

## 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: 19s

Final accuracy: 0.337000 (train) ; 0.528900 (test)

SGD:

Learning rate: 0.010000

Time used: 19s

Final accuracy: 0.592900 (train) ; 0.802800 (test)

SGD:

Learning rate: 0.030000

Time used: 17s

Final accuracy: 0.857700 (train) ; 0.893400 (test)

SGD:

Learning rate: 0.050000

Time used: 17s

Final accuracy: 0.882600 (train) ; 0.906400 (test)

SGD with momentum:

Learning rate: 0.005000

Momentum: 0.900000

Time used: 19s

Final accuracy: 0.885000 (train) ; 0.905000 (test)

SGD with momentum:

Learning rate: 0.010000

Momentum: 0.900000

Time used: 11s

Final accuracy: 0.910500 (train) ; 0.927100 (test)

SGD with momentum:

Learning rate: 0.030000  
Momentum: 0.900000  
Time used: 10s  
Final accuracy: 0.931400 (train) ; 0.949600 (test)

SGD with momentum:  
Learning rate: 0.050000  
Momentum: 0.900000  
Time used: 5s  
Final accuracy: 0.948100 (train) ; 0.965900 (test)

SGD with momentum:  
Learning rate: 0.010000  
Momentum: 0.500000  
Time used: 19s  
Final accuracy: 0.801000 (train) ; 0.876000 (test)

SGD with momentum:  
Learning rate: 0.010000  
Momentum: 0.900000  
Time used: 10s  
Final accuracy: 0.905200 (train) ; 0.922200 (test)

SGD with momentum:  
Learning rate: 0.010000  
Momentum: 0.990000  
Time used: 6s  
Final accuracy: 0.933100 (train) ; 0.957300 (test)

SGD with Nesterov-momentum:  
Learning rate: 0.005000  
Momentum: 0.900000  
Time used: 9s  
Final accuracy: 0.884600 (train) ; 0.907600 (test)

SGD with Nesterov-momentum:  
Learning rate: 0.010000  
Momentum: 0.900000  
Time used: 13s  
Final accuracy: 0.908600 (train) ; 0.924100 (test)

SGD with Nesterov-momentum:  
Learning rate: 0.030000  
Momentum: 0.900000  
Time used: 6s  
Final accuracy: 0.935700 (train) ; 0.951800 (test)

SGD with Nesterov-momentum:  
Learning rate: 0.050000  
Momentum: 0.900000  
Time used: 5s  
Final accuracy: 0.944400 (train) ; 0.959100 (test)

SGD with Nesterov-momentum:  
Learning rate: 0.010000  
Momentum: 0.500000  
Time used: 12s  
Final accuracy: 0.800100 (train) ; 0.875000 (test)

SGD with Nesterov-momentum:  
Learning rate: 0.010000  
Momentum: 0.900000  
Time used: 12s  
Final accuracy: 0.908000 (train) ; 0.923500 (test)

SGD with Nesterov-momentum:  
Learning rate: 0.010000  
Momentum: 0.990000  
Time used: 7s  
Final accuracy: 0.930200 (train) ; 0.963300 (test)

AdaGrad:  
Learning rate: 0.005000  
Initial accumulation: 0.100000  
Time used: 10s  
Final accuracy: 0.734200 (train) ; 0.846300 (test)

AdaGrad:  
Learning rate: 0.010000  
Initial accumulation: 0.100000  
Time used: 13s  
Final accuracy: 0.842700 (train) ; 0.890000 (test)

AdaGrad:  
Learning rate: 0.030000  
Initial accumulation: 0.100000  
Time used: 10s  
Final accuracy: 0.902500 (train) ; 0.920800 (test)

AdaGrad:  
Learning rate: 0.050000  
Initial accumulation: 0.100000  
Time used: 15s  
Final accuracy: 0.916100 (train) ; 0.935100 (test)

AdaGrad:  
Learning rate: 0.010000  
Initial accumulation: 0.100000  
Time used: 14s  
Final accuracy: 0.846600 (train) ; 0.889300 (test)

AdaGrad:  
Learning rate: 0.010000  
Initial accumulation: 0.300000  
Time used: 15s

Final accuracy: 0.771100 (train) ; 0.865700 (test)

AdaGrad:

Learning rate: 0.010000

Initial accumulation: 0.500000

Time used: 18s

Final accuracy: 0.699000 (train) ; 0.834100 (test)

RMSProp:

Learning rate: 0.005000

Initial accumulation: 0.900000

Time used: 4s

Final accuracy: 0.946800 (train) ; 0.968600 (test)

RMSProp:

Learning rate: 0.010000

Initial accumulation: 0.900000

Time used: 9s

Final accuracy: 0.947500 (train) ; 0.967300 (test)

RMSProp:

Learning rate: 0.030000

Initial accumulation: 0.900000

Time used: 22s

Final accuracy: 0.931200 (train) ; 0.949100 (test)

RMSProp:

Learning rate: 0.050000

Initial accumulation: 0.900000

Time used: 7s

Final accuracy: 0.914800 (train) ; 0.934800 (test)

RMSProp:

Learning rate: 0.010000

Initial accumulation: 0.500000

Time used: 10s

Final accuracy: 0.950500 (train) ; 0.964800 (test)

RMSProp:

Learning rate: 0.010000

Initial accumulation: 0.900000

Time used: 21s

Final accuracy: 0.957900 (train) ; 0.960200 (test)

RMSProp:

Learning rate: 0.010000

Initial accumulation: 0.990000

Time used: 8s

Final accuracy: 0.882500 (train) ; 0.941300 (test)

Adam:

Learning rate: 0.005000

Beta1: 0.900000, Beta2: 0.999000

Time used: 11s  
Final accuracy: 0.957000 (train) ; 0.965700 (test)

Adam:

Learning rate: 0.010000  
Beta1: 0.900000, Beta2: 0.999000  
Time used: 5s  
Final accuracy: 0.954200 (train) ; 0.964700 (test)

Adam:

Learning rate: 0.030000  
Beta1: 0.900000, Beta2: 0.999000  
Time used: 2s  
Final accuracy: 0.944400 (train) ; 0.944100 (test)

Adam:

Learning rate: 0.050000  
Beta1: 0.900000, Beta2: 0.999000  
Time used: 11s  
Final accuracy: 0.916200 (train) ; 0.929100 (test)

Adam:

Learning rate: 0.005000  
Beta1: 0.900000, Beta2: 0.999000  
Time used: 5s  
Final accuracy: 0.957000 (train) ; 0.969600 (test)

Adam:

Learning rate: 0.010000  
Beta1: 0.900000, Beta2: 0.999000  
Time used: 4s  
Final accuracy: 0.955500 (train) ; 0.962100 (test)

Adam:

Learning rate: 0.030000  
Beta1: 0.900000, Beta2: 0.999000  
Time used: 17s  
Final accuracy: 0.941200 (train) ; 0.939700 (test)

Adam:

Learning rate: 0.050000  
Beta1: 0.900000, Beta2: 0.999000  
Time used: 12s  
Final accuracy: 0.922100 (train) ; 0.903900 (test)

Adam:

Learning rate: 0.010000  
Beta1: 0.900000, Beta2: 0.999000  
Time used: 3s  
Final accuracy: 0.957000 (train) ; 0.965300 (test)

Adam:

Learning rate: 0.010000

Beta1: 0.900000, Beta2: 0.999000

Time used: 17s

Final accuracy: 0.957500 (train) ; 0.962400 (test)

Adam:

Learning rate: 0.010000

Beta1: 0.900000, Beta2: 0.999000

Time used: 6s

Final accuracy: 0.955900 (train) ; 0.965800 (test)