HW6 report

In this homework, I use a 3 layers network. The numbers of nodes in input, hidden and output layer are seperately 784, 50, 10. The activation functions I used are firsly Relu, secondly softmax, the weights initialized $W_i \sim U[-\frac{\sqrt{6}}{\sqrt{n_{i-1}+n_i}}, \frac{\sqrt{6}}{\sqrt{n_{i-1}+n_i}}]$, learning rate is 0.1. I used SGD as optimizer. The epoch is 50 and mini-batch is 100.

The screenshot of last 20 SGD iterations:

$$= 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots , \qquad -\infty < x < \infty$$

```
Epoch: 0031 cost= 0.039767262
Epoch: 0032 cost= 0.038331559
Epoch: 0033 cost= 0.037152905
Epoch: 0034 cost= 0.035896467
Epoch: 0035 cost= 0.034462831
Epoch: 0036 cost= 0.033326531
Epoch: 0037 cost= 0.032292291
Epoch: 0038 cost= 0.030971272
Epoch: 0039 cost= 0.030077547
Epoch: 0040 cost= 0.029152687
Epoch: 0041 cost= 0.028074718
Epoch: 0042 cost= 0.027294771
Epoch: 0043 cost= 0.026334689
Epoch: 0044 cost= 0.025474672
Epoch: 0045 cost= 0.024632952
Epoch: 0046 cost= 0.023773384
Epoch: 0047 cost= 0.022998822
Epoch: 0048 cost= 0.022287669
Epoch: 0049 cost= 0.021640176
Epoch: 0050 cost= 0.020960004
Optimization Finished!

The accuracy on the testing set:  97.60%
The cross-entropy cost on the testing set: 0.091
```