Report for Assignment 3 ----- Hu Xin xh2390

1. Learning rate:
   From the three experiments(learning rate = 0.01, 0.02,.0.025) , we can see that the experiment 1(learning rate = 0.01) has best performance (high accuracy and low loss). As we know, learning rate tells the optimizer how far to move the weights in the direction opposite of the gradient for a mini-batch. If the learning rate is high, then training may not converge or even diverge. Weight changes can be so big that the optimizer overshoots the minimum and makes the loss worse.

2. Activation
   From the three experiments (relu + softmax, tanh + softmax and relu + sigmoid), we can see that the experiment(relu + softmax) has the best performance. the ReLU is half rectified and when we use relu and softmax which is used for Multiclass, single-label classification. So in our case, relu + softmax is better.

3. Optimizers
   From the three experiments(adam, sgd, nadam), we can see that the experiment (nadam) has the best performance. Adam can be viewed as a combination of RMSprop and momentum, and Nadam is like combines Adam and NAG. And Nadam is slightly better than adam in this test case.

4. Weight
   From the three experiments(zeros, ones, constant(2)), we can see that the experiment (zeros) has the best performance.