深度學習
Deep Learning

Course Introduction

Instructor: 林英嘉 (Ying-Jia Lin)
2025/09/03

Course Website & Slido

114-1

Deep Learning

課程頁面

https://github.com/mcps5601/CGUDL_2025_fall

Slido

DL0902
(Link)

授課教師：林英嘉

長庚大學 人工智慧學系 助理教授 (2025/02 -)

國立清華大學 資訊工程學系 博士後研究員 (2024 - 2025)

國立成功大學 資訊工程學系 博士 (2019 - 2024)

國立陽明大學 生物醫學資訊研究所 碩士 (2017 - 2019)

長庚大學 生物醫學系 學士 (2013 - 2017)

2017 年開始一腳踏進深度學習

以生成式對抗網路進行醫學影像資料擴增

國立陽明大學 / 生物醫學資訊研究所 / 107 / 碩士 / 生命科學學門 / 生物化學學類

研究生:林英嘉

指導教授:鍾翊方

論文種類 : 學術論文

電子全文

國圖紙本論文

被引用:1 點閱:778 評分: 下載:21 書目收藏:0

電腦視覺 (Computer Vision)

基於深度學習的醫學文字探勘與序列標註任務

國立成功大學 / 資訊工程學系 / 112 / 博士 / 工程學門 / 電資工程學類

研究生:林英嘉

指導教授:高宏宇

論文種類 : 學術論文

電子全文

被引用:0 點閱:423 評分: 下載:10 書目收藏:0

自然語言處理 (Natural Language Processing)

Self-driving cars

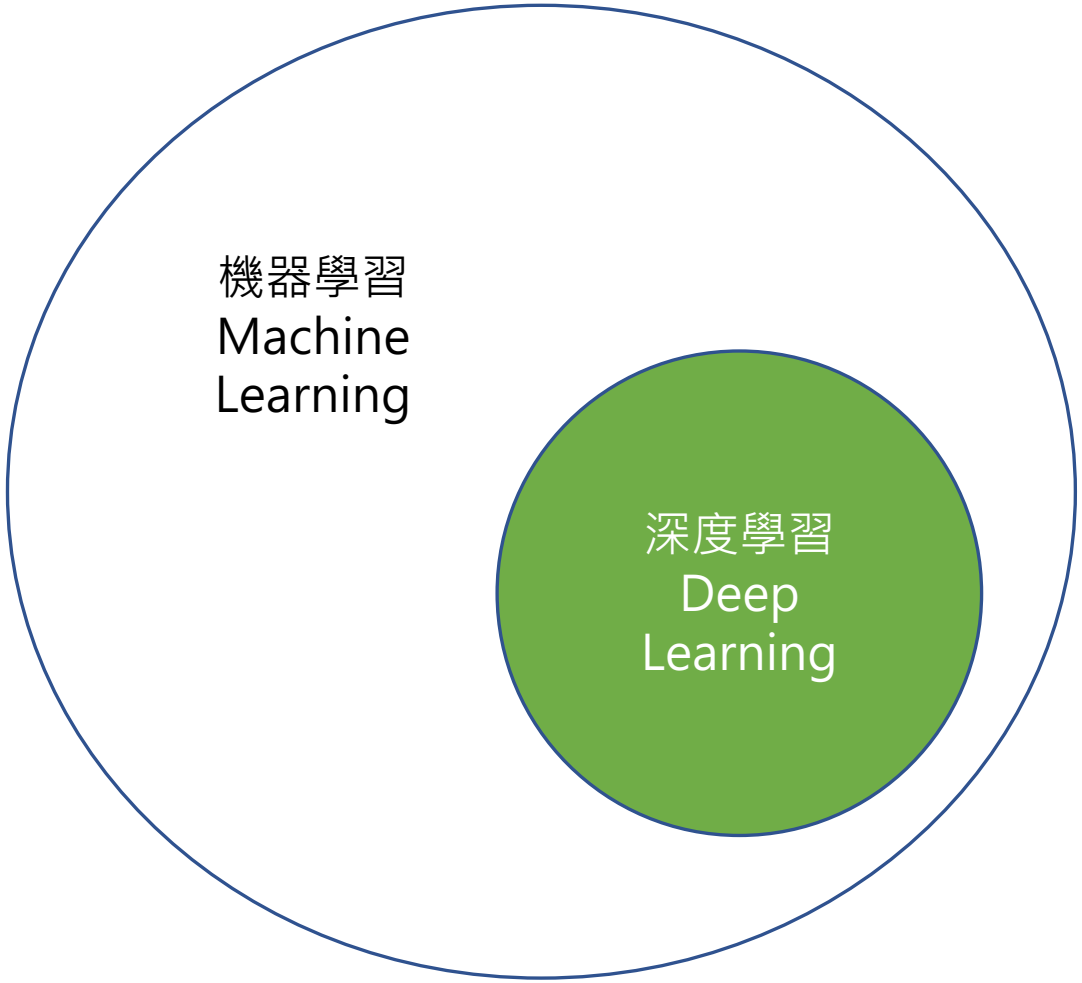
大學to碩班時期

Tesla Self-Driving
https://youtu.be/tlThdr3O5Qo?si=LGyPVqj68AGCdA_L

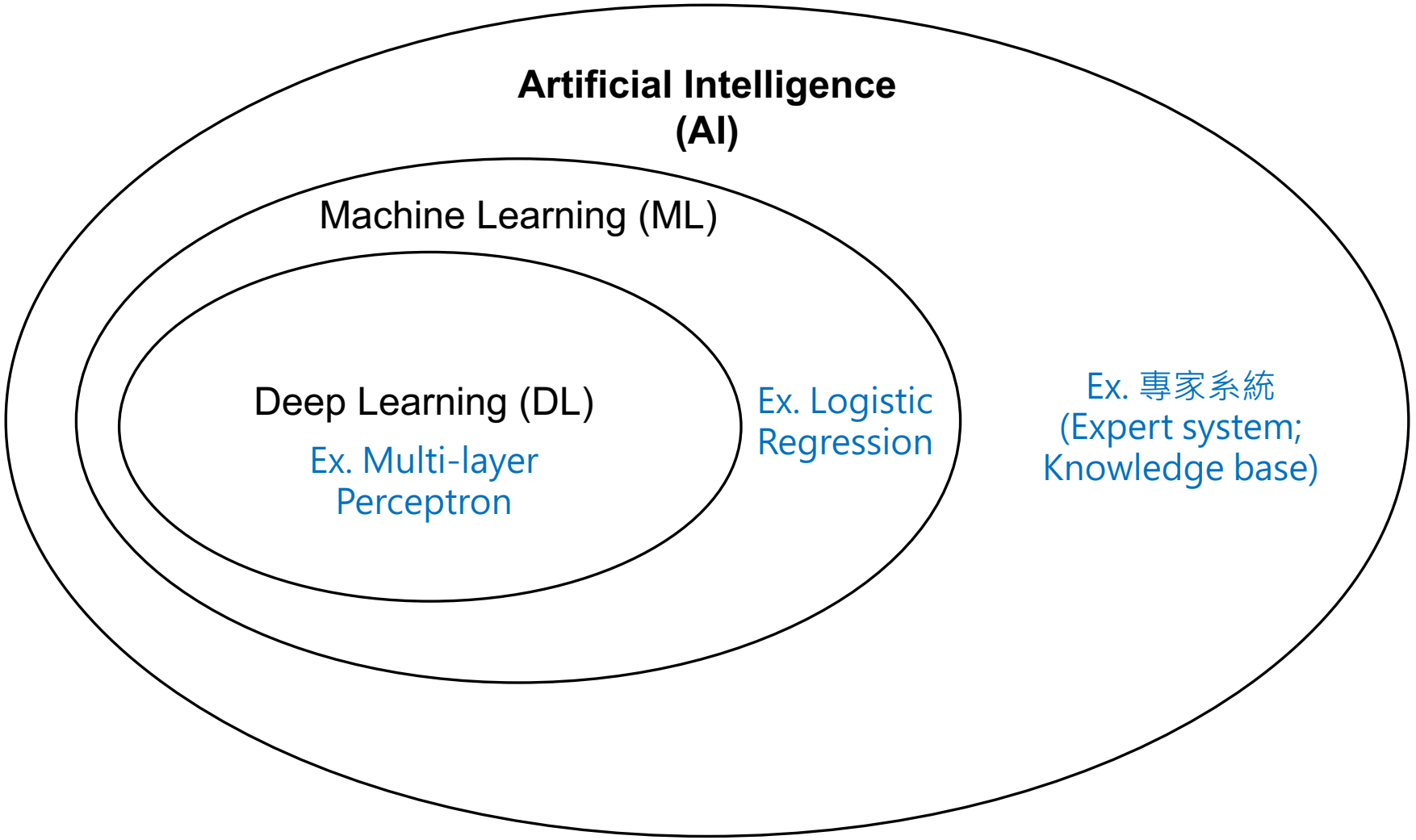
Outline

- Deep Learning 是什麼？
- 深度學習為什麼現在變強了？
- 我們這學期要學什麼？

機器學習與深度學習的關係



機器學習與深度學習的關係



Mycin 專家系統 (Stanford Universty)

```
(defrule 52
 if (site culture is blood)
  (gram organism is neg)
  (morphology organism is rod)
  (burn patient is serious)
 then .4
  (identity organism is pseudomonas))
```

如果滿足這些條件
則有0.4的機率是綠膿桿菌

https://www.britannica.com/technology/expert-system

另一個廣義型專家系統：https://en.akinator.com/

什麼是機器學習？(以電腦視覺為例)

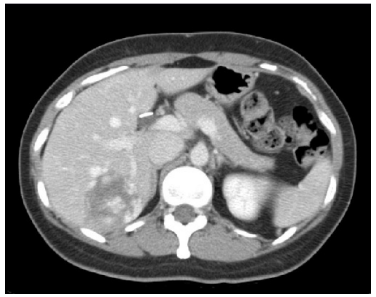
Machine Learning function!

Classifier

Cat ☑

Dog

Apple



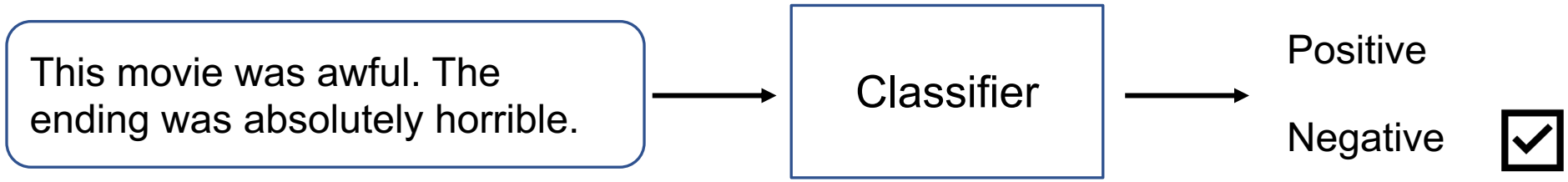
Classifier

Malignant

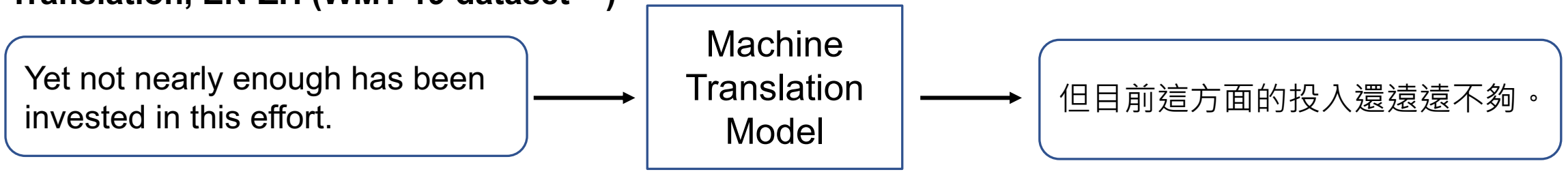
Benign ☑

什麼是機器學習？(以自然語言處理為例)

Sentiment Analysis (IMDb dataset [1])



Translation, EN-ZH (WMT-19 dataset [2])



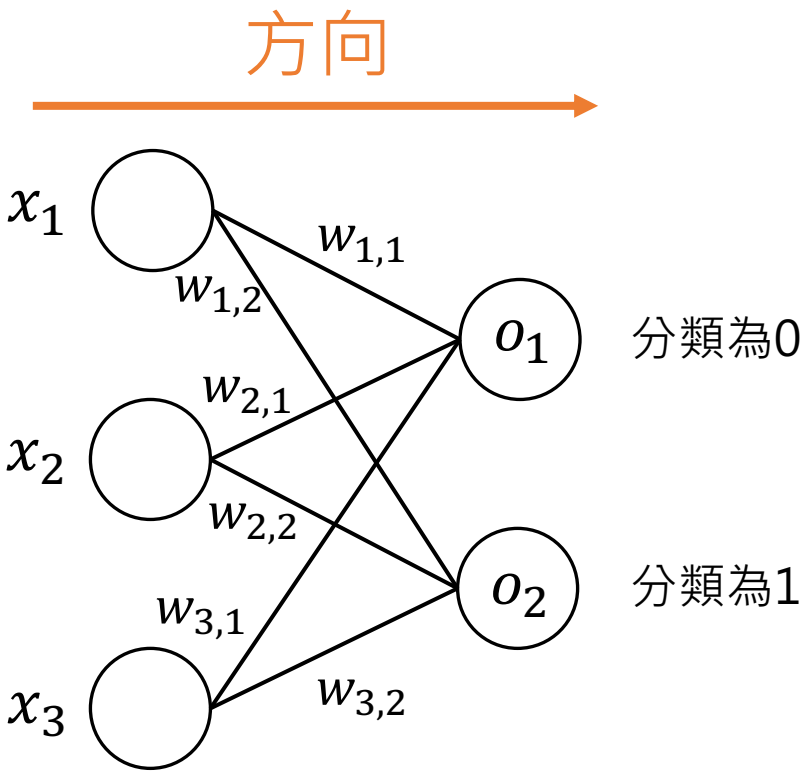
[1] https://huggingface.co/datasets/stanfordnlp/imdb
[2] https://huggingface.co/datasets/wmt/wmt19/viewer/zh-en

什麼是深度學習?

- 「深度學習」是一種機器學習方法，著重於使用多層的類神經網路 (Neural Networks) 來完成任務，例如分類、生成等等。

深度！

簡單的類神經網路：Perceptron (感知機)



x 代表輸入

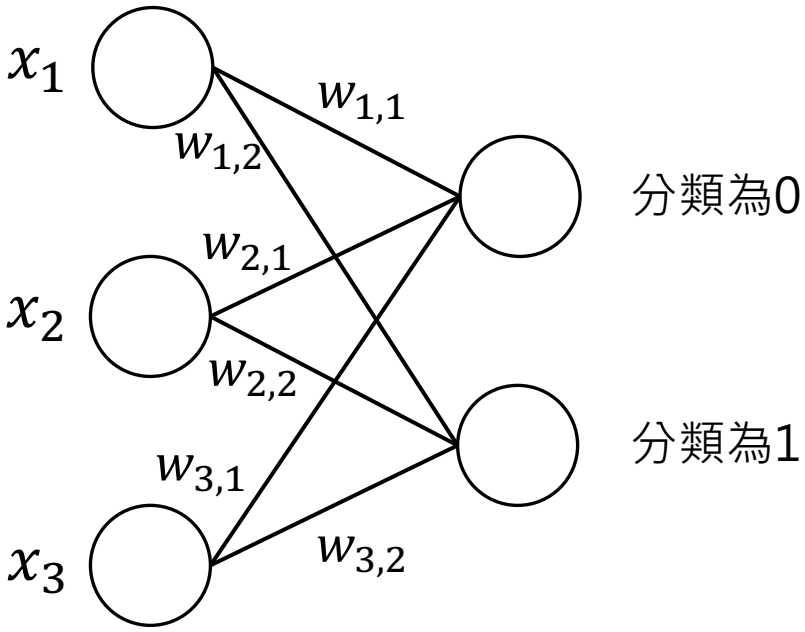
w 代表權重值

代表x有多少比例要被傳至下一層

$$o_1 = x_1 w_{1,1} + x_2 w_{2,1} + x_3 w_{3,1}$$

→ 只有經過一層權重值計算

類神經網路：Perceptron 有點像神經細胞



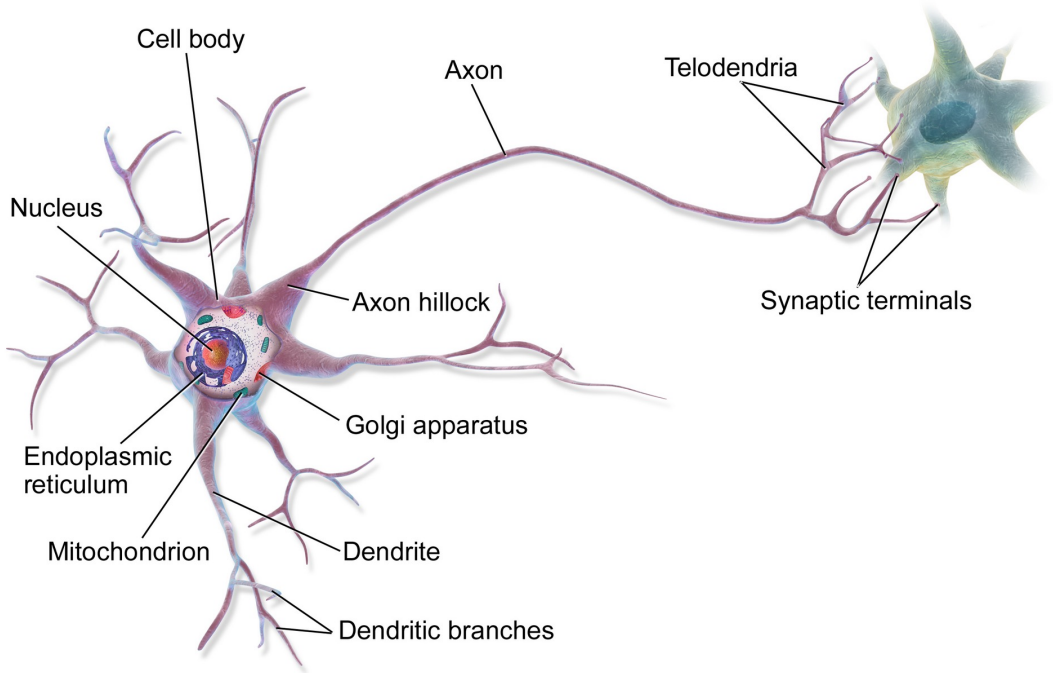
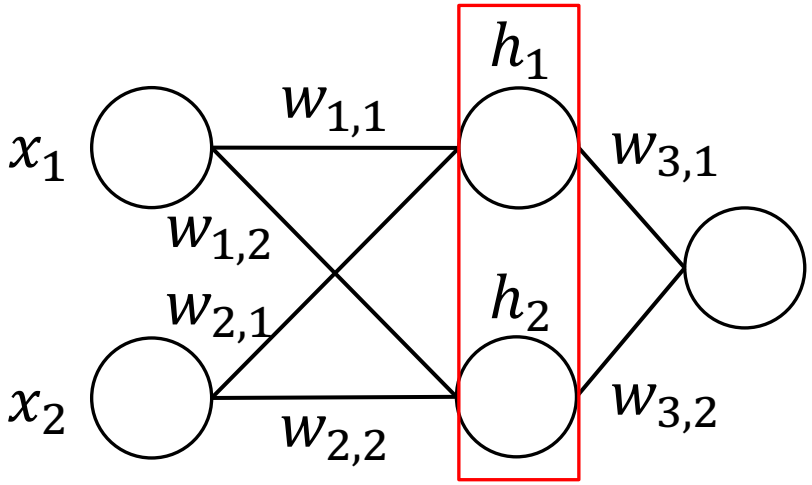


Figure source: https://zh-yue.wikipedia.org/wiki/神經細胞

Multi-layer Perceptron (MLP, 多層感知機)



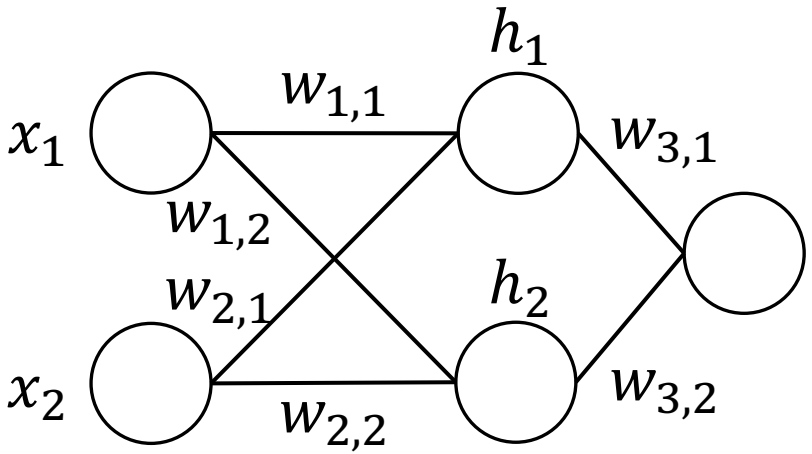
隱藏層：不是輸入層也不是輸出層的層

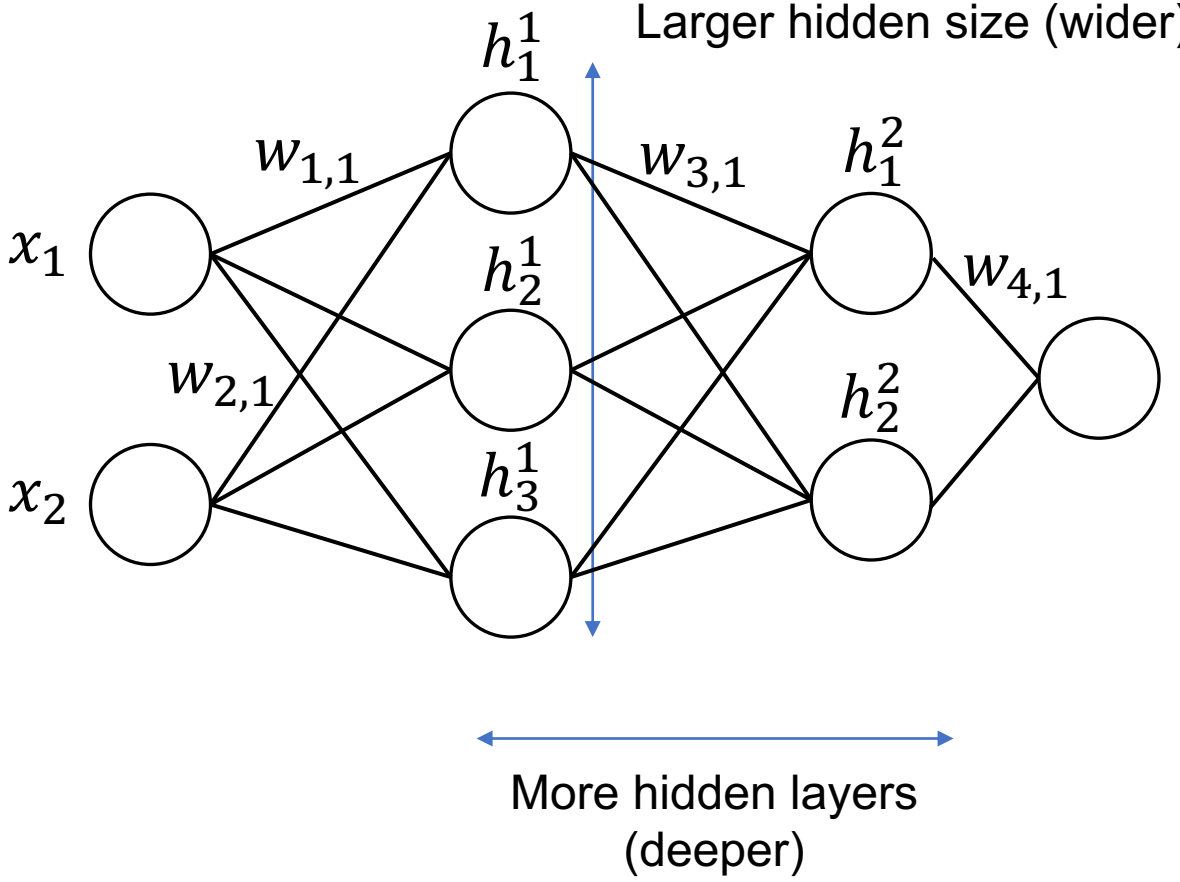
輸入層　　隱藏層 (1st 層)　　輸出層 (2nd 層) ⟶ 經過兩層權重值計算 (可以被稱作多層)

什麼是深度學習?

- 「深度學習」是一種機器學習方法，著重於使用多層的類神經網路 (Neural Networks) 來完成任務，例如分類、生成等等。
- 在深度學習模型中，通常參數數量都非常龐大，這主要來自於隱藏層大小 (hidden size) 與模型層數 (number of hidden layers) 的增加。

Deeper and Wider (加深、加寬)





Deep 真的比較好嗎？

Sun, Chen, et al. "Revisiting unreasonable effectiveness of data in deep learning era." Proceedings of the IEEE international conference on computer vision. 2017.

提到：

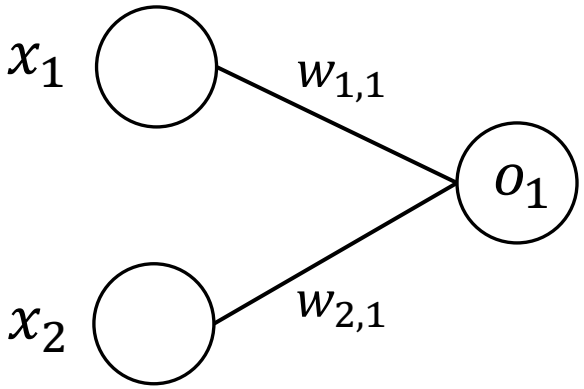
1. 資料量一樣的情況下，模型越大表現越好
2. 在模型一樣的情況下，資料量越多，表現越好

還有蠻多論文：

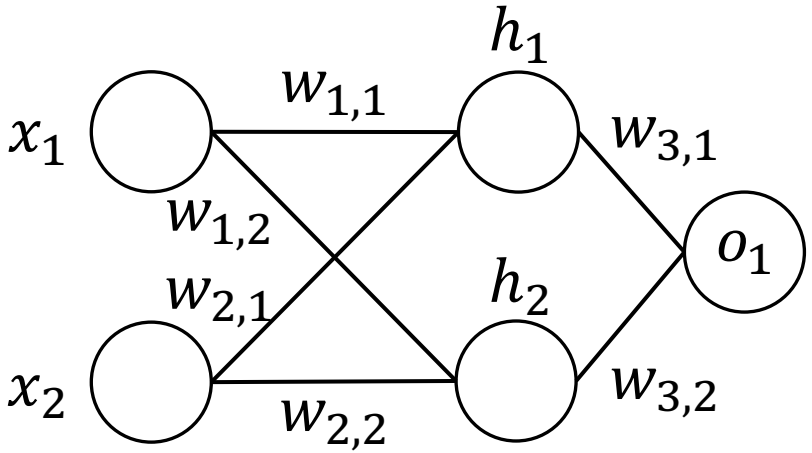
Kaplan, Jared, et al. "Scaling laws for neural language models." arXiv preprint arXiv:2001.08361 (2020).

Rae, Jack W., et al. "Scaling language models: Methods, analysis & insights from training gopher." arXiv preprint arXiv:2112.11446 (2021).

實驗結果以外的「理論」

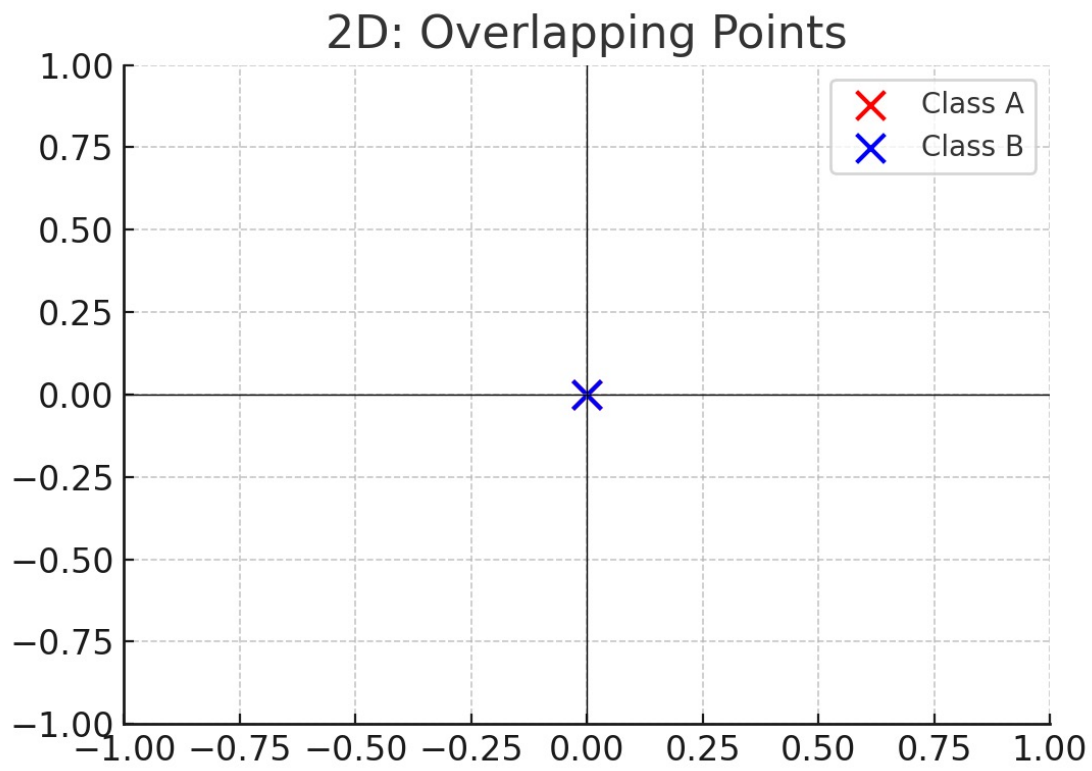


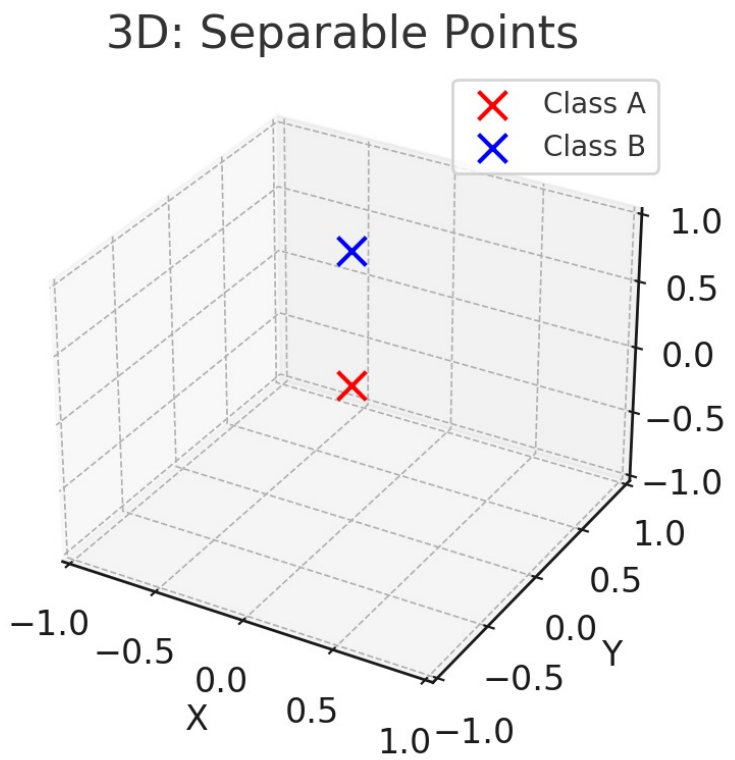
$$o_1 = x_1 w_{1,1} + x_2 w_{2,1}$$



$$o_1 = h_1 w_{3,1} + h_2 w_{3,2}$$
$$= (x_1 w_{1,1} + x_2 w_{2,1}) w_{3,1}$$
$$+ (x_2 w_{2,1} + x_2 w_{2,2}) w_{3,2}$$

實驗結果以外的「理論」


2D: Overlapping Points
Class A
Class B
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
−1.00 −0.75 −0.50 −0.25 0.00 0.25 0.50 0.75 1.00


3D: Separable Points
Class A
Class B
X
Y
−1.0
−0.5
0.0
0.5
1.0

現在的模型有多 Deep? (B: 10億)

Dinov3 🦖

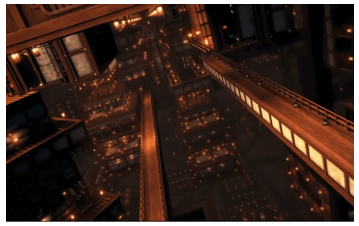
40層
7B

deepseek
R1

61層
685B

GPT-3

96層
175B



深度學習為什麼現在變強了？

Difference between AI, ML, and DL

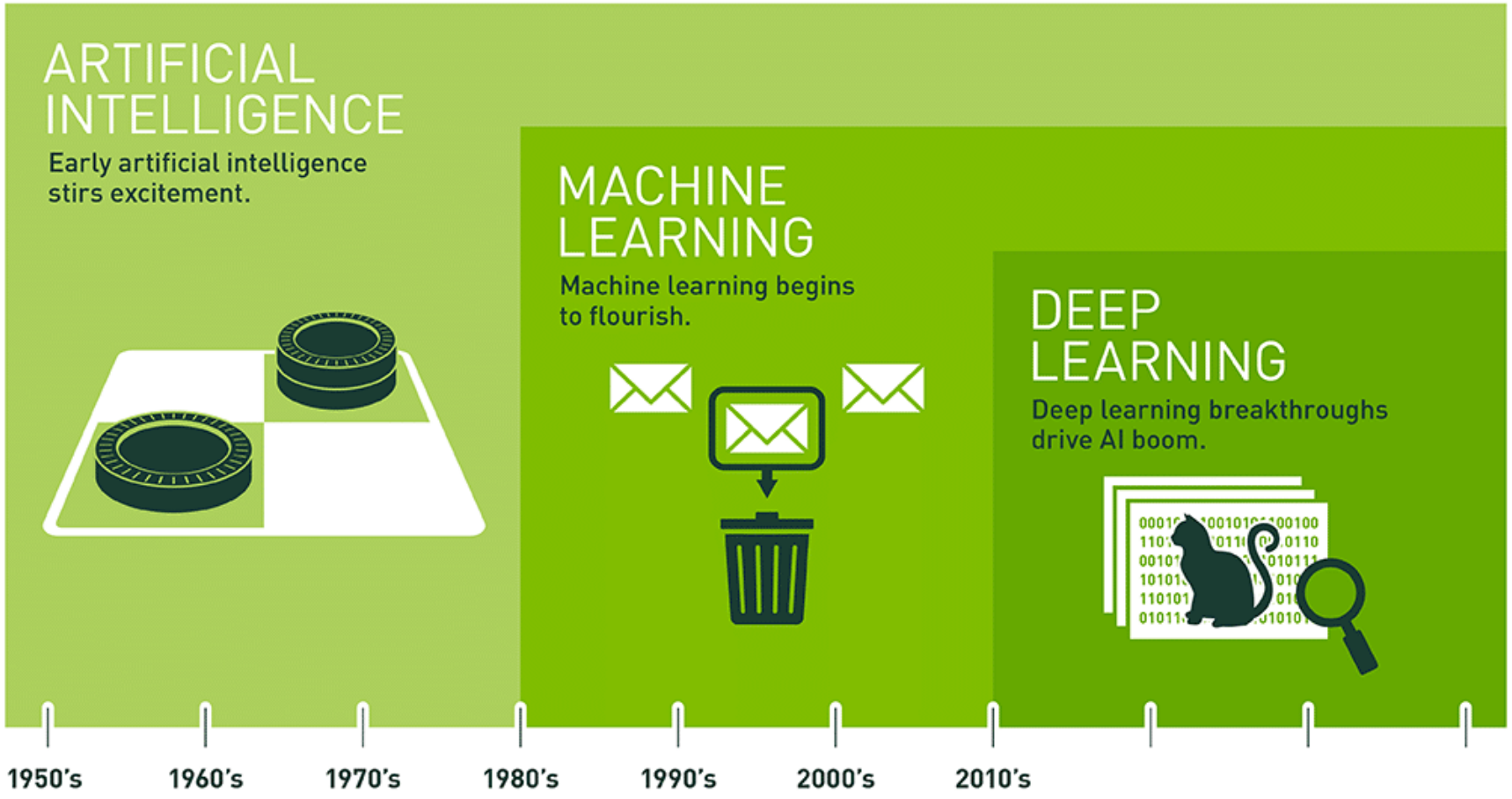
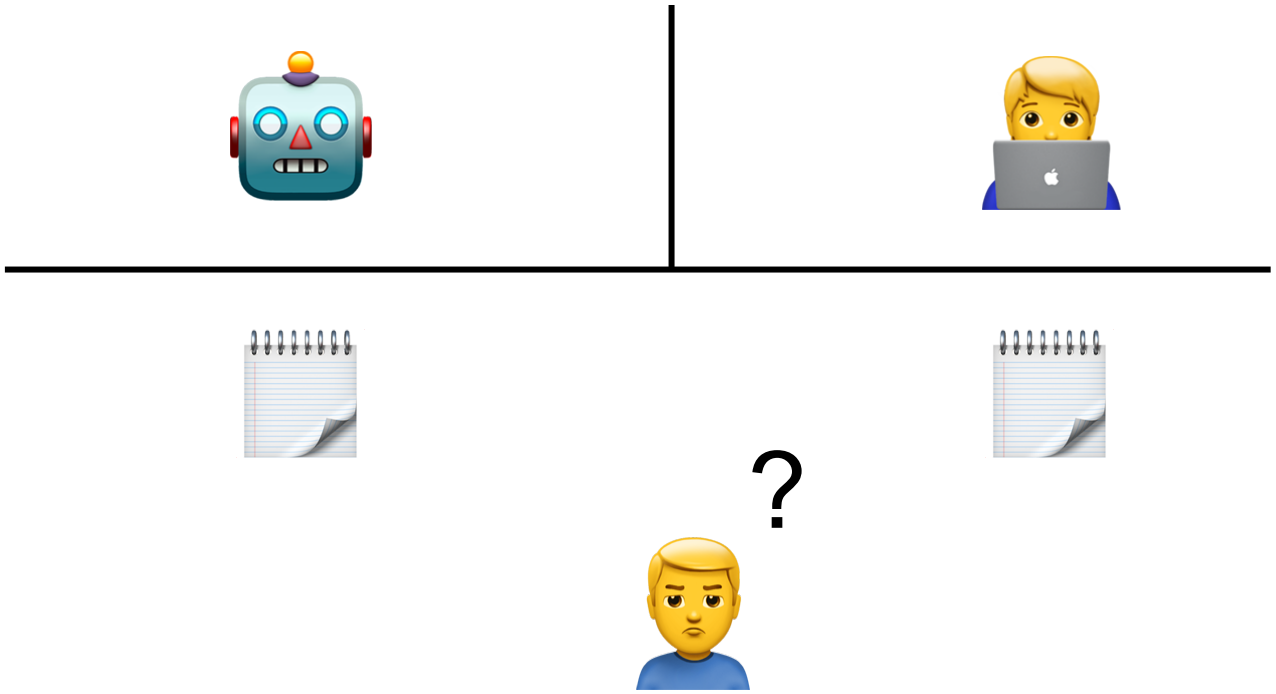


Figure source: https://blogs.nvidia.com/blog/whats-difference-artificial-intelligence-machine-learning-deep-learning-ai/

圖靈測試 (Turing Test)

- Proposed by Alan Turing in 1950
- One of the earliest concepts of artificial intelligence.



The first Perceptron (1958)

Rosenblatt, Frank. "The perceptron: a probabilistic model for information storage and organization in the brain." Psychological review 65.6 (1958): 386.

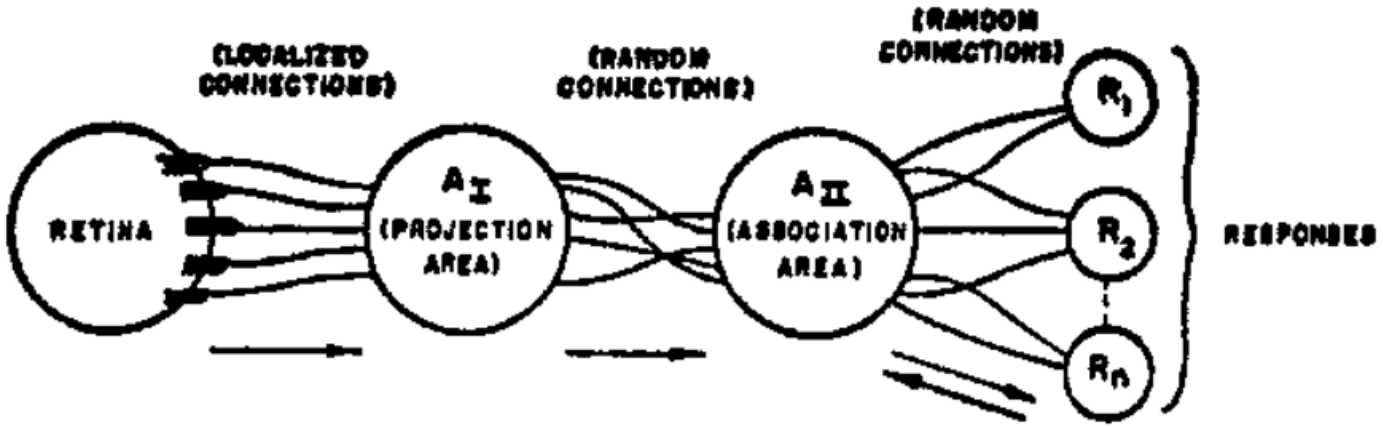


FIG. 1. Organization of a perceptron.

The First Convolutional Neural Networks: Neocognitron (1980)

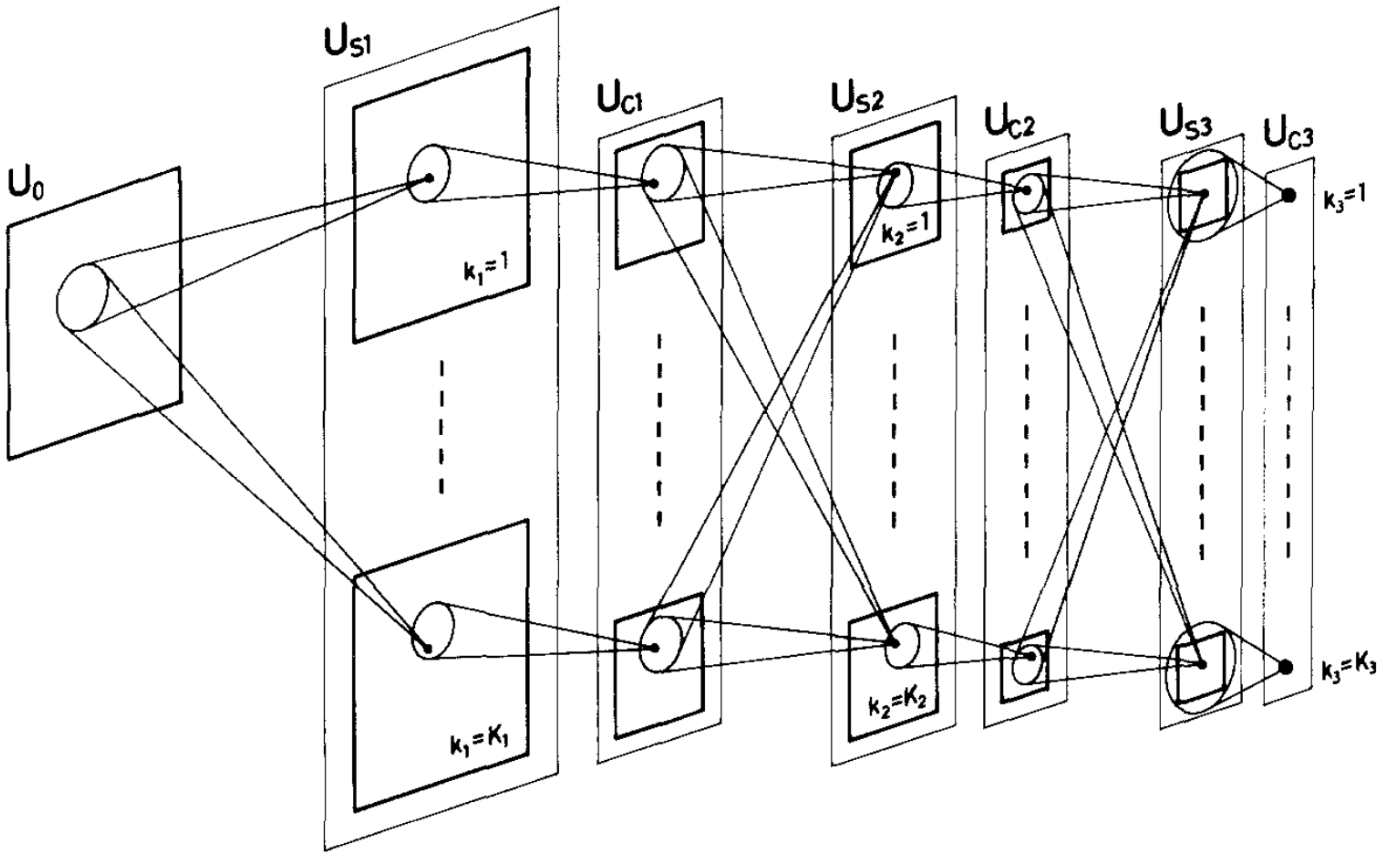


Fig. 2. Schematic diagram illustrating the interconnections between layers in the neocognitron

Fukushima, Kunihiko. "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position." Biological cybernetics 36.4 (1980): 193-202.

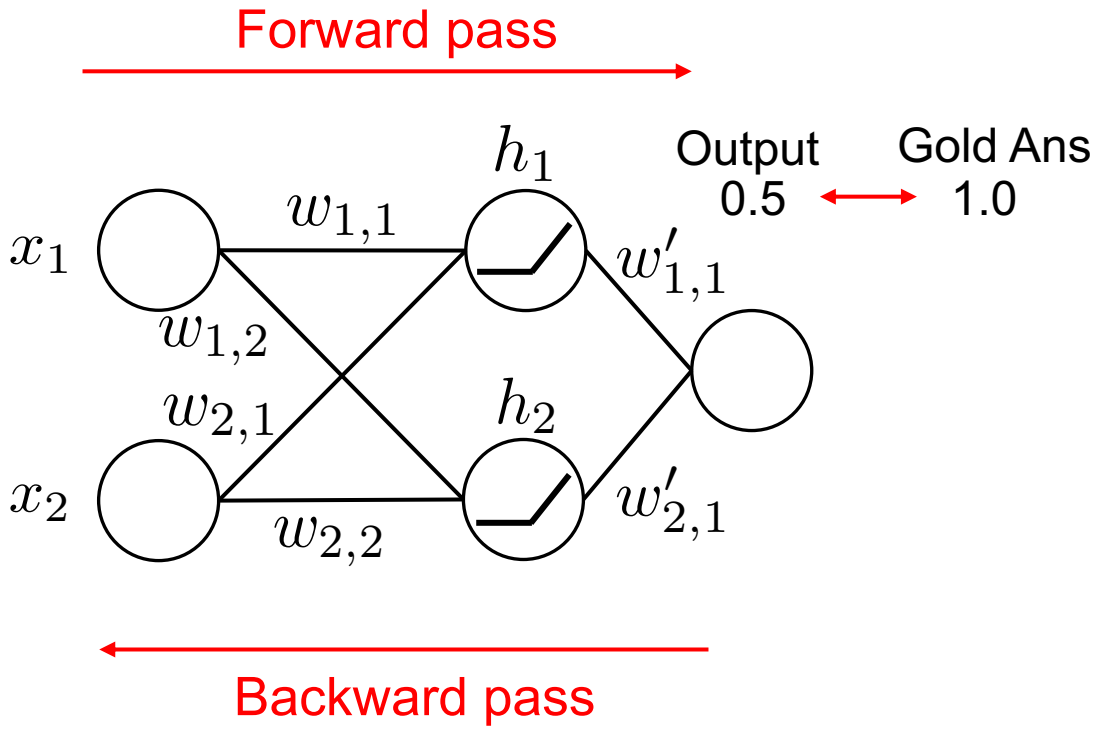
The Back-Propagation Algorithm (1986)

Learning representations by back-propagating errors

David E. Rumelhart*, Geoffrey E. Hinton† & Ronald J. Williams*

* Institute for Cognitive Science, C-015, University of California, San Diego, La Jolla, California 92093, USA
† Department of Computer Science, Carnegie-Mellon University, Pittsburgh, Philadelphia 15213, USA

We describe a new learning procedure, back-propagation, for networks of neurone-like units. The procedure repeatedly adjusts the weights of the connections in the network so as to minimize a measure of the difference between the actual output vector of the net and the desired output vector. As a result of the weight adjustments, internal 'hidden' units which are not part of the input or output come to represent important features of the task domain, and the regularities in the task are captured by the interactions of these units. The ability to create useful new features distinguishes back-propagation from earlier, simpler methods such as the perceptron-convergence procedure[1].



Rumelhart, David E., Geoffrey E. Hinton, and Ronald J. Williams. "Learning representations by back-propagating errors." nature 323.6088 (1986): 533-536.

LeNet-5: Early Convolutional Neural Networks

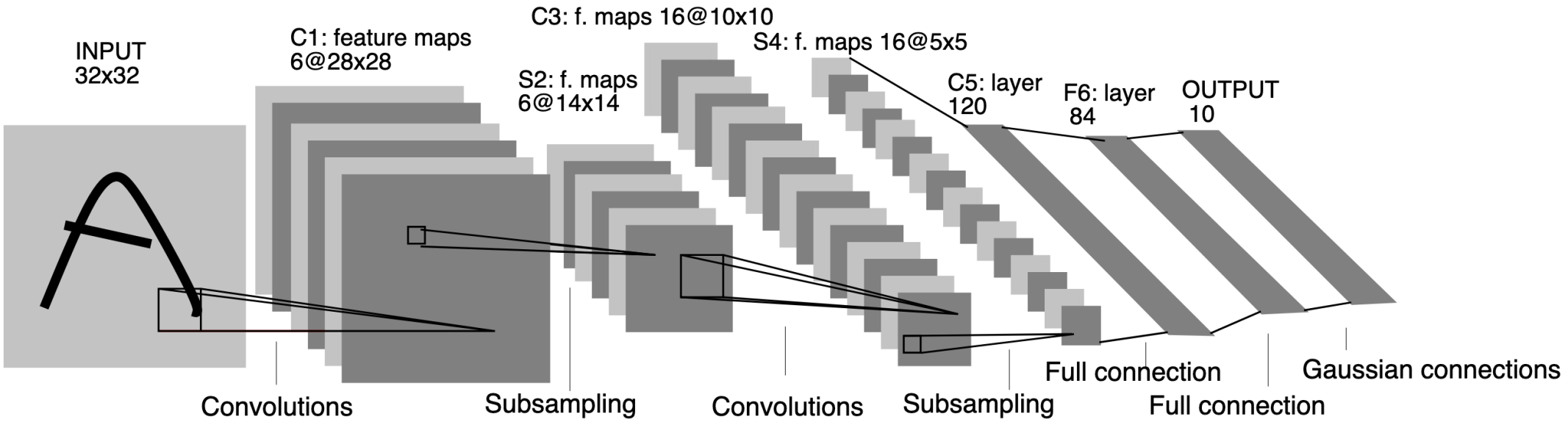
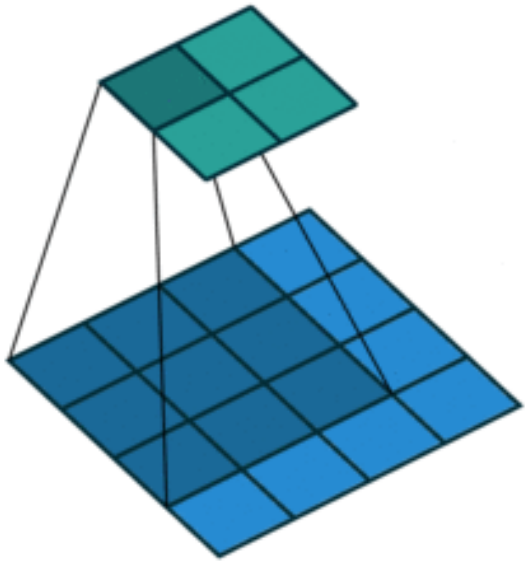


Fig. 2. Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.

Examples of Convolutions

Padding = 0, stride = 1



Padding = 0, stride = 2

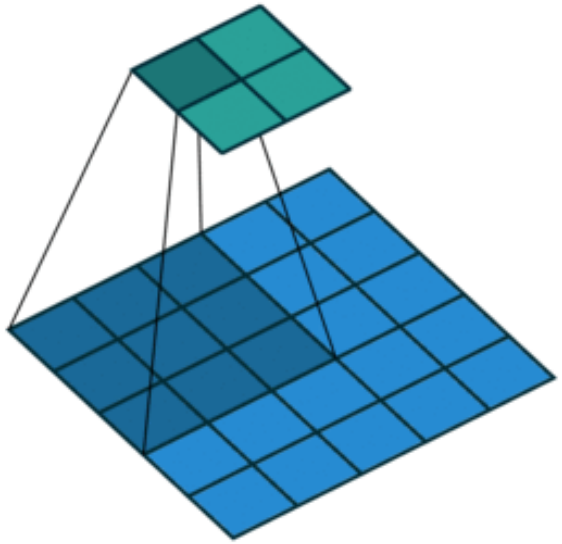


Figure source:
https://hannibunny.github.io/mlbook/neuralnetworks/convolutionDemos.html

ImageNet Competition

Complete name: ImageNet Large Scale Visual Recognition Challenge (ILSVRC)

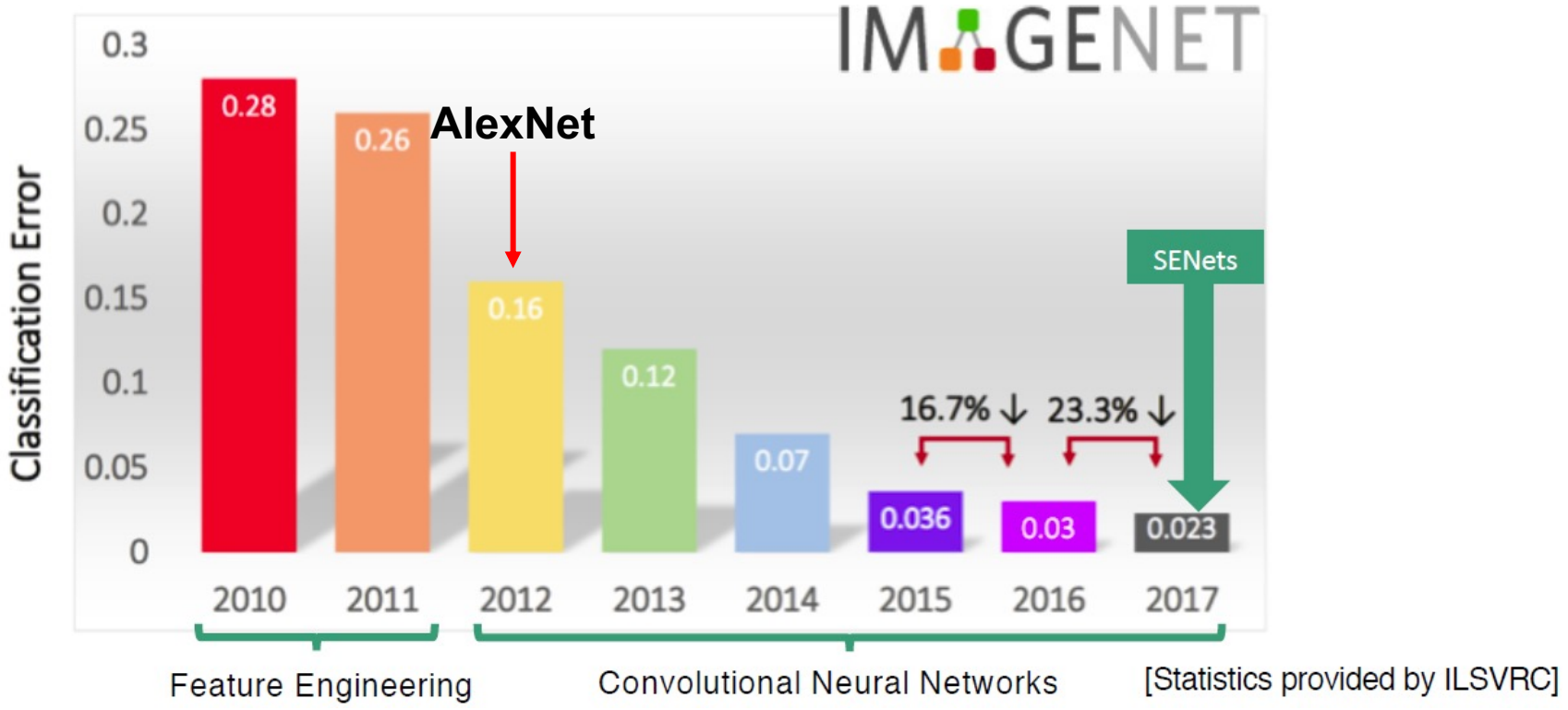


Figure source: https://www.kaggle.com/discussions/getting-started/149448

Growing Trend of GPU Usage during the ImageNet Competition

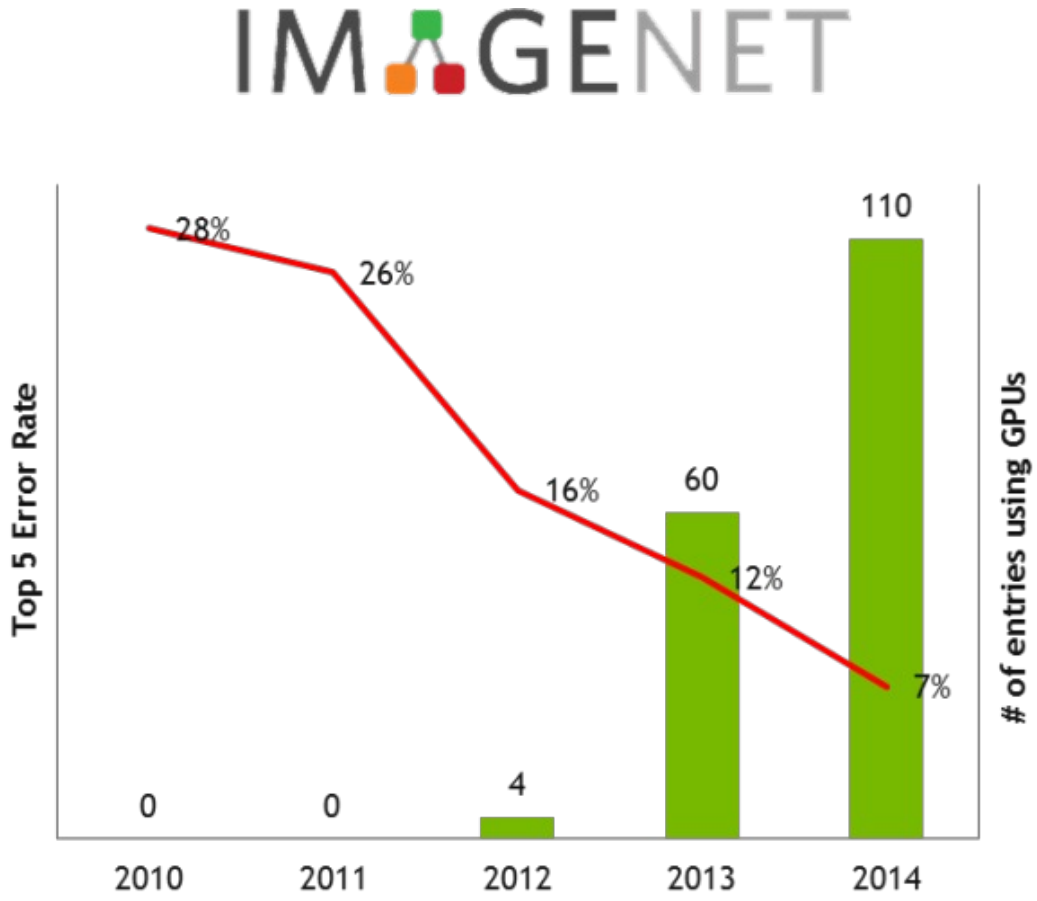
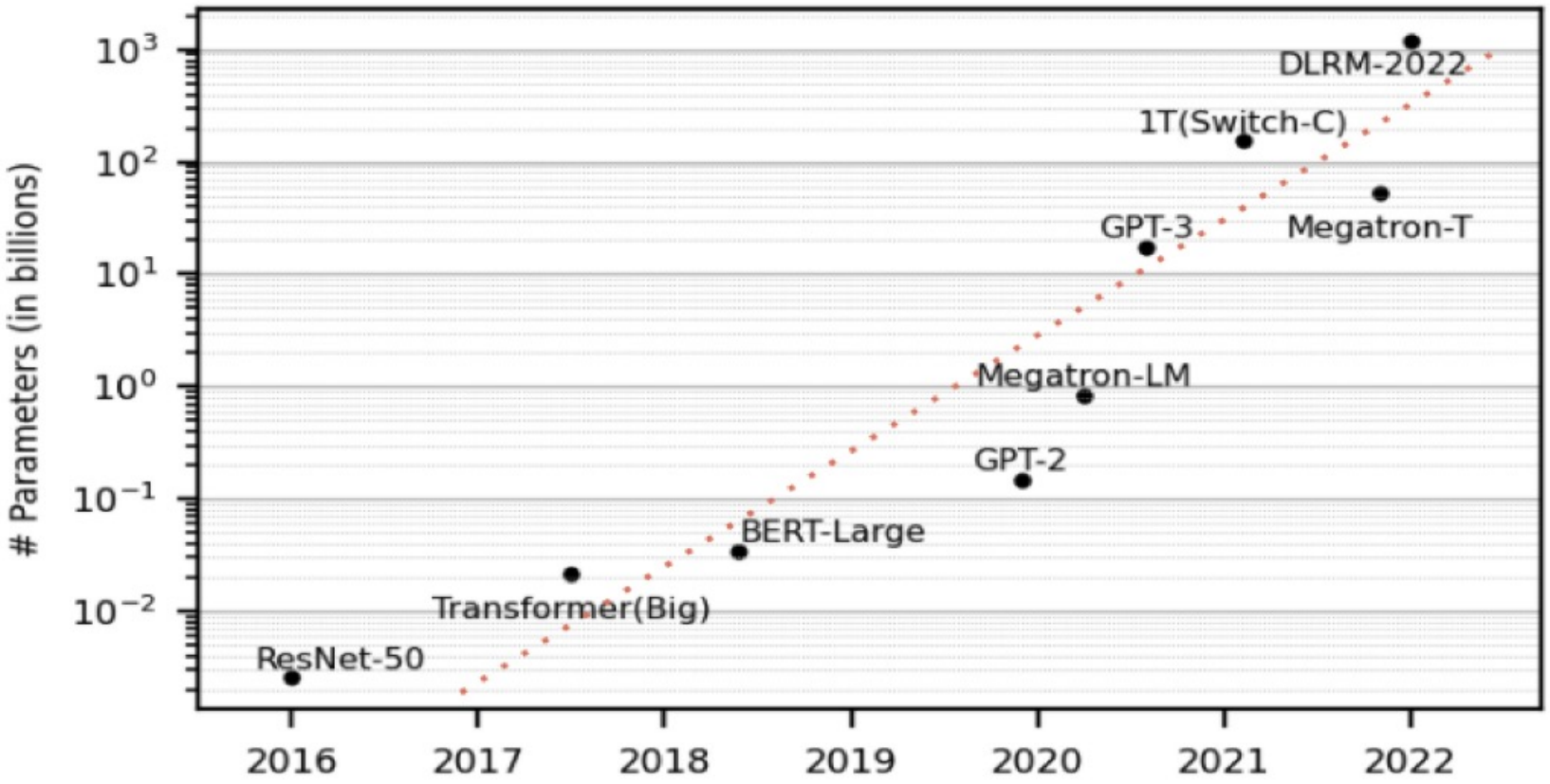


Figure source: https://developer.nvidia.com/blog/nvidia-ibm-cloud-support-imagenet-large-scale-visual-recognition-challenge/

Training a model to play Super Mario

Using reinforcement learning.
Code: https://github.com/jiseongHAN/Super-Mario-RL

Growing Sizes of Deep Learning Models



Wu, Z., Dai, L. Y., Novick, A., Glick, M., Zhu, Z., Rumley, S., ... & Bergman, K. (2023). Peta-scale embedded photonics architecture for distributed deep learning applications. Journal of Lightwave Technology, 41(12), 3737-3749.

The Revolution of ChatGPT

ChatGPT came out in November, 2022.

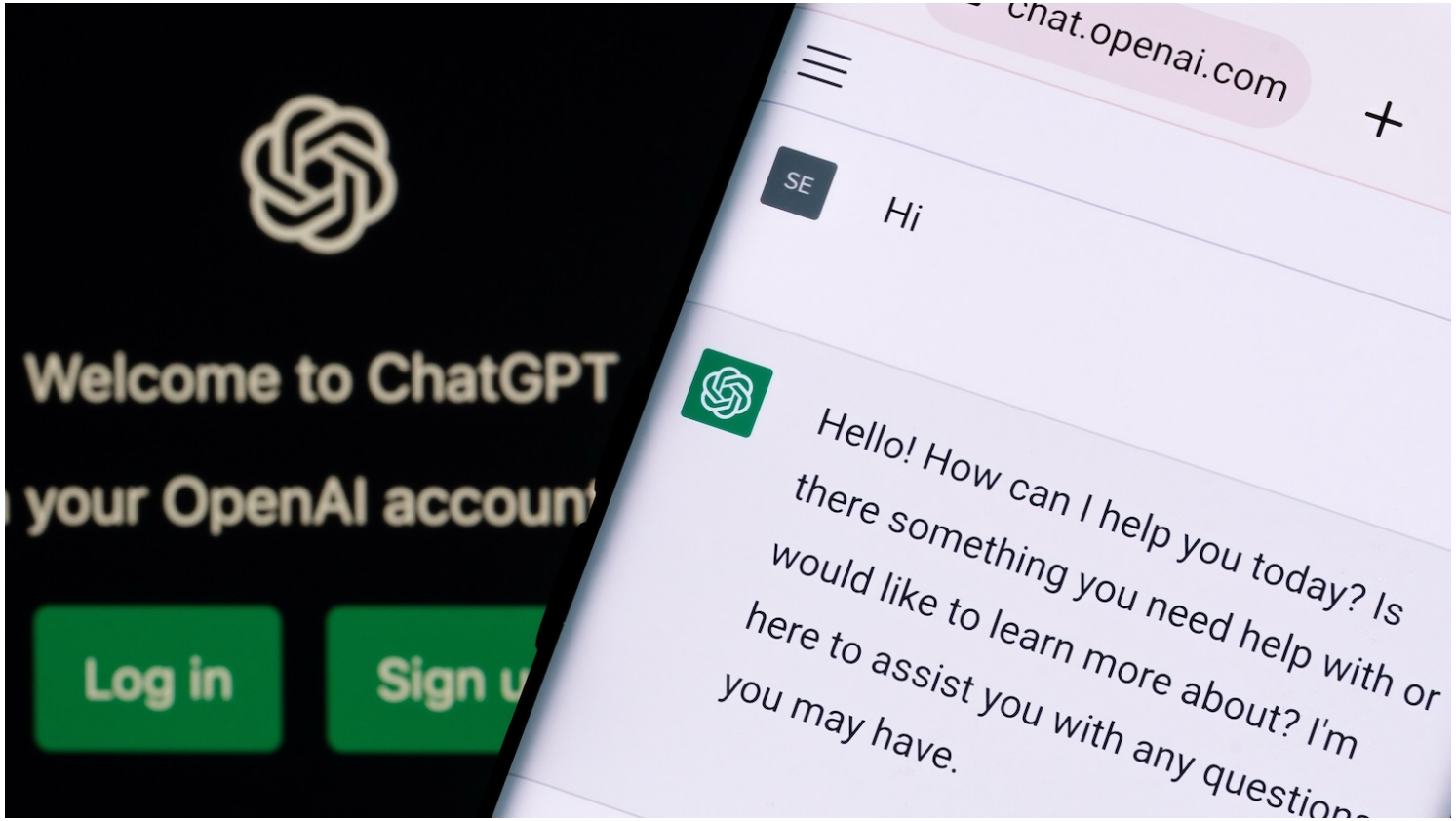


Figure source: https://poole.ncsu.edu/thought-leadership/article/lets-chat-about-chatgpt/
https://openai.com/index/chatgpt/

InstructGPT: the precursor of ChatGPT

Model size: 175B (same as GPT-3)

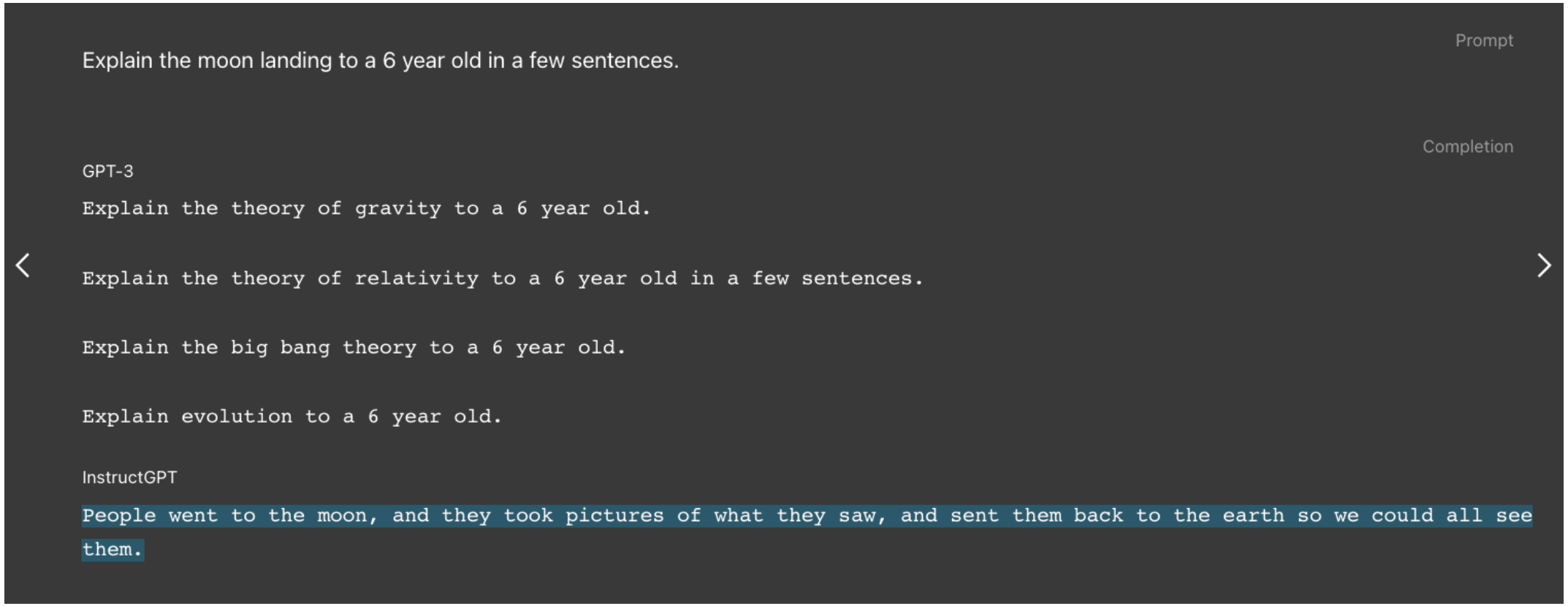


Figure source: https://openai.com/index/instruction-following/
Ouyang, Long, et al. “Training language models to follow instructions with human feedback.” NeurIPS 2022.

深度學習可以減少 Feature Selection

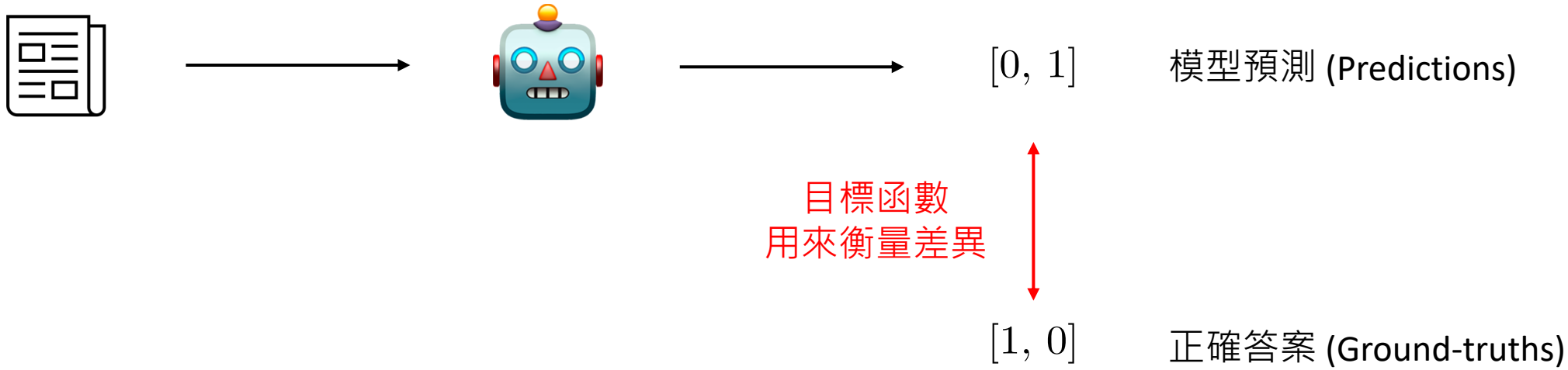
找出重要的特徵

Machine Learning Classifier

Deep Learning Classifier

模型訓練師新手村

所有的機器學習都是在做最佳化



最小化Predictions和Ground-truths之間的差異即為機器學習的目標

目標函數的別名

- 目標函數 (Objective Function)
- 損失函數 (Loss Function)
- 誤差值 (Loss)
- 可能採用「交叉熵」 (Cross entropy)

如何訓練模型：資料集

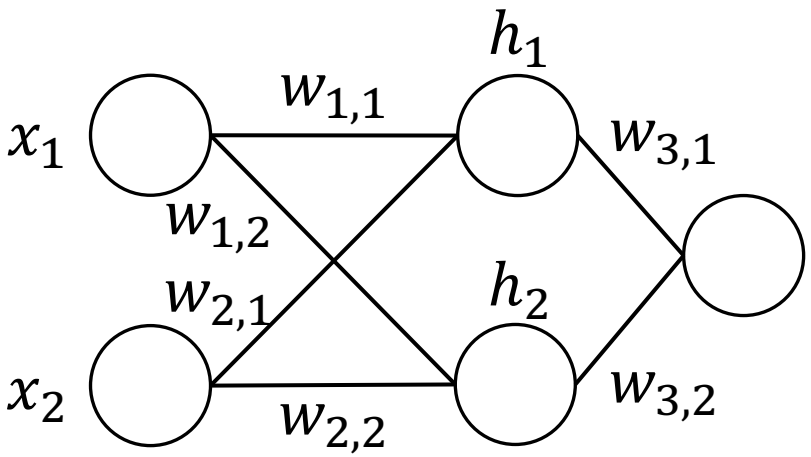
Training Set
(訓練資料集)

Test Set
(測試資料集)

Training Set
(訓練資料集)

Validation Set
(驗證資料集)

如何訓練模型：演算法



模型預測 (Predictions) ⟷ 正確答案 (Ground-truths)

- 透過 Back-Propagation 計算神經網路中的梯度
- 透過 Gradient Descent (梯度下降) 更新權重值 (w跟b)

我們這學期要學什麼？

你會學到：深度學習基礎 (1)

- 神經網路與梯度下降 (含基礎線性代數與微積分)
- 反向傳播法 (Back-propagation)

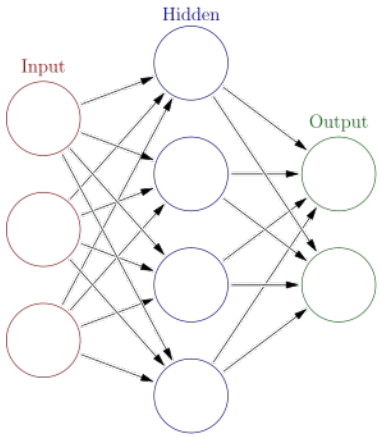
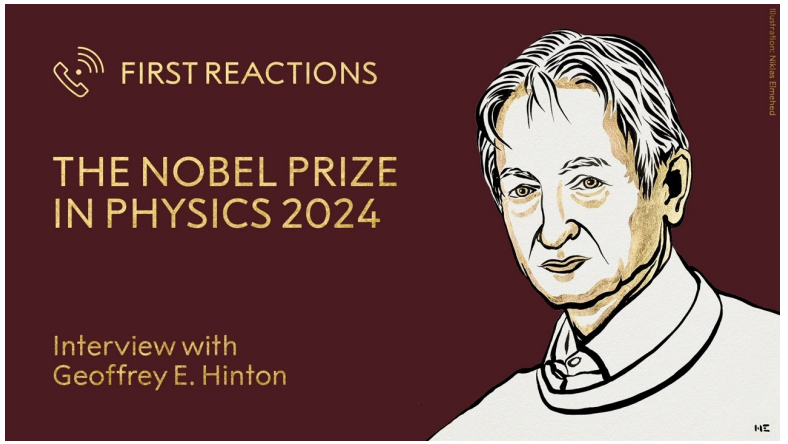


Figure source:
https://en.wikipedia.org/wiki/Neural_network_(machine_learning)



https://www.youtube.com/watch?v=-icD_KmvnnM

你會學到：深度學習基礎 (2)

- 最佳化方法：SGD, Momentum, RMSProp, Adam
- 常見損失函數介紹 (含基礎資訊理論與機率統計)

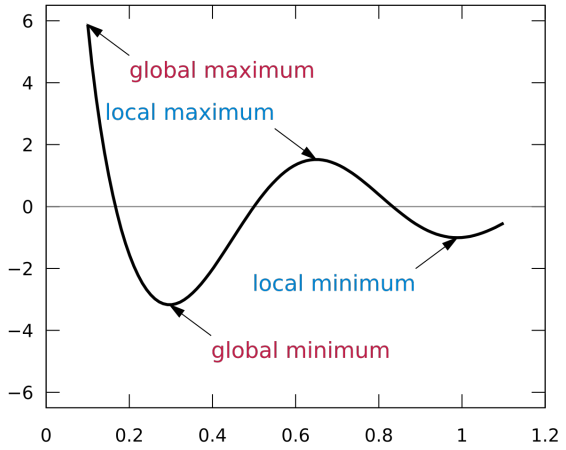


Figure source:
https://en.wikipedia.org/wiki/Maximum_and_minimum

你會學到：深度學習基礎 (3)

PyTorch

- Why PyTorch?

	PyTorch	TensorFlow
Developer	Meta	Google
Number of models on Hugging Face	216,295	14,415

https://huggingface.co/models

你會學到：深度學習應用 (1)

- (基礎) 電腦視覺
  - 卷積神經網路
  - Vision Transformers (ViT)
- 過擬合、正規化、以及模型訓練技巧

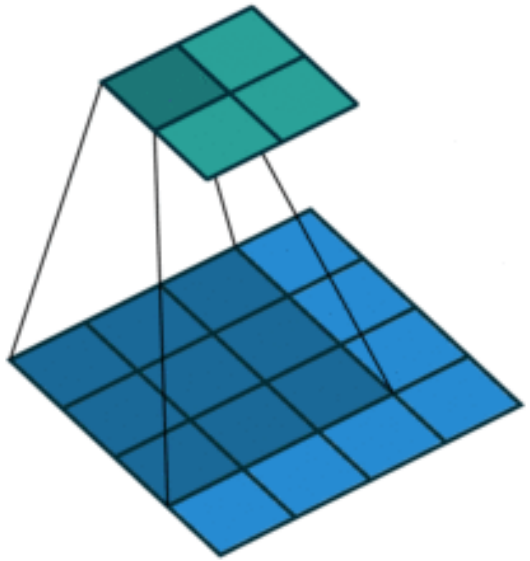
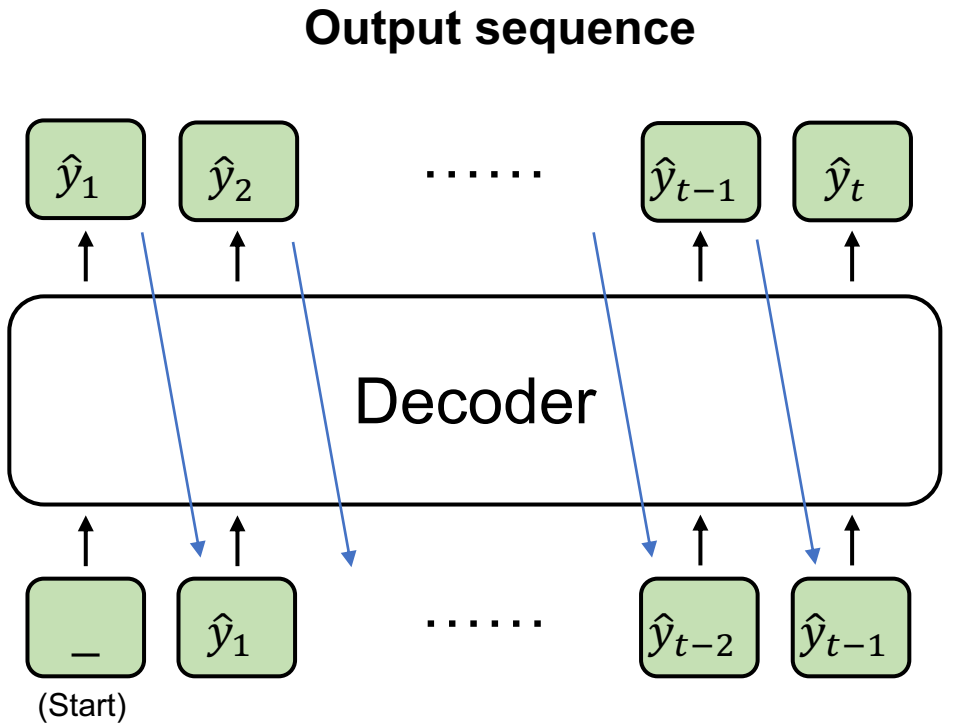


Figure source: https://hannibunny.github.io/mlbook/neuralnetworks/convolutionDemos.html

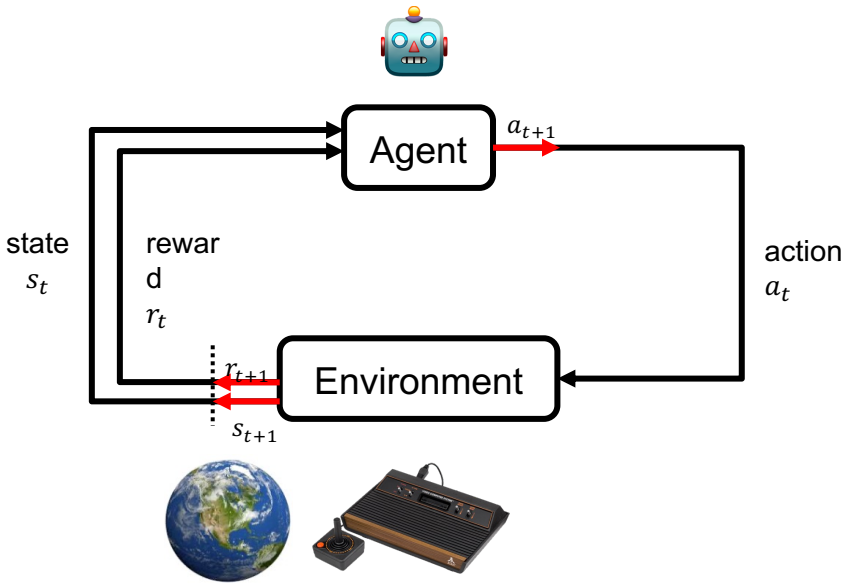
你會學到：深度學習應用 (2)

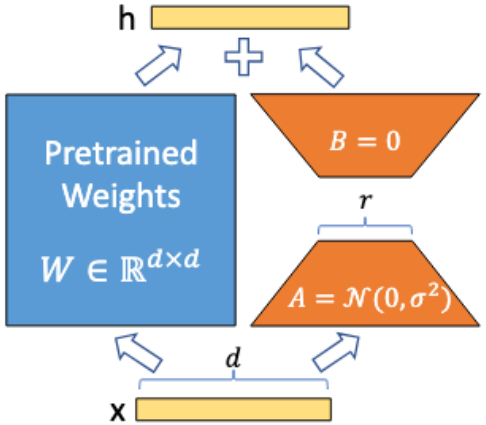
- 自然語言處理
  - RNN與序列建模
  - 自注意力機制模型
  - 自監督式學習與預訓練模型
  - 大型語言模型



你會學到：深度學習進階

- 大模型時代如何有效率訓練模型？
- 圖神經網路
- 強化學習

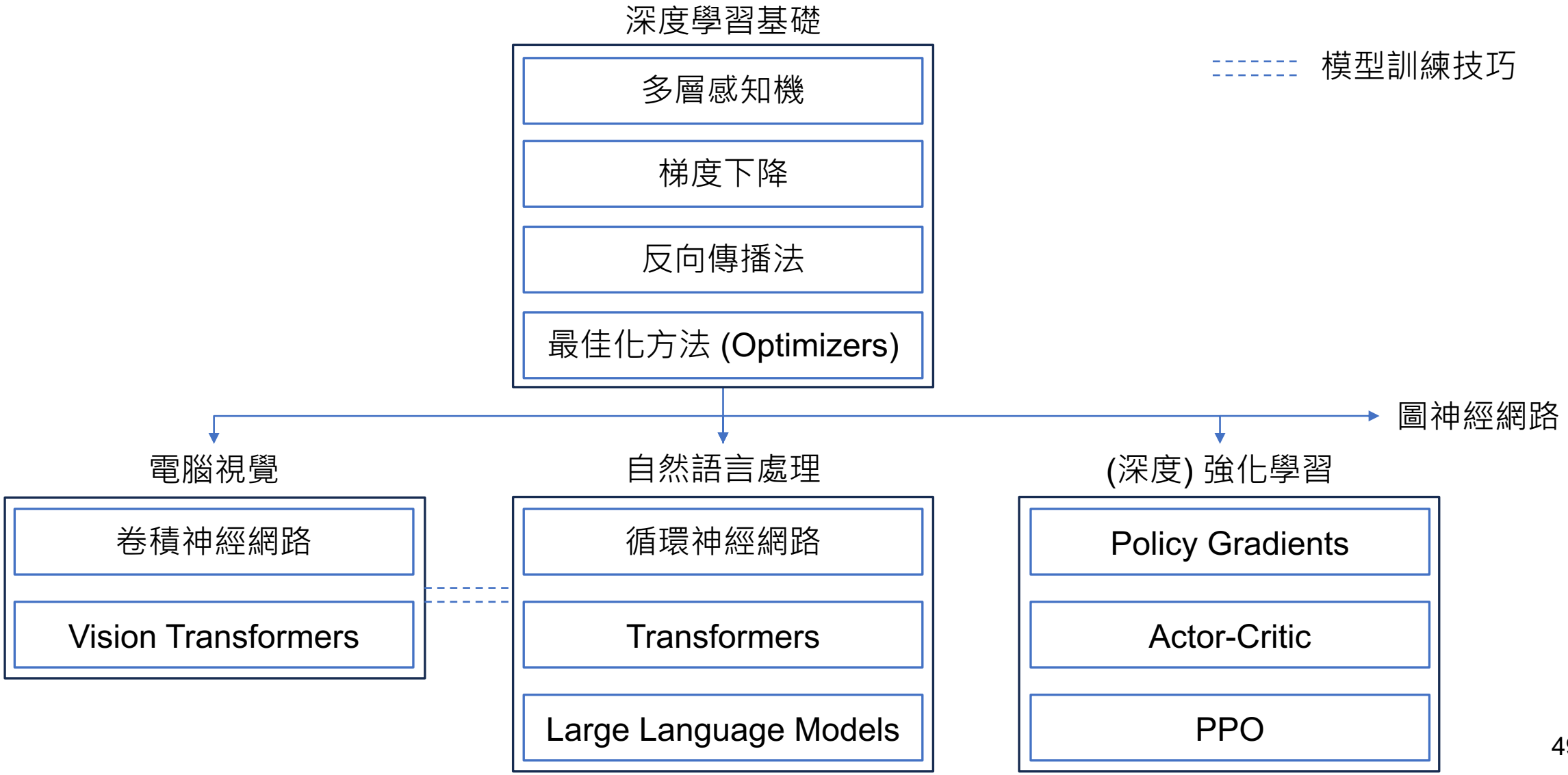




(Figure from the LoRA paper)

Hu, Edward J., et al. "LoRA: Low-Rank Adaptation of Large Language Models." ICLR 2022.

深度學習課程地圖


深度學習基礎
多層感知機
梯度下降
反向傳播法
最佳化方法 (Optimizers)
模型訓練技巧
圖神經網路
電腦視覺
卷積神經網路
Vision Transformers
自然語言處理
循環神經網路
Transformers
Large Language Models
(深度) 強化學習
Policy Gradients
Actor-Critic
PPO

深度學習課程的修課體感

- 並非一個完全連續的故事
- 而是可以針對不同領域有所跳躍

Scoring of this course

- Assignments: 7.5% x 4 times (30% in total)
- Quizzes: 1% x 10 times
- Mid-term exam: 20%
- Final project: 40%

About the Assignments

- 作業一定都會有範例程式碼 (Sample Code)
- 基本上是挖空程式題

作業可不可以用AI？

- 可以使用，但不得直接複製貼上
- 若使用 AI 生成程式碼，須以註解形式說明，例如：

```
class MLP(nn.Module):
   def __init__(self):
      # 以下幾行由gpt-5生成
      # 建構MLP的輸入層、中間層、以及輸出層
      self.layer_1 = nn.Linear(28*28, hidden_size1)
      self.layer_2 = nn.Linear(256, hidden_size2)
      self.cls_layer = nn.Linear(hidden_size2, 10)
      self.relu = nn.ReLU()
```

錯誤使用範例 (1)

請幫我寫一個影像辨識的程式碼，使用手寫字辨識資料集

已思考 25s

給你一個「可直接跑」的 PyTorch 版本，使用 MNIST 手寫數字資料集，完整包含資料下載、CNN 建模、訓練、驗證、儲存最佳模型。

安裝套件

bash 複製程式碼

```
pip install torch torchvision
```

train_mnist.py

python 複製程式碼

```
import os, time
import torch
import torch.nn as nn
import torch.optim as optim
from torchvision import datasets, transforms
from torch.utils.data import DataLoader

# ------ 超參數 ------
EPOCHS = 5
```

不建議的原因：

難以除錯，且下一步通常就是複製貼上執行，這樣的過程學不到東西

錯誤使用範例 (2)

CGU-DL-HW2-CIFAR10-cls.pdf
PDF

假設我現在已經完成程式碼作業，請幫我根據此作業需求，寫出一份報告

好的 ✅ 我已經讀過你提供的作業需求文件 。下面我會幫你草擬一份**完整的報告**（符合評分細則的 30% Report 部分），你可以依照自己的實作結果再補上實際數據。

Homework 2: Image Classification Report

Course: Deep Learning, Chang Gung University

Name / Student ID: （請填寫）

Date: 2025/03/31

不建議的原因：

這樣的過程同樣學不到東西

正確使用範例 (1)：實作細節討論

請問with torch.no_grad()跟model.eval()差在哪？

這兩個東西其實是不同層次的設定：

---

1. `with torch.no_grad()`
   - **作用**：關閉梯度計算。
   - **影響範圍**：包在 `with` 區塊中的所有運算。
   - **用途**：
     - 減少記憶體消耗（因為不用存計算圖）。
     - 加快推論速度（不用做反向傳播相關準備）。
   - **例子**：

python 複製程式碼

```python
with torch.no_grad():
    y = model(x)
```

建議的原因：幫自己釐清觀念

藉由 LLM 學到更細部的知識

正確使用範例 (2)

請問我的程式碼哪裡寫錯了？無法執行

```
python
def load_HF_model(model_name, num_labels=2)
    model = AutoModelForImageClassification
        model_name,
        num_labels=num_labels,  # Pneumonial
        ignore_mismatched_sizes=True, # htt
        problem_type="singel_label_classifi
    )
    return model
```

建議的原因：幫自己釐清觀念

加速程式除錯

你這段程式碼有一個小錯誤：

`problem_type="singel_label_classification"` 這裡拼錯了，正確應該是 `"single_label_classification"`。

修正後的程式碼如下：

期中考是什麼？

- 在電腦教室進行上機考
- 類型為 Debugging (暫定) 以及觀念題
  - 考試程式碼來自 Week4 - Week8
  - LLM 不太會
- 佔學期成績比例 20%

第九週改時間

- 原訂 10/29 13:10 – 16:00
- 改為 10/29 16:10 – 19:00
  - 該週 10/29 13:10 – 16:00 不上課

Final Project 是什麼？

- Final Project 是一種完整的訓練
- 你將自行從頭到尾做完一個題目，包含：
  - 資料前處理 (Data Pre-processing)
  - 模型訓練 (Model Training)
  - 模型評估 (Evaluations) -->了解問題並改進模型
- 每組人數待修課同學總數訂定，原則上1-3人/組

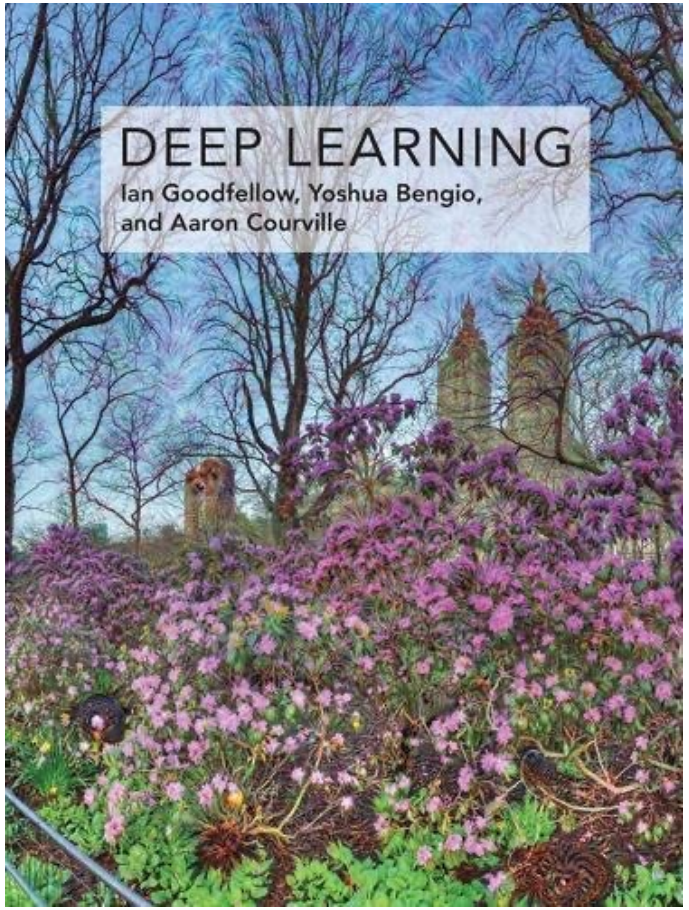
Syllabus

出作業日 /
上個作業截止日

考試

週	主題	程式教學	Note
1	深度學習介紹以及本課程綱要		
2	神經網路與梯度下降 (含基礎線性代數與微積分)		
3	反向傳播法		HW1
4	最佳化方法：SGD, Momentum, RMSProp, Adam	PyTorch Basics	
5	常見損失函數介紹 (含基礎資訊理論與機率統計)	PyTorch Modeling	
6	卷積神經網路	CV in PyTorch (1)	HW2
7	過擬合、正規化、模型訓練技巧、期末專案介紹	CV in PyTorch (2)	
8	自然語言處理：RNN與序列建模	NLP in PyTorch	
9	期中考		上機
10	自注意力機制模型		HW3
11	Vision Transformers (ViT)、自監督式學習與預訓練模型		
12	大型語言模型	🤗 Transformers	
13	大模型時代如何有效率訓練模型？	🤗 Quantization	HW4
14	圖神經網路		
15	強化學習		
16	小組實作成果報告		

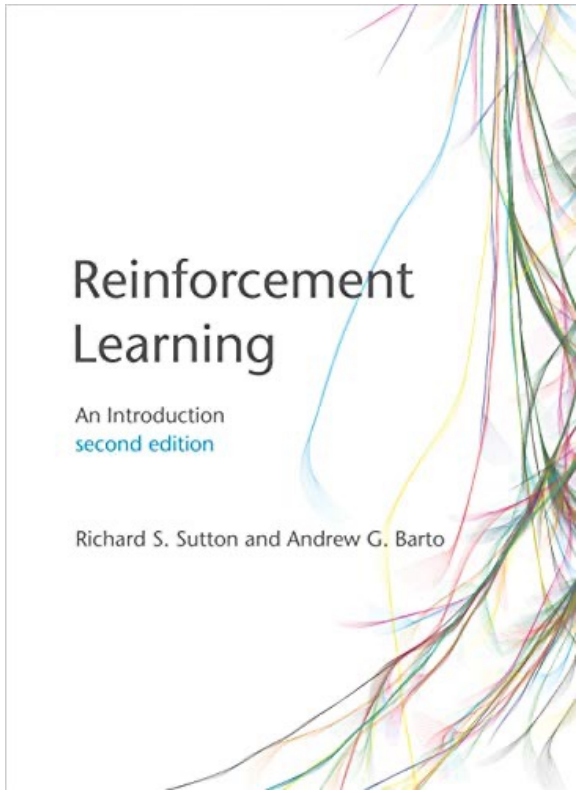
Textbook (1)



Ian Goodfellow, Yoshua Bengio, & Aaron Courville (2016). Deep Learning. MIT Press.

Free Online book:
https://www.deeplearningbook.org/

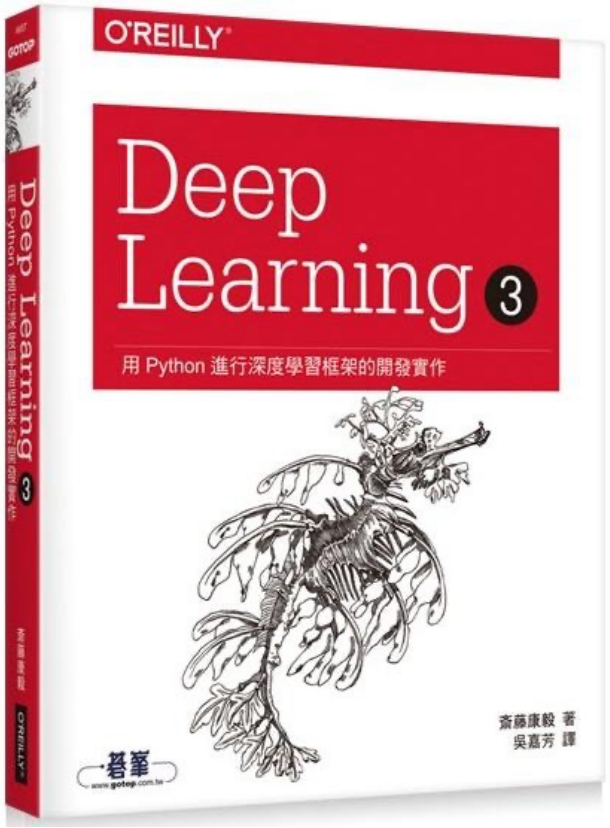
Textbook (2)



Richard S. Sutton and Andrew G. Barto (2018). Reinforcement Learning: An Introduction. MIT Press.

Free Online book: http://incompleteideas.net/book/the-book-2nd.html

Reference book



Deep Learning 3｜用Python進行深度學習框架的開發實作，歐萊禮 (ISBN: 9789865027346)

GitHub of this book:
https://github.com/oreilly-japan/deep-learning-from-scratch-3

Contact Information

Instructor: 林英嘉

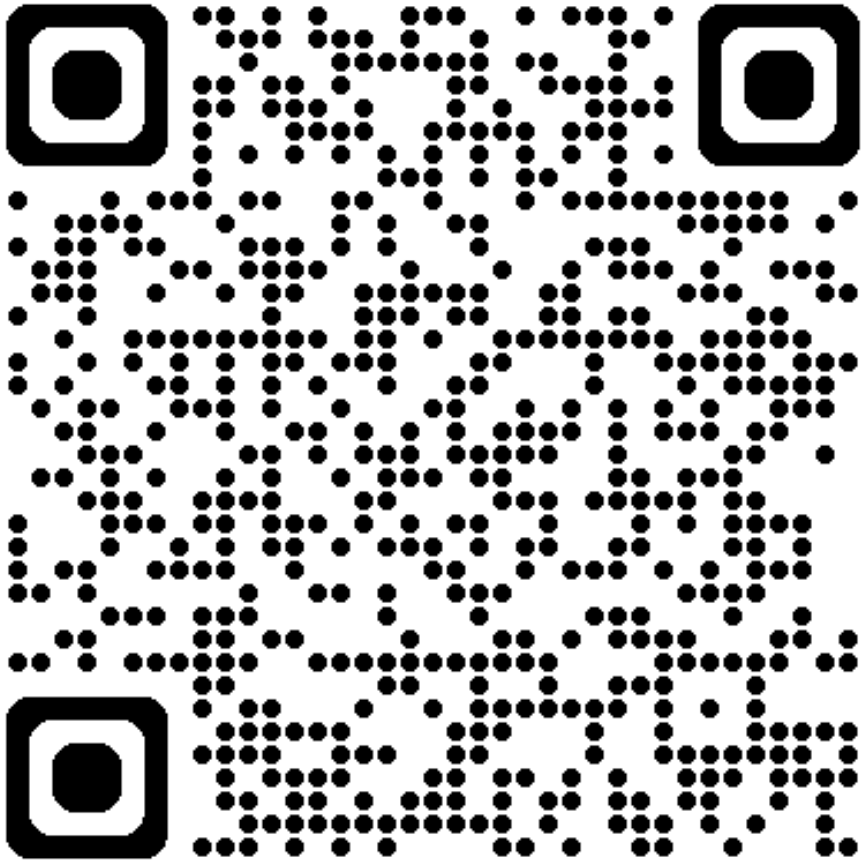
✉️ yjlin@cgu.edu.tw

🕘 Office Hour: 每週星期四 13:00 – 14:00 B1412室

寄信前請在主旨加註記 [深度學習]

Previous Course Page

https://github.com/mcps5601/CGUDL_2025_Spring



Thank you!

Instructor: 林英嘉

✉ yjlin@cgu.edu.tw