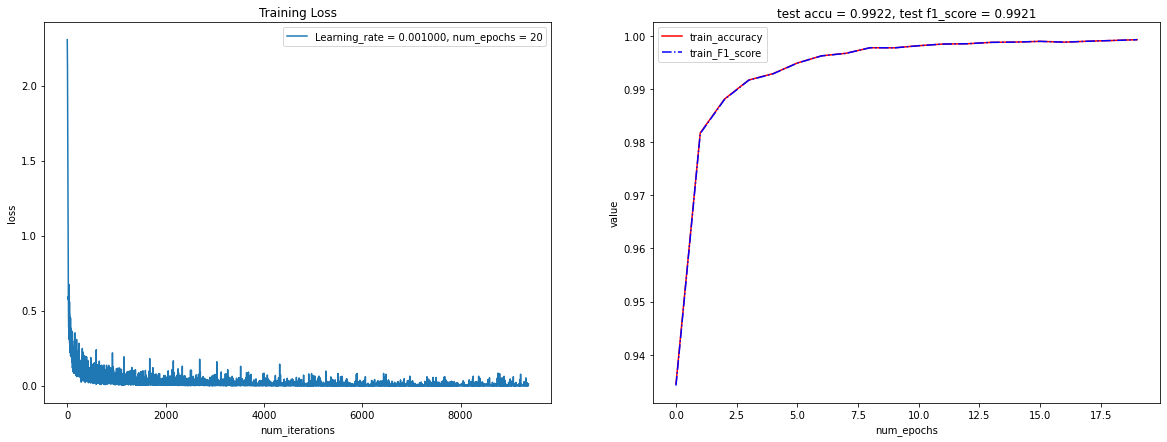
**Report**

自硕21 崔晏菲 2021210976

1. 我选择使用CNN进行MNIST手写数字识别的任务。我设置了三组不同的超参数，来探究batch size和learning rate对CNN模型的影响。

2. 第一组超参数为BATCH\_SIZE = 128, N\_EPOCHS = 20, learning\_rate = 0.001。

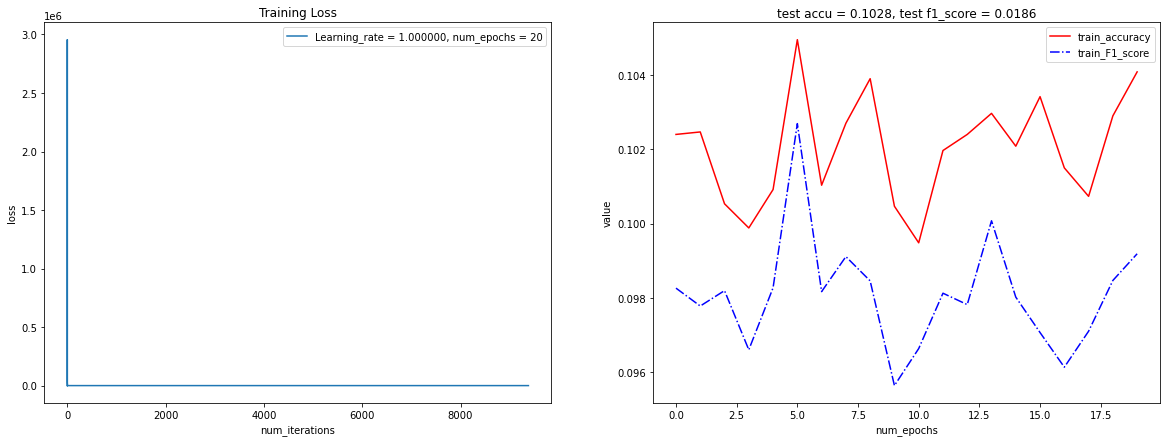
训练和测试结果如下：



可以看到，采取了适当的学习率的batch size，模型收敛的非常好。尤其是准确率和F1 score几乎相同，说明模型对不同样本的识别效果几乎一样好。在梯度下降的过程中，loss呈现总体上下降，但是局部上有较大波动的情况，这是因为我取的batch size非常小。

2. 第二组超参数为BATCH\_SIZE = 128, N\_EPOCHS = 20, learning\_rate = 1。

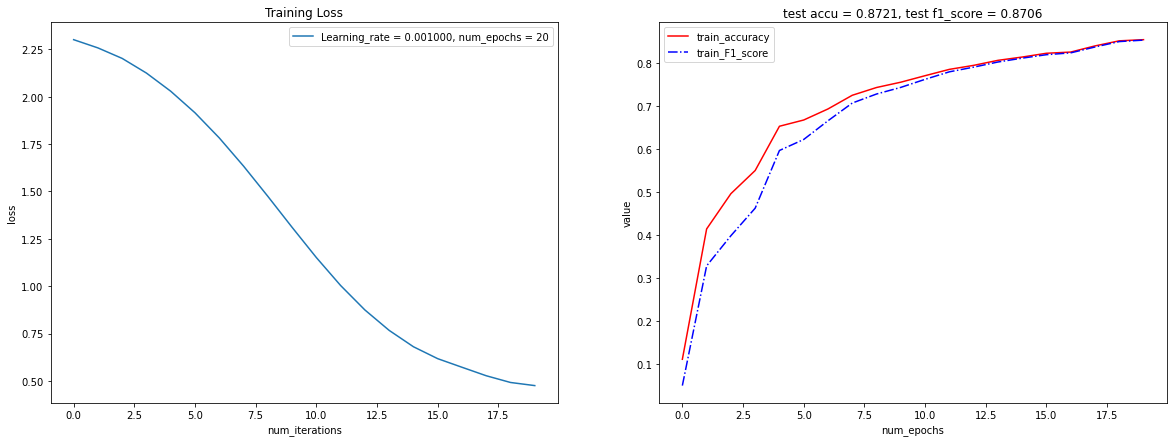
训练和测试结果如下：



可以看到，和第一组超参数相比，第二组仅仅是采取了较大的学习率，但是我们就会发现模型始终没有收敛，准确率和F1 score告诉我们，模型几乎什么也没有学到，就是在瞎猜。这说明，深度学习模型要采取较小的学习率，不然很难收敛到局部最优。

3. 第三组超参数为BATCH\_SIZE = 65536, N\_EPOCHS = 20, learning\_rate = 0.001。

训练和测试结果如下：



可以看到，和第一组超参数相比，第三组仅仅是采取了较大的batch size，这相当于从nini batch SGD变成了GD，我们可以看到，loss下降的曲线光滑了很多，但是收敛速度变慢，20个epoch远远不能让模型收敛。这证明，batch size也不是越大越好，更大的batch size应该匹配更大的学习率和N\_EPOCHS。