

Programming Exercise

Problem 1

Hyper Parameters:

Learning Rate=0.1

Number of epochs=200

Batch size =784

Hidden layers = [128,128]

Sigmoid function is used as activation function.

Results:

For 200 epochs:

Training loss = 0.219

Training error = 0.0646

Validation Loss = 0.2746

Validation error = 0.082

Test Loss = 0.285

Test error = 0.0843

```
epoch: 199
training loss: 0.21921003481259824
training error: 0.0646
validation loss: 0.27461911089040125
validation error: 0.082
test loss: 0.28575270933317304
test error: 0.0843
```

