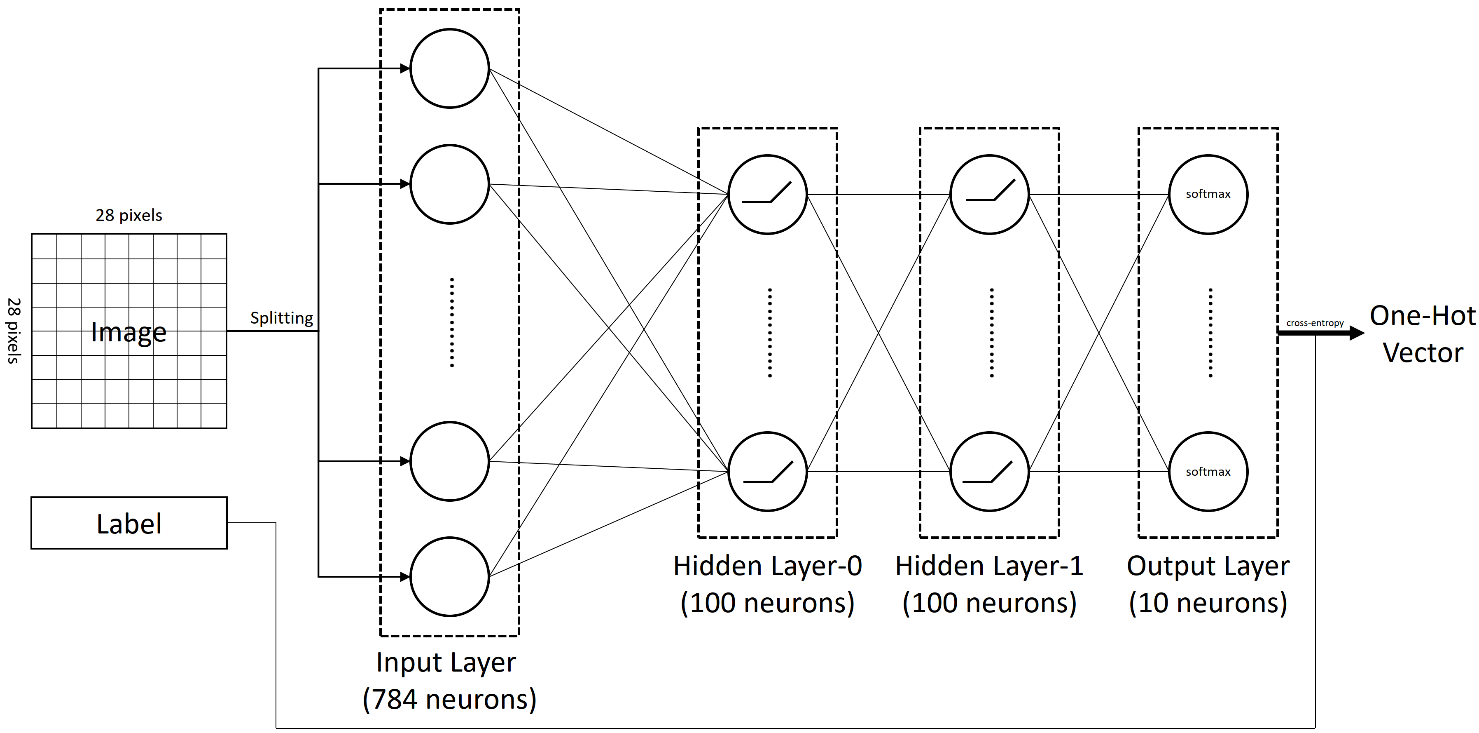
Homework 1

10910COM 526000 Deep Learning　王傳鈞　109062631

第一題

關於我的neural network（NN）架構，可以參考下圖：



圖一：架構圖（我的NN不含任何bias參數）

因為輸入的訓練資料集（training set）是六萬張的28x28畫素照片，所以NN的輸入層（input layer）總共有784個節點。經過一些的文獻查找，我發現過往的研究結果大多只需要一至二層的隱藏層（hidden layer），就可以得到不錯的正確性（accuracy）。另外，我的NN對於每張輸入的照片，會輸出一個one-hot向量，向量當中的每一個元素值分別代表NN對於該圖片會是某個數字的機率大小，例如one-hot向量「index = 0的元素」就是該圖片可能是「數字零」的機率。

經過大約138.04秒之後，就可以得到使用上圖架構訓練出的NN模型，預測測試資料集（testing set）有accuracy約為90.2003%。

第二題

* Forward propagation

基本上就是拿資料，放入NN model當中，然後取得預測值，因此，流程就如同圖一的架構圖所示：圖片放入input layer之後，經過每一層的neurons進行矩陣乘法與ReLU函數轉換，進入最後一層的output layer做矩陣乘法、softmax函數轉換，最後再與label一起計算cross-entropy，產生one-hot向量。

* Backward propagation

因為output layer與hidden layer的activation function不同，所以進行backward propagation時所使用的偏微分函數也不同。（以下的代表位於某層的第i個neuron，把前一層節點的輸出值經過矩陣運算之後的值）

* + Output layer
  + Hidden layer

第三題

因為本次作業使用的資料集不是太複雜，所以NN模型經過約10次的epoch之後，對於training loss和validation loss的變動，都趨向於收斂到某個定值，所以以下的圖表就以epoch總數為10的情況來繪製。

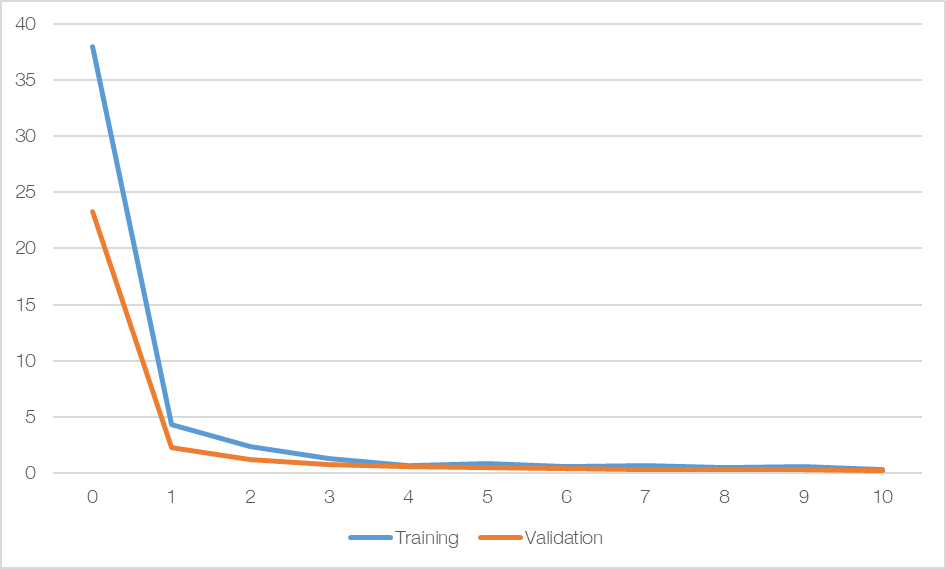


Figure: Total loss on training set & validation set

(X-axis: the number of epoch, Y-axis: C-E loss value)

第四題

一般來說，適度地增加NN的層數，以及每一層neuron的個數，可以讓NN抽取出training data的更多特徵（feature），進而訓練出更好的模型，對於testing data的預測正確率（accuracy）越高。然而，這不代表增加NN的層數與每一層neuron的個數，就可以無止盡地保持增加預測testing data的accuracy。

抽取過多training data的features，會讓訓練模型朝向過度學習（overfitting）的趨勢，造成我們得到模型在training data有極低的accuracy，但是testing data的accuracy卻沒有同步下降或甚至突然暴增。

在這次的作業當中，MNIST資料集雖然是784維的向量，但是因為每一筆資料其實有許多的地方都是零（對應到圖片檔的純黑像素），所以真正需要抽取的feature其實不多。根據我的驗證發現：假設每一層hidden layer的neuron個數都是50個，使用二層hidden layer與使用三層去訓練出來的模型，對於testing data的accuracy只有相差不到1%。由此可見，增加NN的層數並不會帶來很大的改進幅度，甚至可能因為收取太多的feature，而陷入overfitting。

第五題

對於訓練出一個多層NN（deep neural network）來說，越多的資料加入訓練過程，就可以得到越不容易overfitting的模型；但是，蒐集資料往往是很困難、昂貴的，所以目前手上握有的資料就顯得彌足珍貴。

因為我們只能利用已握有的資料來評估是否模型遭遇overfitting，並沒有訓練玩模型之後就會憑空多出其他testing data；因此，不到最後關頭是不能輕易使用testing data參與訓練或是評估accuracy。

在這樣子的概念之下，我們決定如何挑選特定的hyper-parameters時（例如：進行幾次epoch的mini-batch SGD才足夠），就只能依靠validation set。先把training data分成兩堆—validation set和training set，並讓validation set暫時扮演testing data的角色，不參與訓練、只用來評估accuracy。如此一來，我們就可以盡可能地推延testing data進入倒訓練模型的時間點，盡可能的善用每一筆data隱藏的特徵。