Introduction of this course

李宏毅

Hung-yi Lee

Welcome our TAs

林資偉

盧柏儒

方為

宋昀蓁

賴顗安

沈家豪

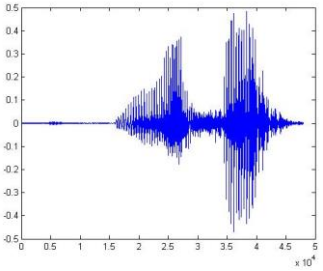
林賢進

敖家維

吳柏瑜

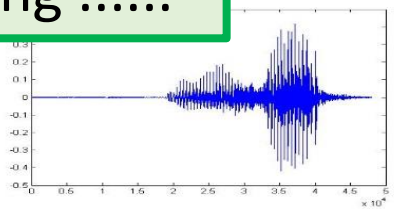
What are we going to learn?

What is Machine Learning?

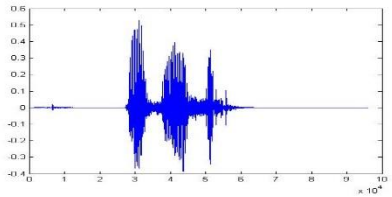


You said “Hello”

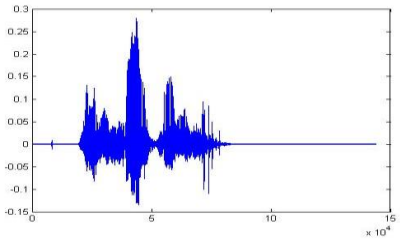
Learning ……



“Hi”



“How are you”



“Good bye”

A large amount of audio data

You write the program for learning.

What is Machine Learning?

Learning .....

This is “cat”

You write the program for learning.

“monkey”

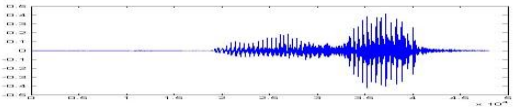
“cat”

“dog”

A large amount of images

Machine Learning ≈ Looking for a Function

- Speech Recognition

$f($  $)=$ “How are you”

- Image Recognition

$f($  $)=$ “Cat”

- Playing Go

$f($  $)=$ “5-5” (next move)

- Dialogue System

$f($ “Hi” $)=$ “Hello”

(what the user said) (system response)

Framework

Image Recognition:

$f($  $)=$ “cat”

A set of function

Model

$f_1, f_2 \cdots$

$f_1($  $)=$ “cat”

$f_2($  $)=$ “money”

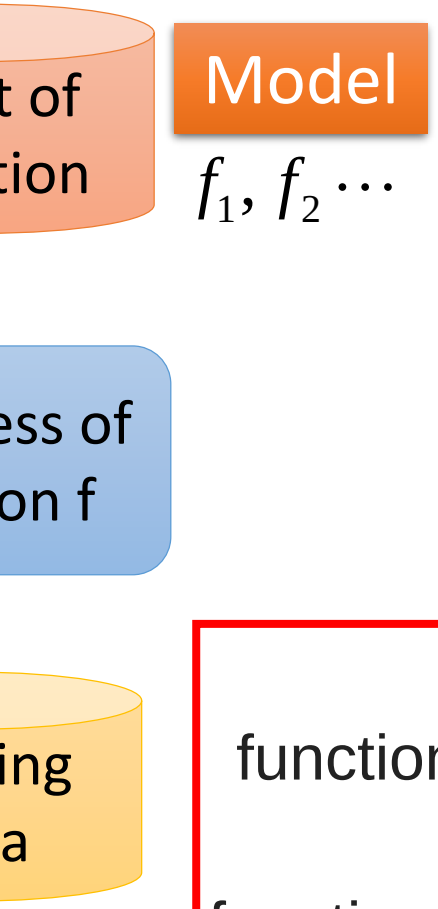
$f_1($  $)=$ “dog”

$f_2($  $)=$ “snake”

Framework

Image Recognition:

$f($  $) =$ "cat"



$f_1($  $) =$ "cat"

$f_2($  $) =$ "money"

Better!

$f_1($  $) =$ "dog"

$f_2($  $) =$ "snake"

function input:   

function output: "monkey" "cat" "dog"

Framework

Image Recognition:

$f($  $) =$ "cat"

Training

Testing

Step 1

A set of function

Model

$f_1, f_2 \cdots$

Step 2

Goodness of function f

Step 3

Pick the "Best" Function

f^*

Using f^*

"cat"

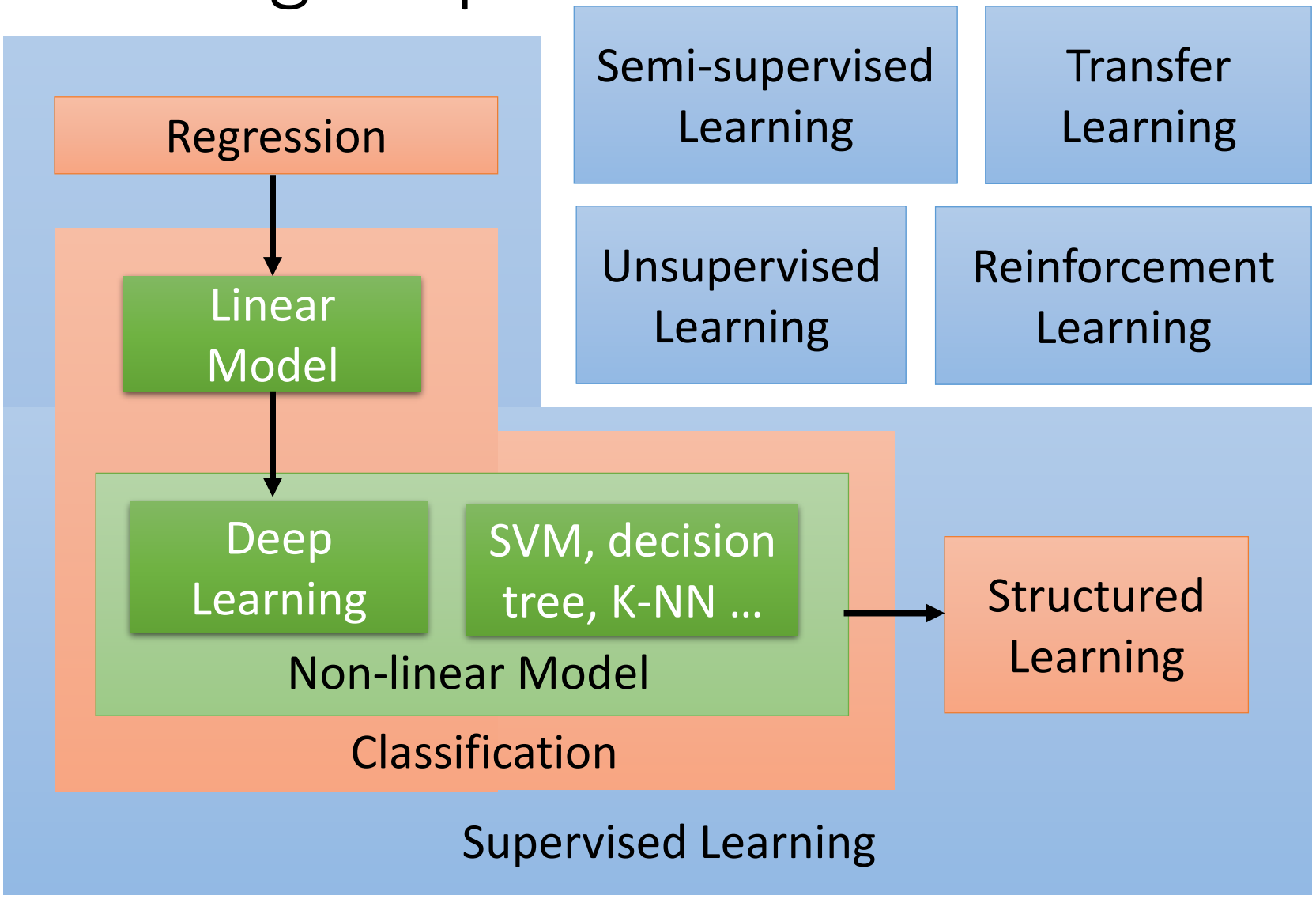
Training Data

"monkey"

"cat"

"dog"

Learning Map

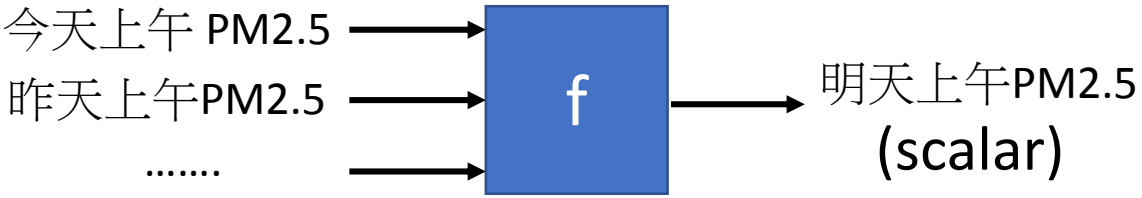


Learning Map

Regression

The output of the target function f is "scalar".

預測 PM2.5



HW1

Training Data:

Input:

9/01 上午 PM2.5 = 63　9/02 上午 PM2.5 = 65

Output:

9/03 上午 PM2.5 = 100

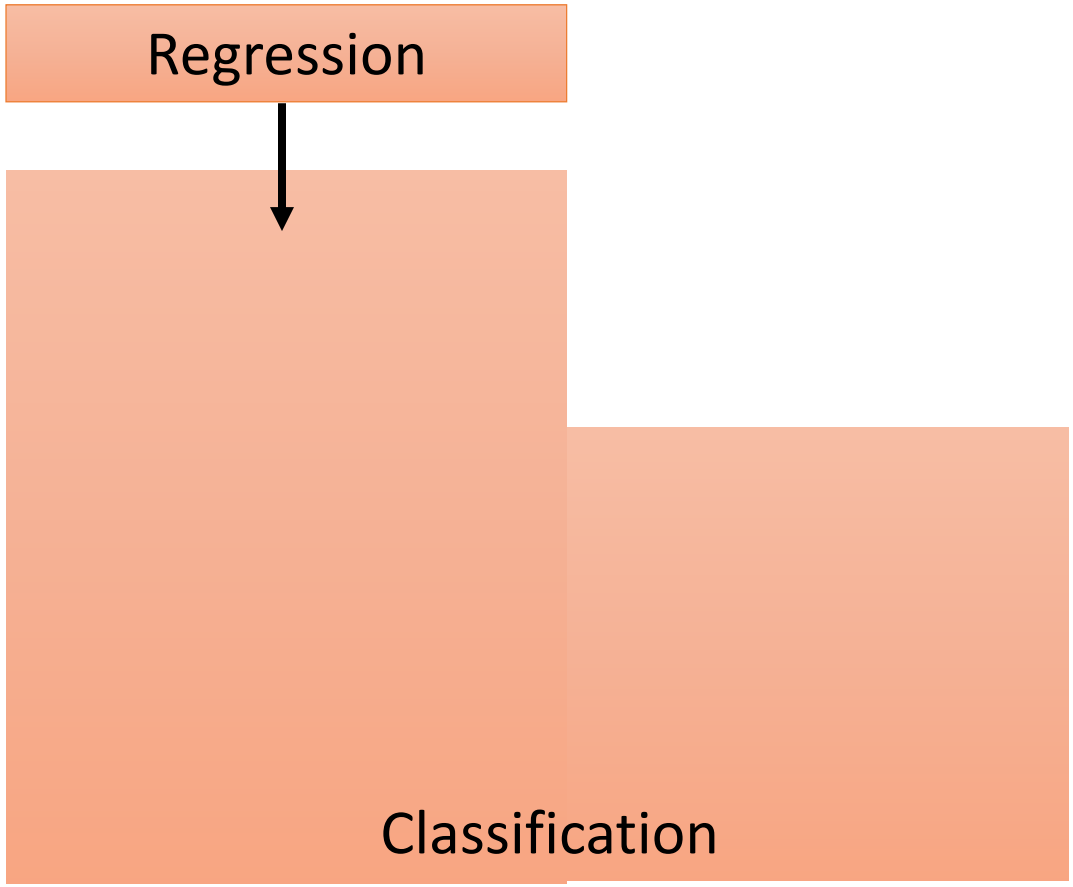
Input:

9/12 上午 PM2.5 = 30　9/13 上午 PM2.5 = 25

Output:

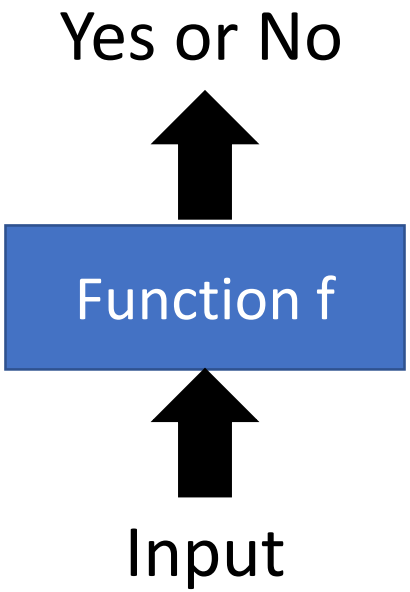
9/14 上午 PM2.5 = 20

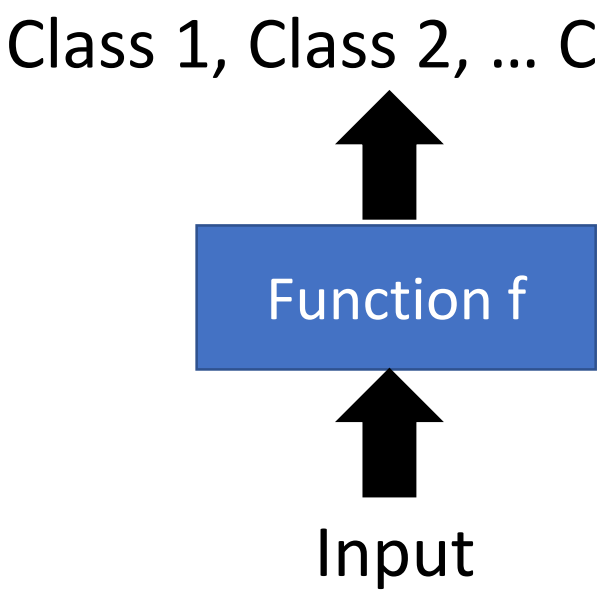
Learning Map



Classification

- Binary Classification
- Multi-class Classification





Binary Classification

Spam filtering

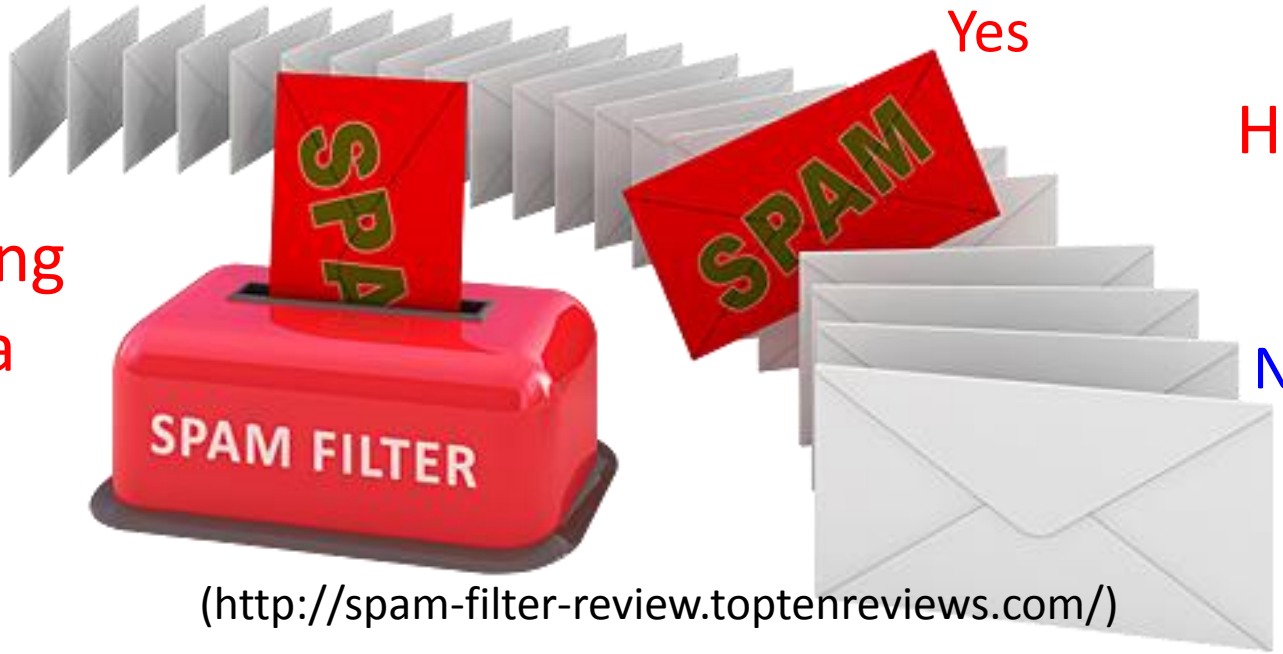
Function

Yes/No

Yes

HW2

Training Data



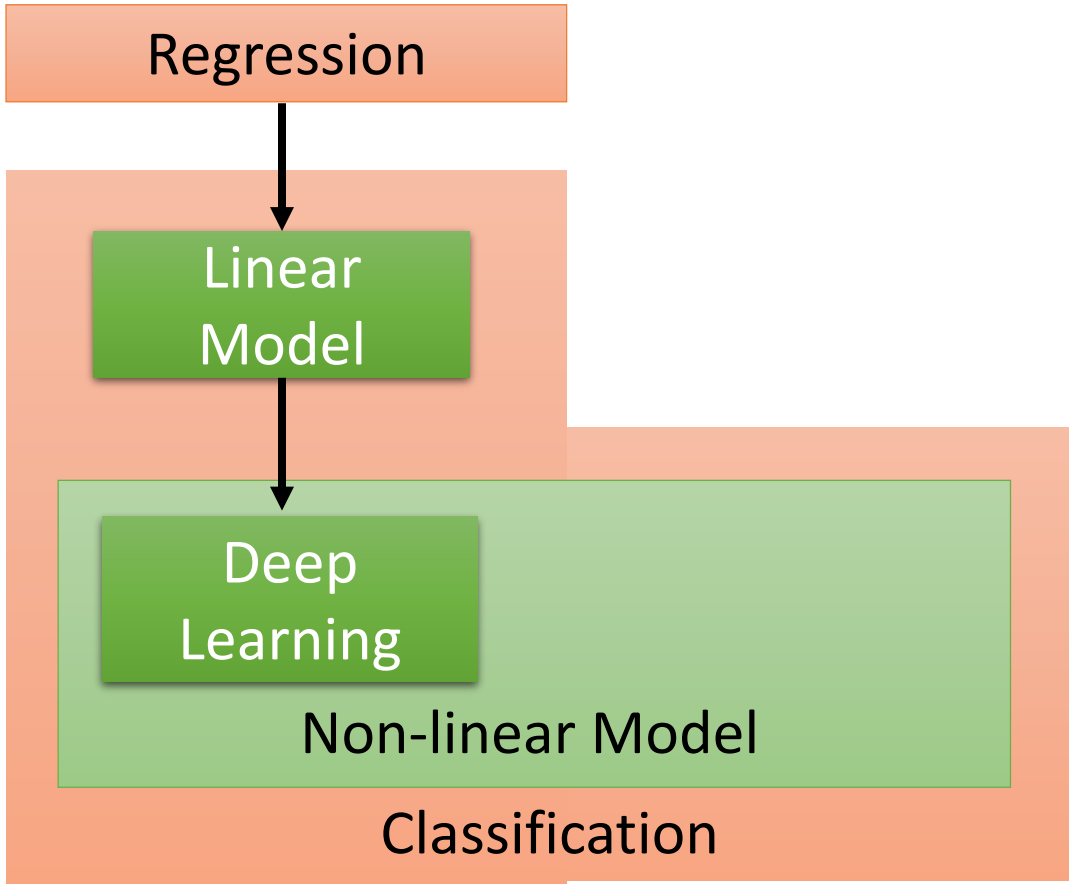
No

(http://spam-filter-review.toptenreviews.com/)

Multi-class Classification

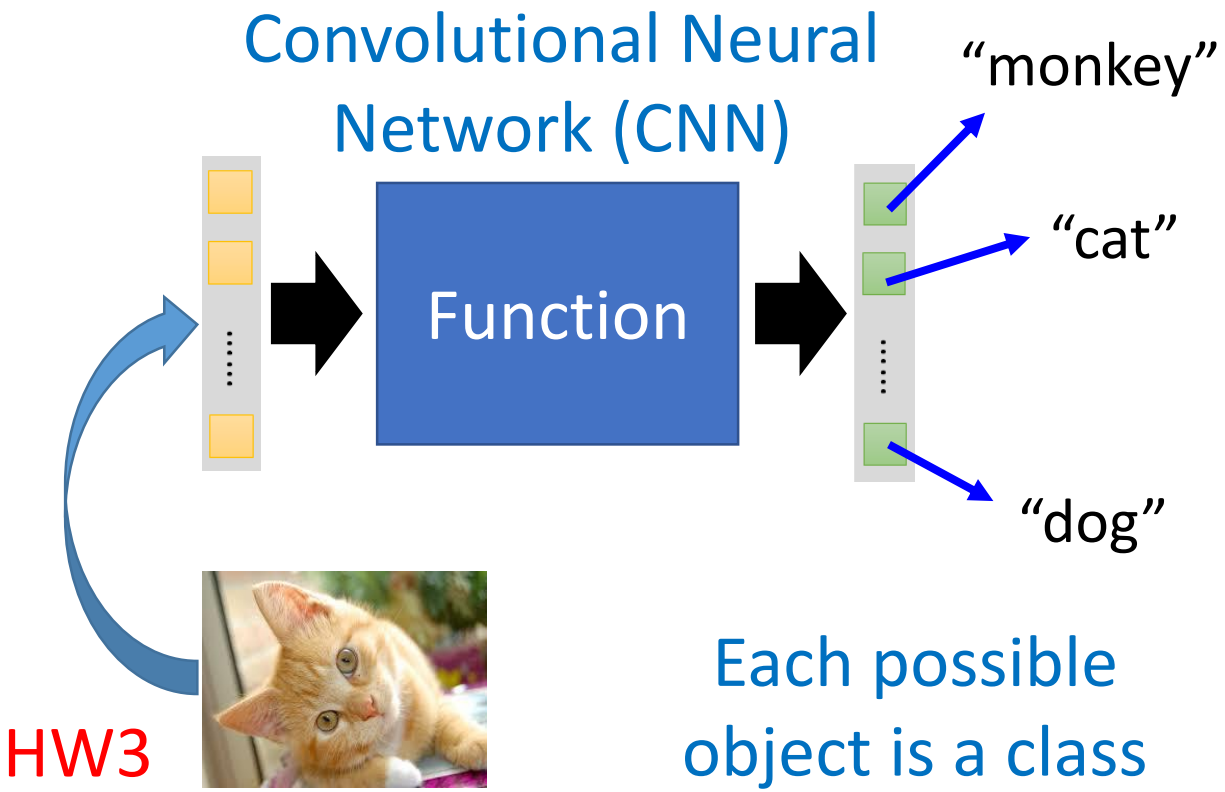
Document Classification

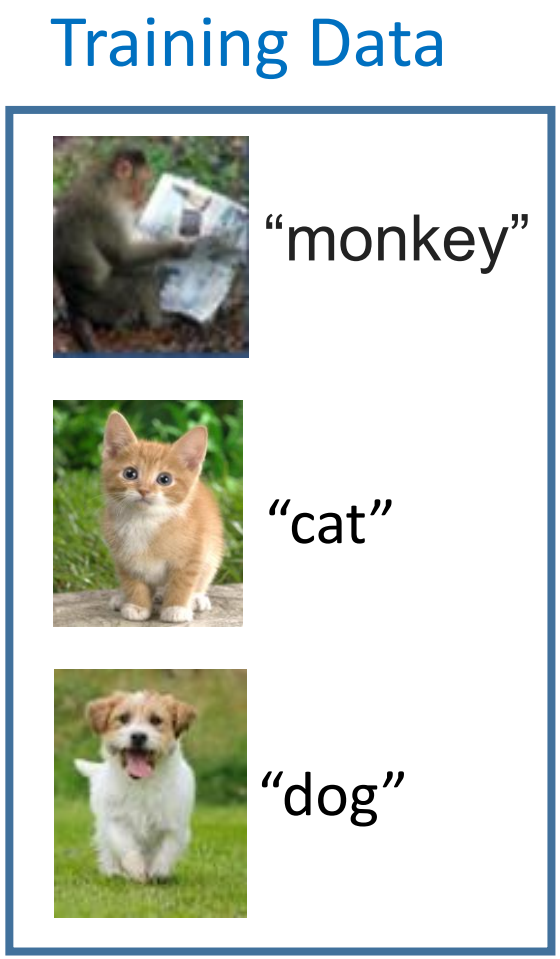
Learning Map



Classification - Deep Learning

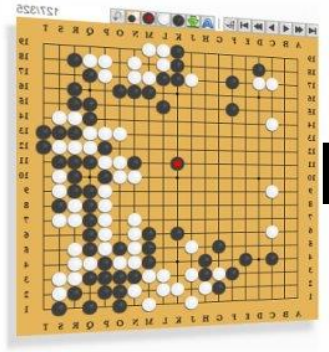
- Image Recognition





Classification - Deep Learning

- Playing GO



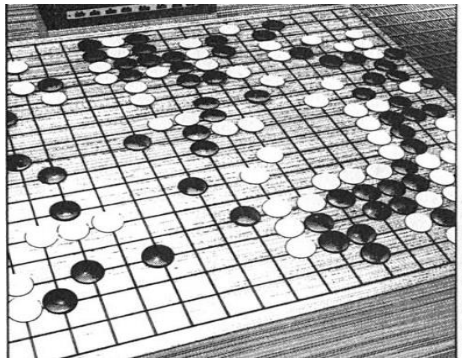
Function

Next move

Each position is a class

(19 x 19 classes)

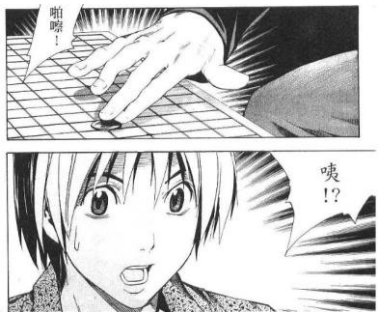
Training Data



一堆棋譜

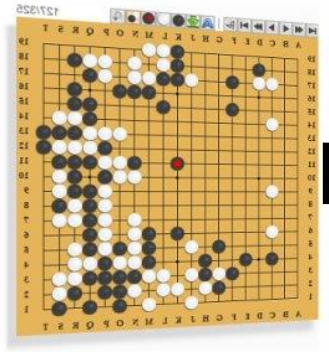
進藤光 v.s. 社清春

黑: 5之五 ⟶ 白: 天元 ⟶ 黑: 五之5



Classification - Deep Learning

- Playing GO



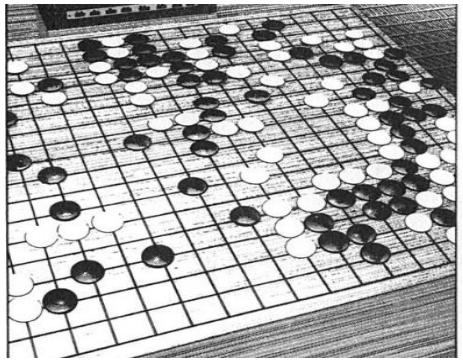
Function

Next move

Each position is a class

(19 x 19 classes)

Training Data



一堆棋譜

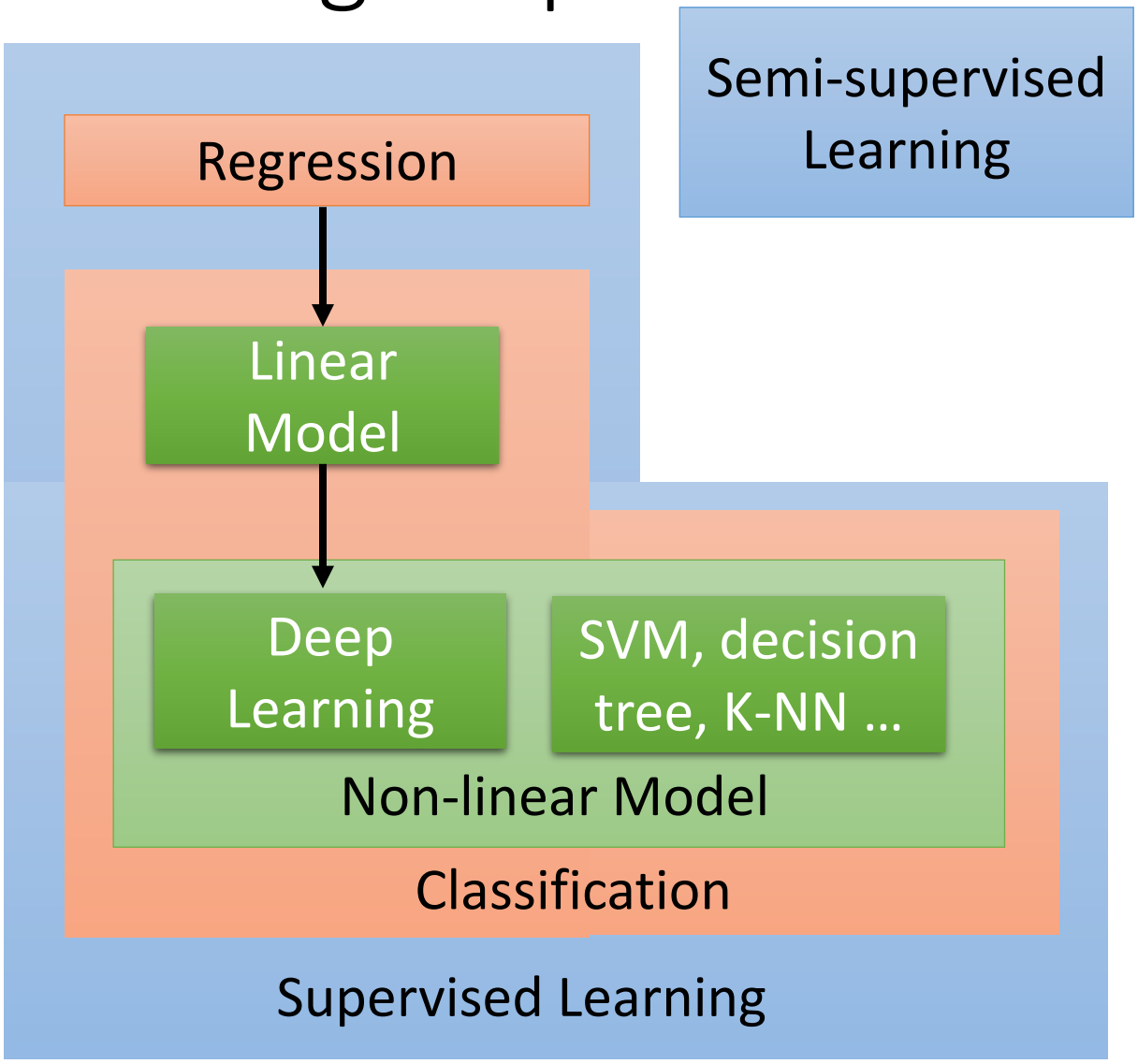
進藤光 v.s. 社清春

黑: 5之五 → 白: 天元 → 黑: 五之5

Input:
黑: 5之五 → Output:
天元

Input:
黑: 5之五、白: 天元 → Output:
五之5

Learning Map



Training Data:

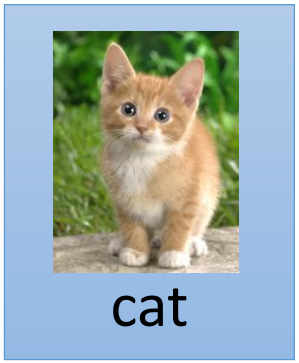
Input/output pair of target function

Function output = label

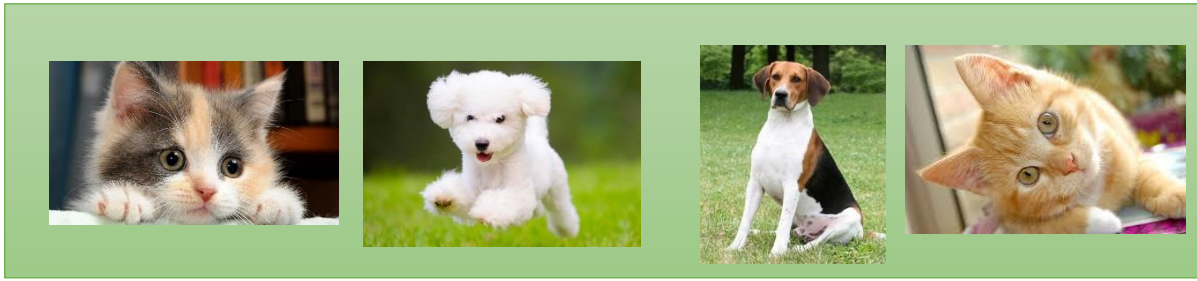
Semi-supervised Learning

For example, recognizing cats and dogs

Labelled data

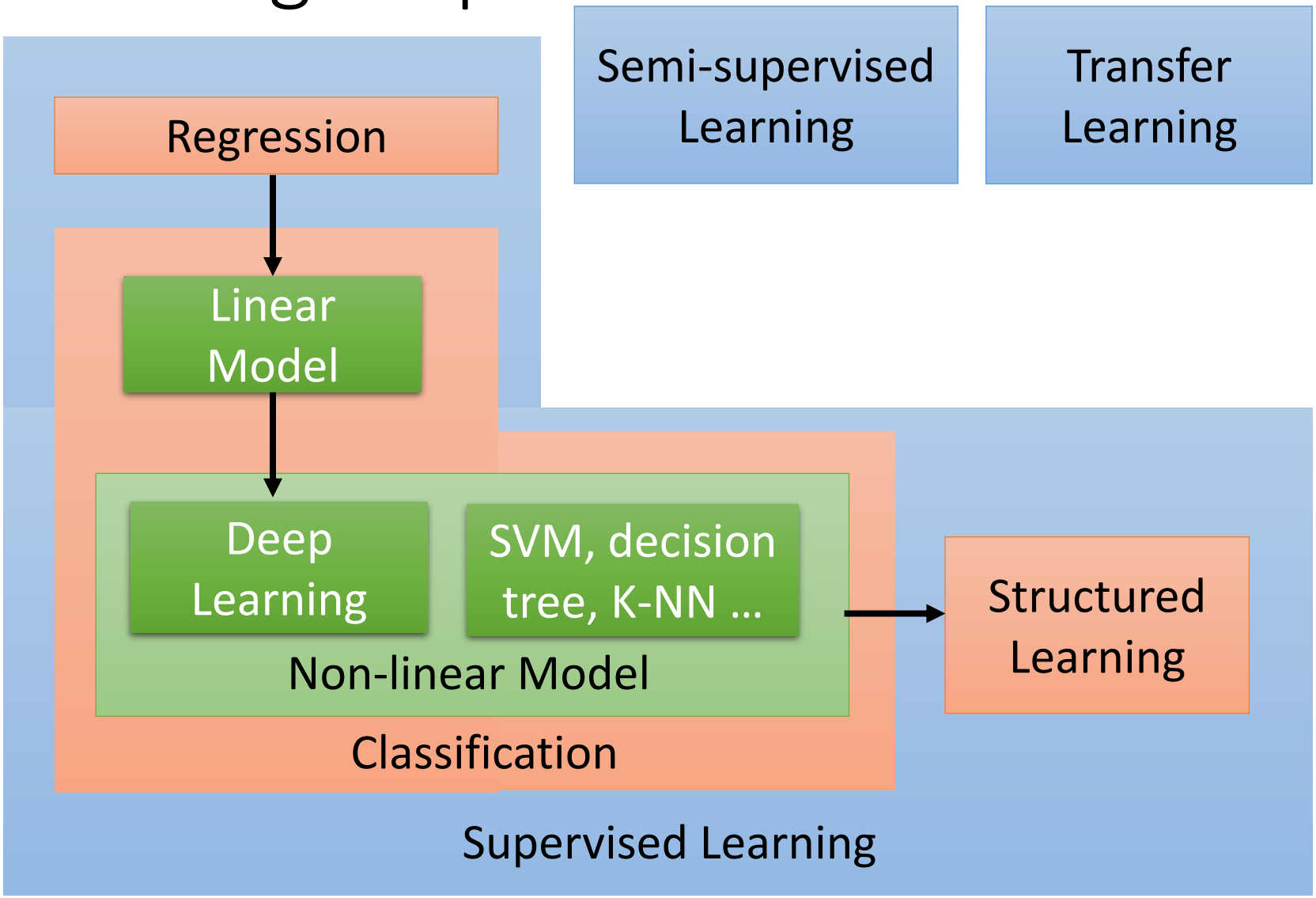


Unlabeled data



(Images of cats and dogs)

Learning Map



Transfer Learning

For example, recognizing cats and dogs

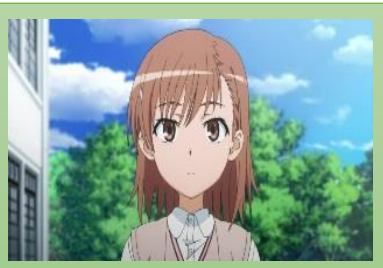
Labelled data

cat

dog

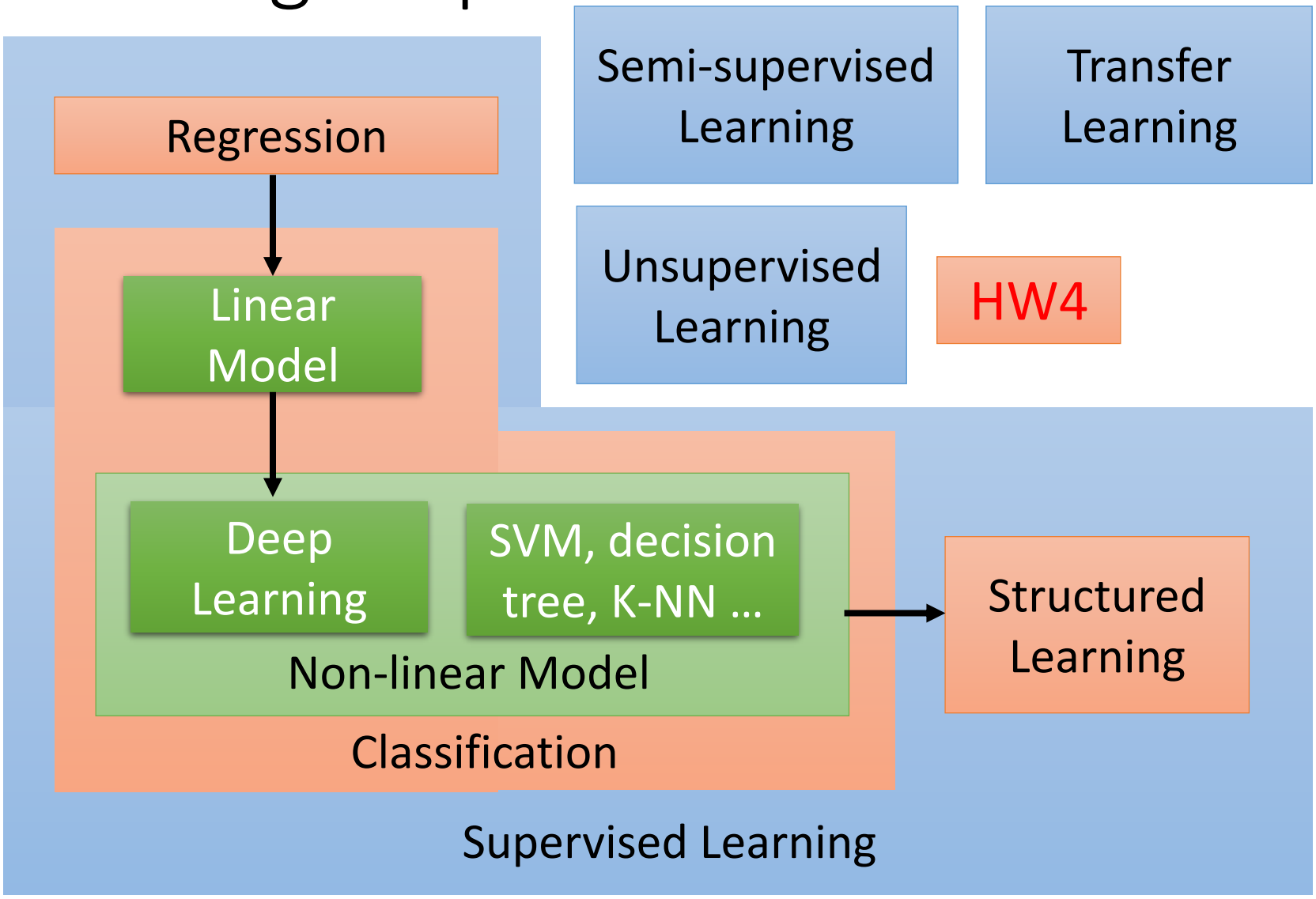
elephant

Haruhi



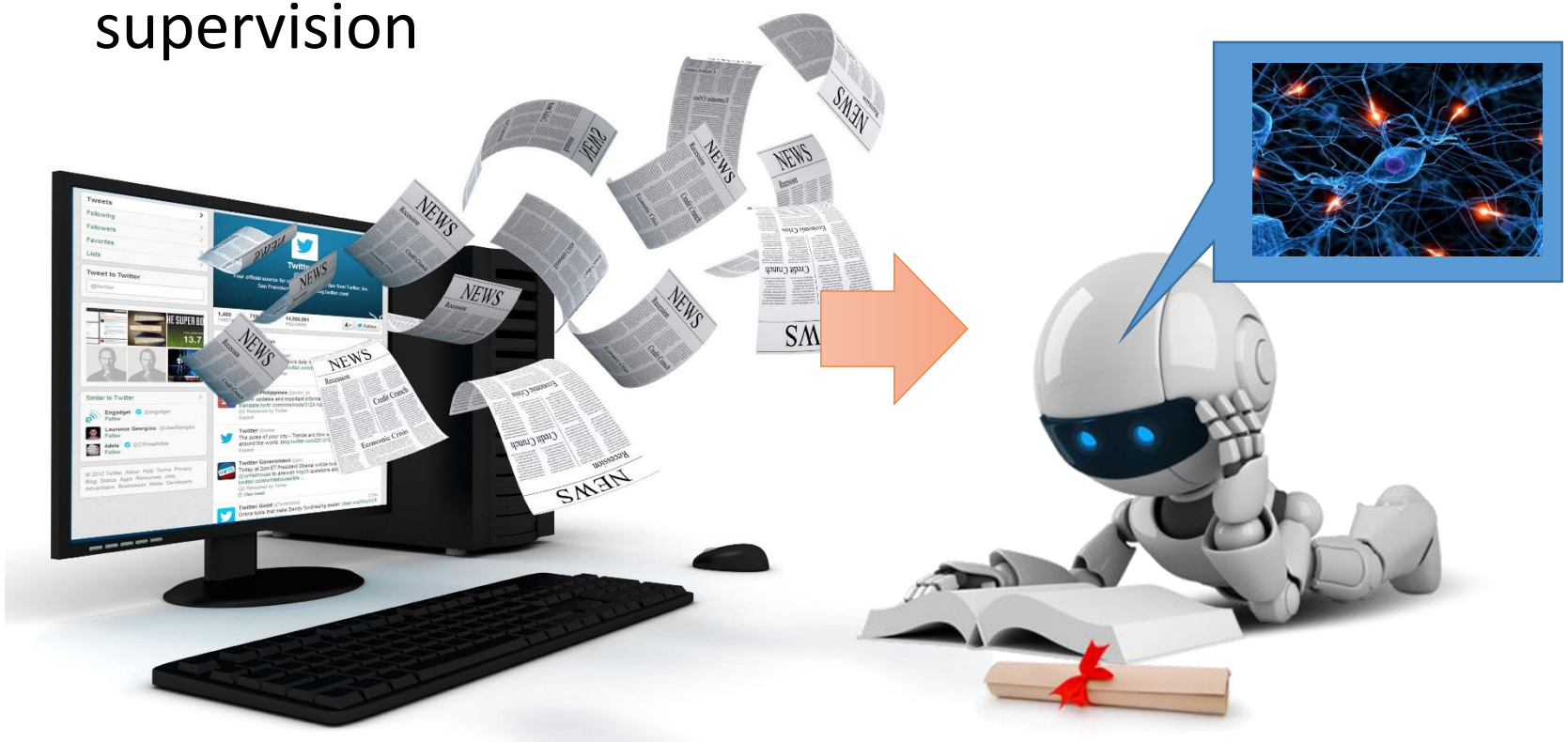
Data not related to the task considered
(can be either labeled or unlabeled)

Learning Map



Unsupervised Learning

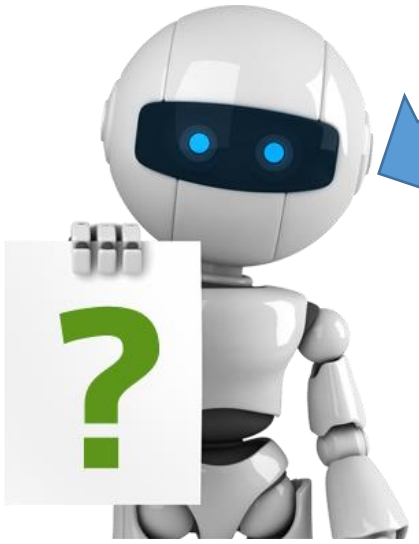
- Machine Reading: Machine learns the meaning of words from reading a lot of documents without supervision



http://top-breaking-news.com/

Unsupervised Learning

Ref: https://openai.com/blog/generative-models/



Draw something!

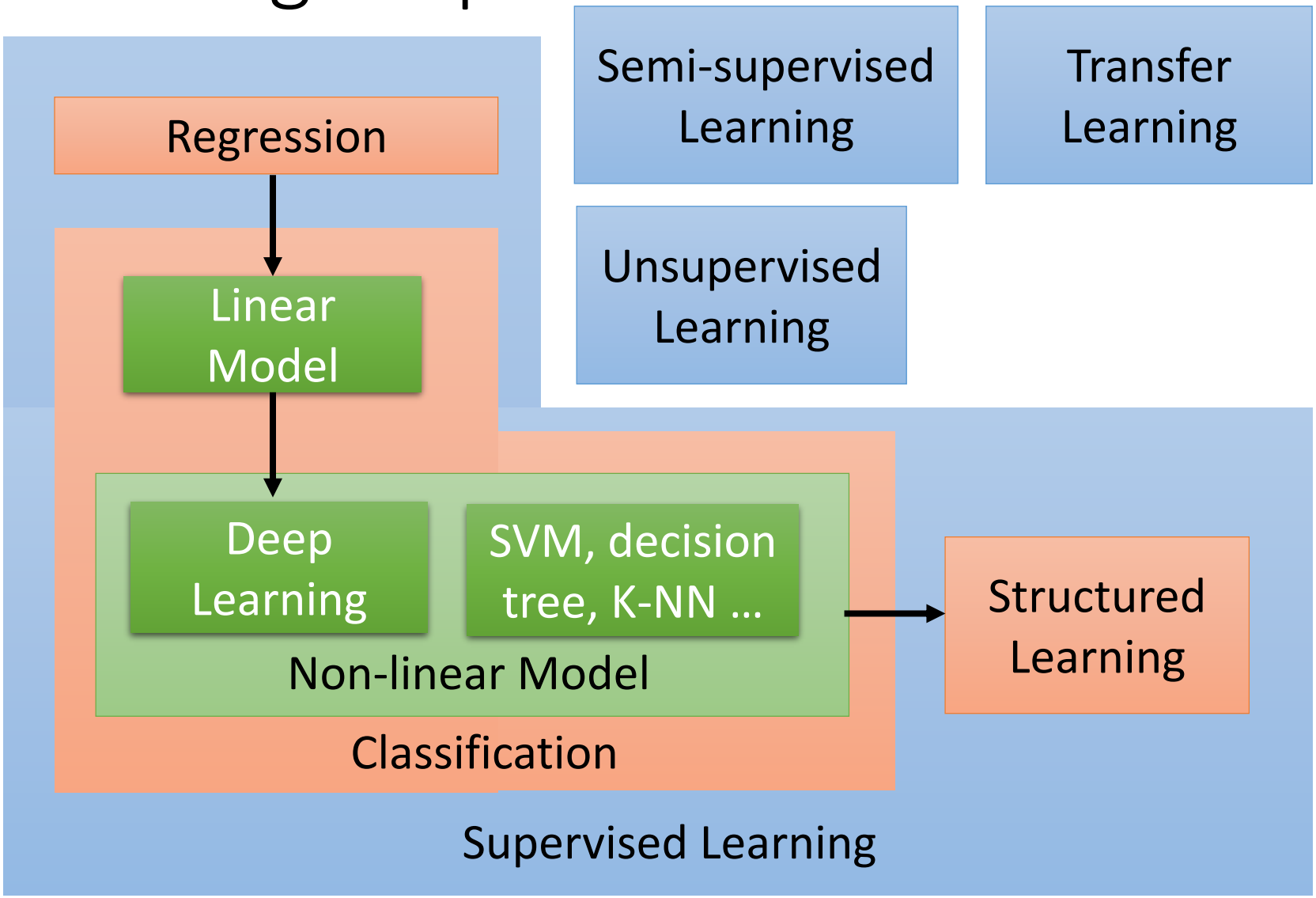
Unsupervised Learning

Machine listens to lots of audio book

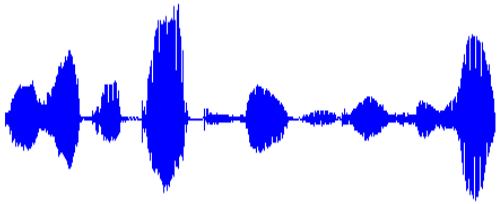
[Chung & Lee, INTERSPEECH 2016]

How about machine watch TV?

Learning Map



Structured Learning - Beyond Classification



“大家好，歡迎大家來修機器學習”

Speech Recognition

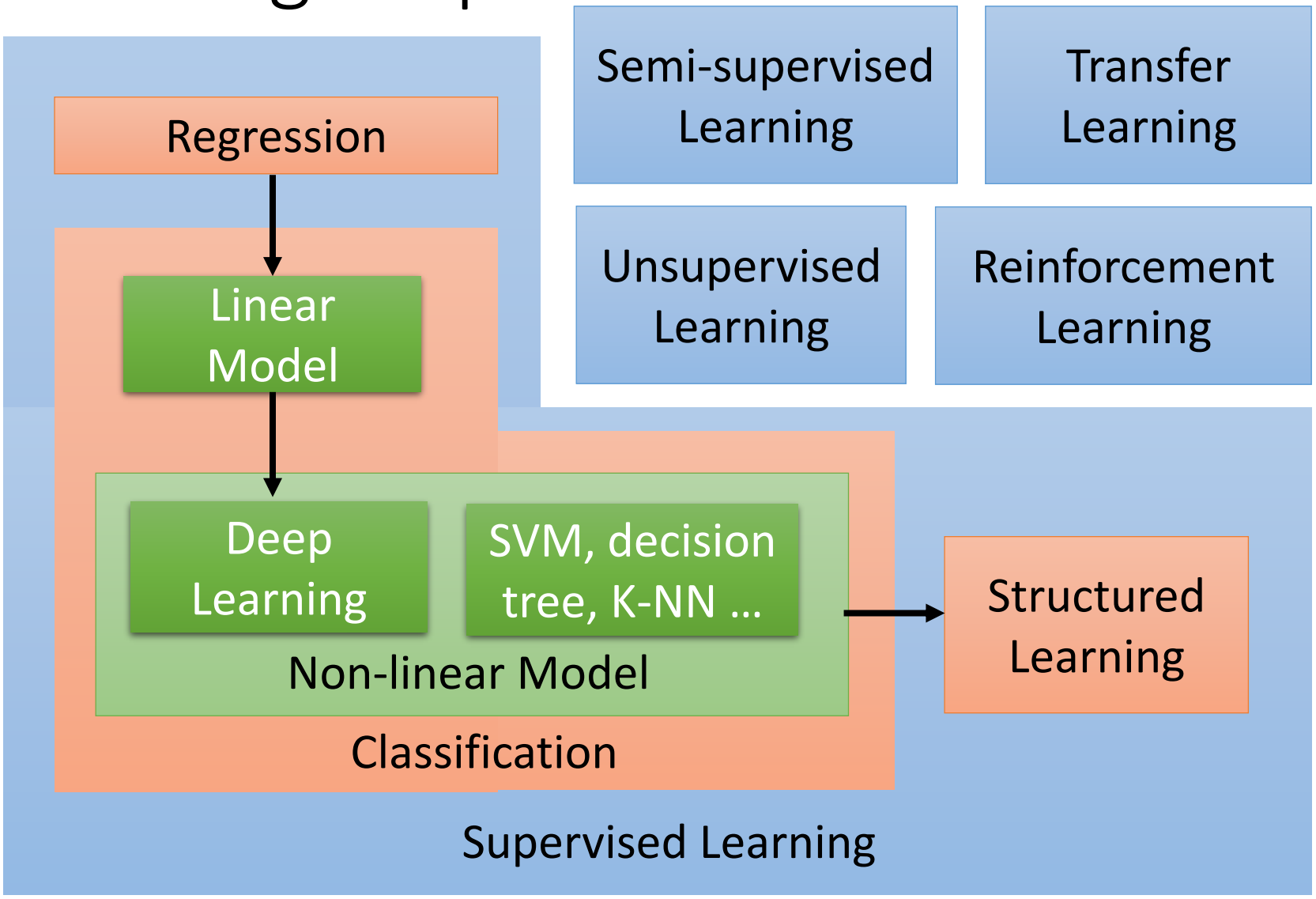
“機器學習”

“Machine Learning”

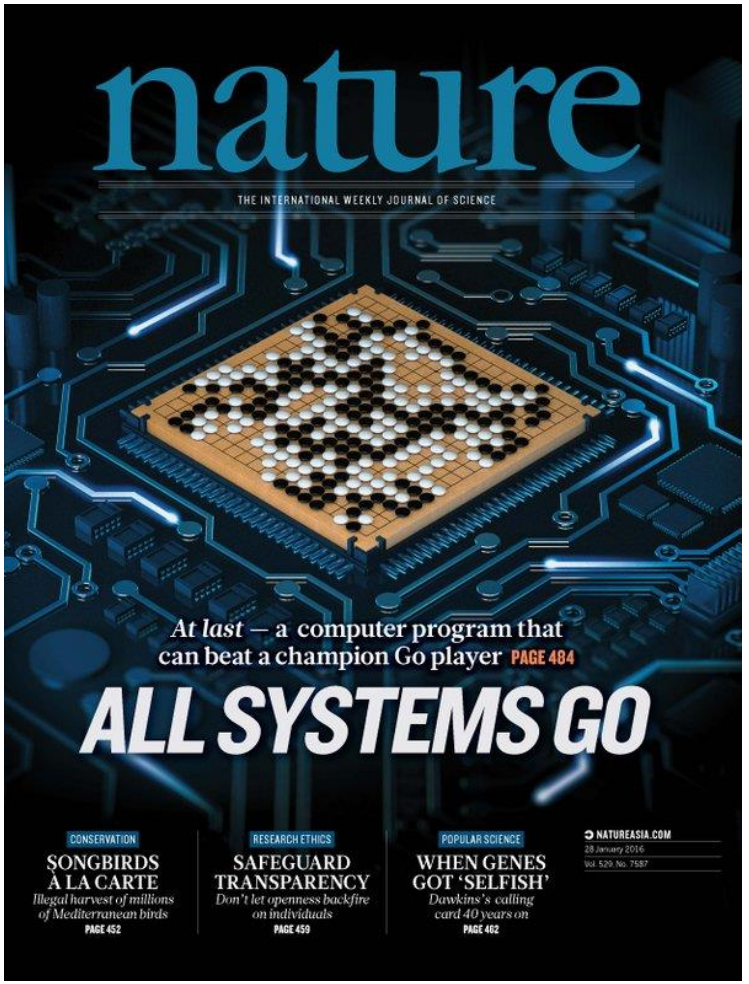
Machine Translation

人臉辨識

Learning Map



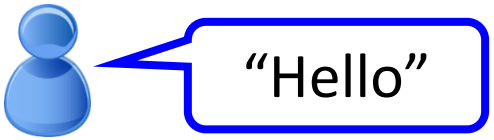
Reinforcement Learning



Supervised v.s. Reinforcement

- Supervised

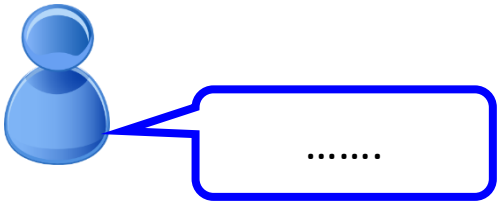
Learning from teacher

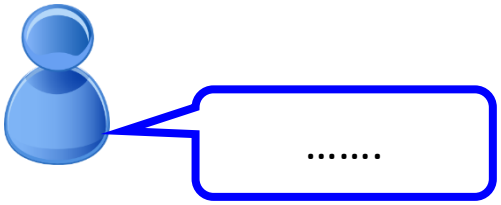


Say "Hi"

Say "Good bye"

- Reinforcement

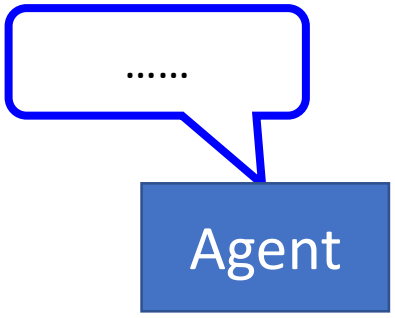




.....

Bad

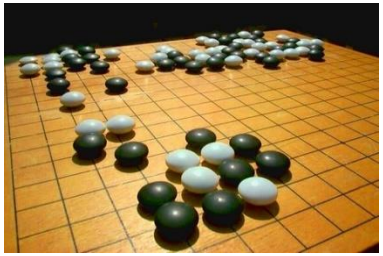
Learning from critics



Supervised v.s. Reinforcement

- Supervised:

Next move:
"5-5"

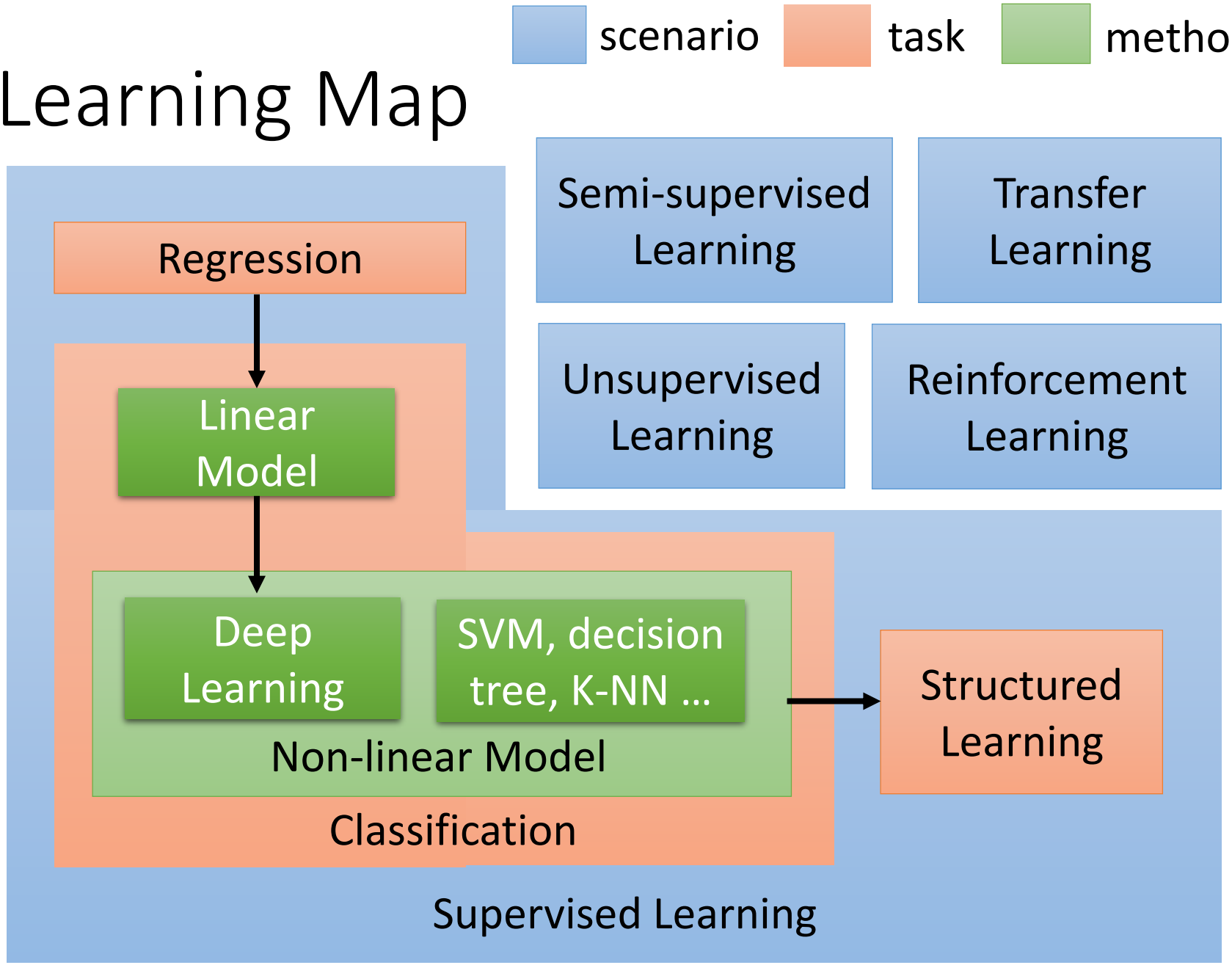


Next move:
"3-3"

- Reinforcement Learning

First move → …… many moves …… → Win!

Alpha Go is supervised learning + reinforcement learning.

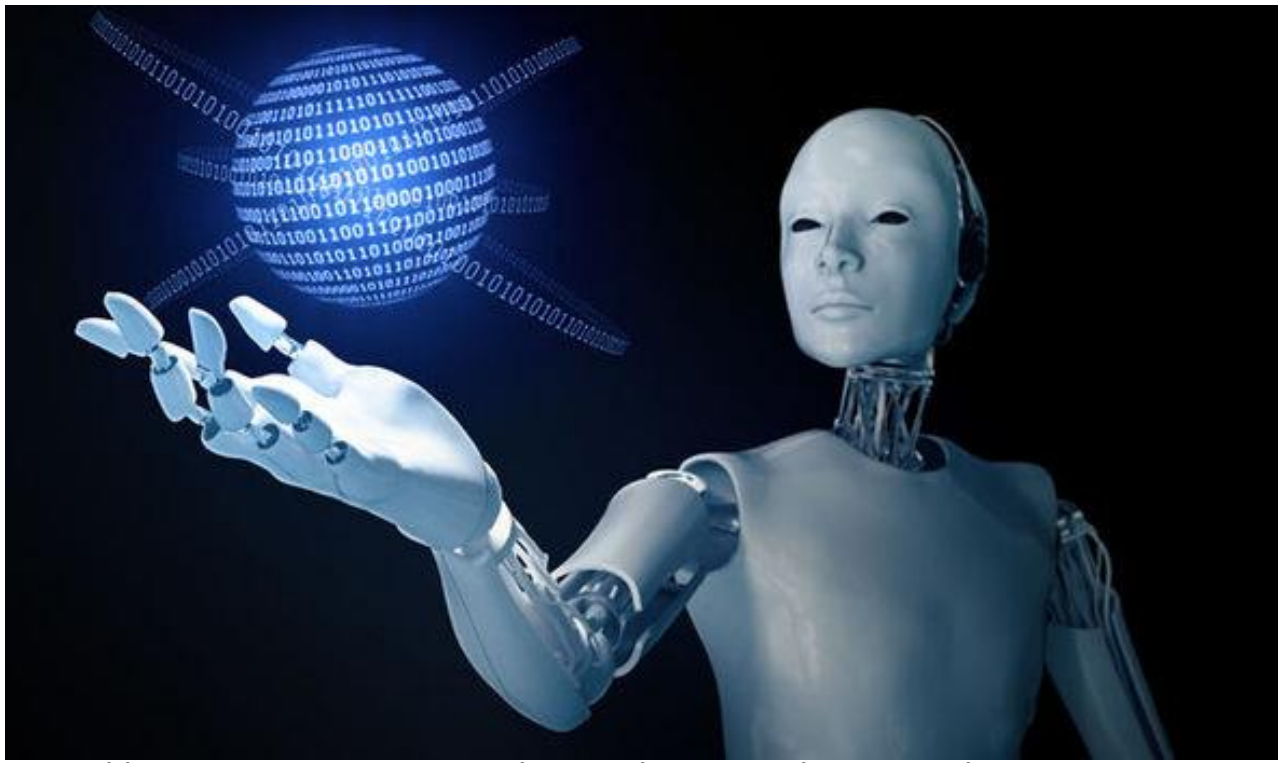

scenario
task
method
Learning Map
Semi-supervised Learning
Transfer Learning
Unsupervised Learning
Reinforcement Learning
Regression
Linear Model
Deep Learning
SVM, decision tree, K-NN …
Non-linear Model
Classification
Structured Learning
Supervised Learning

Why I need to learn Machine Learning?

AI 即將取代多數的工作?

- New Job in AI Age → **AI 訓練師**

(機器學習專家、資料科學家)



http://www.express.co.uk/news/science/651202/First-step-towards-The-Terminator-becoming-reality-AI-beats-champ-of-world-s-oldest-game

AI 訓練師

機器不是自己會學嗎？
為什麼需要 AI 訓練師

戰鬥是寶可夢在打，
為什麼需要寶可夢訓練師？

AI 訓練師

寶可夢訓練師

- 寶可夢訓練師要挑選適合的寶可夢來戰鬥
  - 寶可夢有不同的屬性
- 召喚出來的寶可夢不一定聽話
  - E.g. 小智的噴火龍
  - 需要足夠的經驗

AI 訓練師

- 在 step 1，AI訓練師要挑選合適的模型
  - 不同模型適合處理不同的問題
- 不一定能在 step 3 找出 best function
  - E.g. Deep Learning
  - 需要足夠的經驗

http://www.gvm.com.tw/webonly_content_10787.html

AI 訓練師

- 厲害的 AI ， AI 訓練師功不可沒
- 讓我們一起朝 AI 訓練師之路邁進

Policy

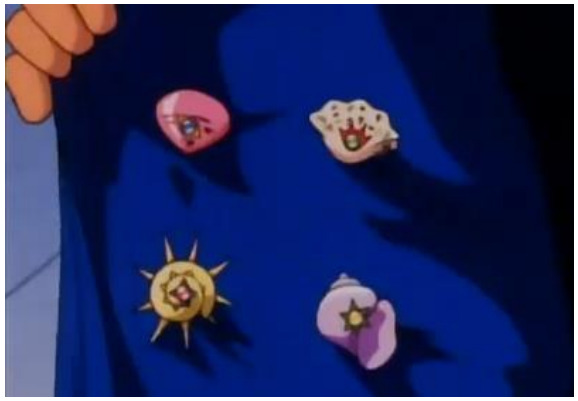
上課教材

- 以後上課會錄音
- 上課投影片和錄音會放到 ceiba 和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_ML16.html

FB 社團

- 社團: Machine Learning (2016, Fall)
  - https://www.facebook.com/groups/1774853276101781/
- 有問題可以直接在 FB社團上發問
  - 如果有同學知道答案請幫忙回答
- 有想法也可以在 FB社團上發言
- 會紀錄好的問題、答案、留言，期末會加分

評量方式



- 不點名、不考試
- 作業：沒有分組、每個人都要繳交
  - 作業一 (10%)：9/30 – 10/14 (二週)
  - 作業二 (10%)：10/14 – 10/28 (二週)
  - 作業三 (10%)：10/28 – 11/18 (三週)
  - 作業四 (10%)：11/18 – 12/02 (二週)
- 期末專題 (60%)：分組進行
  - 以比賽方式進行

評量方式 - 期末專題

- 11/18 公告
- 2 ~ 4人一組
- 進行方式：會公告幾個可能的題目給同學們選擇
  - Intrusion Detection比賽
  - Fintech 比賽 (規劃中)
  - 指定的 Kaggle 比賽
- 最後會有組內互評

加簽

- 助教會公告作業 0 ，今天晚上 10:00 前完成
  - 作業 0 跟機器學習無關，只是測驗基礎程式能力
- 如果可以解決空間的問題，完成作業 0 就加簽，助教會公告授權碼取得方式
  - 但如果無法解決，就不得不有所篩選
  - 這學期沒修到也不用太難過，未來還會再開