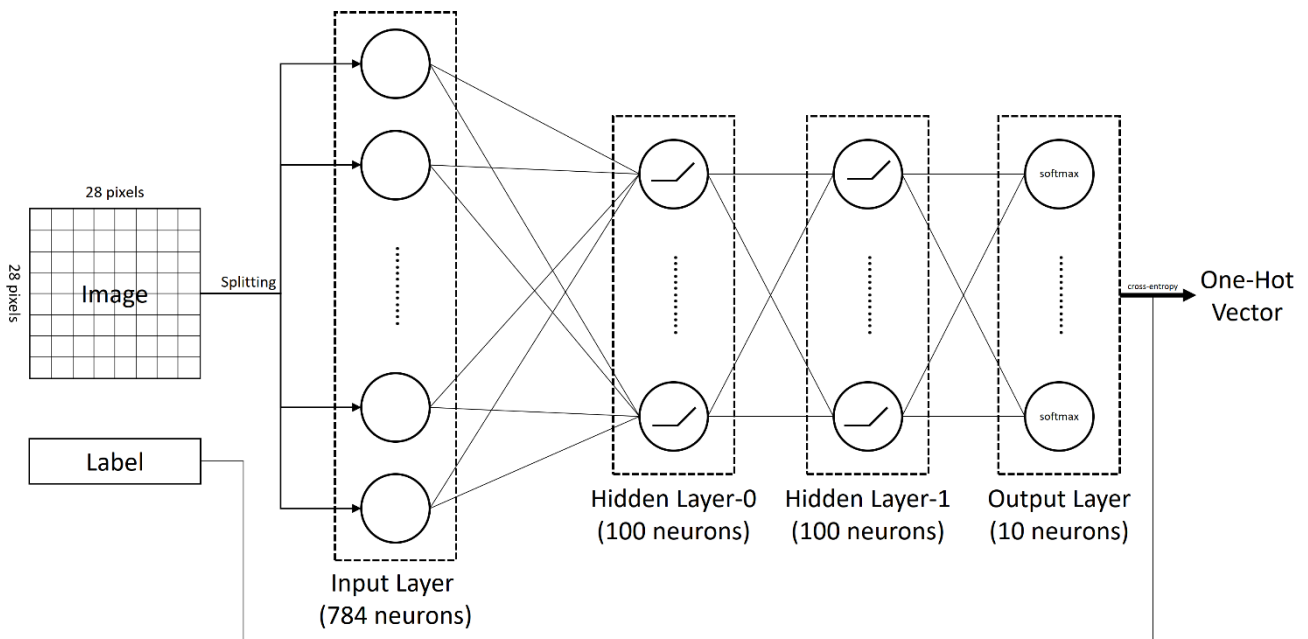


Homework 1

10910COM 526000 Deep Learning 王傳鈞 109062631

第一題

關於我的 neural network 架構，可以參考下圖：



圖一：架構圖

因為輸入的訓練資料集 (training set) 是六萬張的 28x28 畫素照片，所以 neural network 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 neural network 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 neural network 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。

第二題

- Forward propagation

基本上就是拿資料，放入 neural network 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 z_i 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

$$\frac{\partial Loss}{\partial z_i} = \text{softmax}(z_i) - \mathbb{I}(i == \text{label}) \text{ for } i \in \{0, \dots, 9\}$$

$$\mathbb{I}(i == \text{label}) = \begin{cases} 1, & \text{if } i \text{ is equal to label} \\ 0, & \text{otherwise} \end{cases}$$

- Hidden layer

$$\frac{\partial Loss}{\partial z_i} = \mathbb{I}(z_i > 0)$$
$$\mathbb{I}(z_i > 0) = \begin{cases} 1, & \text{if } z_i \text{ is greater than one} \\ 0, & \text{otherwise} \end{cases}$$

第三題

一般來說，適度地增加 neural network 的層數，以及每一層 neuron 的個數，可以讓 neural network 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

第四題

一般來說，適度地增加 neural network 的層數，以及每一層 neuron 的個數，可以讓 neural network 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 neural network 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 100 個，使用一層 hidden layer 與使用二層去訓練出來的模型，對於 testing data 的 accuracy 只有相差不到 3%。由此可見，增加 neural network 的層數並不會帶來很大的改進幅度。

第五題

對於訓練出一個 deep neural network 來說，越多的資料參與訓練的過程，就可以得到出現 overfitting 機會越低的模型；但是，蒐集資料往往是很困難的，所以目前握有的資料就顯得彌足珍貴。我們唯一可以利用已握有的資料來評估是否模型遭遇 overfitting 的方式，就只有透過比較模型對於 training data accuracy 與 testing data accuracy 之間的差異。因此，不到最後關頭是不能輕易使用 testing data。

在這樣子的概念之下，我們決定到底要經歷幾次的 epoch 來使用 mini-batch SGD 更新模型參數才算足夠，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，意即不參與訓練只用來評估 accuracy。如此一來，就只會最後一次總評估模型預測能力時，才真正動用到 testing data。