Introduction of this course

李宏毅
Hung-yi Lee

Welcome our TAs

TA 信箱：ntu.mldsta@gmail.com

TAs

劉浩然


Grand Theater 大劇院 GT
Rehearsal Room 3 排練室3
Costume Shop 服裝修補室
Laundry Room 洗衣房

劉記良

TAs

孫凡耕

吳思霖

TAs

宋易霖

賴至得

TAs

陳力維

陳佳佑

TAs

陳冠宇

劉達融 (大助教)

What are we going to learn?

本課程內容和《機器學習》沒有重疊

課程名稱解釋

機器學習
及其深層與結構化

Machine Learning
and having it Deep and Structured

Method

Deep Learning 可以解決一切 …

遇到問題，~~用 c4~~ 就對了！

用 deep learning “硬train一發”

萬事皆可 train ……

Practice of Deep Learning

- Previous machine learning developers
  - Carefully design your algorithm
  - Theoretically know its performance
- Deep learning
  - Try first
  - Many results contradict our intuition
  - Find some reasons to explain what we observed
  - More like chemistry
  - Or even worse ……

Practice of Deep Learning

- Even a simple model can be hard to train ……

Practice of Deep Learning

- Interesting facts ……

From: Boris
To: Ali

"On Friday, someone on another team changed the default rounding mode of some Tensorflow internals (from "truncate toward 0" to "round to even").

Our training broke. Our error rate went from <25% error to ~99.97% error (on a standard 0-1 binary loss)."

Ali Rahimi, Test of Time Award, NIPS 2017

Theory of Deep Learning

Theory 1: Expressiveness

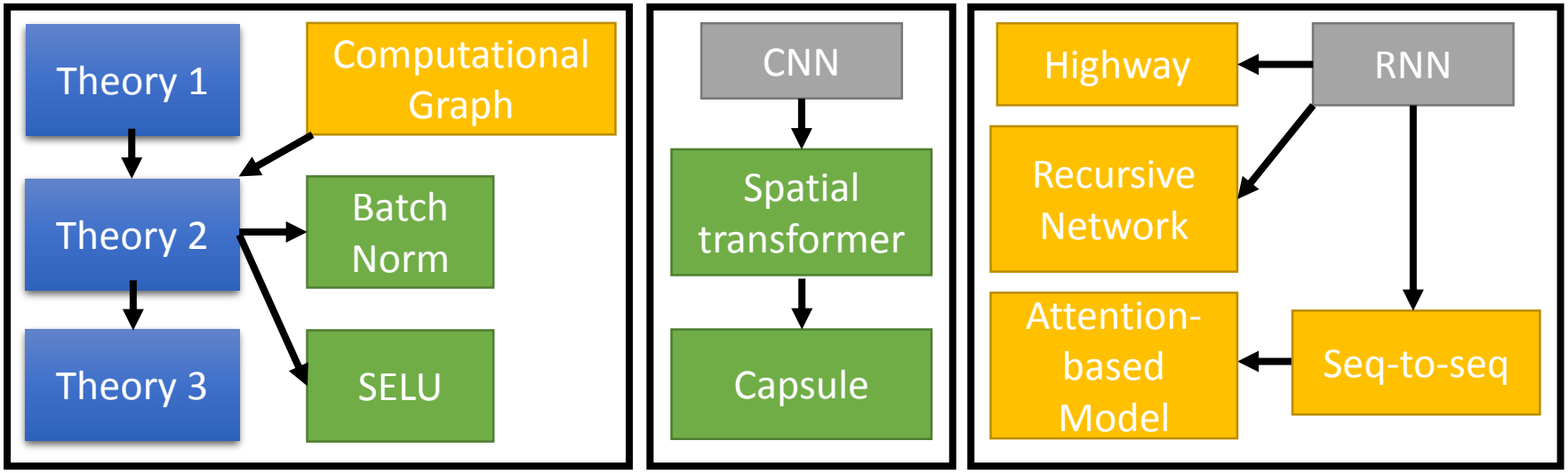
- A network structure defines a function set
- Is deep better than shallow?

Theory 2: Optimization

- How can we optimize by gradient descent?
- There are local minima ……

Theory 3: Generalization

- Why deep network does not overfit?
- Although it can ……


Theory 1
Computational Graph
Theory 2
Batch Norm
Theory 3
SELU
CNN
Spatial transformer
Capsule
Highway
RNN
Recursive Network
Attention-based Model
Seq-to-seq

課程名稱解釋

機器學習
及其深層與結構化

Machine Learning
and having it Deep and Structured

Task

Structured Learning

Machine learning is to find a function f

$$f : X \rightarrow Y$$

Regression: output a scalar

Classification: output a "class" (one-hot vector)

Class 1 Class 2 Class 3

Structured Learning/Prediction: output a sequence, a matrix, a graph, a tree ……

Output is composed of components with dependency

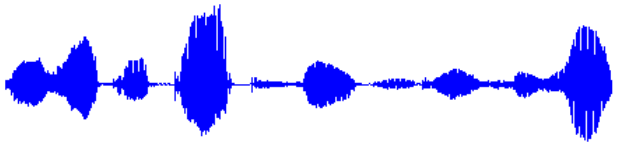
Output Sequence $f: X \rightarrow Y$

Machine Translation

X：“機器學習及其深層與結構化”
(sentence of language 1)

Y：“Machine learning and having it deep and structured”
(sentence of language 2)

Speech Recognition

X： 
(speech)

Y：感謝大家來上課”
(transcription)

Chat-bot

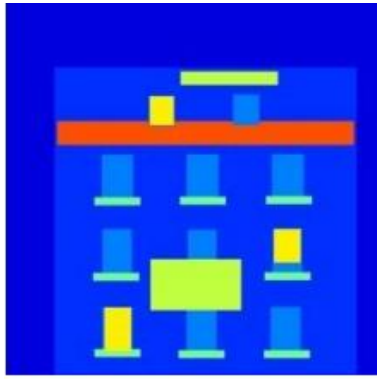
X： “How are you?”
(what a user says)

Y： “I’m fine.”
(response of machine)

Output Matrix

$$f : X \rightarrow Y$$

Image to Image

X : 

Y : 

Colorization:

Ref: https://arxiv.org/pdf/1611.07004v1.pdf

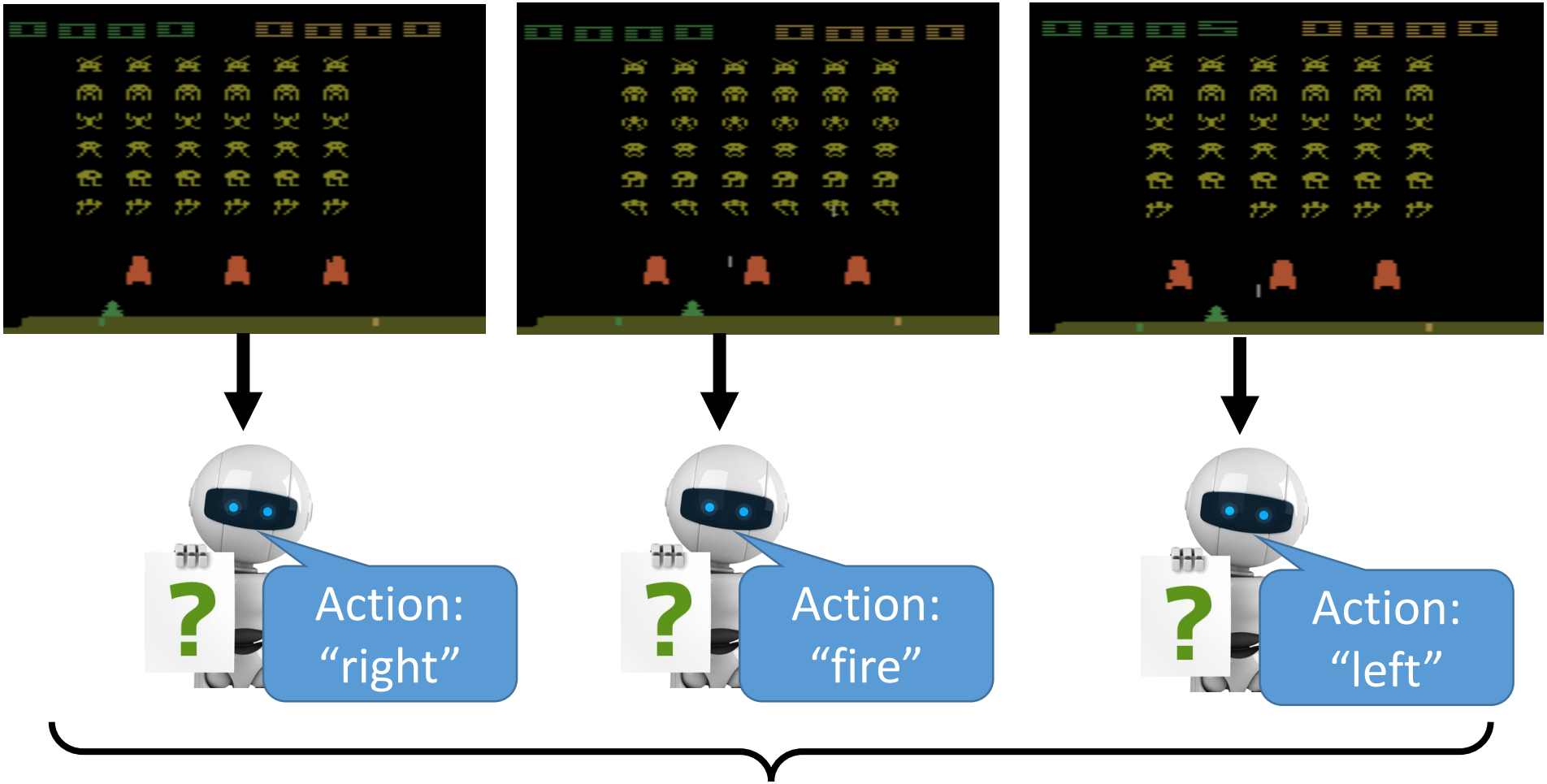
Text to Image

X : "this white and yellow flower have thin white petals and a round yellow stamen"

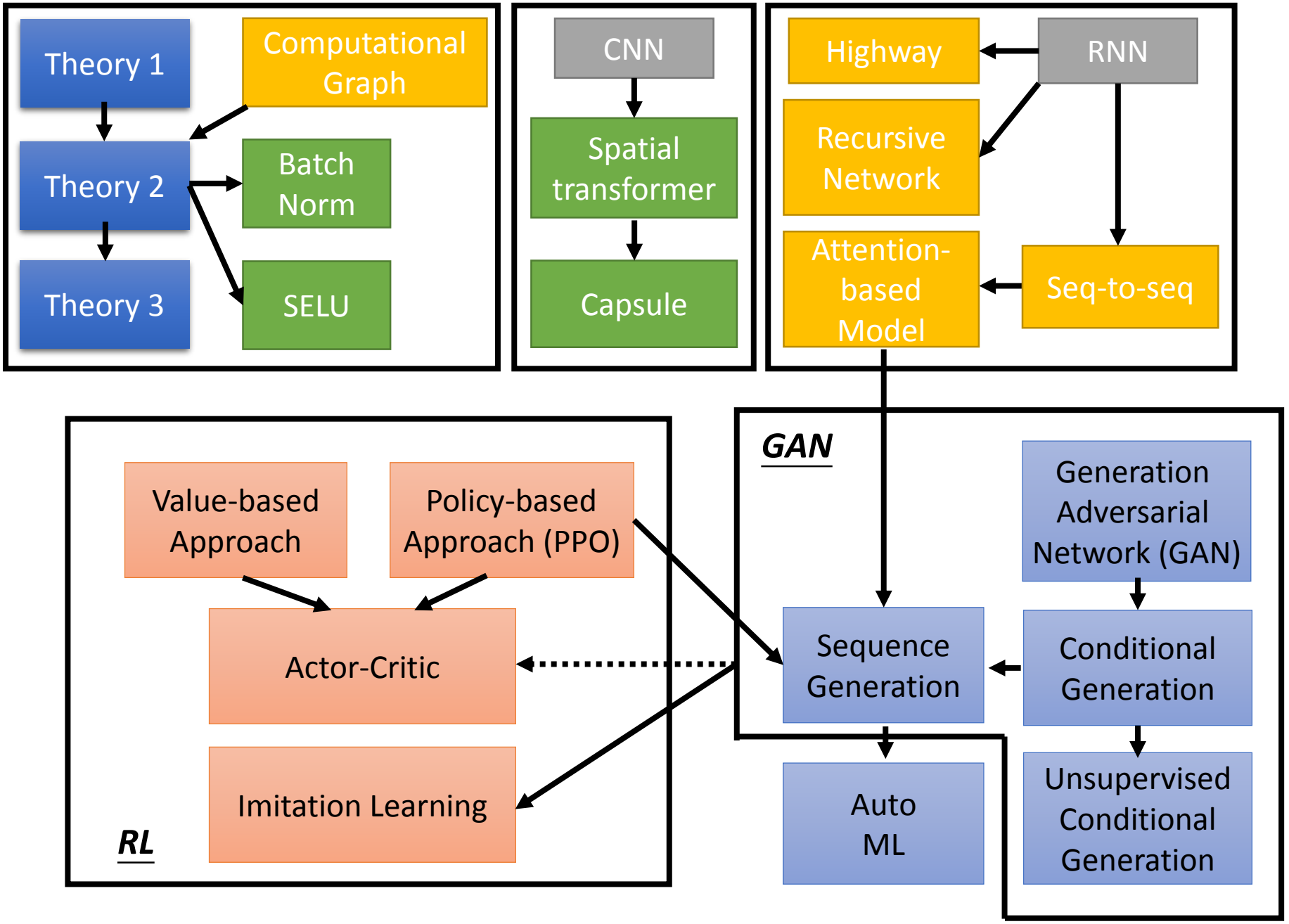
Y : 

ref: https://arxiv.org/pdf/1605.05396.pdf

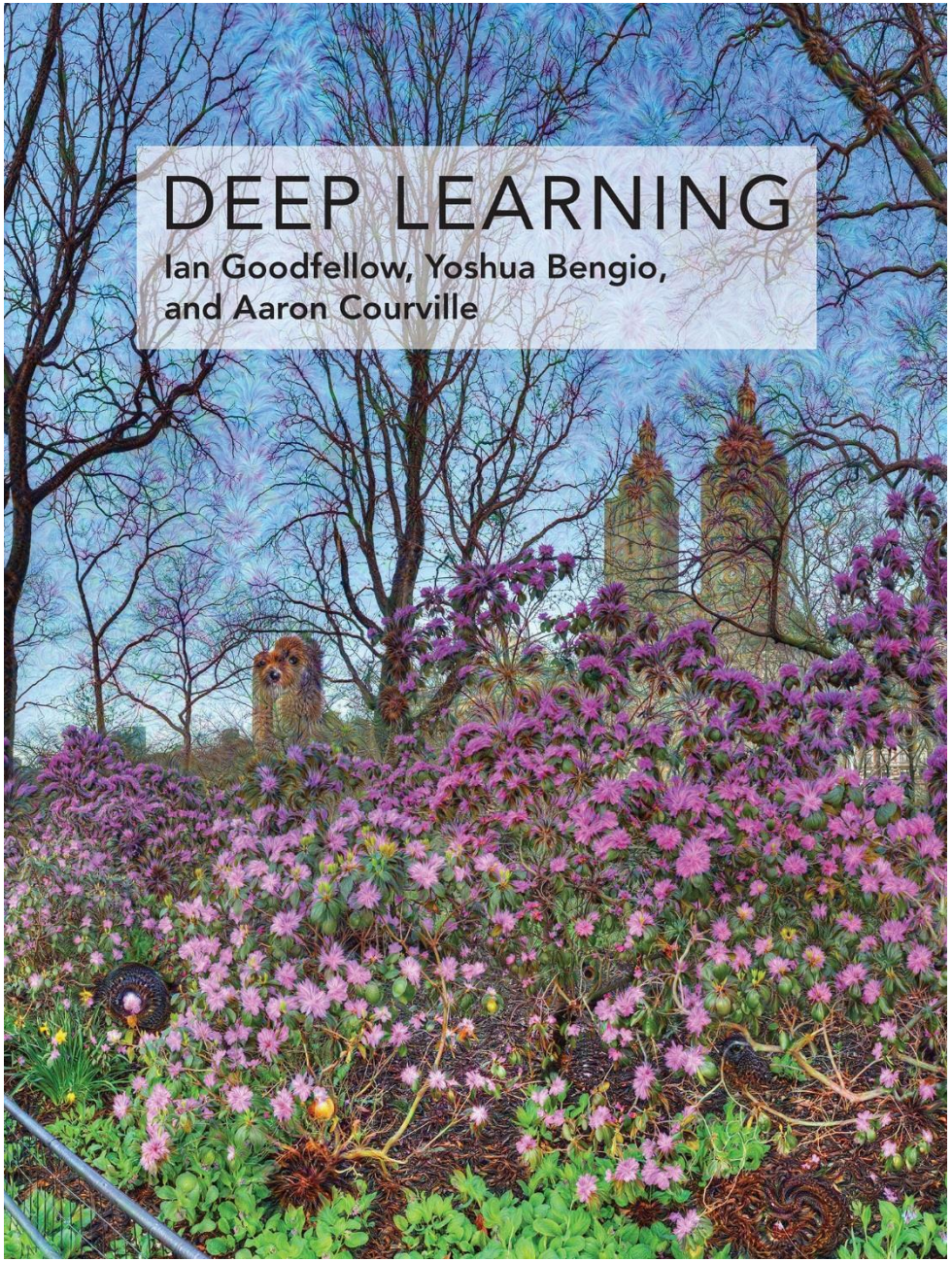
Reinforcement Learning



Regression,
Classification


Theory 1
Computational Graph
Theory 2
Batch Norm
Theory 3
SELU
CNN
Spatial transformer
Capsule
Highway
RNN
Recursive Network
Attention-based Model
Seq-to-seq
GAN
Value-based Approach
Policy-based Approach (PPO)
Actor-Critic
Imitation Learning
RL
Generation Adversarial Network (GAN)
Sequence Generation
Conditional Generation
Auto ML
Unsupervised Conditional Generation

參考書籍



Original image:
http://www.danielambrosi.com/Grand-Format-Collection/i-jbhqVhS/A

http://www.deeplearningbook.org/

Schedule

四次作業

- HW1：深度學習流言終結者
  - 1-1: Deep is better than shallow?
  - 1-2: Is local minima an issue?
  - 1-3: Is deep learning generalizable?
- HW2：Seq-to-seq model
  - 2-1: Video caption generation
  - 2-2: Chat-bot (option)

四次作業

- HW3: Generative Adversarial Network (GAN)
  - 3-1: Generation
  - 3-2: Conditional Generation
  - 3-3: Unsupervised Conditional Generation (option)
- HW4: Reinforcement learning
  - 4-1: Policy gradient
  - 4-2: Q-learning
  - 4-3: Actor-critic

週次	日期	單元主題
第1週	3/02	Course Introduction
第2週	3/09	Theory I: Is deep structure better than shallow one? (HW1-1 released)
第3週	3/16	Theory II: Why local minima is not a problem? (HW1-2 released)
第4週	3/23	Theory III: Why large network does not overfit (even though theoretically it should)? (HW1-3 released)
第5週	3/30	Sequence-to-sequence Model (HW2-1 released)
第6週	4/06 (放假)	按學校行事曆放假一週 (HW1 due)
第7週	4/13	Special Network Structures (HW2-2 released)
第8週	4/20	期中考前一週放假
第9週	4/27 (期中考週)	作業一表現優異同學上台分享
第10週	5/04	Deep Generative Model (HW2 due, HW3-1 released)
第11週	5/11	Deep Generative Model (HW3-2 released)
第12週	5/18	Deep Generative Model (HW3-3 released)
第13週	5/25	Deep Generative Model
第14週	6/01	Deep Reinforcement Learning (HW3 due, HW4-1 released)
第15週	6/08	Deep Reinforcement Learning (HW4-2 released)
第16週	6/15	Deep Reinforcement Learning (HW4-3 released)
第17週	6/22	期末考前一週放假
第18週	6/29 (期末考週)	作業二、三、四表現優異同學上台分享 (HW4 due)

https://ceiba.ntu.edu.tw/modules/index.php?csn=8e8d96&default_fun=syllabus

Policy

評量方式

- 不點名、不考試
- 作業一 (25%)、作業二 (25%)、作業三 (25%)、作業四 (25%)

成績是相對的

成績是相對的

成績是相對的

組隊

- 每組 1 ~ 3 人，其中一人為組長
- 以組為單位進行所有作業
- 隊伍登記方法之後和作業一一起說明
- 組內互評：學期結束前會有組內互評，會影響成績

上課方式

- 上課投影片和錄音會放到李宏毅的個人網頁上
  - 李宏毅的個人網頁：http://speech.ee.ntu.edu.tw/~tlkagk/courses_MLSD18.html
- 社團: “Machine learning and having it deep and structured (2018 spring)”
  - https://www.facebook.com/groups/907901649378100/
  - 有問題歡迎直接在 FB社團上發問
  - 如果有同學知道答案歡迎回答
- 有任何和機器學習相關的想法都可以在 FB社團上發言

提醒

- 深度學習必要之惡：訓練過程需要時間、調參數需要時間 ……
  - 作業早點開始
    - 死線前爆氣沒有什麼幫助
  - 健全的心靈
    - 試著調適等待過程的焦慮
  - 運算資源
    - 會有 MS Azure 的贊助，但是自備運算資源更好
- 請勿有作弊行為 (例如：抄襲同學的作業和程式、抄襲前人的作業和程式等等)
  - 請遵守各作業的規定
  - **如有規定未盡之處，則依照一般社會常識**

需要的基礎能力和知識

- 本課程的定位為機器學習進階課程
- 程式能力：能夠使用某一個深度學習框架 (e.g. Tensorflow, pyTorch)
  - 使用 Keras 無法完成所有的作業
  - 本學期不會教深度學習框架的使用，請自學
    - Tensorflow
      - https://fgc.stpi.narl.org.tw/activity/videoDetail/4b1141305d9cd231015d9d07dbe1002a
      - https://fgc.stpi.narl.org.tw/activity/videoDetail/4b1141305d9cd231015d9d0852c5002b
      - https://fgc.stpi.narl.org.tw/activity/videoDetail/4b1141305d9cd231015d9d08fb62002d
    - pyTorch
      - https://fgc.stpi.narl.org.tw/activity/videoDetail/4b1141305d9cd231015d9d0992ef0030

需要的基礎能力和知識

- 基礎知識：希望聽課同學具備深度學習的基礎知識
- 《機器學習》錄影
  - DNN: https://www.youtube.com/watch?v=Dr-WRlEFefw
  - Tips for DNN: https://www.youtube.com/watch?v=xki61j7z-30
  - CNN: https://www.youtube.com/watch?v=FrKWiRv254g
  - RNN (Part 1): https://www.youtube.com/watch?v=xCGidAeyS4M
  - RNN (Part 2): https://www.youtube.com/watch?v=xCGidAeyS4M
  - Why Deep: https://www.youtube.com/watch?v=XsC9byQkUH8
  - Auto-encoder: https://www.youtube.com/watch?v=Tk5B4seA-AU
  - Deep generative model (Part 1): https://www.youtube.com/watch?v=YNUek8ioAJk
  - Deep generative model (Part 2): https://www.youtube.com/watch?v=8zomhgKrsmQ
  - Reinforcement Learning: https://www.youtube.com/watch?v=W8XF3ME8G2I

加簽

- 等一下助教會公告作業 0
  - 作業零其實是《機器學習》這門課的作業
  - 變相擋修《機器學習》
  - 可以用任何機器學習方法完成，不限深度學習，只要達到要求的正確率就行
  - 助教不會改作業零的程式
  - 以個人為單位完成
- 本課程預期修課人數為 40 ~ 80 人 ，上限為 120 人
  - 如果完成作業零人數超過上限，則按照完成作業的時間排序
  - 完成作業 0 後，授權碼取得方式另行公告