

# 媒体与认知 第一次作业

无 81 马啸阳 2018011054

2020 年 3 月 22 日

## 1 选择题

1. B
2. A
3. A
4. A
5. C

## 2 实验结果

自动评判程序的结果如图 1 所示。

调整参数的实验中，总共试验了六组参数，分别为（以下网络结构以数组表示，包含输入层、隐藏层、输出层节点数目）：

- (1) 原始结构：[784, 32, 32, 32, 10]
- (2) 增加网络节点：[784, 256, 128, 32, 10]
- (3) 增加网络层数：[784, 256, 256, 32, 32, 32, 10]
- (4) 增加网络节点并在倒数第二个隐藏层使用 ReLU 激活：[784, 256, 128, 32, 10]
- (5) 减少网络层数：[784, 256, 32]
- (6) 减少网络层数并增大动量：[784, 256, 32], Momentum = 0.92

```
.....
Run time: 0.43268918991088867
```

TASK	SCORE
Sigmoid Non-Linearity (forward)	2
Sigmoid Non-Linearity (derivative)	2
Tanh Non-Linearity (forward)	2
Tanh Non-Linearity (derivative)	2
ReLU Non-Linearity (forward)	2
ReLU Non-Linearity (derivative)	2
Softmax Cross Entropy (forward)	2
Softmax Cross Entropy (derivative)	2
Batch Normalization (training time)	10
Batch Normalization (inference time)	5
Linear Layer (Forward)	2
Linear Layer (Backward)	2
Linear Classifier (forward)	2
Linear Classifier (backward)	2
Linear Classifier (step)	1
Single Hidden Layer (forward)	5
Single Hidden Layer (backward)	5
N Hidden Layer (forward) 1	5
N Hidden Layer (forward) 2	5
N Hidden Layer (forward) 3	5
N Hidden Layer (backward) 1	5
N Hidden Layer (backward) 2	5
N Hidden Layer (backward) 3	5
Momentum	5
TOTAL SCORE	85

图 1: 自动评判程序运行结果

Training loss, training error, validation loss, validation error 的结果图表如图 2 与表 1 所示, 图中从上至下依次对应上述六次训练参数。表格中为六次训练迭代 200 个 epoch 后的结果比较。

MLP 结构	Train loss	Train error	Validation loss	Validation error
原始结构	0.4507	0.1397	0.4000	0.1212
增加节点	0.3389	0.1010	0.3945	0.1218
增加层数	0.3660	0.1075	0.4554	0.1364
增加节点 + ReLU	0.3826	0.1165	0.4188	0.1294
减少层数	0.2681	0.0664	0.4105	0.0950
减少层数 + 增大动量	0.1642	0.0426	0.3465	0.0802

表 1: MLP 训练结果比较

其中可以看见, 增加节点、增加层数都有利于在训练集上获得更好的结果, 但可能产生过拟合, 从而在验证集上表现不如更简单的网络。ReLU 激活函数区别不明显, 而在实际训练中还多次发生数值计算溢出, 不如 sigmoid 将数值限制在一定范围内。而在 MNIST 数据集上, 实际上更简单的网络达到了更好的效果, 只用一个隐层的 MLP 表现的效果最好, 无论是在训练集还是验证集上都比更复杂的网络更优。而增大动量项可以使得收敛更快, 但是进一步增大则可能导致不收敛。

表现最好的最后一个网络在验证集上达到了 91.98% 的准确率。

### 3 实验总结

本实验总体而言较为简单, 实现上没有遇到大的困难。其中有对于交叉熵求导的部分第一次实现虽然通过了自动评测, 但由于没有进行数值上减去最大项的调整, 数值计算多次出现了溢出, 后来进行了调整。其它困难大都通过阶段性输出变量维度与数值进行调试成功。MLP 方面网上可查阅的资料较多, 因而实现起来可参考的疑难解答也较为充分, 容易定位并解决前人编写程序时遇到的类似问题。感谢助教对于实验代码的整理与提示, 但也希望以后能够加入对变量的更多说明。

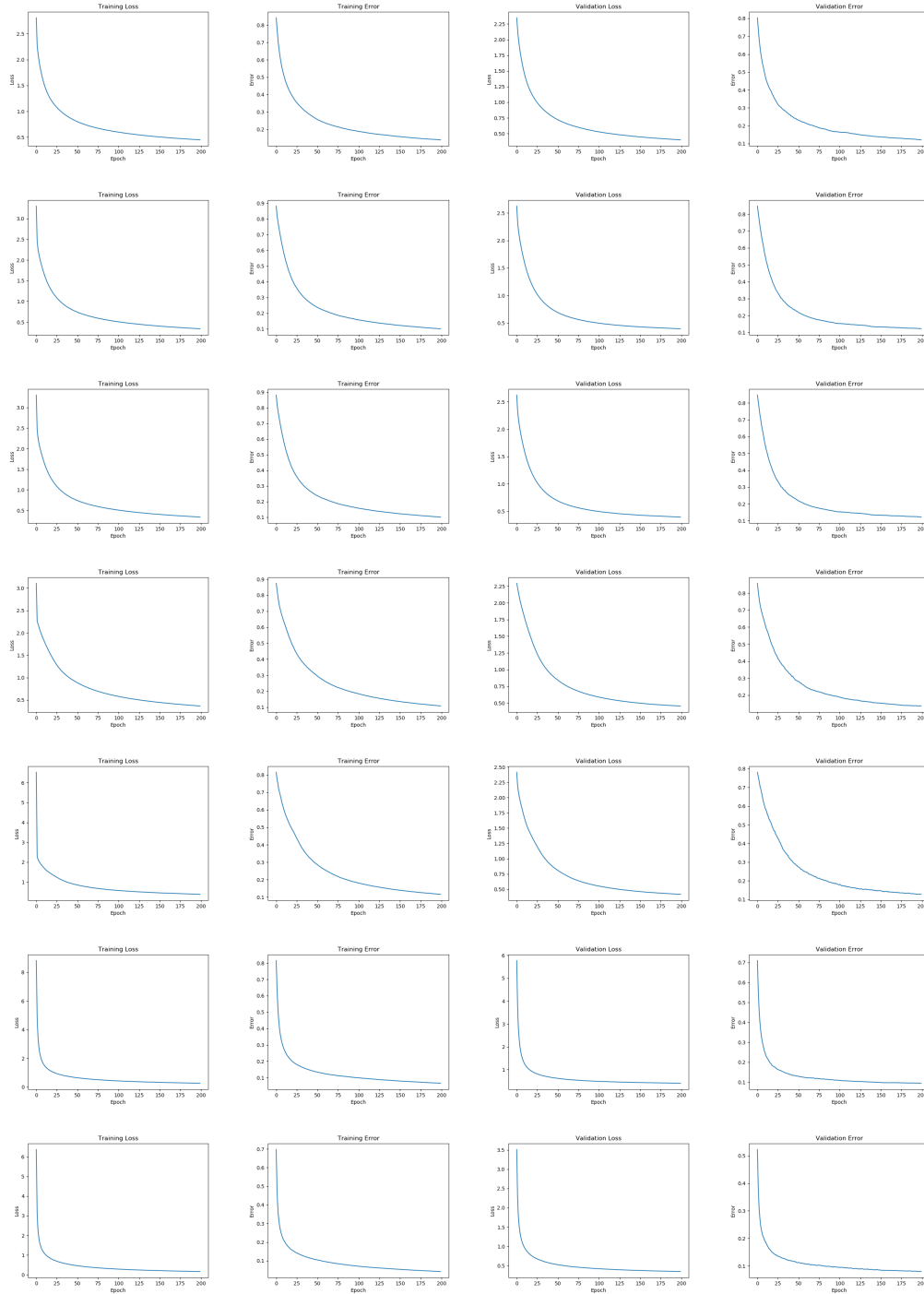


图 2: MLP 训练结果图