一、用 numpy 实现训练 MLP 网络识别手写数字 MNIST 数据集

隐层使用的激活函数为 relu 函数，输出层使用的激活函数为 softmax 函数，损失函数为交叉熵函数。

将隐层数量由一层增加为两层，优化方式为动量优化。

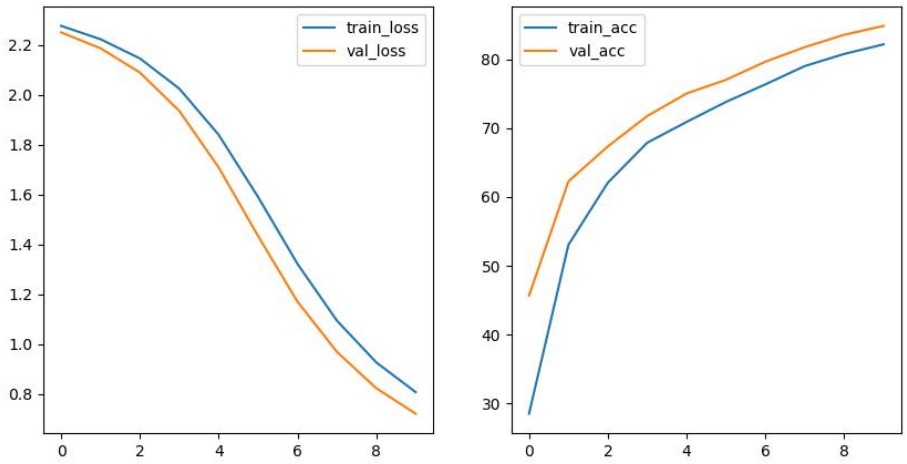
```
100%|| 782/782 [00:05<00:00, 140.51it/s]
epoch:1 acc:0.9266
100%|| 782/782 [00:06<00:00, 126.06it/s]
epoch:2 acc:0.9705
100%|| 782/782 [00:06<00:00, 114.98it/s]
epoch:3 acc:0.9824
100%|| 782/782 [00:09<00:00, 86.68it/s]
epoch:4 acc:0.9881
100%|| 782/782 [00:08<00:00, 88.35it/s]
epoch:5 acc:0.9926
100%|| 782/782 [00:09<00:00, 80.97it/s]
epoch:6 acc:0.9940
100%|| 782/782 [00:09<00:00, 78.31it/s]
epoch:7 acc:0.9950
100%|| 782/782 [00:11<00:00, 70.48it/s]
epoch:8 acc:0.9955
100%|| 782/782 [00:12<00:00, 65.05it/s]
epoch:9 acc:0.9964
100%|| 782/782 [00:11<00:00, 70.43it/s]
epoch:10 acc:0.9965
val_acc:0.9747
```

二、使用 pytorch 训练 MNIST 数据集的 MLP 模型

增加了一层隐藏层，将第二，第三层隐藏层的结点数设定为 600。

未做任何改变的情况下，原模型输出的结果为：

```
Validation set: Average loss: 0.7206, Accuracy: 8484/10000 (84.84%)
```

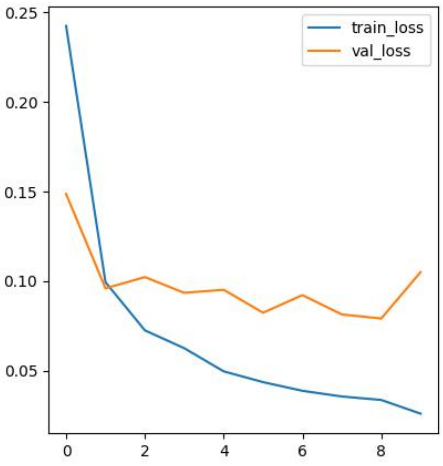


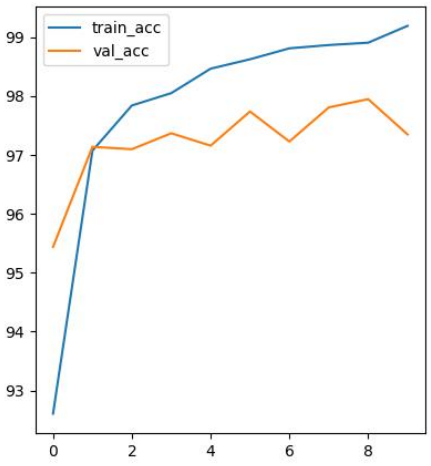
将优化器换为 Adam 后，输出结果为：

```
Train Epoch: 0 [0/50000 (0%)]   Loss: 2.302586
Train Epoch: 0 [6400/50000 (13%)]       Loss: 0.312006
Train Epoch: 0 [12800/50000 (26%)]      Loss: 0.195957
Train Epoch: 0 [19200/50000 (38%)]      Loss: 0.234844
Train Epoch: 0 [25600/50000 (51%)]      Loss: 0.219701
Train Epoch: 0 [32000/50000 (64%)]      Loss: 0.027843
Train Epoch: 0 [38400/50000 (77%)]      Loss: 0.215573
Train Epoch: 0 [44800/50000 (90%)]      Loss: 0.124487
Train set: Average loss: 0.2425, Accuracy: 46300/50000 (92.60%)
Validation set: Average loss: 0.1487, Accuracy: 9544/10000 (95.44%)
Train Epoch: 1 [0/50000 (0%)]   Loss: 0.043932
Train Epoch: 1 [6400/50000 (13%)]       Loss: 0.104224
Train Epoch: 1 [12800/50000 (26%)]      Loss: 0.099468
Train Epoch: 1 [19200/50000 (38%)]      Loss: 0.054329
Train Epoch: 1 [25600/50000 (51%)]      Loss: 0.015165
Train Epoch: 1 [32000/50000 (64%)]      Loss: 0.163292
Train Epoch: 1 [38400/50000 (77%)]      Loss: 0.110413
Train Epoch: 1 [44800/50000 (90%)]      Loss: 0.107171
Train set: Average loss: 0.0994, Accuracy: 48539/50000 (97.08%)
Validation set: Average loss: 0.0959, Accuracy: 9714/10000 (97.14%)
Train Epoch: 2 [0/50000 (0%)]   Loss: 0.144913
Train Epoch: 2 [6400/50000 (13%)]       Loss: 0.072262
Train Epoch: 2 [12800/50000 (26%)]      Loss: 0.086510
Train Epoch: 2 [19200/50000 (38%)]      Loss: 0.039476
Train Epoch: 2 [25600/50000 (51%)]      Loss: 0.127109
Train Epoch: 2 [32000/50000 (64%)]      Loss: 0.135080
Train Epoch: 2 [38400/50000 (77%)]      Loss: 0.064859
Train Epoch: 2 [44800/50000 (90%)]      Loss: 0.013757
Train set: Average loss: 0.0725, Accuracy: 48923/50000 (97.85%)
Validation set: Average loss: 0.1022, Accuracy: 9710/10000 (97.10%)
Train Epoch: 3 [0/50000 (0%)]   Loss: 0.013397
Train Epoch: 3 [6400/50000 (13%)]       Loss: 0.146767
Train Epoch: 3 [12800/50000 (26%)]      Loss: 0.124745
Train Epoch: 3 [19200/50000 (38%)]      Loss: 0.002939
Train Epoch: 3 [25600/50000 (51%)]      Loss: 0.037004
Train Epoch: 3 [32000/50000 (64%)]      Loss: 0.134108
Train Epoch: 3 [38400/50000 (77%)]      Loss: 0.179564
Train Epoch: 3 [44800/50000 (90%)]      Loss: 0.011275
Train set: Average loss: 0.0625, Accuracy: 49025/50000 (98.05%)
Validation set: Average loss: 0.0934, Accuracy: 9737/10000 (97.37%)
```

```
Train Epoch: 4 [0/50000 (0%)]   Loss: 0.045182
Train Epoch: 4 [6400/50000 (13%)]       Loss: 0.019786
Train Epoch: 4 [12800/50000 (26%)]      Loss: 0.153844
Train Epoch: 4 [19200/50000 (38%)]      Loss: 0.008211
Train Epoch: 4 [25600/50000 (51%)]      Loss: 0.029722
Train Epoch: 4 [32000/50000 (64%)]      Loss: 0.068703
Train Epoch: 4 [38400/50000 (77%)]      Loss: 0.077965
Train Epoch: 4 [44800/50000 (90%)]      Loss: 0.034186
Train set: Average loss: 0.0496, Accuracy: 49236/50000 (98.47%)
Validation set: Average loss: 0.0951, Accuracy: 9716/10000 (97.16%)
Train Epoch: 5 [0/50000 (0%)]   Loss: 0.018937
Train Epoch: 5 [6400/50000 (13%)]       Loss: 0.148570
Train Epoch: 5 [12800/50000 (26%)]      Loss: 0.007334
Train Epoch: 5 [19200/50000 (38%)]      Loss: 0.006720
Train Epoch: 5 [25600/50000 (51%)]      Loss: 0.013298
Train Epoch: 5 [32000/50000 (64%)]      Loss: 0.058239
Train Epoch: 5 [38400/50000 (77%)]      Loss: 0.110809
Train Epoch: 5 [44800/50000 (90%)]      Loss: 0.100379
Train set: Average loss: 0.0436, Accuracy: 49313/50000 (98.63%)
Validation set: Average loss: 0.0823, Accuracy: 9774/10000 (97.74%)
Train Epoch: 6 [0/50000 (0%)]   Loss: 0.010123
Train Epoch: 6 [6400/50000 (13%)]       Loss: 0.080519
Train Epoch: 6 [12800/50000 (26%)]      Loss: 0.004560
Train Epoch: 6 [19200/50000 (38%)]      Loss: 0.028330
Train Epoch: 6 [25600/50000 (51%)]      Loss: 0.034455
Train Epoch: 6 [32000/50000 (64%)]      Loss: 0.045414
Train Epoch: 6 [38400/50000 (77%)]      Loss: 0.007485
Train Epoch: 6 [44800/50000 (90%)]      Loss: 0.022289
Train set: Average loss: 0.0387, Accuracy: 49406/50000 (98.81%)
Validation set: Average loss: 0.0921, Accuracy: 9723/10000 (97.23%)
Train Epoch: 7 [0/50000 (0%)]   Loss: 0.216683
Train Epoch: 7 [6400/50000 (13%)]       Loss: 0.115690
Train Epoch: 7 [12800/50000 (26%)]      Loss: 0.009640
Train Epoch: 7 [19200/50000 (38%)]      Loss: 0.039049
Train Epoch: 7 [25600/50000 (51%)]      Loss: 0.098935
Train Epoch: 7 [32000/50000 (64%)]      Loss: 0.054180
Train Epoch: 7 [38400/50000 (77%)]      Loss: 0.005111
Train Epoch: 7 [44800/50000 (90%)]      Loss: 0.014005
Train set: Average loss: 0.0355, Accuracy: 49437/50000 (98.87%)
Validation set: Average loss: 0.0814, Accuracy: 9781/10000 (97.81%)
```

```
Train Epoch: 8 [0/50000 (0%)]   Loss: 0.025170
Train Epoch: 8 [6400/50000 (13%)]       Loss: 0.057043
Train Epoch: 8 [12800/50000 (26%)]      Loss: 0.018282
Train Epoch: 8 [19200/50000 (38%)]      Loss: 0.032111
Train Epoch: 8 [25600/50000 (51%)]      Loss: 0.021720
Train Epoch: 8 [32000/50000 (64%)]      Loss: 0.146865
Train Epoch: 8 [38400/50000 (77%)]      Loss: 0.039516
Train Epoch: 8 [44800/50000 (90%)]      Loss: 0.017740
Train set: Average loss: 0.0336, Accuracy: 49454/50000 (98.91%)
Validation set: Average loss: 0.0791, Accuracy: 9795/10000 (97.95%)
Train Epoch: 9 [0/50000 (0%)]   Loss: 0.022227
Train Epoch: 9 [6400/50000 (13%)]       Loss: 0.001074
Train Epoch: 9 [12800/50000 (26%)]      Loss: 0.036389
Train Epoch: 9 [19200/50000 (38%)]      Loss: 0.070689
Train Epoch: 9 [25600/50000 (51%)]      Loss: 0.000902
Train Epoch: 9 [32000/50000 (64%)]      Loss: 0.109149
Train Epoch: 9 [38400/50000 (77%)]      Loss: 0.091017
Train Epoch: 9 [44800/50000 (90%)]      Loss: 0.113384
Train set: Average loss: 0.0259, Accuracy: 49597/50000 (99.19%)
Validation set: Average loss: 0.1050, Accuracy: 9735/10000 (97.35%)
```



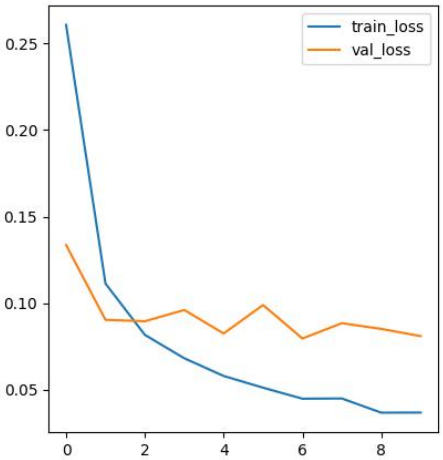


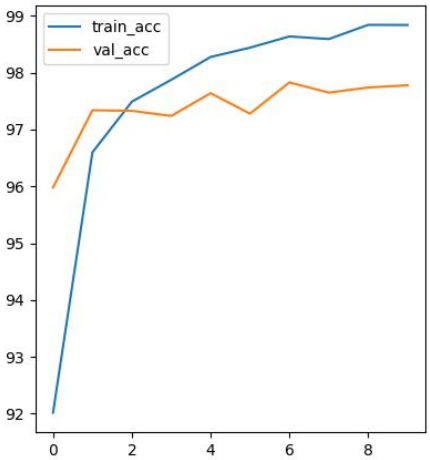
增加了 dropout 层后，输出结果为：

```
Train Epoch: 0 [0/50000 (0%)]   Loss: 2.300390
Train Epoch: 0 [6400/50000 (13%)]       Loss: 0.318626
Train Epoch: 0 [12800/50000 (26%)]      Loss: 0.139931
Train Epoch: 0 [19200/50000 (38%)]      Loss: 0.212153
Train Epoch: 0 [25600/50000 (51%)]      Loss: 0.161944
Train Epoch: 0 [32000/50000 (64%)]      Loss: 0.089960
Train Epoch: 0 [38400/50000 (77%)]      Loss: 0.179400
Train Epoch: 0 [44800/50000 (90%)]      Loss: 0.132125
Train set: Average loss: 0.2608, Accuracy: 46004/50000 (92.01%)
Validation set: Average loss: 0.1336, Accuracy: 9598/10000 (95.98%)
Train Epoch: 1 [0/50000 (0%)]   Loss: 0.100581
Train Epoch: 1 [6400/50000 (13%)]       Loss: 0.028668
Train Epoch: 1 [12800/50000 (26%)]      Loss: 0.010605
Train Epoch: 1 [19200/50000 (38%)]      Loss: 0.071069
Train Epoch: 1 [25600/50000 (51%)]      Loss: 0.024823
Train Epoch: 1 [32000/50000 (64%)]      Loss: 0.206532
Train Epoch: 1 [38400/50000 (77%)]      Loss: 0.046682
Train Epoch: 1 [44800/50000 (90%)]      Loss: 0.070899
Train set: Average loss: 0.1113, Accuracy: 48304/50000 (96.61%)
Validation set: Average loss: 0.0903, Accuracy: 9734/10000 (97.34%)
Train Epoch: 2 [0/50000 (0%)]   Loss: 0.112442
Train Epoch: 2 [6400/50000 (13%)]       Loss: 0.080660
Train Epoch: 2 [12800/50000 (26%)]      Loss: 0.178221
Train Epoch: 2 [19200/50000 (38%)]      Loss: 0.065268
Train Epoch: 2 [25600/50000 (51%)]      Loss: 0.043029
Train Epoch: 2 [32000/50000 (64%)]      Loss: 0.127605
Train Epoch: 2 [38400/50000 (77%)]      Loss: 0.027510
Train Epoch: 2 [44800/50000 (90%)]      Loss: 0.047636
Train set: Average loss: 0.0816, Accuracy: 48744/50000 (97.49%)
Validation set: Average loss: 0.0895, Accuracy: 9733/10000 (97.33%)
```

```
Train Epoch: 3 [0/50000 (0%)]   Loss: 0.417667
Train Epoch: 3 [6400/50000 (13%)]       Loss: 0.023677
Train Epoch: 3 [12800/50000 (26%)]      Loss: 0.037086
Train Epoch: 3 [19200/50000 (38%)]      Loss: 0.007513
Train Epoch: 3 [25600/50000 (51%)]      Loss: 0.064009
Train Epoch: 3 [32000/50000 (64%)]      Loss: 0.051299
Train Epoch: 3 [38400/50000 (77%)]      Loss: 0.254050
Train Epoch: 3 [44800/50000 (90%)]      Loss: 0.171137
Train set: Average loss: 0.0681, Accuracy: 48940/50000 (97.88%)
Validation set: Average loss: 0.0960, Accuracy: 9724/10000 (97.24%)
Train Epoch: 4 [0/50000 (0%)]   Loss: 0.010041
Train Epoch: 4 [6400/50000 (13%)]       Loss: 0.066590
Train Epoch: 4 [12800/50000 (26%)]      Loss: 0.052145
Train Epoch: 4 [19200/50000 (38%)]      Loss: 0.054052
Train Epoch: 4 [25600/50000 (51%)]      Loss: 0.078898
Train Epoch: 4 [32000/50000 (64%)]      Loss: 0.062900
Train Epoch: 4 [38400/50000 (77%)]      Loss: 0.006414
Train Epoch: 4 [44800/50000 (90%)]      Loss: 0.037791
Train set: Average loss: 0.0579, Accuracy: 49141/50000 (98.28%)
Validation set: Average loss: 0.0824, Accuracy: 9764/10000 (97.64%)
Train Epoch: 5 [0/50000 (0%)]   Loss: 0.010588
Train Epoch: 5 [6400/50000 (13%)]       Loss: 0.003770
Train Epoch: 5 [12800/50000 (26%)]      Loss: 0.123502
Train Epoch: 5 [19200/50000 (38%)]      Loss: 0.051454
Train Epoch: 5 [25600/50000 (51%)]      Loss: 0.090252
Train Epoch: 5 [32000/50000 (64%)]      Loss: 0.002646
Train Epoch: 5 [38400/50000 (77%)]      Loss: 0.056229
Train Epoch: 5 [44800/50000 (90%)]      Loss: 0.031119
Train set: Average loss: 0.0511, Accuracy: 49220/50000 (98.44%)
Validation set: Average loss: 0.0989, Accuracy: 9728/10000 (97.28%)
```

```
Train Epoch: 6 [0/50000 (0%)]   Loss: 0.103393
Train Epoch: 6 [6400/50000 (13%)]       Loss: 0.038933
Train Epoch: 6 [12800/50000 (26%)]      Loss: 0.045153
Train Epoch: 6 [19200/50000 (38%)]      Loss: 0.047706
Train Epoch: 6 [25600/50000 (51%)]      Loss: 0.017920
Train Epoch: 6 [32000/50000 (64%)]      Loss: 0.057514
Train Epoch: 6 [38400/50000 (77%)]      Loss: 0.024461
Train Epoch: 6 [44800/50000 (90%)]      Loss: 0.039293
Train set: Average loss: 0.0447, Accuracy: 49319/50000 (98.64%)
Validation set: Average loss: 0.0795, Accuracy: 9783/10000 (97.83%)
Train Epoch: 7 [0/50000 (0%)]   Loss: 0.116898
Train Epoch: 7 [6400/50000 (13%)]       Loss: 0.026799
Train Epoch: 7 [12800/50000 (26%)]      Loss: 0.027299
Train Epoch: 7 [19200/50000 (38%)]      Loss: 0.248506
Train Epoch: 7 [25600/50000 (51%)]      Loss: 0.019778
Train Epoch: 7 [32000/50000 (64%)]      Loss: 0.009230
Train Epoch: 7 [38400/50000 (77%)]      Loss: 0.055417
Train Epoch: 7 [44800/50000 (90%)]      Loss: 0.005042
Train set: Average loss: 0.0449, Accuracy: 49296/50000 (98.59%)
Validation set: Average loss: 0.0884, Accuracy: 9765/10000 (97.65%)
Train Epoch: 8 [0/50000 (0%)]   Loss: 0.007174
Train Epoch: 8 [6400/50000 (13%)]       Loss: 0.003142
Train Epoch: 8 [12800/50000 (26%)]      Loss: 0.083335
Train Epoch: 8 [19200/50000 (38%)]      Loss: 0.088105
Train Epoch: 8 [25600/50000 (51%)]      Loss: 0.061650
Train Epoch: 9 [0/50000 (0%)]   Loss: 0.057930
Train Epoch: 9 [6400/50000 (13%)]       Loss: 0.028368
Train Epoch: 9 [12800/50000 (26%)]      Loss: 0.005201
Train Epoch: 9 [19200/50000 (38%)]      Loss: 0.010235
Train Epoch: 9 [25600/50000 (51%)]      Loss: 0.002015
Train Epoch: 9 [32000/50000 (64%)]      Loss: 0.048895
Train Epoch: 9 [38400/50000 (77%)]      Loss: 0.031350
Train Epoch: 9 [44800/50000 (90%)]      Loss: 0.022727
Train set: Average loss: 0.0367, Accuracy: 49420/50000 (98.84%)
Validation set: Average loss: 0.0809, Accuracy: 9778/10000 (97.78%)
```





从输出结果上来看，与未使用 dropout 的模型差别不大。