

Lab 1 結報

姓名：賴昱凱

學號：111511141

1. 請敘述在課堂上實作之神經網路的架構

```
Model: "model"
```

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[(None, 784)]	0
dense (Dense)	(None, 512)	401920
dense_1 (Dense)	(None, 256)	131328
dense_2 (Dense)	(None, 128)	32896
dense_3 (Dense)	(None, 64)	8256
dense_4 (Dense)	(None, 10)	650

```
=====  
Total params: 575,050  
Trainable params: 575,050  
Non-trainable params: 0
```

本模型是由 4 層 hidden layer 以及 input、output layer，共 6 層所組成的 DNN 模型，DNN 模型的全稱為 Deep Neural Network，顧名思義代表所有的節點都與前後層的所有節點相連，利用權重做相乘 $a_j = \sum_{i=1}^D W_{ji}x_i + w_0$ 往下一層前進，因此整個每層共有 $D_{in} \times D_{out} + bias$ 個參數，以本模型為例：

$$784 \times 512 + 512 = 401920$$

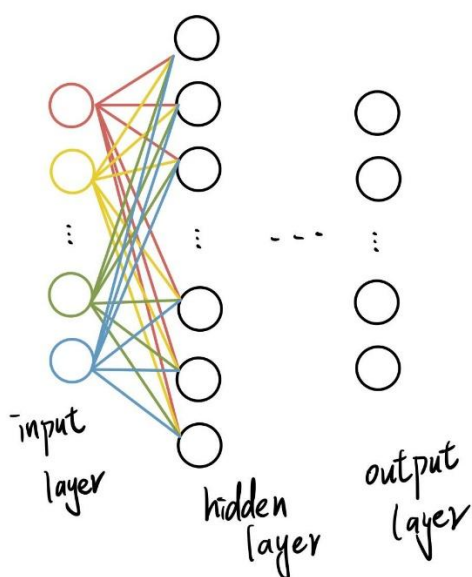
$$512 \times 256 + 256 = 131328$$

$$256 \times 128 + 128 = 32896$$

$$128 \times 64 + 64 = 8256$$

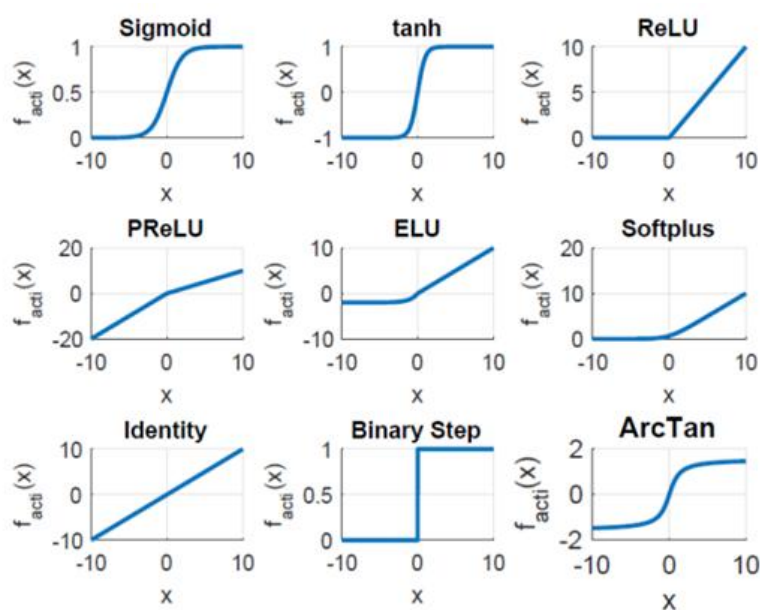
$$64 \times 10 + 10 = 650$$

全部相加即為本模型的總參數：575050



除了 linear 以外，還有一個部分稱為 activation function，他是非線性的 layer，因為若整個 model 都是 linear 的 layer，會很難處理現實中的問題，因為現實中的問題基本上都不會是 linear 的，因此需要在 model 中加上 activation function 來讓模型更加靈活。

本模型皆使用 ReLU 這個最常見的 activation function，他會將正數原封不動的送給下一層，而遇到負數就將其設為 0，這是一種最不耗計算資源的 non-linear 方式。下圖還有更多常見的 activation function。



本模型的輸出層為有 10 個節點的神經層，最後一層利用 softmax 作為 activation function，利用指數化確保輸出皆為正數，再讓所有輸出的值都介

於 0~1 且相加為 1，代表各種類的可能性，最高的那個就代表此次模型預測的結果，這是幾乎所有分類問題處理輸出最基本的方式。

Softmax:

$$\hat{y}_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}}$$

2. 請敘述 Loss Function 與 Optimizer 的作用

Loss Function:

衡量模型預測值與真實值之間的差距，通常來說 loss function 的目標是值越小越好，代表與真實值的差距越小。常用的 loss function 有：

Mean Square Error

$$E_{MSE}(y, \hat{y}) = ||y - \hat{y}||^2$$

由方程式可知，他是直接計算模型輸出與真實值的距離，因此非常適合用於回歸問題，例如氣溫變化等。

Cross Entropy

$$E_{CE}(y, \hat{y}) = - \sum_{i=1}^{D_{out}} y_i \log \hat{y}_i$$

通常 y 會是一個 N 維的 one-hot 編碼 (N 代表分類數量)，也就是只有一個值為 1，其他皆為 0，這個 1 就代表 data 的分類；而 \hat{y} 會是一個機率分佈，代表模型認為此 data 為各種類的機率為何。

根據方程式我們可以發現，當預測答案為正解的機率很高時，式子所算出的 loss 會很低，例如 $-1 \times \log(0.99) \approx 0.01$ ，而其他值都會是 0；反之當正解被預測得很低時，式子將會算出十分高的 loss，例如 $-1 \times \log(0.1) = 1$ 。因此 Cross Entropy 十分適合應用在分類問題上。

Optimizer:

Optimizer 的作用是根據 Loss Function 的結果來調整模型的參數，讓模型預測得越來越好，最終目的是找到一組參數使 loss 最小化。可以把 Loss 想像成地形圖，Optimizer 的功能就是設計一種算法，根據地形圖的斜率（梯度）讓 loss 往山谷走，達到地形的最低點。

常見的 Optimizer 有 SGD、Adam 等，SGD 就是單純的觀察梯度進行參數的更新，而 Adam 等其他更複雜的 Optimizer 還會根據動量或其他因素動態更新參數，讓模型訓練收斂更快更穩定。

3. 心得

本次的實驗是 module 2 的第一次課程，先簡單的講解無線通訊以及人工智慧最基本的知識，讓我們了解課程圍繞的主題。而本次的實驗內容是很基礎的手寫數字分類，讓我們熟悉環境以及 model training 相關程式語法的使用。基本上我們都很快的完成作業，希望可以在未來的課程中順利完成並學習到 AI 在無線通訊的應用。