Deep Learning
Assignment 1
7 points
1. Linear Softmax
a. [2 marks]
The output is as following:
0.9198
```
[[ 952    0    4    2    0    8    9    1    4    0]
 [   0 1106    2    2    0    2    4    2   17    0]
 [   4    6  934    8    6    5    9   11   38   11]
 [   3    1   27  898    0   42    1   10   19    9]
 [   1    1   10    1  867    2   11    3   14   72]
 [   8    3    6   26    5  792   10    5   30    7]
 [   8    3   10    1    8   18  903    2    5    0]
 [   3    5   22    8    5    2    0  940    3   40]
 [   4    4    7   21    7   29    9   10  876    7]
 [   8    5    2   10   12   10    0   19   13  930]]
```

Here are the misclassification with number of occurences>20:
2->8(38), 3->2(27), 3->5(42), 4->9(72), 5->3(26), 5->8(30), 7->2(22), 7->9(40), 8->3(21), 8->5(29).

We can see there is high rate for misclassification on 2->8, 3->5, 4->9, 5->8, 7->9, 8->5 etc. For example 4->9 is very common misclassification due to handwriting on 4. the same as 7->9 etc.
2 has many different handwriting style that is easily to be misclassfied as 8.

b. [1 mark]
Given that there are 28 x 28 = 784 inputs and 10 outputs,
total trainable params = (784+1)*10 = 7850 trainable params.

2. Two-Layer Neural Network
a. [1 mark]
After improving the code, the output is as following:
0.9789
```
[[ 968    0    2    0    0    4    1    2    2    1]
 [   0 1128    1    1    0    1    2    1    1    0]
 [   4    0 1005    7    5    1    1    4    4    1]
 [   0    0    3  995    1    1    0    2    4    4]
 [   1    0    3    1  960    1    5    1    1    9]
 [   5    0    1   12    1  860    6    0    5    2]
 [   5    2    0    1    3    4  943    0    0    0]
 [   0    2   10    2    3    0    1 1000    4    6]
 [   4    1    0    7    3    2    3    4  948    2]
 [   0    2    0    4    9    4    1    5    2  982]]
```

The top five misclassifications: 5->3, 7->2, 9->4, 4->9, 8->3
Yes, these are the kinds of errors that humans might also make if digits were poorly written by taking a look at the images on chapter 6 of http://neuralnetworksanddeeplearning.com/chap6.html

b. [1 mark]

Assuming an MLP with 100 hidden nodes, calculate the total number of weights in (i) the hidden layer and (ii) the output layer, of the MLP model.
(784+1)*100 + (100+1)*10 = 79510 trainable params

c. [1 mark]
n_hidden=100: 0.9798, 0.9793, 0.9798
n_hidden=50: 0.972, 0.9728, 0.971
n_hidden=20: 0.9535, 0.9541, 0.9535
I also tried n_hidden=200 for once, with accuracy=0.9822.
We can see when n_hidden value decrease, the accuracy decrease too.

3. [1 mark]
Write the equations for each of these activation functions:
sigmoid, tanh, ReLU

sigmoid: $S(z)=1/(1+\exp(-z))$

tanh: $\tanh(z)=(\exp(z)-\exp(-z))/(\exp(z)+\exp(-z))/$

RuLU: $R(z) = \max(0, z)$