**HW1**

**109062580李佳陽**

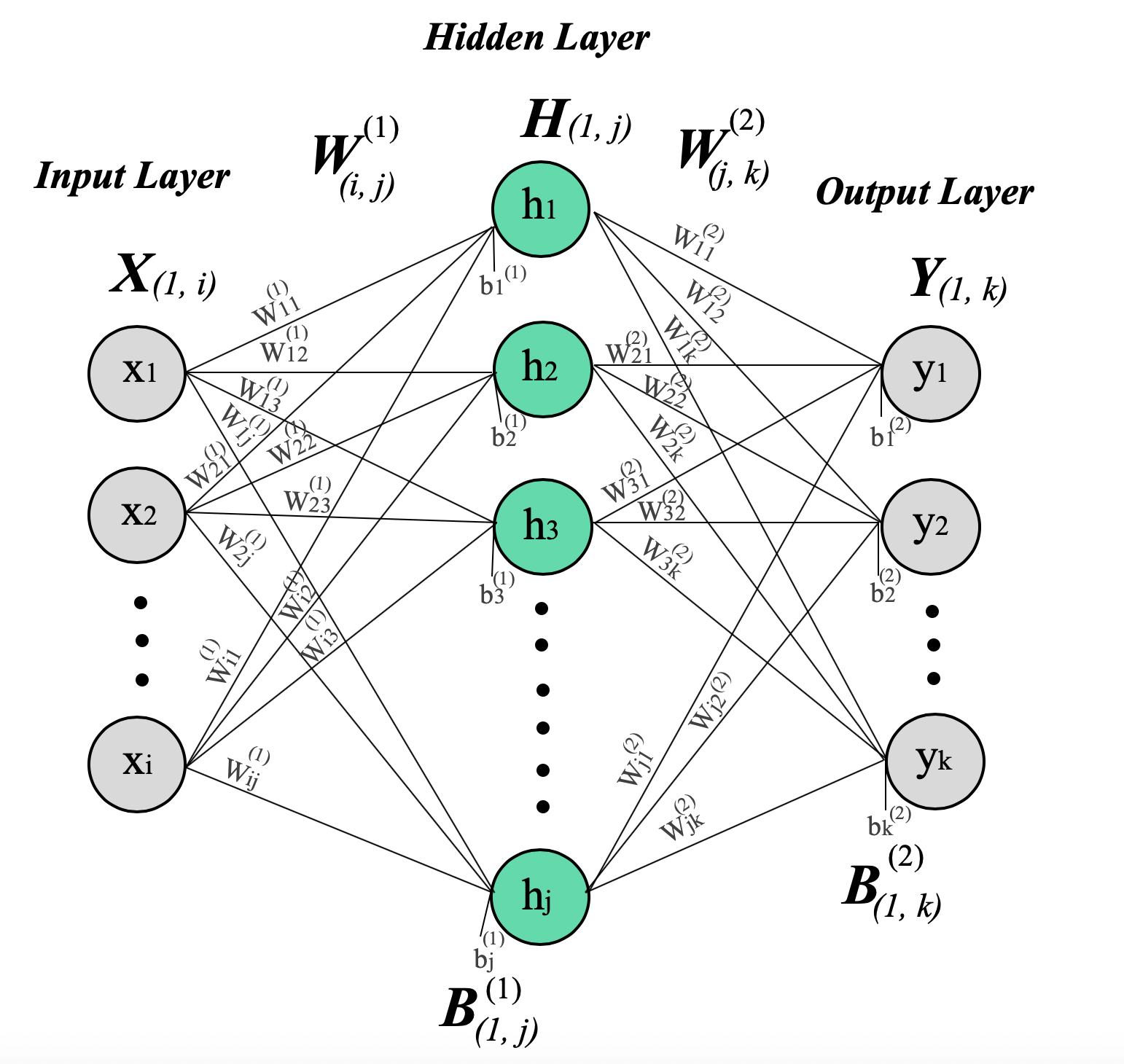
1. Model architecture:使用3層，結構如下圖

Input layer: 784 units

Hidden layer: 300 units

Output layer: 10 units

其中Hidden layer為Dense layer(全連接層)，輸出透過relu作為下一層輸入;Output layer為Dense layer(全連接層)，輸出透過softmax輸出10個class的機率。



Testing accuracy:98.03%

1. Forward pass:

Input layer:x

Input layer -> Hidden layer, z1 = w1 \* x + b1

ReLU layer, a1 = ReLU(z1)

Hidden -> output layer, z2 = w2 \* a1 + b2

Output layer, a2 = Softmax(z2)

Backward pass:

According to the chain rule.

Derivative of z2: d\_z2 = predict y – ground truth y

Gradients at last layer: d\_w2 = d\_z2 \* a1, d\_b2 = d\_z2

Back propagate through first layer:

d\_a1 = w2 \* d\_z2, d\_z1 = d\_a1 \* Relu ' (z1)

Gradients at first layer: d\_w1 = d\_z1 \* a0, d\_b1 = d\_z1

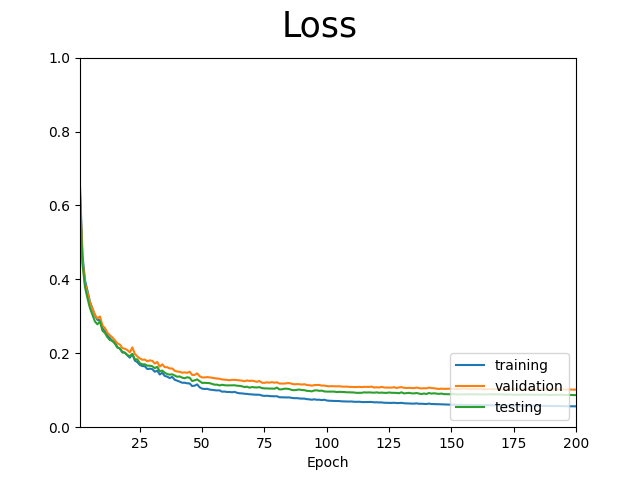
Update network parameters according to update rule from Stochastic Gradient Descent.

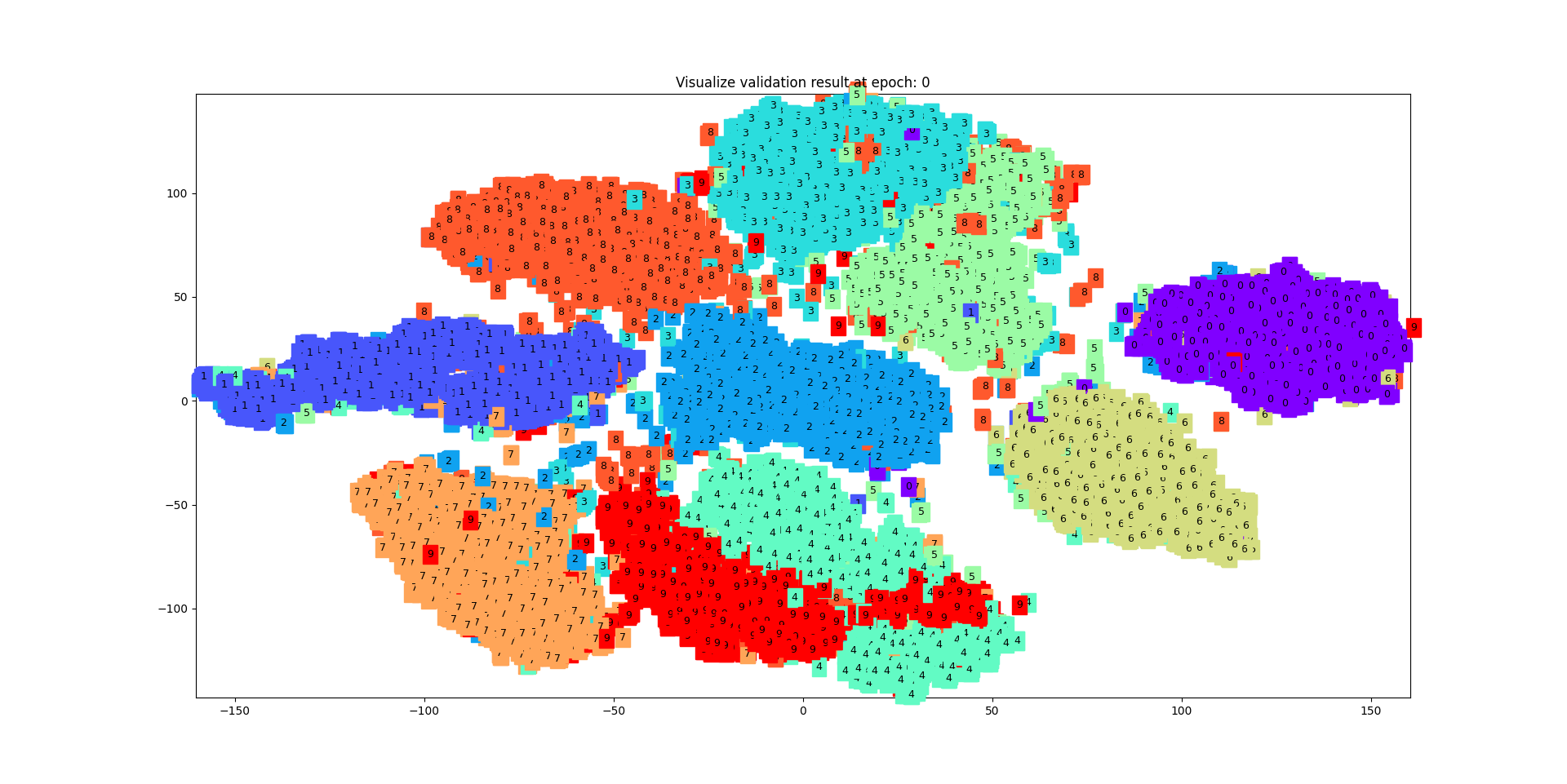
w1 = w1 – learning rate \* d\_w1

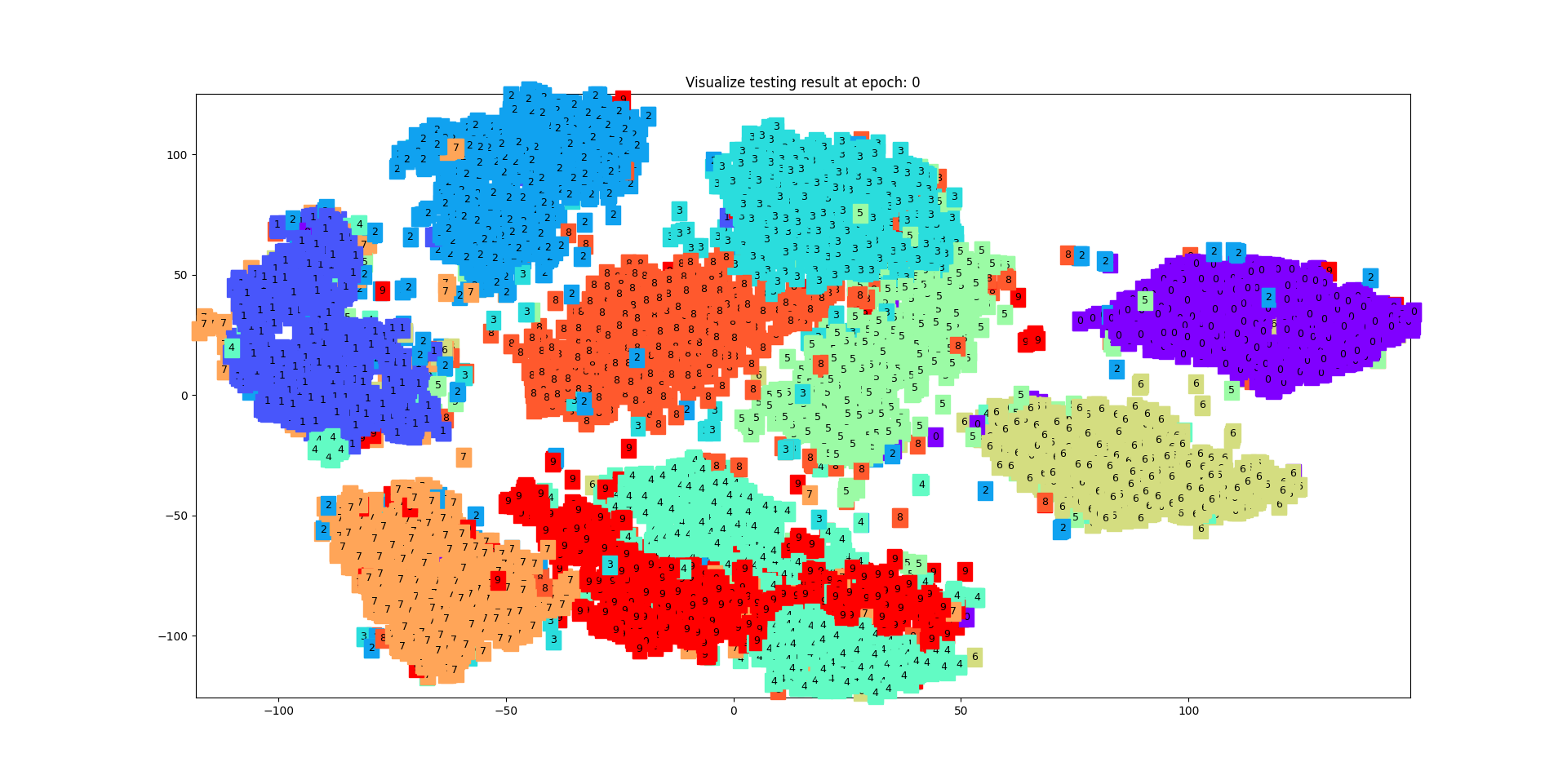
b1 = b1 – learning rate \* d\_b1

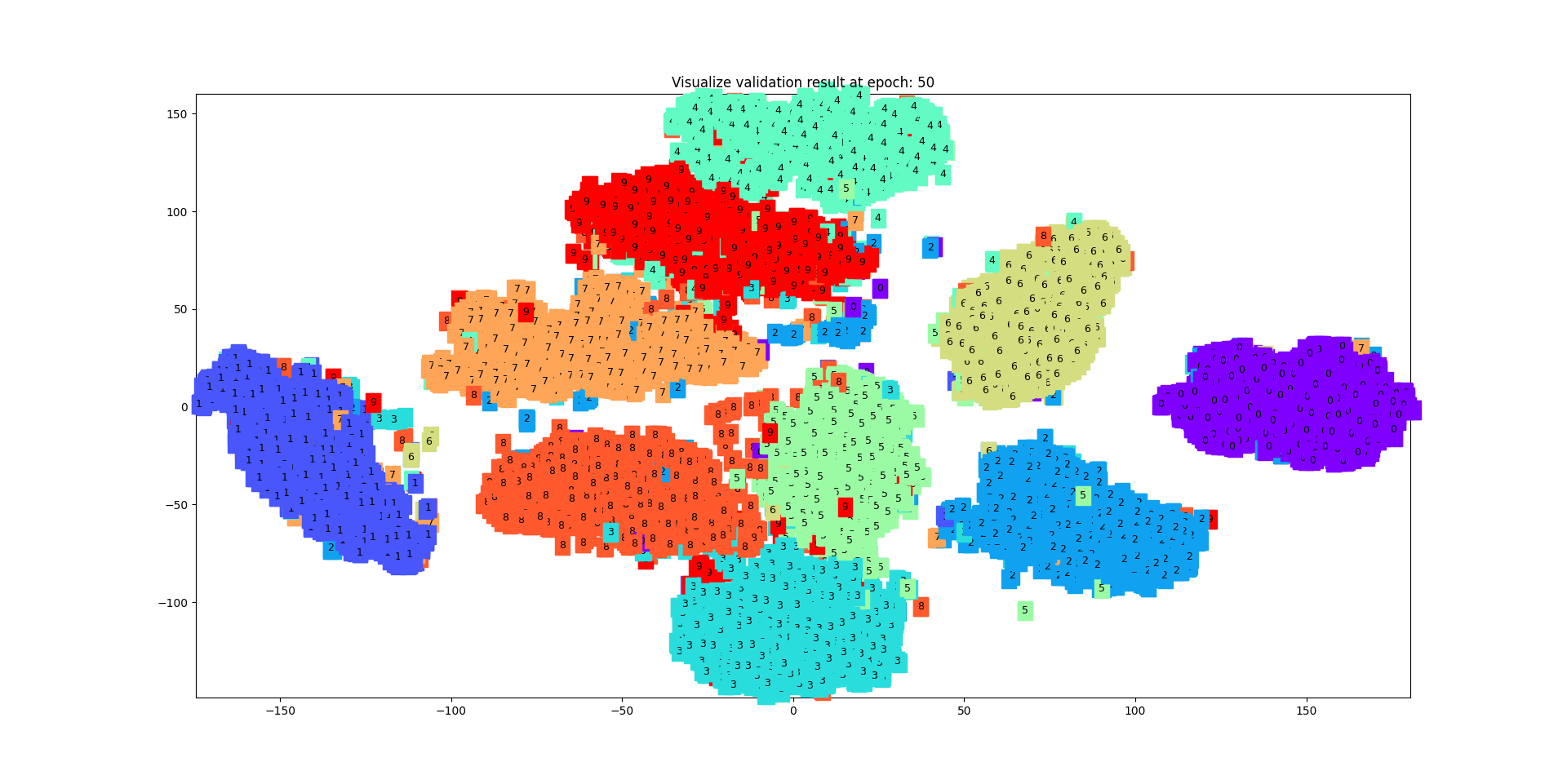
w2 = w2 – learning rate \* d\_w2

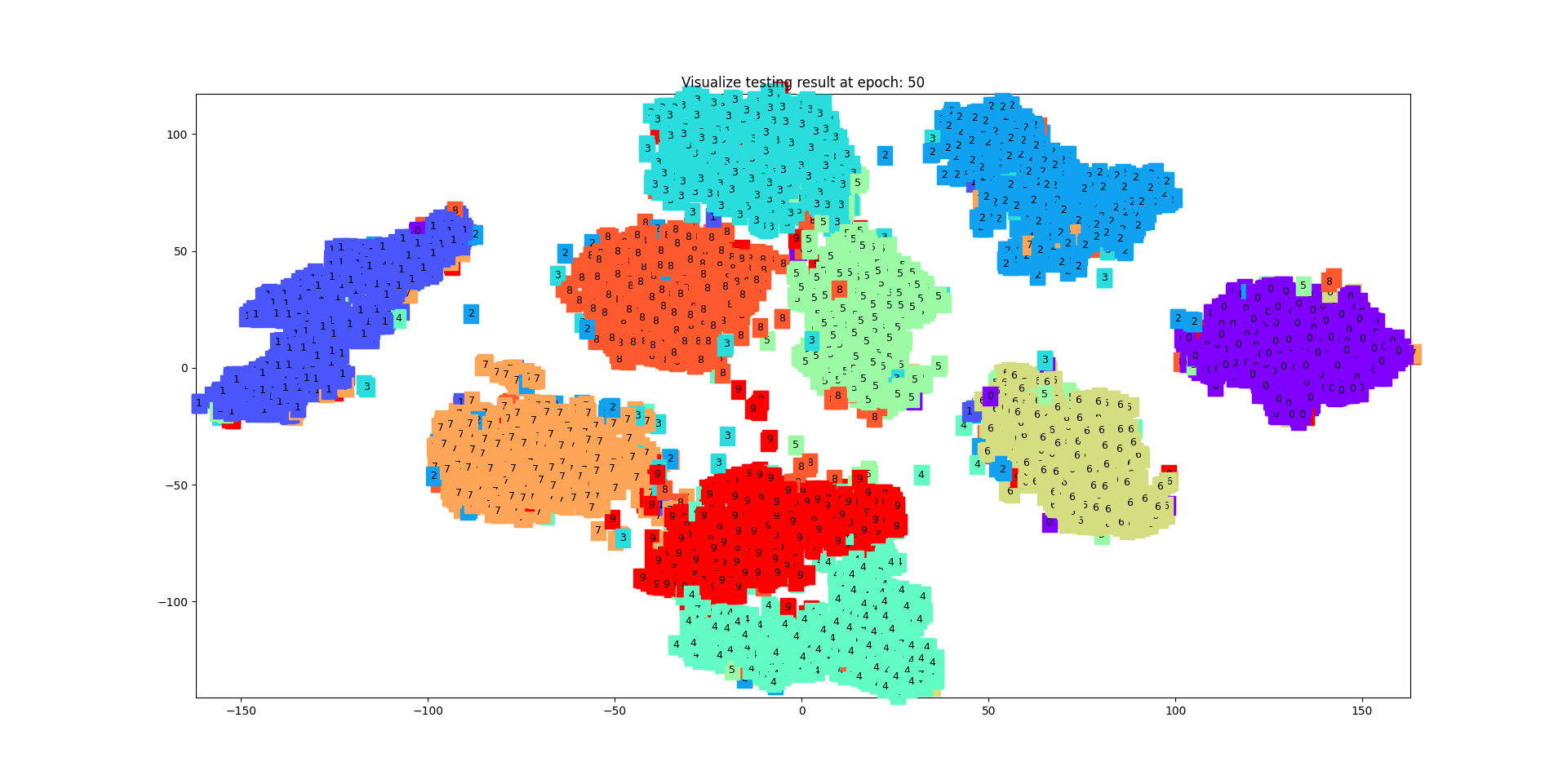
b2 = b2 – learning rate \* d\_b2

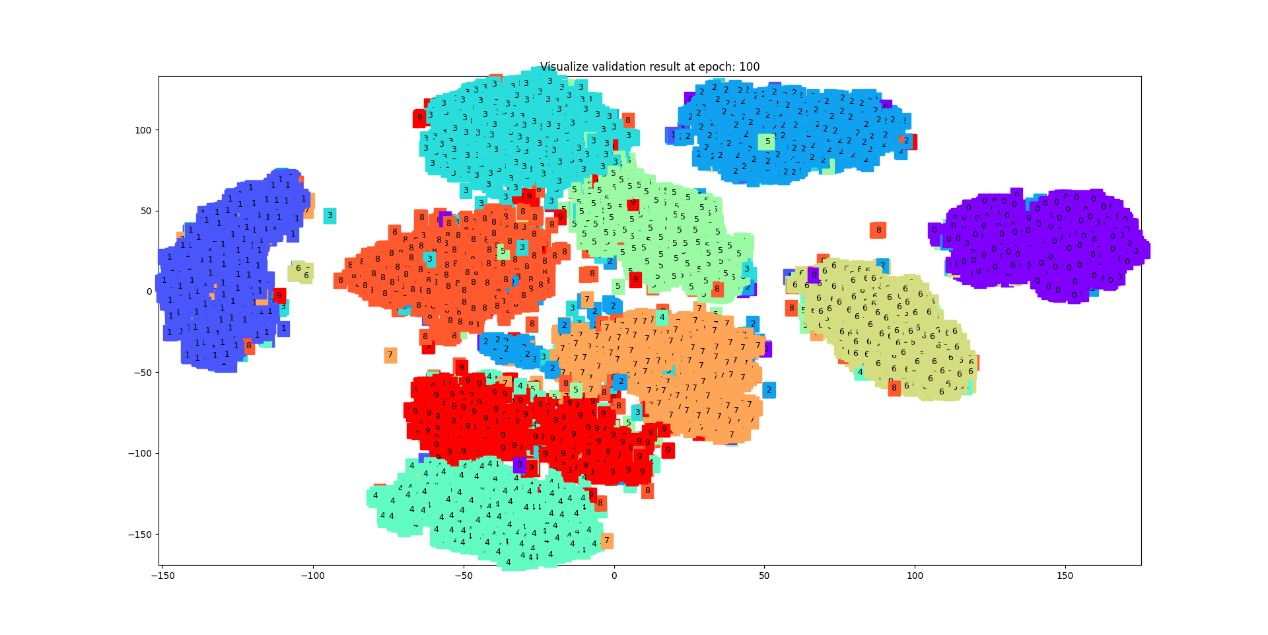
1. Training loss, validation loss and testing loss，其中，每個epoch的loss都是對training data 全部迭代優化後所計算出。
2. 當使用更深及更多的神經元時，需要更多的data去做訓練，而深度越深可能發生Vanishing gradient problem，在backward propagation會沒有梯度傳遞，導致model的參數無法更新。
3. Validation 的步驟最主要是在training過程使用，用來驗證model是否發生overfitting。
4. t-SNE results(Bonus):epoch 0的結果為未訓練的結果，之後，每個epoch的結果為對training data全部迭代優化後的結果。若傳入資料量大，程式需執行很久。

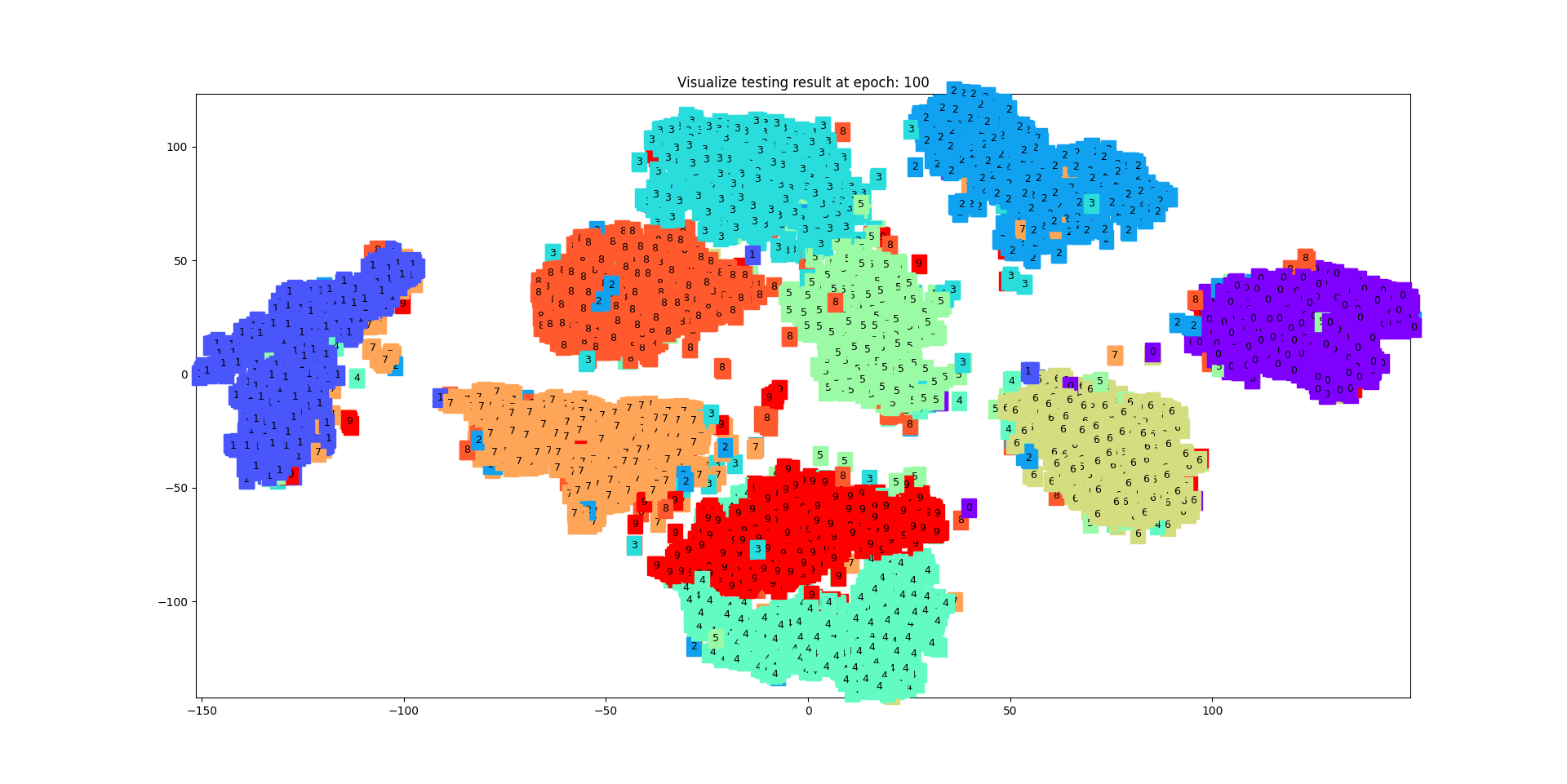
validation at epoch 0:

testing at epoch 0:

validation at epoch 50:

testing at epoch 50:

validation at epoch 100:

testing at epoch 100:

從以上的結果可發現，本model對於4和9較易分辨錯誤。