

# 4105931 機器學習 (Machine Learning)

姓名: 蕭 標 立  
學號: 605546001

**Midterm**  
總分 100 分  
中文作答

得分
42

1. a) (2%) 試解釋什麼是'機器學習'? b) (5%) 試舉出五種可以應用機器學習的實際問題?

a. 使用資料計算假設, 最大似然結果, 在大量資料中找出 outlier, 並做分類或機率預測。  
b. 圖片分類, 語音辨識, 情緒語句分析, 文件分類, 人臉辨識

2. (8%) 有鑑於理工科的男生都不太會穿搭, 所以我做了下面這樣的事情。我首先收集了 2018 年整年的紐約時裝周服飾照片, 然後依照影像處理的作法, 擷取了顏色、花紋、材質三種特徵, 每一張影像將三種特徵串接成一種新的特徵表示法, 然後分群成 100 群, 接著我拿出我自己穿搭的照片, 取出這種新的特徵表示法, 跟 100 群每一群的群中心算距離, 假設距離小於預設的門閥值 50, 我就得一分, 看看最終總得分的高低, 來決定我的穿搭帥不帥。試以 'Type of Learning' 的角度, 說明這個方法是屬於什麼樣的機器學習方法?

8 監督式學習, = 不分類。

3. (7%) 當使用某一個已經分好訓練、和測試資料的 dataset, 如何利用分成三包的交叉驗證技巧, 訓練一個好的模型? 請從訓練、驗證到最終的測試, 詳細列出全部的步驟。

5 訓練資料 → 訓練 → 模型 → 測試資料 → 驗證 → 最終測試

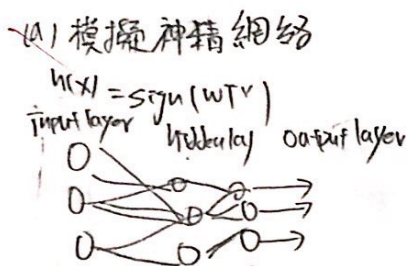
4. (7%) 如果我想訓練一個 "人生勝利組預測模型", 知道一個人到 60 歲的時候是溫拿(winner)還是魯蛇(loser), 試說明如何設計這樣的模型? 從資料、特徵、模型、損失函數、訓練、到測試, 說明整個步驟。

資料 → 特徵 → 模型 → 損失函數 → 訓練 → 模型  
取得人生勝利組是基本資料  
特徵: 學歷、收入、存款  
模型: Linear classification  
損失函數: input  
訓練: 調整參數  
測試: 得出是否勝利組

5. a) (2%) 試說明什麼是 Perceptron? b) (2%) 試說明 Perceptron Learning Algorithm 中, 尋找分類錯誤的  $x$ , 為什麼可以用公式  $\text{sign}(w_i^T x_{n(i)}) \neq y_{n(i)}$ ? c) (4%) 試說明  $w_{t+1}$  的更新為什麼可以用  $w_t + y_{n(i)} x_{n(i)}$  取代?

d) (2%) 保證能讓 Perceptron Learning Algorithm 的停止條件是什麼? 如何能保證演算法一定停的下來? e) (2%) 試比較 Perceptron Learning Algorithm 與 Pocket Algorithm 相同與相異之處為何?

f) (4%) 試說明若資料維度是  $d$ , 為什麼 Perceptron Hypothesis  $h(x)$  的維度是  $d+1$ ?



b) = 元分類  
c) 滑槽增加偏差值調整模型。

d) 必須正確答案, 得到結果才信下

e) 相同處理分類問題

相異 pocket 可區分非線性

f) 包含自己 input  $h(x)$  以  $d+1$ 。  
資料  $d = \text{input} \times w$ , 包含 input 與權重



6. 假設 dataset  $X$  有五筆資料  $x_1, x_2, \dots, x_5$ ，其資料維度為 2，每一筆資料的答案為  $y_1, y_2, \dots, y_5$ ，a) (2%) 試寫出要求出  $y_1, y_2, \dots, y_5$  的 regression model 公式解。b) (3%) 假設  $X^T X$  可逆，試寫出公式解中每一個矩陣或向量的維度。c) (2%) 假設 dataset 每一筆資料維度為 10，利用這個 dataset 求出的線性回歸模型，共會有多少的參數？

-4 
$$E(w) = \frac{1}{N} (XW - Y)^T (XW - Y) = \frac{1}{N} (W^T X^T X W - 2W^T X^T Y + Y^T Y)$$
 (A)  $10 \times 2 = 20$

(b) 
$$W_{min} = \frac{(X^T X)^{-1} X^T Y}{\text{psub} - \text{mm} \text{ } X^{-1}}$$

7. a) (3%) 試說明 PLA 演算法跟 Gradient Decent 演算法，在表示  $w_{t+1}$  的解時，是如何寫成相同的形式。b) (2%) Logistic Regression 演算法中，若資料筆數是 10，資料維度是 2，label 維度是 1，則

$\nabla E_{in}(w_t)$  維度是多少？c) (2%) 承上題， $\theta(-y_n w_n^T x_n)$  的維度是多少？20

8. a) (2%) 說明深度學習架構中，一個節點、一層與多層，其功能差異為何？b) (4%) 試以數學證明多次線性組合的結果還是線性。c) (2%) 如何引入非線性的機制到深度學習模型中？d) (3%) 2-4-6-2 NNet 的模型參數總共有多少個？e) (4%) 詳細說明 Backpropagation Algorithm 每一個步驟的意義。

節點 = 只有一次 input, 產生結果  
一層 = 多個 input, 加入權重, 產生結果

多層 = 多個 input, 加入多層權重, 產生結果

13 
$$h(x) = \text{sign} \left( \sum_{i=1}^d w_i x_i - \text{threshold} \right)$$
  
$$= \text{sign} \left( \sum_{i=1}^d w_i x_i + (-\text{threshold}) \cdot (+1) \right)$$
  
$$= \text{sign} \left( \sum_{i=1}^d w_i x_i \right) = \text{sign}(w^T x)$$

2-4-6-2  
(d)  $2+1=3$   
 $3 \times 4 = 12$   
 $12 \times 6 = 72$   
 $2+3=5$   
 $72+5=77$

9. a) (3%) 試說明 Convolutional Neural Network 與一般類神經網路架構的相同處為何？b) (3%) 與一般類神經網路相比，CNN 如何降低模型的參數量？c) (4%) 訓練一般類神經網路模型時，要學習的參數是什麼？訓練 CNN 時，要學習的參數是什麼？d) (2%) 何謂 mini-batch？e) (2%) 說明 epoch, iteration 與 batch size 的關係。

(a) 均有 Input, hidden, output layer, CNN 有 Filter。

(e) epoch = 整個 dataset 跑一輪

batch size: 一批次的資料，計算梯度平均。

(b) 透過 Filter，做 convolution 得 feature map，再透過 Max Pooling 找最大 response。

(c) 一般: Loss function.  
CNN: feature map

(d) 每批次訓練資料數目較小。

10. a) (5%) 請提出兩個方法，能增進你對課程的理解。b) (5%) 寫出對本課程的建議事項。

(a) (1) 公開推薦，上 youtube 看其他老師的教法

(b) 檢討 hw 時間以簡單說明一下 code 流程。

(2) 上完課後複習所學