**National Sun Yat-sen University**

**Department of Electrical Engineering**

**112-2 機器學習系統設計實務與應用**

**HW2 Keras神經網路應用**

**姓名: 胡庭翊**

**學號: B103015006**

**組別: 你說的隊**

1. Dataset MNIST

以下會針對TensorFlow官方提供的手寫數字資料集MNIST建構不同的NN及CNN架構訓練，並且從不同可能變因探討架構之間不同有何影響。

* 1. 對照組

首先先就範本內的初始架構做預測：

* + 1. 網路架構

|  |  |
| --- | --- |
| NN | CNN |
|  |  |

表一、網路架構

* + - 1. NN網路架構分析

這裡的NN模型首先是先將訓練與測試資料從三圍轉為四維，接著再使用網路架構，將三維的輸入轉換成一維後，使用ReLU做為激勵函數增加類神經網路內的非線性特徵，再使用Softmax放置在hidden layer的最後一層。

ReLU的優點是比起其他函數更能解決梯度消失的問題，計算量小較節省資源，以及其全有全無的特性更接近生物神經網路的模型。其中，梯度消失指的是當建立深層的類神經網路時，在接近輸出端的類神經元更新速度與接近輸入端的類神經元速度不一致，而導致接近輸入層的網路無法作用，實際上整個網路架構只有少數層運作，失去深度學習的意義。

Softmax 回歸是使用Softmax 運算使得最後一層輸出的機率分佈總和為 1，將最後一層的所有節點輸出都通過指數函數，再將結果相加作為分母、個別輸出作為分子，通常都會放置在類神經網路的最後一層。

* + - 1. CNN網路架構分析

至於CNN模型結構則是由兩層卷積層加上池化層，後面再加上NN組成，使用Keras高階封裝的神經框架，搭建網路透過tensorflow運作。

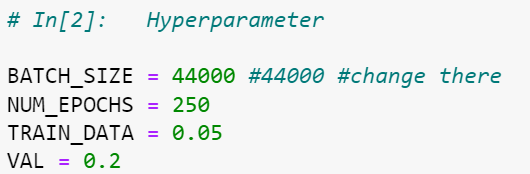
範本中首先經過卷積層Conv2D(8, (5, 5), strides = (1, 1), activation = 'relu', name = 'conv1',padding='same')(X\_input)，使用8個5x5的filter配合固定的權重做到影像處理，提取特徵。並且將strides設為(1, 1)，意指卷積在圖像上更新的xy方向移動部署是所有的點逐格掃描。同時，除了增加ReLU激勵函式外，也使用padding補邊方法，針對輸入的維度做擴增，使得最後的輸出結果與原輸入大小相同。

池化層的部分MaxPooling2D((2, 2), name='max\_pool1\_W1')(W1)使用最大值方法 maxpool-pooling，保留每個2x2資訊中的最大值，再將其餘資訊丟棄，將資訊壓縮簡化以增加模型的收斂及穩定性。

最後再經過一次的卷積層Conv2D(16, (3, 3), strides = (1, 1), activation = 'relu', name = 'conv2')(W1) ，使用16層3x3的filter做影像處理。除了filter的層數與大小，還有這次沒有使用padding之外，其餘都與前面的卷積層相仿，而後續的NN架構亦類似於上面的NN架構。

* + 1. 超參數設置

超參數指的是訓練神經網路時需要手動設定的參數。這裡，我將每次迭代時從訓練集中提取，送入神經網路的資料數量設為44000，提取其中的0.05作為訓練資料的比例，其中的0.2作為驗證的資料比例，進行總計250次的迭代訓練。也就是說，共有2200份資料做為資料集，8800份資料作為驗證集。



圖一、超參數設置

* + 1. 準確度
       1. 訓練

|  |  |
| --- | --- |
| NN |  |
| CNN |  |
| NN、CNN比較 |  |

表二、訓練準確度

損失函數介於0到1之間，用以評估兩個機率分配有多接近。而訓練模型的目的便是讓損失函數最小化。

從結果可知，NN測試表現損失函數較大(0.41491)，而CNN的損失函數較小(0.17929)，這也反映到了測試結果的正確性。另外，從結果圖可以發現CNN模型在做跌代時數值也有微幅的震盪，以訓練精準度而言，CNN模型可以更精確地尋找損失函數的最低值，不會在遇到local minima時便誤以為是global minima，然後停止繼續搜尋。

* + - 1. 測試

|  |  |
| --- | --- |
| NN |  |
| CNN |  |

表二、測試準確度

從train以及

* 1. Activation Function
  2. Layer of Network
  3. Loss
  4. Optimizer

1. 高鐵驗證碼
   1. 網路架構

圖一、網路架構

* 1. 超參數設置
  2. 預處理方式
  3. 準確度
     1. 訓練
     2. 測試
  4. 分析與討論