**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 , 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:



