4105931 機器學習 (Machine Learning)

得分

姓名:_____

學號:_____

Final Exam

總分 100 分

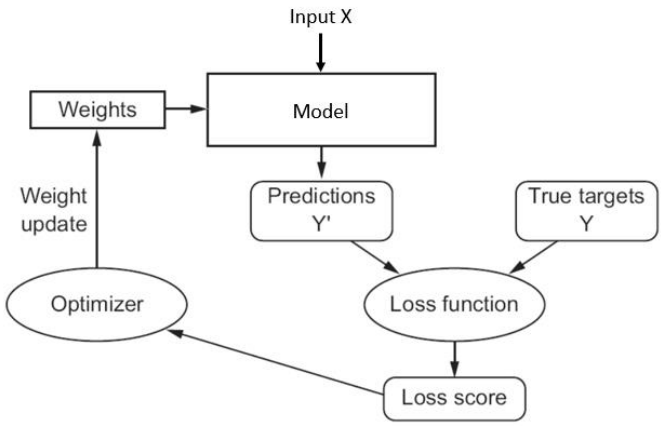
請用**中文作答**、並依題號順序作答於 A4 白紙，若題目有疑問可在 Slido 發問

1. a) (3%)試說明深度學習的三個工作原理。b) (3%)說明這三個工作原理如何形成一個閉迴路的運作。

(a) (各 1 分)

1. 神經網路是由其權重來參數化
   - 神經網絡中每層實現的變換由其權重來參數化，神經網路的所有層找到其權重值，使得該網路能夠將每個輸入與其目標答案正確地一一對應。
2. 損失函數的任務
   - 控制神經網路的輸出，就需要能夠衡量該輸出與預期值之間的距離。
   - 損失函數的輸入是網路預測值與真實目標
3. 利用“損失值”重新調整模型權重
   - 深度學習的做法是利用這個距離值作為反饋信號來對權重值進行微調，以降低當前資料的損失值

(b) (依完整度給分)



深度學習模型利用初始化參數由輸入資料預測輸出(工作一)，並由損失函數來評估模型預測值的好壞(工作二)。如果預測值與真實目標差距過大，可以利用優化器及反向傳播演算法來更新模型的參數(工作三)，並利用更新後的模型來輸出新的預測值(工作一)。

這樣的閉迴路會在預測值與真實目標差距最小或是在可以接受的範圍內才會停止循環。

2. a) (3%)試說明深度學習模型中為何需要加入 activation function。b) (4%)說明 Sigmoid、Hyperbolic Tangent、Rectified Linear Unit、Leaky Rectified Linear Unit 四種 activation function 的優缺點及相同相異處。

(a) (3 分)

類神經網路是以線性的方式組合運算，隱藏層以及輸出層皆是將上層之結果輸入，並以線性組合計算，作為這一層的輸出，使得輸出與輸入只存在著線性關係，因為多次線性組合後還是為一種線性組合，若不使用 activation function，假設全部都是線性，結果出來也會線性，也就是說等於等比例放大或縮小而已，所以要用 activation function，但若所有問題皆屬於非線性問題，無使用非線性之 activation function，則類神經網路訓練出之模型便失去意義。選擇可微分之函數，因為在 backpropagation 運算需要進行一次微分計算，微分後結果會與輸入有關且堆疊多層 layer 變得可行。

(b) (每一 function 各 1 分)

Activation Function	**優點**	**缺點**
Sigmoid	平滑漸變、明確的預測	梯度消失、中心不在 0、計算複雜

Activation Function	優點	缺點
Hyperbolic Tangent	中心在 0、平滑漸變、明確的預測	梯度消失、計算複雜
Rectified Linear Unit	計算簡單	因負值時沒有梯度因此部分神經元無法更新
Leaky Rectified Linear Unit	在負值時可更新、計算簡單、非線性	負值的梯度差不多

- Sigmoid、Hyperbolic Tangent 都為平滑漸變函數，但計算複雜且有梯度消失問題，有時避免使用。
- ReLU、Leaky ReLU 為計算簡單的非線性函數，如遇到 dead neurons 時可使用 Leaky ReLU 解決。

3. (8%)請從下列**四個面向**，解釋深度學習模型不是黑盒子: 核心精神、幾何意義、數學運算、特徵表示。
   - 核心精神(2 分)
     - 深度學習的核心在於有意義的變換數據，學習輸入數據更有用的資料表示法，新的表示法讓數據更接近預期輸出。例如透過座標變換、線性投影、平移、非線性操作等等，根據任務將數據轉化為更加有用的表示。
   - 幾何意義(2 分)
     - 神經網路由一系列運算組成，這些運算就是在高維空間上做一連串的幾何轉換。
   - 數學運算(2 分)
     - 深度學習模型在計算神經網路各層的權重參數與資料做一系列相乘相加；實際上代表模習訓練時的 Pattern Extraction，測試時看看測試資料有沒有符合這些 pattern。
   - 特徵表示(2 分)
     - 把模型辨認的負擔分散到各個節點、各個層上面，每一個地方學一點的特徵表示法，再將每個地方的特徵表示集合起來辨認。

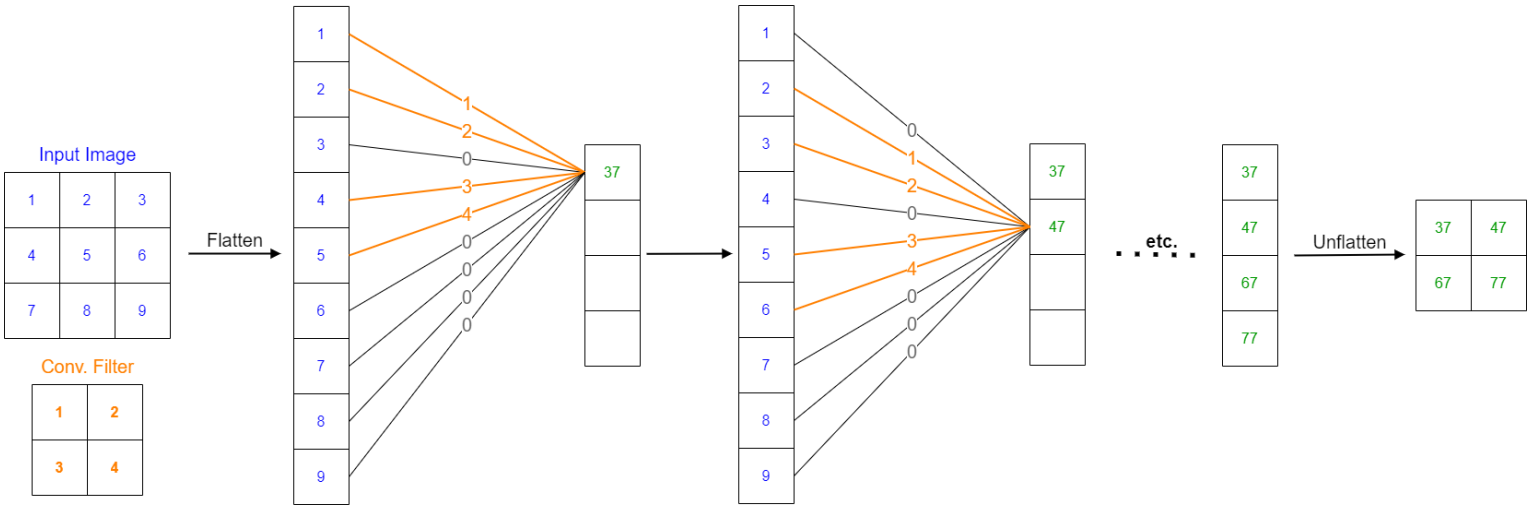
4. (6%)請詳細說明反向傳播演算法的詳細步驟。
   - 反向傳播算法主要由兩個階段組成：激勵傳播與權重更新。
     1. 第 1 階段：激勵傳播
        - 每次迭代中的傳播環節包含兩步：
          1. （前向傳播階段）將輸入資料送入網絡以獲得模型輸出；
          2. （反向傳播階段）利用損失函數來計算模型輸出與真實標籤(Ground Truth)的誤差。
     2. 第 2 階段：權重更新
        - 對於每個節點上的權重，按照以下步驟進行更新：
          1. 將模型輸出和損失誤差相乘，從而獲得權重的梯度；
          2. 將這個梯度乘上一個倍率因子並取負值後加到權重上。
             - 梯度的方向指明了誤差擴大的方向，因此在更新權重的時候需要對其取負值，從而減小權重引起的誤差。
   - 第 1 和第 2 階段可以反覆循環迭代，直到損失誤差達到滿意的預定的目標範圍為止。

5. a) (3%)試說明影像處理中，影像捲積運算的方式。b) (4%)說明 CNN 的捲積運算如何變成與 DNN 的全連結架構的運算方式。c) (5%)與 DNN 相比，CNN 模型在處理影像時的優勢為何?d) (4%)CNN 模型在架構上可以視為哪兩部分?兩部分在結構大小上為何會呈現一大一小的趨勢?

(a) 用 filter 對相對應位置像素做「elementwise product」，且加總輸出成一數值，並且 filter 會根據 stride 移動位置(左到右和上到下都移動)做同樣的動作直到最後遍歷整張影像得到卷積結果。

(b)

- 將輸入影像攤平成一維的向量，此向量為對應到全連接層的輸入層。
- 卷積核對應到全連接層的連線，對應方式以下圖為例。
- 以下為輸入圖片大小為 3*3*1，conv(input channel=1, output channel=1, kernel size=2, stride=1, padding=0) 轉換成 linear(input size=3*3*1, output size=2*2)的範例說明圖。



(c)

- 降低參數量
  - 相比於 DNN 的全連結架構，CNN 減少連線數量，降低模型整體參數
  - 卷積運算對於一個節點來說，只要 9 條線連接(9 個參數)
  - 對於多個節點，權重共享(每個節點都使用相同的 9 個參數)
- 執行時間加速
  - DNN 模型的訓練、測試的執行時間隨之降低
- 結合影像特性
  - 採用影像區域相似、平滑變化的特性於模型架構中
  - 用以學習下一層的節點或特徵表示
- 不再需要專家設計卷積核
  - 模型自行學習什麼樣的卷積核，能達成分類的目的。

(d)

CNN 模型由「特徵表示法的學習」及「全連結作為分類器」組成。因為當特徵學習的效果很好的情況下，分類器的部分只需要幾層全連接層就可以得到很好的分類。

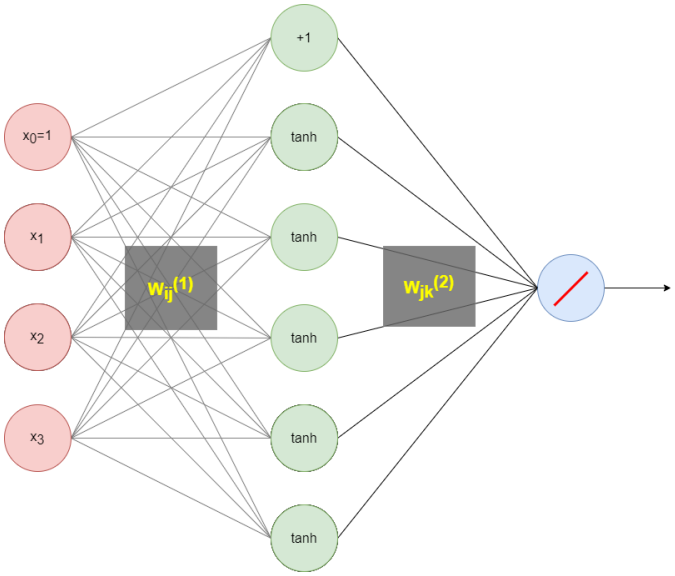
6. a)(9%)請依照 DNN、CNN、LSTM 三種模型，說明他們分別適合解什麼樣的問題?並依據三種模型，每個模型舉兩個實際應用的問題。b) (6%)舉例並畫出 DNN、CNN、LSTM 三種模型架構的例子(例如 3-5-1 的 DNN…)。c) (6%)請在 b)中的範例，說明三種模型架構中，模型要學習的參數是什麼?並在圖中標記出要學習參數的位置。

(a)

- DNN
  - 主要針對一維向量的數據型資料進行分類或回歸問題的分析。
  - 實際應用
    - 房價評估
    - 保險用戶風險分析
- CNN
  - 主要針對二維影像作影像分析的任務，也可以應用在其他維度的資料上(一維、或三維)。
  - 實際應用
    - 蘭花分類
    - 語意分割
- LSTM
  - 適合時序性任務，當資料在時間狀態下保有連續性或是對前一個狀態有依賴關係，那麼可使用 LSTM 去找出資料在時間的潛在空間的關係。
  - 實際應用
    - 預測發電量
    - 天氣預測

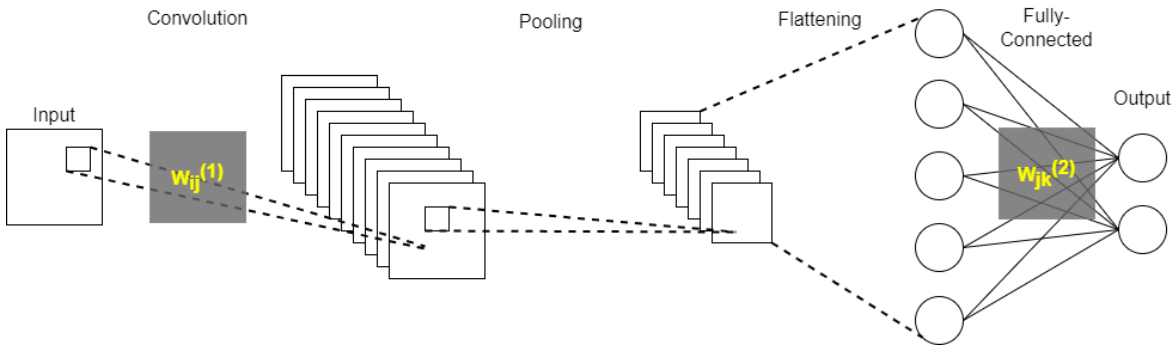
(b)(c)

- DNN



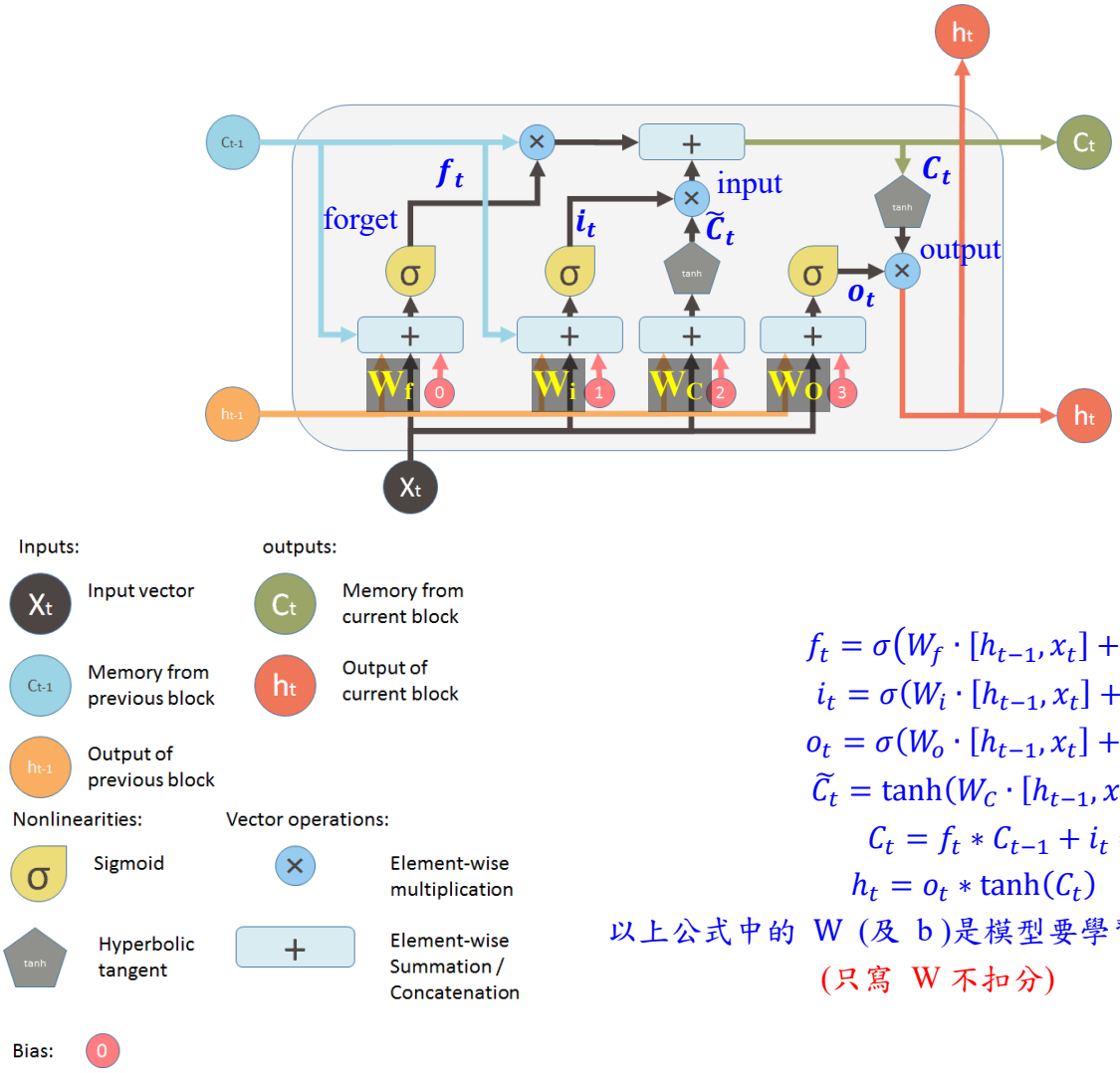
灰色區塊 $w_{ij}^{(1)}$、$w_{jk}^{(2)}$ 為 Fully-Connected Layer 的參數，也是模型要學習的參數。

- CNN



灰色區塊 $w_{ij}^{(1)}$、$w_{jk}^{(2)}$ 分別為 Convolution 及 Fully-Connected Layer 的參數，也是模型要學習的參數。

- LSTM

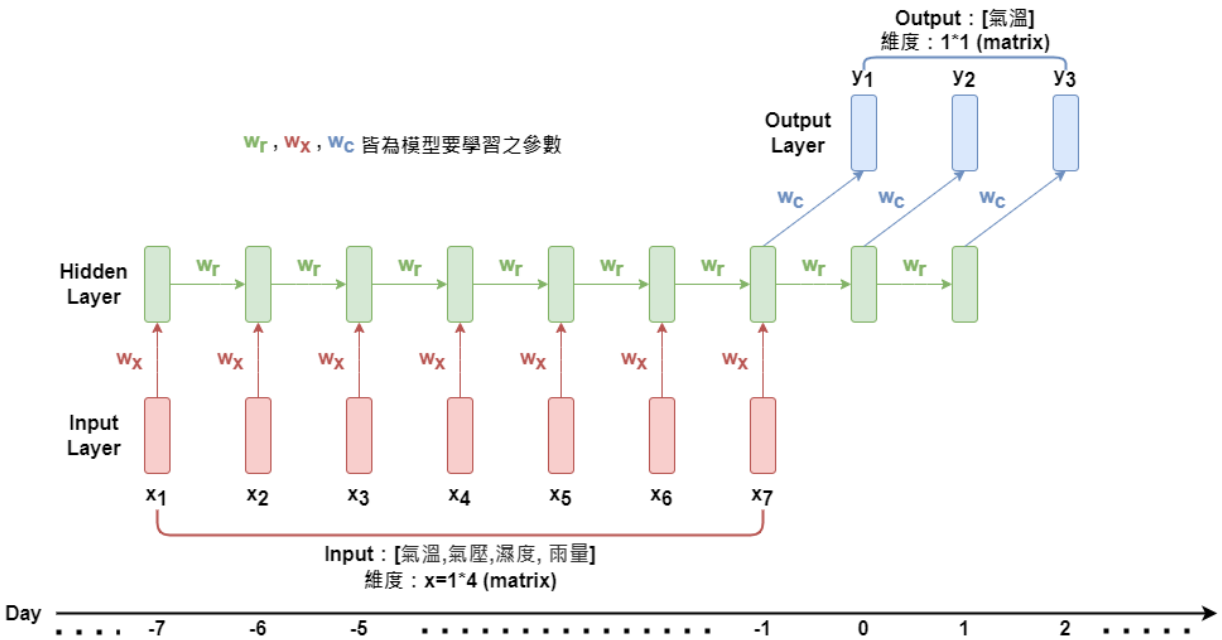


$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$
$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$
$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$
$$\widetilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$
$$C_t = f_t * C_{t-1} + i_t * \widetilde{C}_t$$
$$h_t = o_t * \tanh(C_t)$$

以上公式中的 W (及 b)是模型要學習的參數。
(只寫 W 不扣分)

7. a)(5%) 假設欲使用過去七天的氣溫、氣壓、濕度、雨量的資料，預測接下來三天的氣溫，請說明這是屬於 RNN-based Models 中哪一種類型的模型? b) (5%) 試為 a)問題，繪出一個量身打造，並在時間軸上展開的 RNN 模型架構圖，請在圖上標出**輸入**、**輸出的結果，以及其維度**。

- 多對多模型(many to many) (3 分)
  - 因為使用過去七天的資料(包含氣溫、氣壓、濕度、雨量)，而我們期望輸出為接下來三天的資料 (包含氣溫) (2 分)
- 一個物件 (輸入維度、輸出維度、參數位置、架構圖、整體性) (各 1 分)



8. a) (6%) 說明生成對抗網路中，生成器與判別器的功能分別為何?其輸入輸出分別是什麼? b) (6%) 說明若不使用判別器，單獨使用生成器想要完成原本生成對抗網路的功能，該怎麼作?會遇到什麼困難? c) (6%) 說明若不使用生成器，單獨使用判別器想要完成原本生成對抗網路的功能，該怎麼作?將遇到什麼困難? d) (2%) 如果把生成器、判別器連接起來，當作一個大的 CNN 網路，並利用過去訓練 CNN 的方式訓練，在生成時會遇到什麼問題? e) (3%) 試解釋生成對抗網路損失函數 min-max 形式的設計原理及其意義。f) (3%) 試說明生成對抗網路模型的訓練方式。

(a)

- 生成器: 藉由輸入一個 vector，經過神經網路做 upsample 後產生出一張影像 (2 分)
  - 輸入: A vector (normal distribution, uniform distribution…)
  - 輸出: 一張影像 (1 分)
- 判別器: 判斷由生成器產生的影像是否真實 (2 分)
  - 輸入: 一張由生成器生成的影像
  - 輸出: 0-1 之間的分數，分數越高代表越有可能是 real (1 分)

(b)

- 先準備多張影像以及其對應的 vector，訓練使生成器能透過 vector 生成影像。(3 分)
- 困難: (3 分)
  - 不容易產生影像對應的 vector，資料難以蒐集
  - 生成器會獨立生成個個像素點，無法從全局把握不同像素點的依賴關係

(c)

- 先訓練好一個判讀器 D(x)，並窮舉所有可能的 x，被判別為真實即為較真實的影像。(3 分)
- 困難: (3 分)
  - 在訓練一個好的判別器 D 需要大量的 Real 影像以及高質量的 Fake 影像，不容易取得
  - 窮舉所有可能的 x 會需要大量的資源來計算。

(d) (2 分)

若使用過去訓練 CNN 的方式訓練，在訓練時同時更新生成器及判別器的方式訓練，有可能模型只調整 CNN 模型最後一層的參數，讓最終得到高分，但前面生成器的輸出，卻無法輸出一張正確的影像。

(e) (只寫公式 1 分)

- Loss function：$\min_G \max_D V(D,G) = E_{x\sim p_{data(x)}}[\log D(x)] + E_{z\sim p_{z(z)}}\left[\log\left(1-D\left(G(z)\right)\right)\right]$
- 訓練判別器判別器 D (1 分)
  - D 判斷真實影像，判斷真實影像，$\log D(x)$ 取最大值，使 D 判別真實影像得到高分；D 判斷生成影像，為了使 $\log\left(1-D\left(G(z)\right)\right)$ 的值越大，$D\left(G(z)\right)$ 需要越小越好，$1-D\left(G(z)\right)$才會越大，$D\left(G(z)\right)$ 越小代表 D 判別生成影像得到低分。
- 訓練生成器 G (1 分)
  - 使$\log\left(1-D\left(G(z)\right)\right)$越小越好，所以$D\left(G(z)\right)$接近 1，也就是生成器 G 產生的生成影像要讓判別器 D 認為是真實影像，越像越好。

(f) (3 分)

- 訓練時，先固定生成器的參數，以機率分布生成 m 個 vectors，輸入進生成器產生 m 張 fake 影像，並與準備好的 real 影像來訓練判別器。
- 接著轉為固定判別器參數，再以機率分布生成 m 個 vectors 丟進生成器中，訓練生成器使生成的影像在判別器中能被判別為 real。
- 重複以上步驟直到生成器產生的影像接近真實影像，判別器無法分辨出生成器產生的假影像。