Practical 3

Zheng Zhaoheng

1. The code now evaluates its performance on the test set after every 10 minibatches, and the plot with test loss and training loss is shown below:

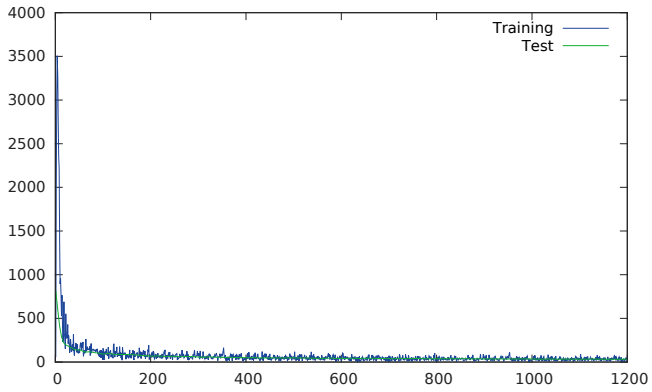

Figure 1: Loss

And the code doing this work is:

```
if i % 10 == 0 then
    local test_outputs = model:forward(test.data)
    local test_loss = criterion:forward(test_outputs, test.labels)
    test_losses[#test_losses + 1] = test_loss;
end
gnuplot.pdffigure("loss.pdf")
gnuplot.plot({torch.range(1, #losses),
    torch.Tensor(losses),
         '-'},      -- Training loss
    {torch.linspace(1, #losses, #test_losses),
    torch.Tensor(test_losses),
    '~'})       -- Test loss
gnuplot.plotflush()
```

This code calculate the test loss every 10 minibatches and plot both the test loss ans the training loss on the same plot.

2. In order to find a configuration that works well on the MNIST dataset, I tried the three optimizers and tuned their parameters respectively. In all cases, I set the size of batch to 100 and train the model with all the 60000 training images through 2 epochs.

   This table below shows the best result in my experiments on all three models:

	Test Set	Training Set
lbfgs	9.01%	9.26%
sgd	8.69%	8.58%
adagrad	9.22%	8.98%

Table 1: Classfication Error

And the parameters of these 3 optimizers are:

```
if opt.optimization == 'lbfgs' then
  optimState = {
    learningRate = 1,
    maxIter = 2,
    nCorrection = 10
  }
  optimMethod = optim.lbfgs
elseif opt.optimization == 'sgd' then
  optimState = {
    learningRate = 1e-2,
    weightDecay = 0,
    momentum = 0.8,
    learningRateDecay = 1e-2
  }
  optimMethod = optim.sgd
elseif opt.optimization == 'adagrad' then
  optimState = {
    learningRate = 2,
  }
  optimMethod = optim.adagrad
end
```

I finally pick the SGD optimizer trained with parameters above as the final model because it has the lowest training set and test classification error rate. But the ADAGRAD optimizer has only one parameters to tune. Thus, the ADAGRAD optimizer is the one easier to configure than others.