第 2 次作业报告

张家治（1601110508）

针对 mnist 的学习问题，构建神经网络，其中包含一层 ReLU 隐层（100 个神经元），输出层为 softmax 层，并采用交叉熵作为损失函数。

尝试使用不同的优化算法训练网络，并测试网络的能力。训练都是基于随机 batch 的，batch 大小为 100，以精确度变化小于 0.001 作为停止条件。测试的网络能力包括：训练用时、训练集准确率、测试集准确率。

针对普通 SGC 优化算法，测试了不同学习率的效果。针对其他的优化算法（包括 Momentum、Nesterov、AdaGrad、RMSProp、Adam），不仅测试了不同学习率的效果，还测试了不同的主要超参数的学习效果。全部测试结果如下。

SGD:
    Learning rate: 0.005000
    Time used: 9s
    Final accuracy: 0.896345 (train) ; 0.901300 (test)
SGD:
    Learning rate: 0.010000
    Time used: 9s
    Final accuracy: 0.912455 (train) ; 0.917100 (test)
SGD:
    Learning rate: 0.030000
    Time used: 9s
    Final accuracy: 0.949000 (train) ; 0.946900 (test)
SGD:
    Learning rate: 0.050000
    Time used: 13s
    Final accuracy: 0.972091 (train) ; 0.966100 (test)
SGD with momentum:
    Learning rate: 0.005000
    Momentum: 0.900000
    Time used: 12s
    Final accuracy: 0.968145 (train) ; 0.963200 (test)
SGD with momentum:
    Learning rate: 0.010000
    Momentum: 0.900000
    Time used: 9s
    Final accuracy: 0.977418 (train) ; 0.971000 (test)
SGD with momentum:

Learning rate: 0.030000
Momentum: 0.900000
Time used: 6s
Final accuracy: 0.986691 (train) ; 0.975200 (test)
SGD with momentum:
Learning rate: 0.050000
Momentum: 0.900000
Time used: 8s
Final accuracy: 0.995800 (train) ; 0.979000 (test)
SGD with momentum:
Learning rate: 0.010000
Momentum: 0.500000
Time used: 12s
Final accuracy: 0.942364 (train) ; 0.940900 (test)
SGD with momentum:
Learning rate: 0.010000
Momentum: 0.900000
Time used: 8s
Final accuracy: 0.975291 (train) ; 0.967100 (test)
SGD with momentum:
Learning rate: 0.010000
Momentum: 0.990000
Time used: 5s
Final accuracy: 0.988145 (train) ; 0.973600 (test)
SGD with Nesterov-momentum:
Learning rate: 0.005000
Momentum: 0.900000
Time used: 11s
Final accuracy: 0.967818 (train) ; 0.963200 (test)
SGD with Nesterov-momentum:
Learning rate: 0.010000
Momentum: 0.900000
Time used: 8s
Final accuracy: 0.977709 (train) ; 0.969500 (test)
SGD with Nesterov-momentum:
Learning rate: 0.030000
Momentum: 0.900000
Time used: 3s
Final accuracy: 0.977727 (train) ; 0.970000 (test)
SGD with Nesterov-momentum:
Learning rate: 0.050000
Momentum: 0.900000
Time used: 2s
Final accuracy: 0.982345 (train) ; 0.970600 (test)

SGD with Nesterov-momentum:
    Learning rate: 0.010000
    Momentum: 0.500000
    Time used: 10s
    Final accuracy: 0.937527 (train) ; 0.937400 (test)
SGD with Nesterov-momentum:
    Learning rate: 0.010000
    Momentum: 0.900000
    Time used: 5s
    Final accuracy: 0.965982 (train) ; 0.960000 (test)
SGD with Nesterov-momentum:
    Learning rate: 0.010000
    Momentum: 0.990000
    Time used: 4s
    Final accuracy: 0.991036 (train) ; 0.972500 (test)
AdaGrad:
    Learning rate: 0.005000
    Initial accumulation: 0.100000
    Time used: 12s
    Final accuracy: 0.923345 (train) ; 0.926300 (test)
AdaGrad:
    Learning rate: 0.010000
    Initial accumulation: 0.100000
    Time used: 15s
    Final accuracy: 0.953727 (train) ; 0.952200 (test)
AdaGrad:
    Learning rate: 0.030000
    Initial accumulation: 0.100000
    Time used: 8s
    Final accuracy: 0.971855 (train) ; 0.965500 (test)
AdaGrad:
    Learning rate: 0.050000
    Initial accumulation: 0.100000
    Time used: 6s
    Final accuracy: 0.974655 (train) ; 0.968700 (test)
AdaGrad:
    Learning rate: 0.010000
    Initial accumulation: 0.100000
    Time used: 9s
    Final accuracy: 0.939527 (train) ; 0.938900 (test)
AdaGrad:
    Learning rate: 0.010000
    Initial accumulation: 0.300000
    Time used: 7s

Final accuracy: 0.919745 (train) ; 0.921300 (test)
AdaGrad:
Learning rate: 0.010000
Initial accumulation: 0.500000
Time used: 12s
Final accuracy: 0.929527 (train) ; 0.930400 (test)
RMSProp:
Learning rate: 0.005000
Initial accumulation: 0.900000
Time used: 7s
Final accuracy: 0.996000 (train) ; 0.976900 (test)
RMSProp:
Learning rate: 0.010000
Initial accumulation: 0.900000
Time used: 10s
Final accuracy: 0.991400 (train) ; 0.971700 (test)
RMSProp:
Learning rate: 0.030000
Initial accumulation: 0.900000
Time used: 2s
Final accuracy: 0.953855 (train) ; 0.947100 (test)
RMSProp:
Learning rate: 0.050000
Initial accumulation: 0.900000
Time used: 9s
Final accuracy: 0.950855 (train) ; 0.937900 (test)
RMSProp:
Learning rate: 0.010000
Initial accumulation: 0.500000
Time used: 3s
Final accuracy: 0.978782 (train) ; 0.968200 (test)
RMSProp:
Learning rate: 0.010000
Initial accumulation: 0.900000
Time used: 8s
Final accuracy: 0.990927 (train) ; 0.973800 (test)
RMSProp:
Learning rate: 0.010000
Initial accumulation: 0.990000
Time used: 6s
Final accuracy: 0.984164 (train) ; 0.968700 (test)
Adam:
Learning rate: 0.005000
Beta1: 0.900000, Beta2: 0.999000

Time used: 4s
Final accuracy: 0.988945 (train) ; 0.974000 (test)
Adam:
Learning rate: 0.010000
Beta1: 0.900000, Beta2: 0.999000
Time used: 13s
Final accuracy: 0.989636 (train) ; 0.970700 (test)
Adam:
Learning rate: 0.030000
Beta1: 0.900000, Beta2: 0.999000
Time used: 1s
Final accuracy: 0.944782 (train) ; 0.936200 (test)
Adam:
Learning rate: 0.050000
Beta1: 0.900000, Beta2: 0.999000
Time used: 2s
Final accuracy: 0.936855 (train) ; 0.928200 (test)
Adam:
Learning rate: 0.005000
Beta1: 0.900000, Beta2: 0.999000
Time used: 10s
Final accuracy: 0.993709 (train) ; 0.972600 (test)
Adam:
Learning rate: 0.010000
Beta1: 0.900000, Beta2: 0.999000
Time used: 6s
Final accuracy: 0.980327 (train) ; 0.966300 (test)
Adam:
Learning rate: 0.030000
Beta1: 0.900000, Beta2: 0.999000
Time used: 19s
Final accuracy: 0.971127 (train) ; 0.952600 (test)
Adam:
Learning rate: 0.050000
Beta1: 0.900000, Beta2: 0.999000
Time used: 3s
Final accuracy: 0.926182 (train) ; 0.920000 (test)
Adam:
Learning rate: 0.010000
Beta1: 0.900000, Beta2: 0.999000
Time used: 9s
Final accuracy: 0.988873 (train) ; 0.972300 (test)
Adam:
Learning rate: 0.010000

Beta1: 0.900000, Beta2: 0.999000
Time used: 3s
Final accuracy: 0.981800 (train) ; 0.969400 (test)

Adam:
Learning rate: 0.010000
Beta1: 0.900000, Beta2: 0.999000
Time used: 3s
Final accuracy: 0.978745 (train) ; 0.965200 (test)