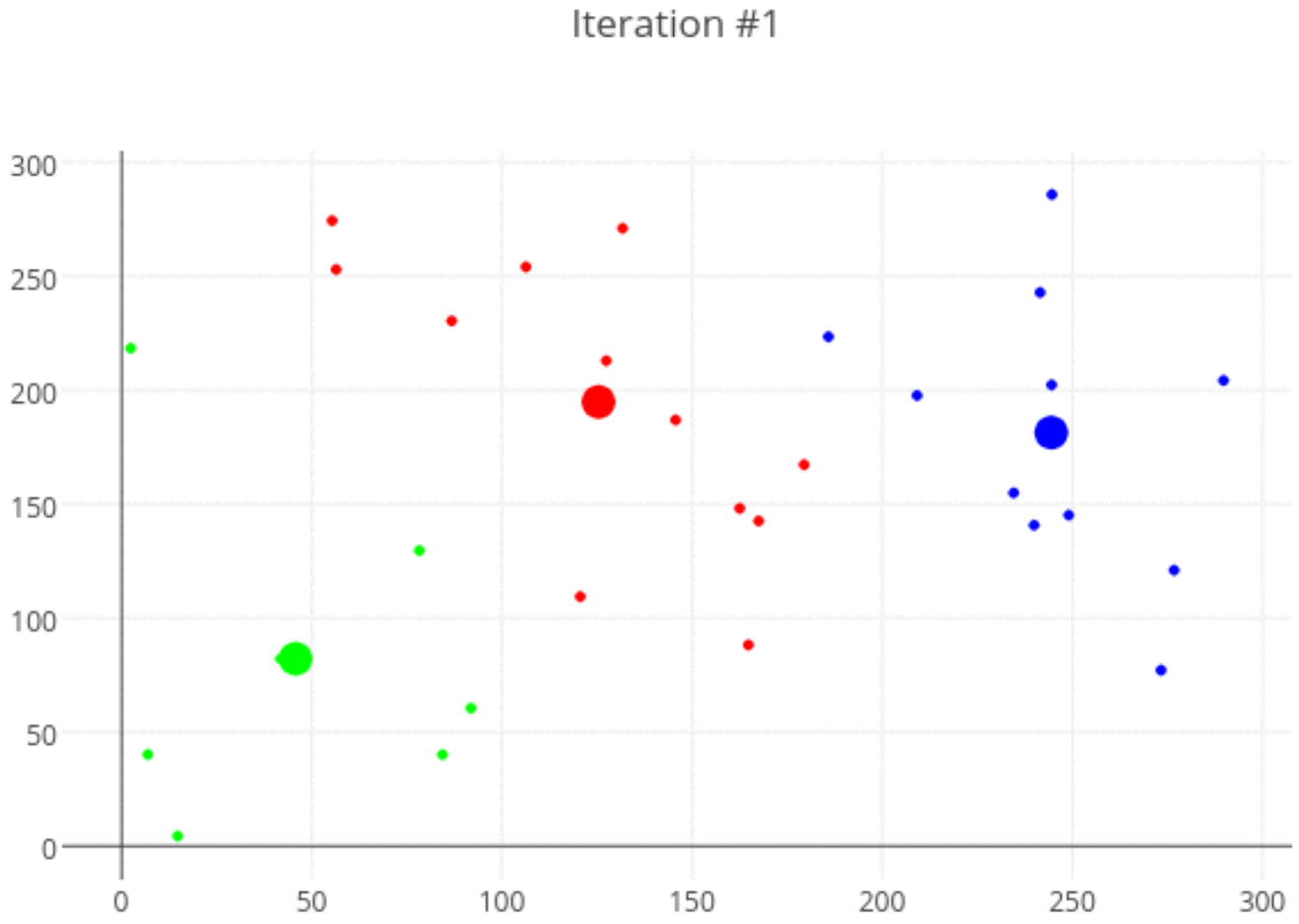


Kernel implicitly maps from 2D to 3D,
making problem linearly separable

Unsupervised Learning

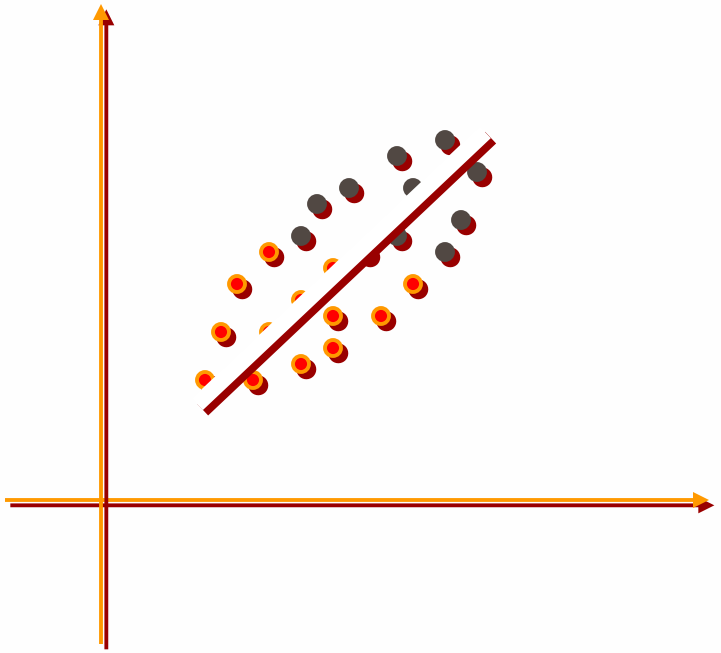
- 代表資料沒有T (答案) 可以預測
  - 從資料本身來分析是否有群聚效應
  - 因此，通常Unsupervised learning的演算法，會叫做分群 (Clustering)
    - 反之，Supervised learning演算法會叫做分類 (Classification)
- Raw data domain可能有太多雜訊，不好分
  - 擷取資料的“特徵”，再進行分群
  - 非監督式學習特徵：
    - 從資料本身
    - PCA (Principle component analysis)，AutoEncoder等

Traditional k-means Clustering



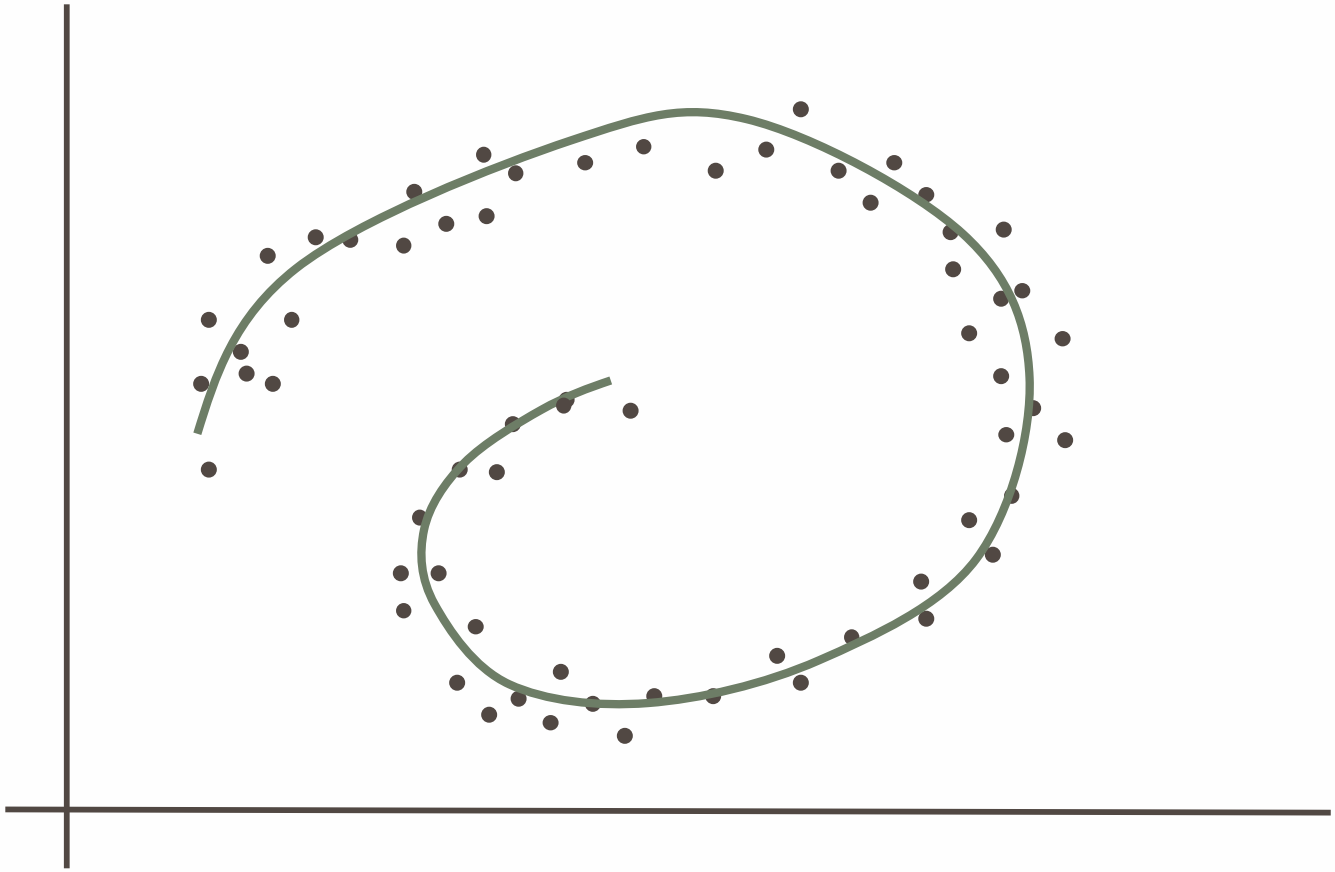
Principal Component Analysis (PCA)

- PCA seeks a projection that best represents the data in a least-squares sense
  - 簡單說，找保留最大能量的方向！



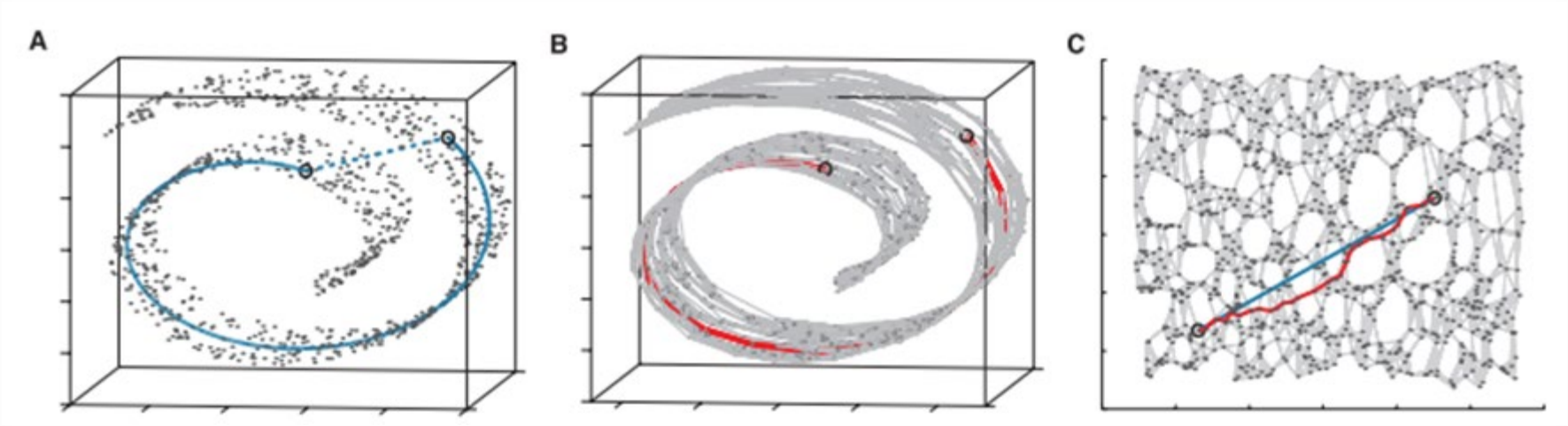
PCA reduces the dimensionality of feature space by restricting attention to those directions along which the scatter of the cloud is greatest.

當然這也會有其他問題：非線性資料



非線性降維法：ISOMAP

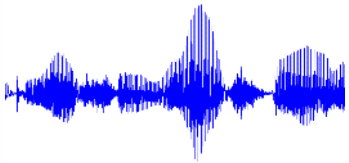
- Piece-wise distance meansurements



機器學習的一些總結

- We have learned
  - Supervised learning vs. Unsupervised learning
    - 有標準答案 & 沒有標準答案的任務
  - Supervised learning
    - Classification
      - SVM, Linear nearest neighbor, Perceptron
  - Unsupervised learning
    - Clustering
      - k-means, EM-algorithm, AutoEncoder, PCA
- Deep learning:
  - 成熟的：Supervised learning (i.e., AlexNet)
  - 熱門的：Unsupervised learning (i.e., GAN)

分類問題的基本模型

input		output
images/video		Label: “猩猩大笑” Suggest tags Image search …
audio		Speech recognition Music classification Speaker identification …
text		Web search Anti-spam Machine translation …

分類問題的基本模

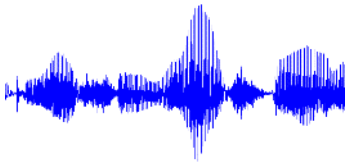
取特徵通常是最花時間、也是最難的一部份

input

output

images/video

audio



text

ML → Label: “猩猩大笑”
Suggest tags
Image search
…

ML → Speech recognition
Music classification
Speaker identification
…

ML → Web search
Anti-spam
Machine translation
…

分類問題的基本模型

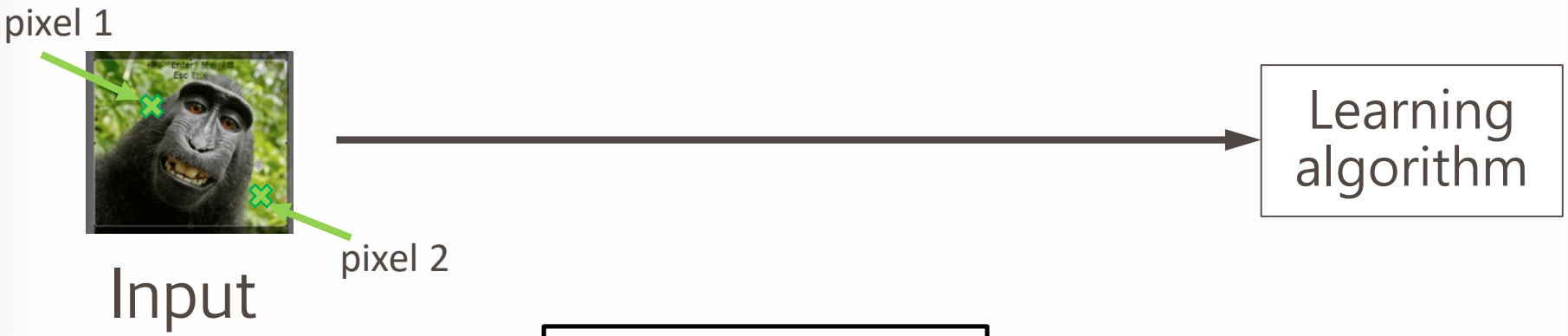
但為什麼取特徵很難?

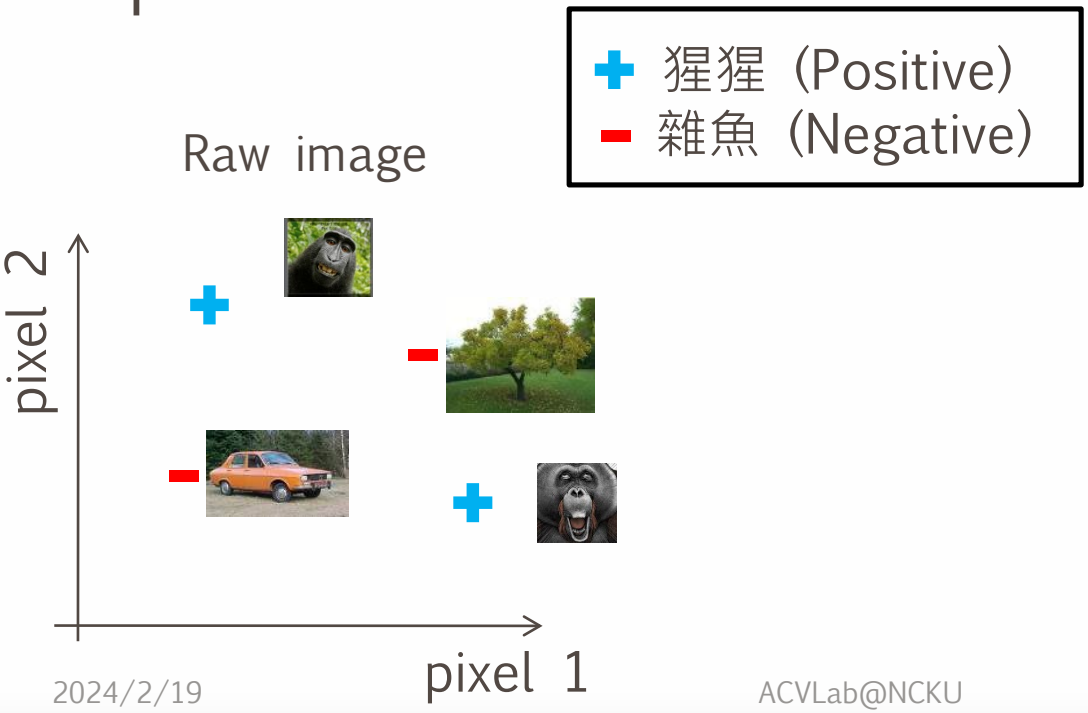
You see this:

But the camera sees this:

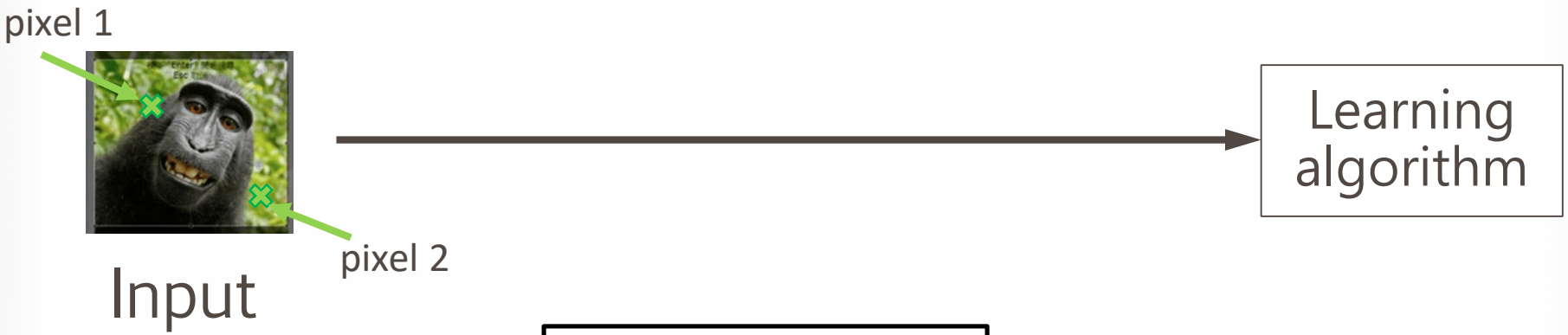
Index	0	1	2	3	4	5	6	7	8	9
0	1	2	3	4	5	6	7	8	9	10
1	11	12	13	14	15	16	17	18	19	20
2	21	22	23	24	25	26	27	28	29	30
3	31	32	33	34	35	36	37	38	39	40
4	41	42	43	44	45	46	47	48	49	50
5	51	52	53	54	55	56	57	58	59	60
6	61	62	63	64	65	66	67	68	69	70
7	71	72	73	74	75	76	77	78	79	80
8	81	82	83	84	85	86	87	88	89	90
9	91	92	93	94	95	96	97	98	99	100

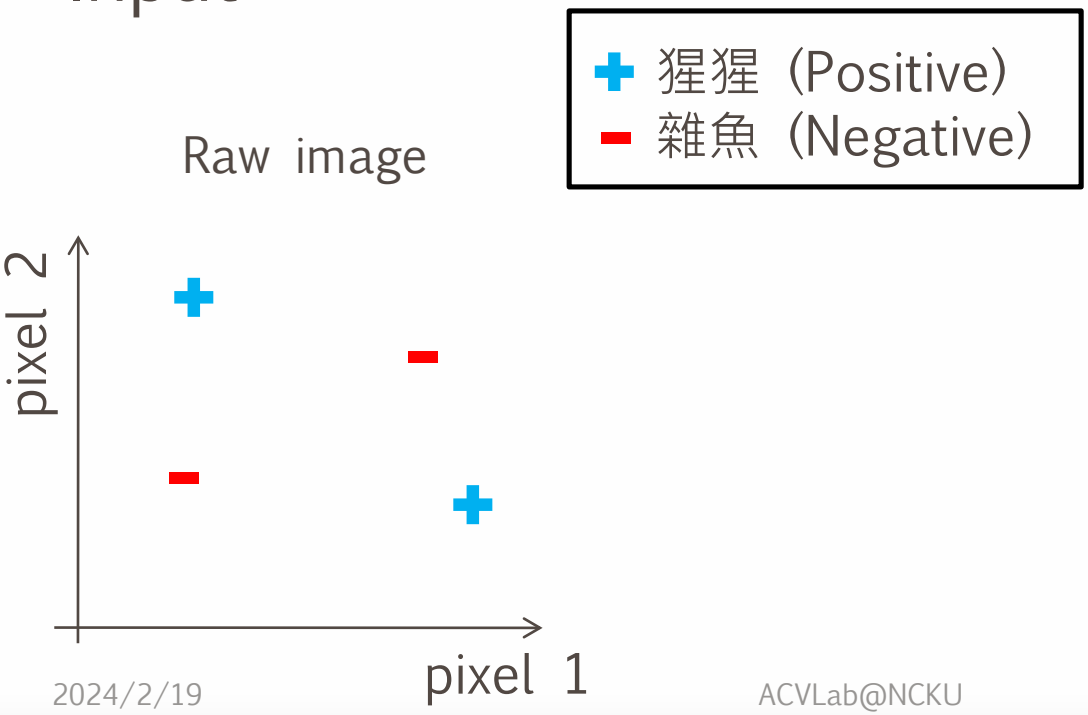
特徵取法I: 每個像素都取



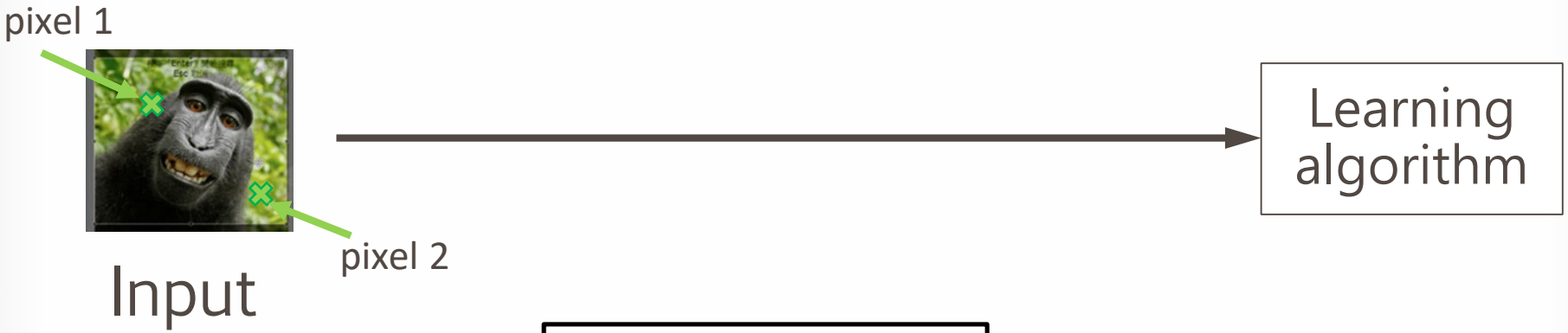


特徵取法I: 每個像素都取

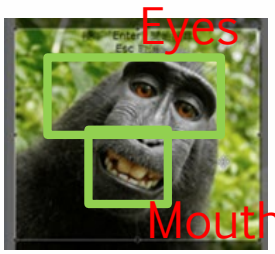




特徵取法I: 每個像素都取



利用較好的特徵來表達



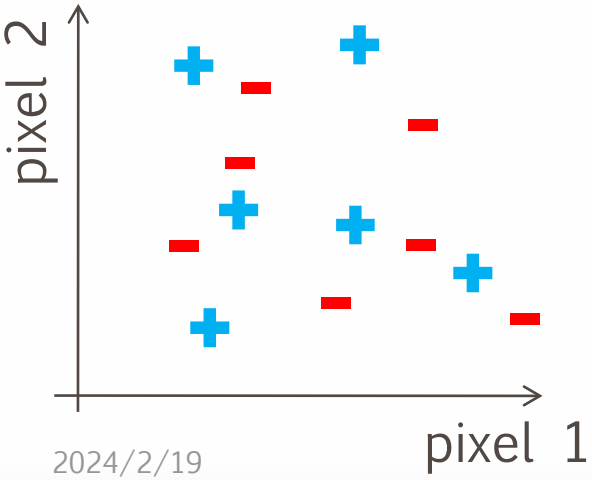
Input

Feature representation

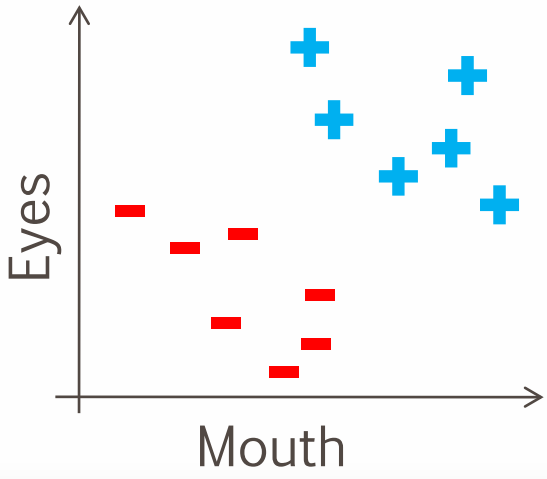
Learning algorithm

猩猩 (Positive)
雜魚(Negative)

Raw image

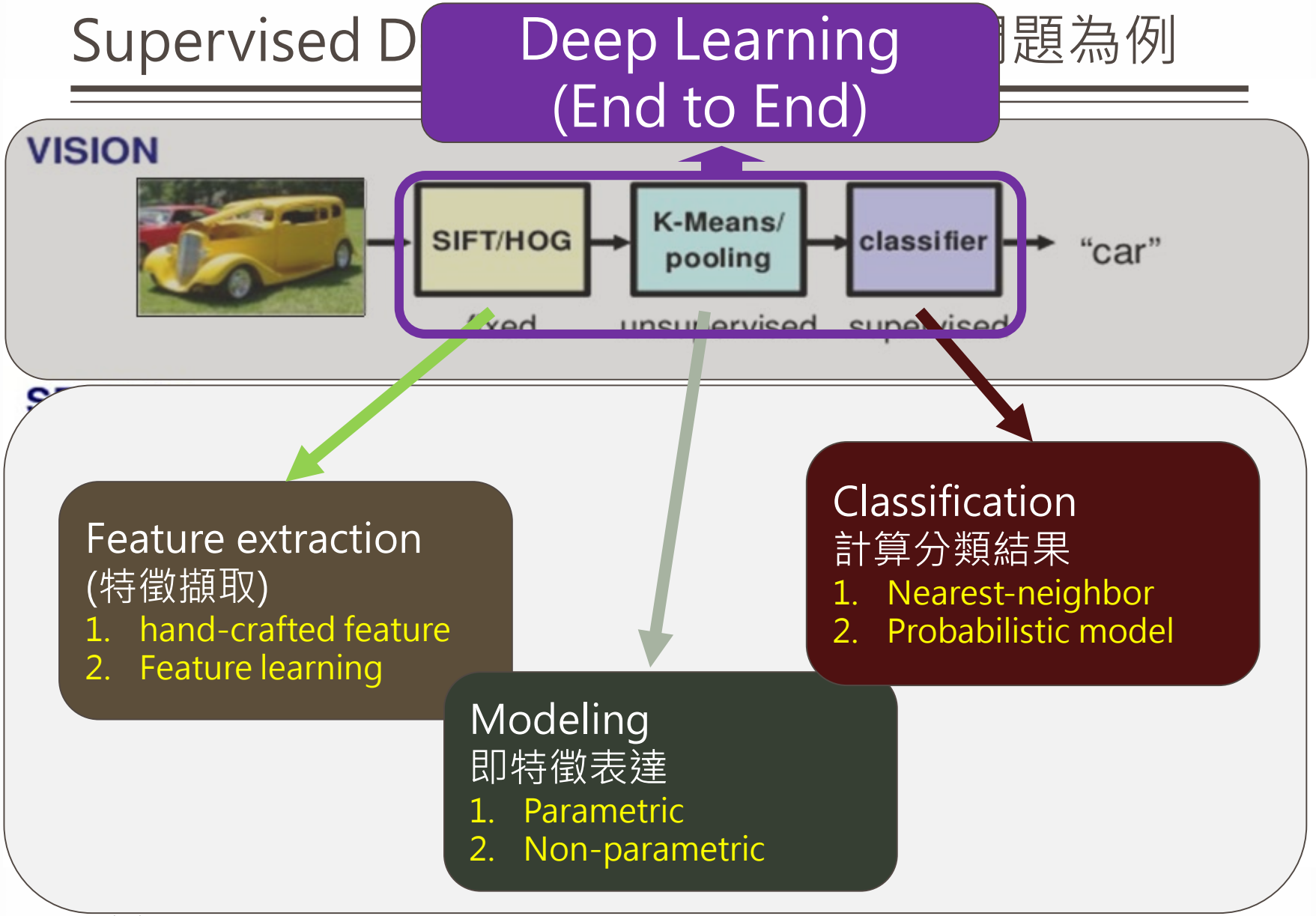


Features



DEEP LEARNING? 深度學習?

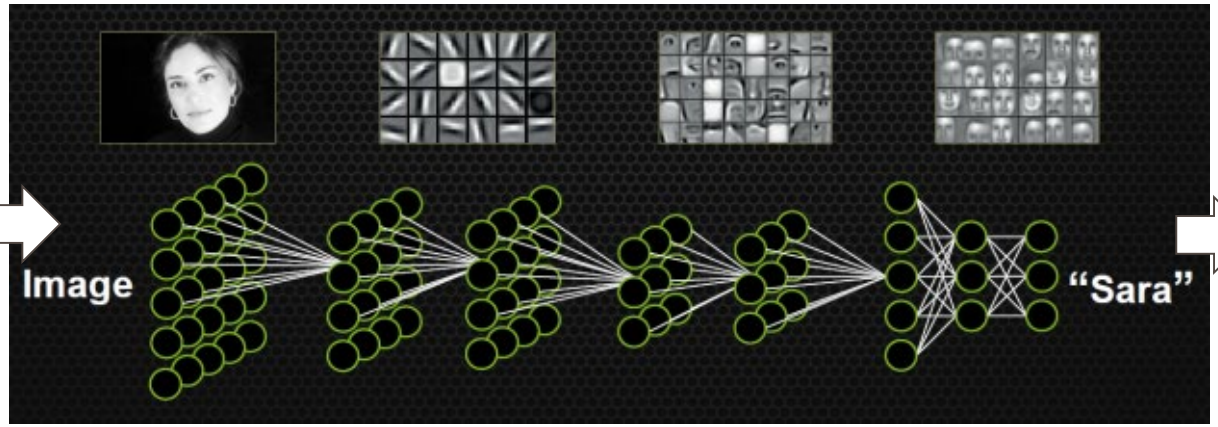
Supervised D 題為例



範例：深度學習網路

- Simple example

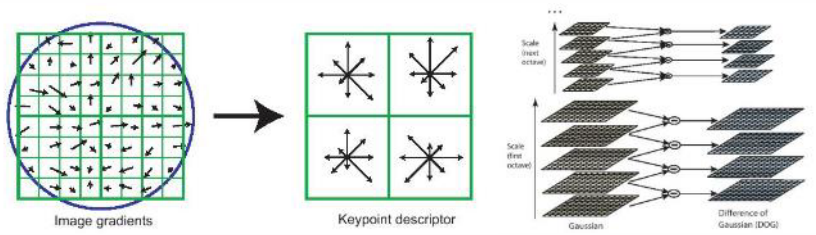
餵資料
(Data)



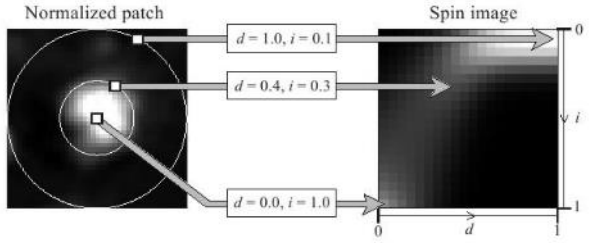
吐答案
(Label)

- 中間隱藏層 (Hidden layer) 可能有/沒有意義
  - 理解為用不同的 Function 對資料的 特徵表達 (Feature representation)

過去數年間所發展的各種特徵



SIFT



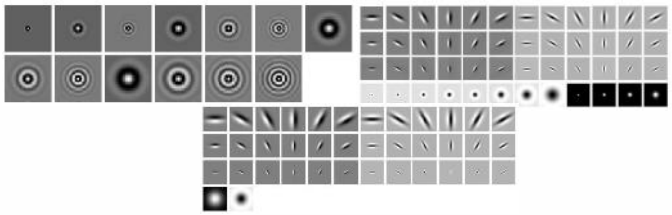
Spin image

找出適合資料的特徵表達，不僅僅耗時費力，更需要該領域的專家才能分析出適合的特徵…

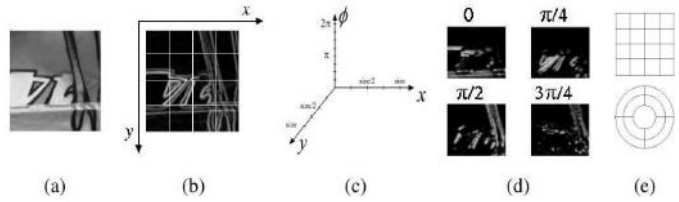
深度學習說：免！我自己來

HoG

RIFT

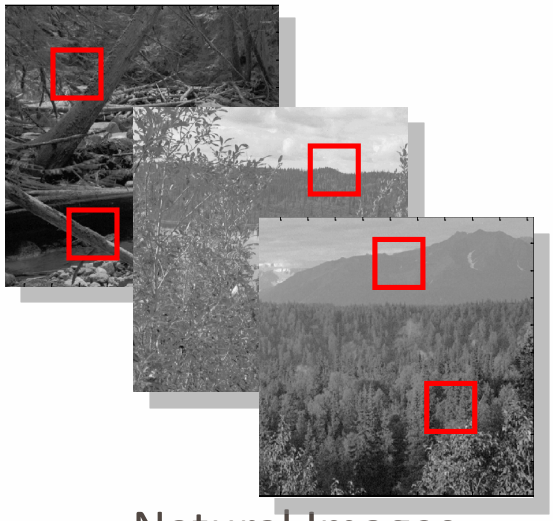


Textons

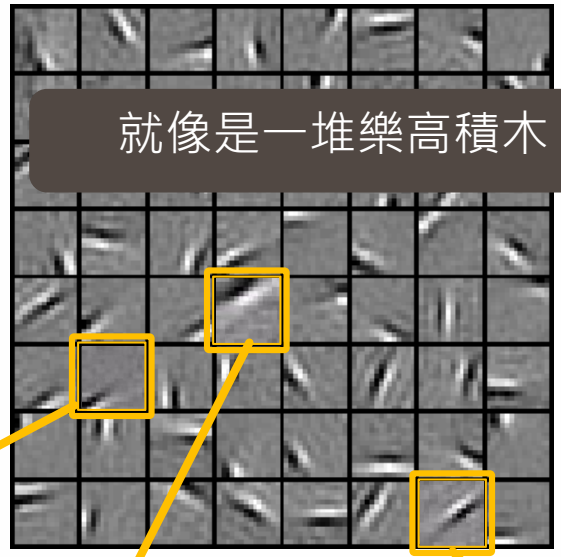


GLOH

特徵表達之於電腦視覺任務

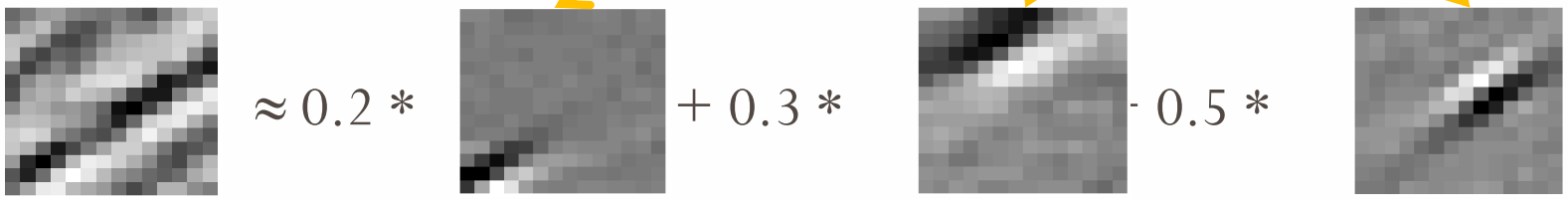


Natural Images



Learned bases ($\phi_1, \ldots, \phi_{64}$): "Edges"

Test example



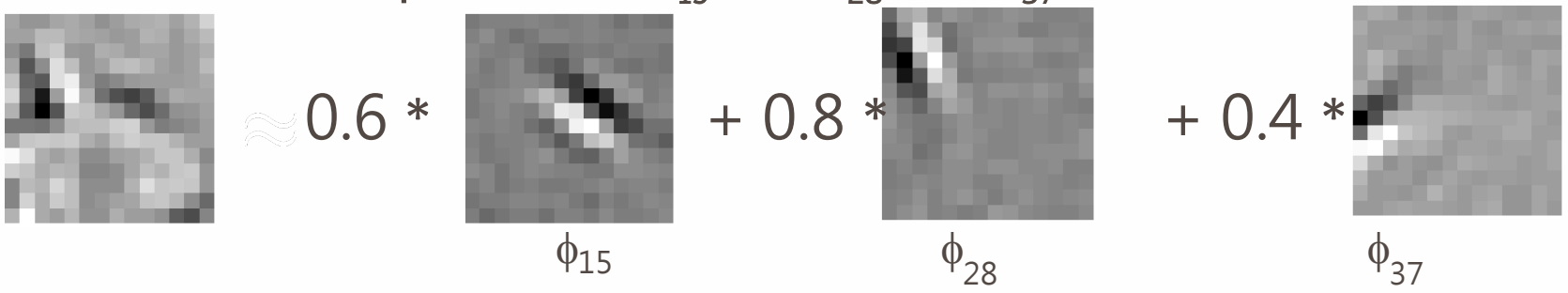
$$x \approx 0.2 * \phi_{36} + 0.3 * \phi_{42} + 0.5 * \phi_{63}$$

$$[a_1, \ldots, a_{64}] = [0, 0, \ldots, 0, \mathbf{0.2}, 0, \ldots, 0, \mathbf{0.3}, 0, \ldots, 0, \mathbf{0.5}, 0]$$

(feature representation)

不同資料其特徵表達不同

Represent as: [a_{15}=0.6, a_{28}=0.8, a_{37} = 0.4]



Represent as: [a_5=1.3, a_{18}=0.9, a_{29} = 0.3]

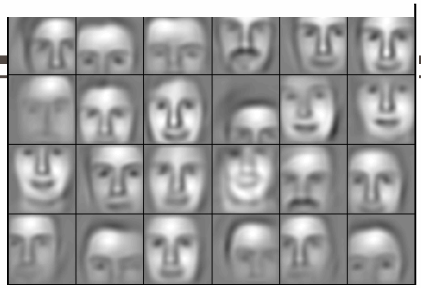
透過特徵表達的數字，

可以較容易分別不同資料的異同

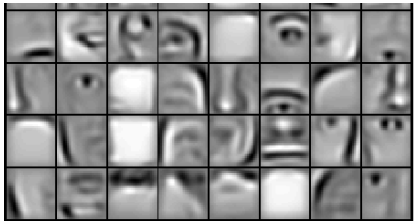
深度學習的特徵表達

有比過去還要多、還要複雜的樂高積木，可以對付更複雜的資料!!

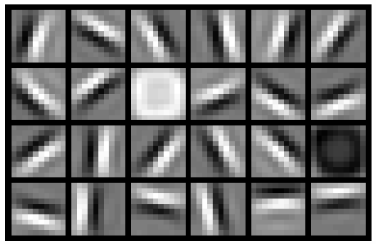
Training images



object models



object parts (combination of edges)

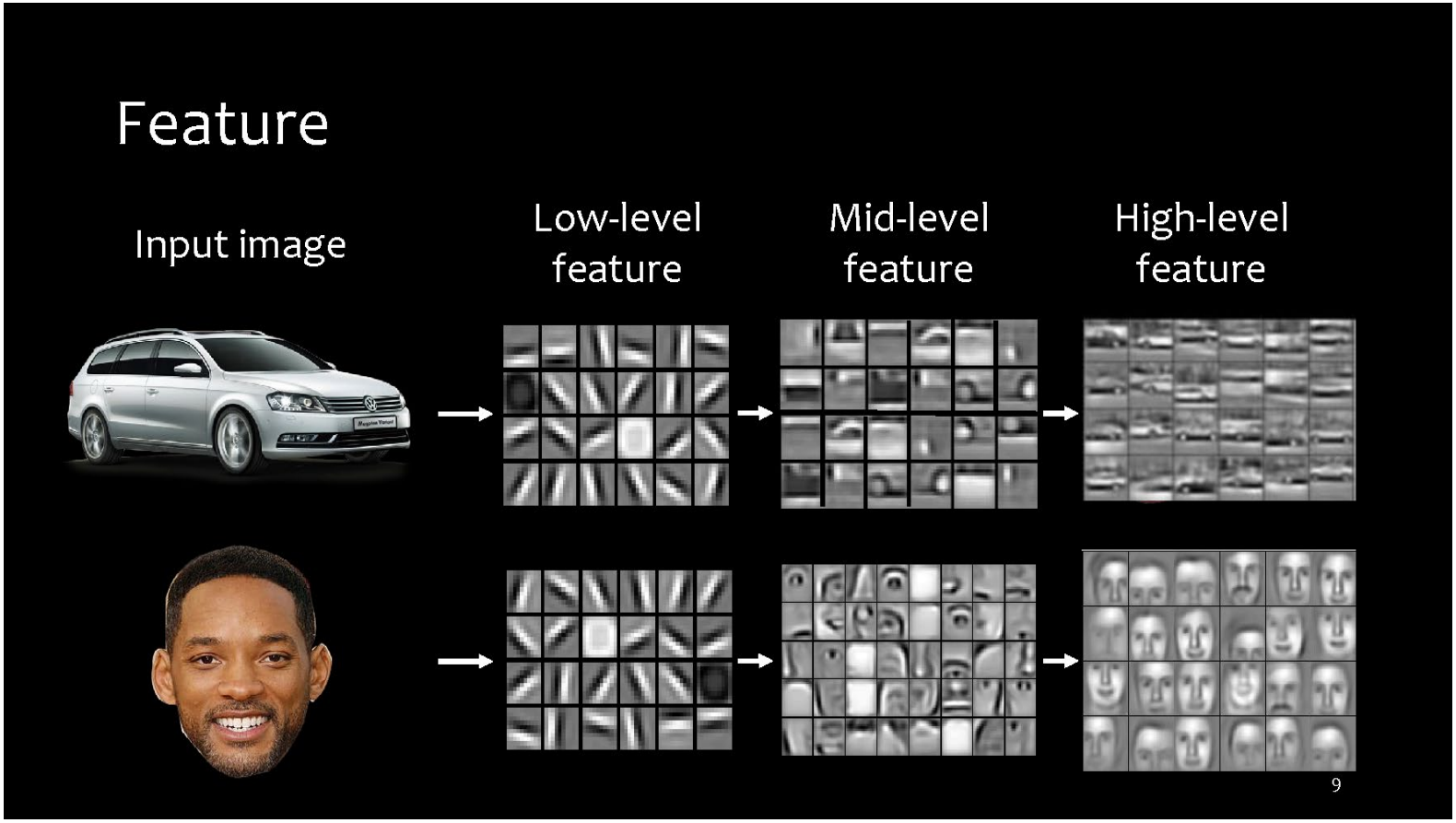


edges

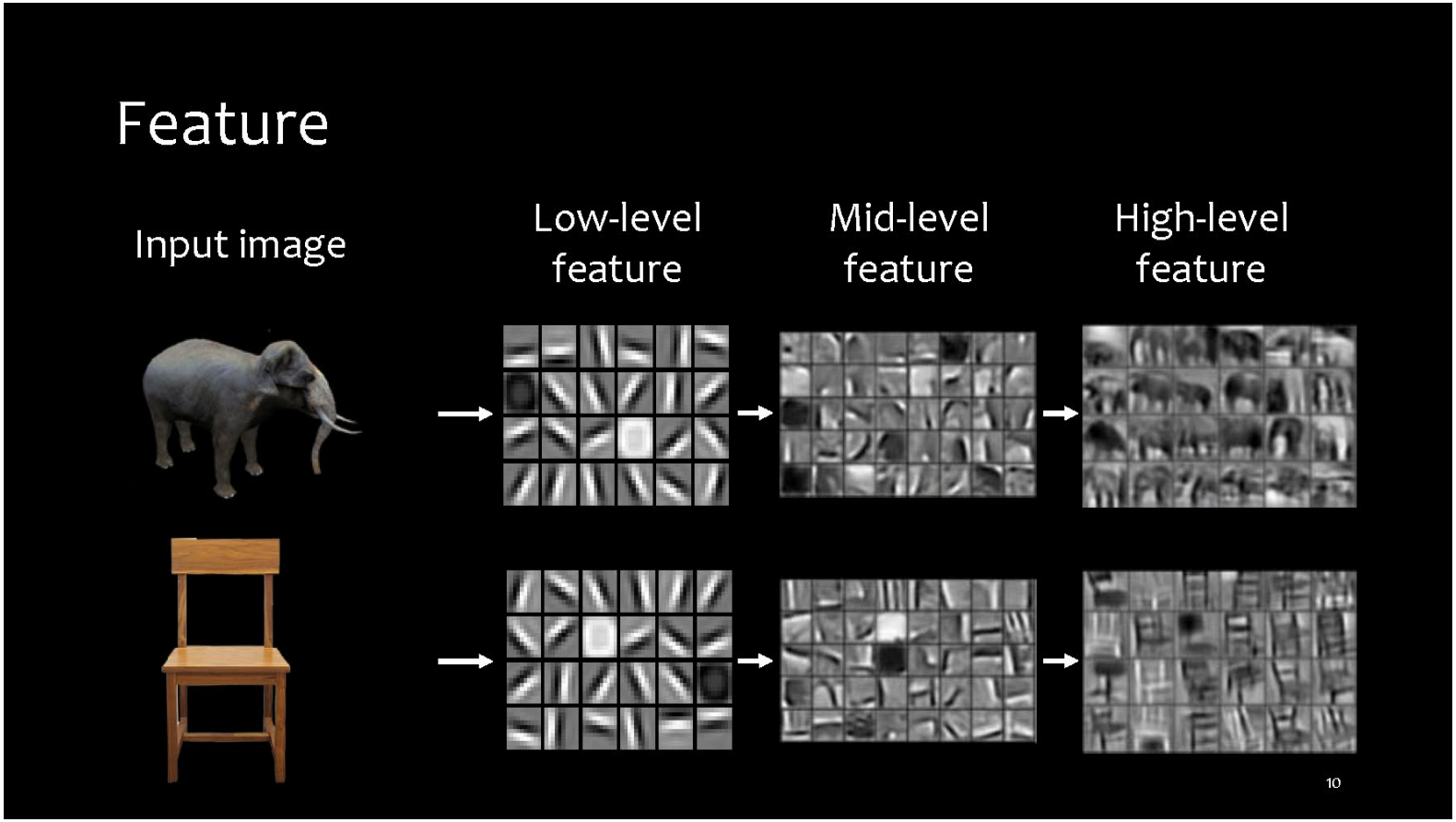
pixels

[Honglak Lee

階層式特徵表達 (Hierarchical Feature)



階層式特徵表達 (Hierarchical Feature)

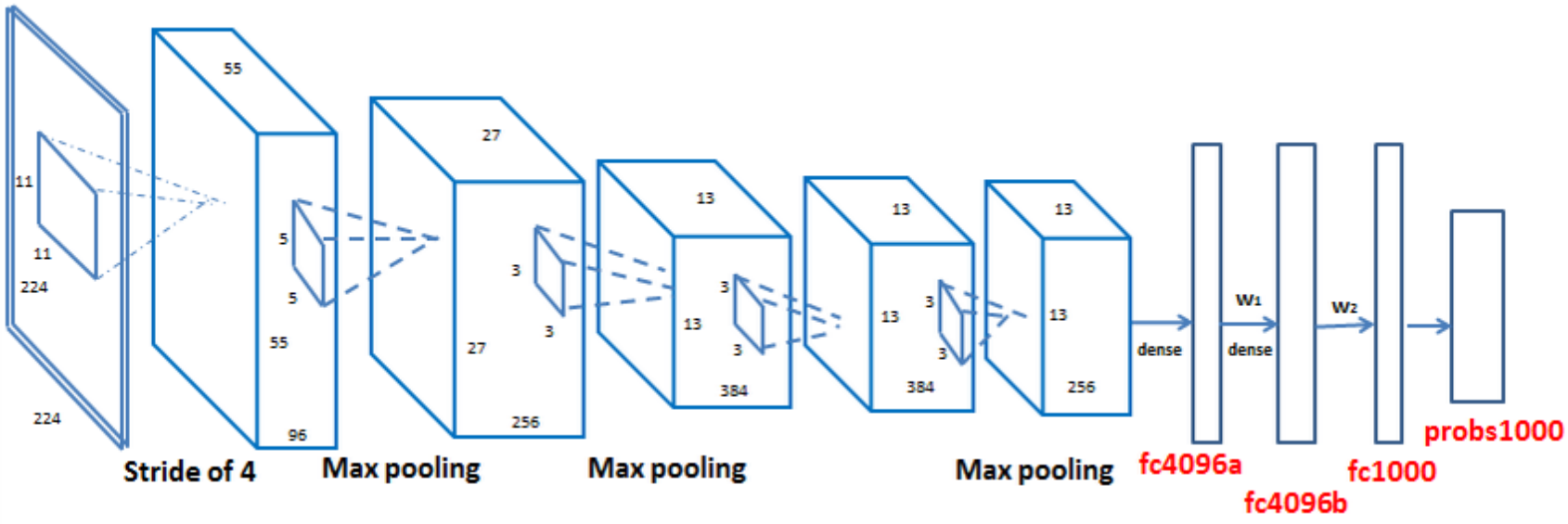


目前深度學習的技術?

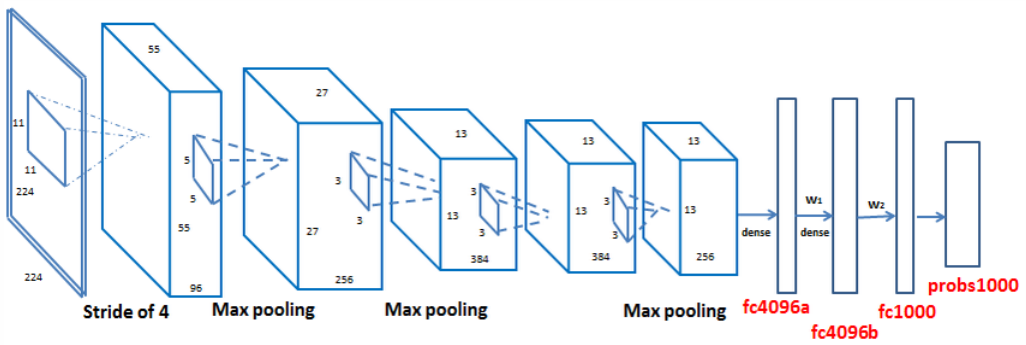
Convolutional Neural Networks

AlexNet (2012, Hinton)

- Winner of ILSVRC 2012
  - Much better than traditional ones!
- So far we know that
  - Deep learning >> machine learning



CNN的經典Models



- 2013年 AlexNet: 8 layers (9 layers)
- 2016年 Residual Net / DenseNet: up to 152 layers…
- 2017年 Stochastic depth Net: up to 1000 layers...
- 2021年 ResNet variants & Network Architecture Search (NAS)
- 2023? Visual Transformer, Mixed Models, Multimodal modal, generative models…