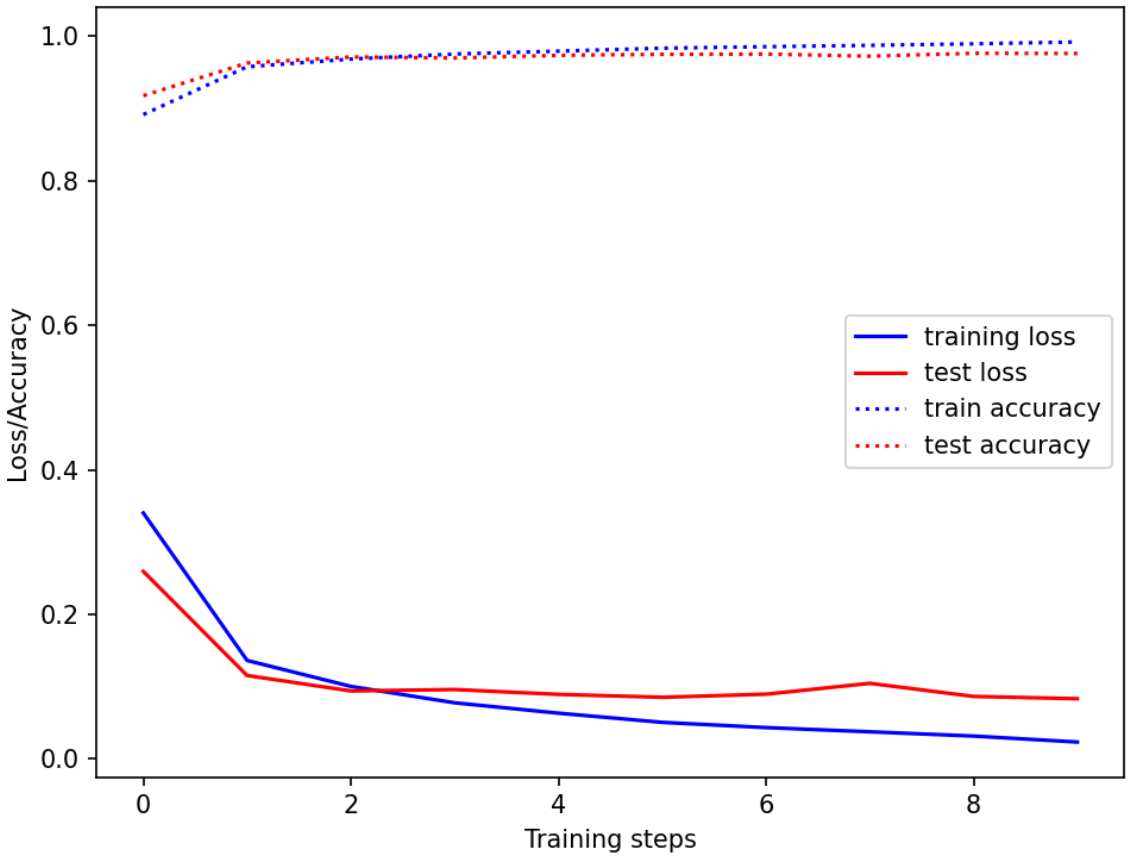
Homework 2 - Task 3 - Adjustments to Hyperparameters

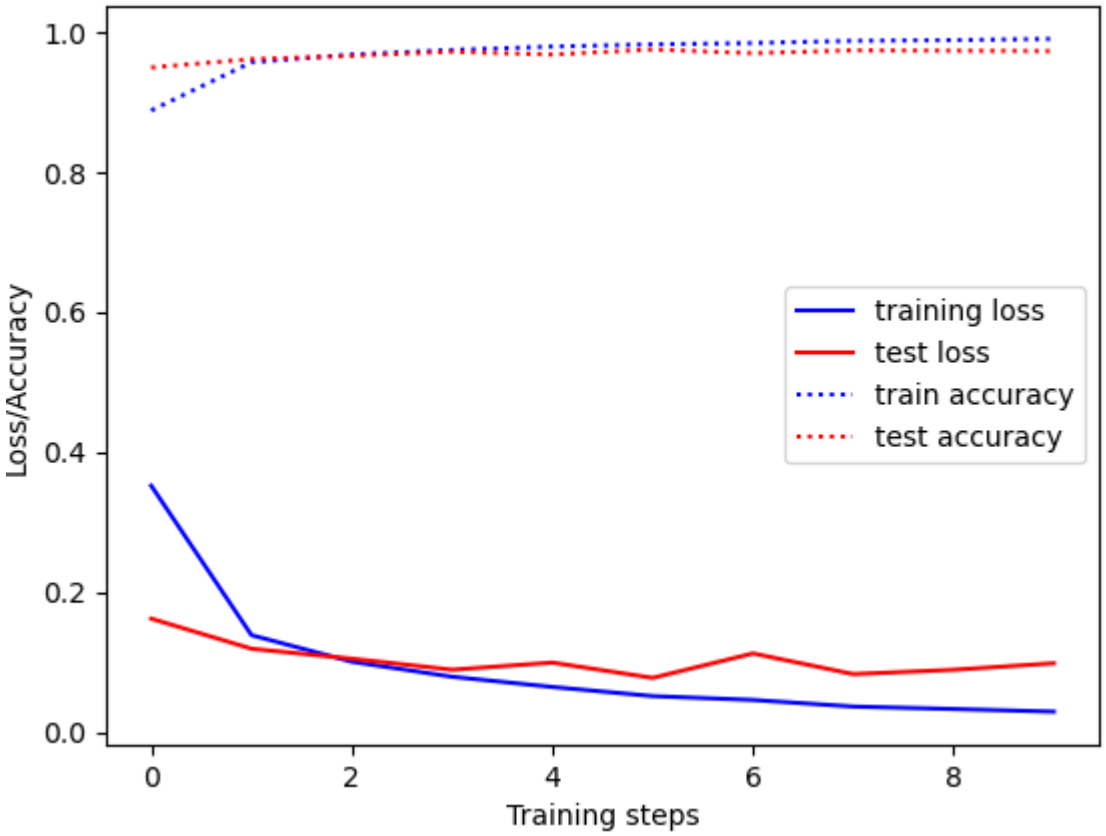
Standard:



Epochs: 10
Learning Rate: 0.1
Batchsize: 32
Number of Hidden Layers: 3
Size of Layers: 256, 256, 256, 10
Optimizer: SGD with momentum 0

Comment:
While training loss and accuracy are optimized continuously throughout the epochs, the testing scores do not seem to change much after the second training step. This could mean that the MLP is overfitting on the training data.

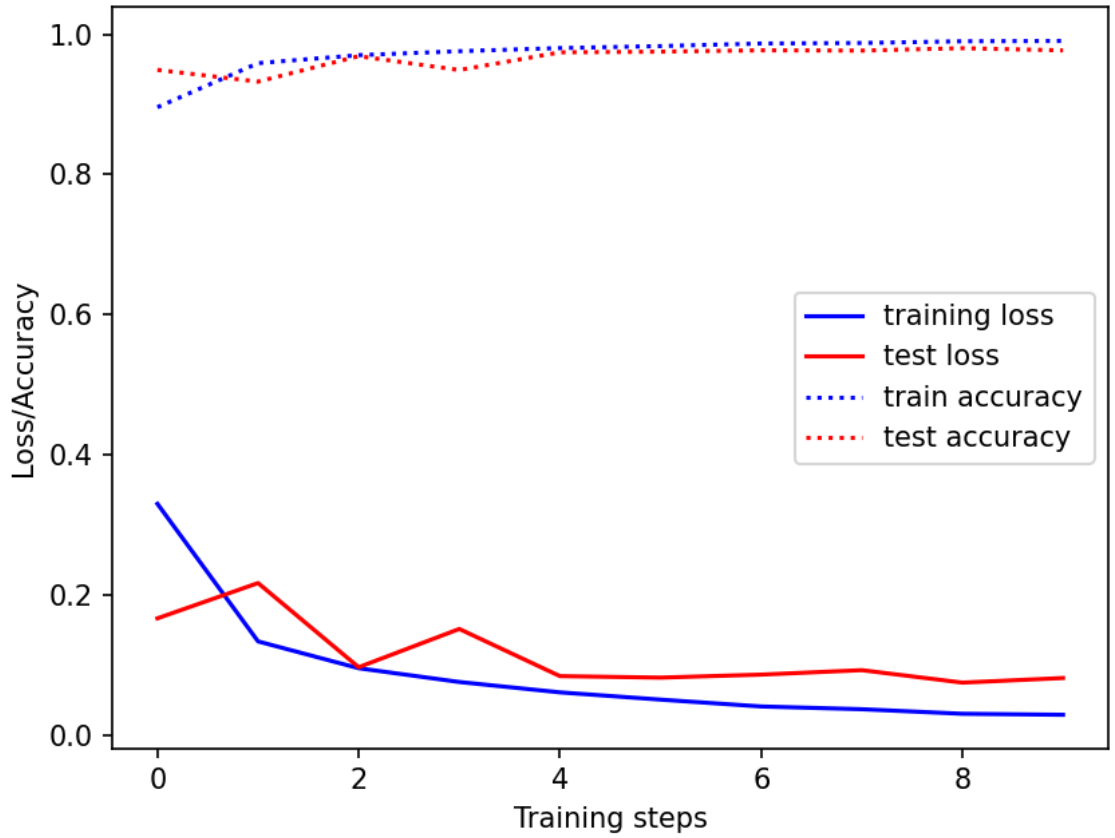
Adjusted Learning Rate and Optimizer Momentum:



Epochs: 10
Learning Rate: 0.5
Batchsize: 32
Number of Hidden Layers: 3
Size of Layers: 256, 256, 256, 10
Optimizer: SGD with momentum 0.3

Comment:
Adjusting the learning rate and optimizer function does not seem to significantly change the model's performance. While the test loss is less severe early on, this decrease still stagnates fairly quickly, although there are some added inconsistencies. The overfitting issue from the standard version seems to persist.

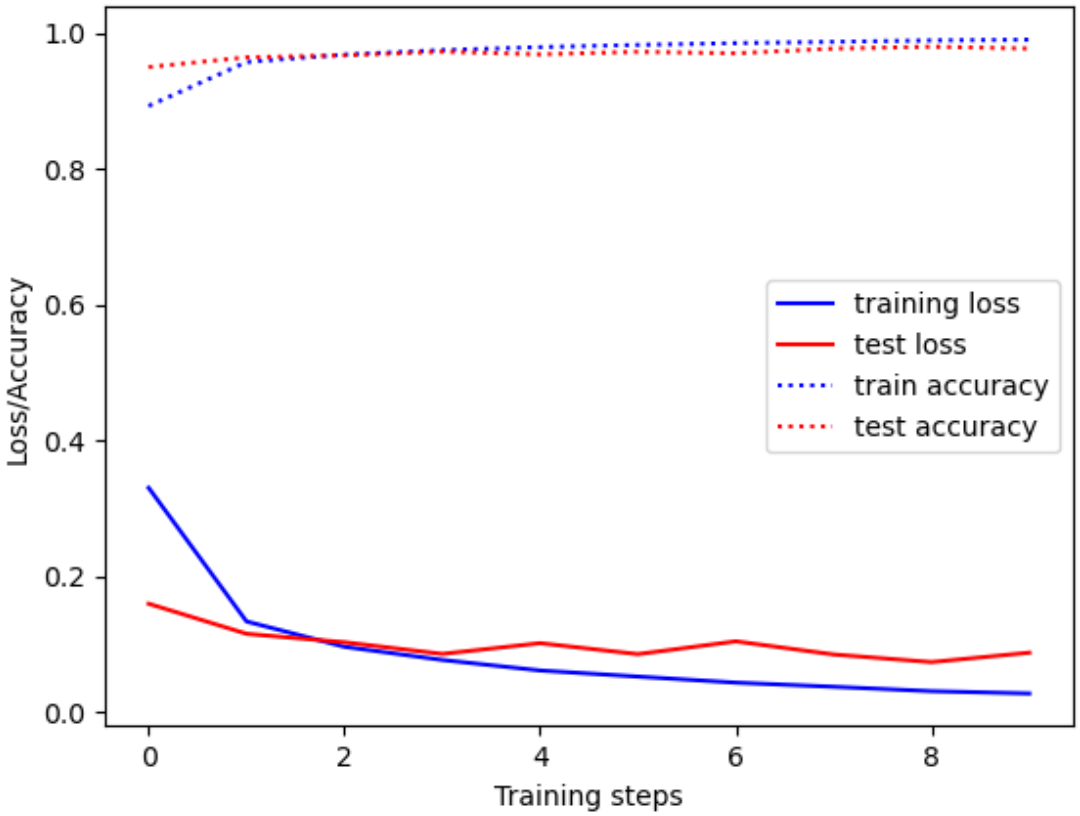
Smaller Hidden Layers



Epochs: 10
Learning Rate: 0.1
Batchsize: 32
Number of Hidden Layers: 2
Size of Layers: 256, 128, 10
Optimizer: SGD with momentum 0

Comment:
While the end result is similar to the first configuration again, the model performs worse during the first few epochs, as shown in the jumps of test loss and accuracy. However, that only 2 hidden layers with less perceptrons achieve a similar goal as 3 hidden layers means we could have saved some computing power during the initial setup.

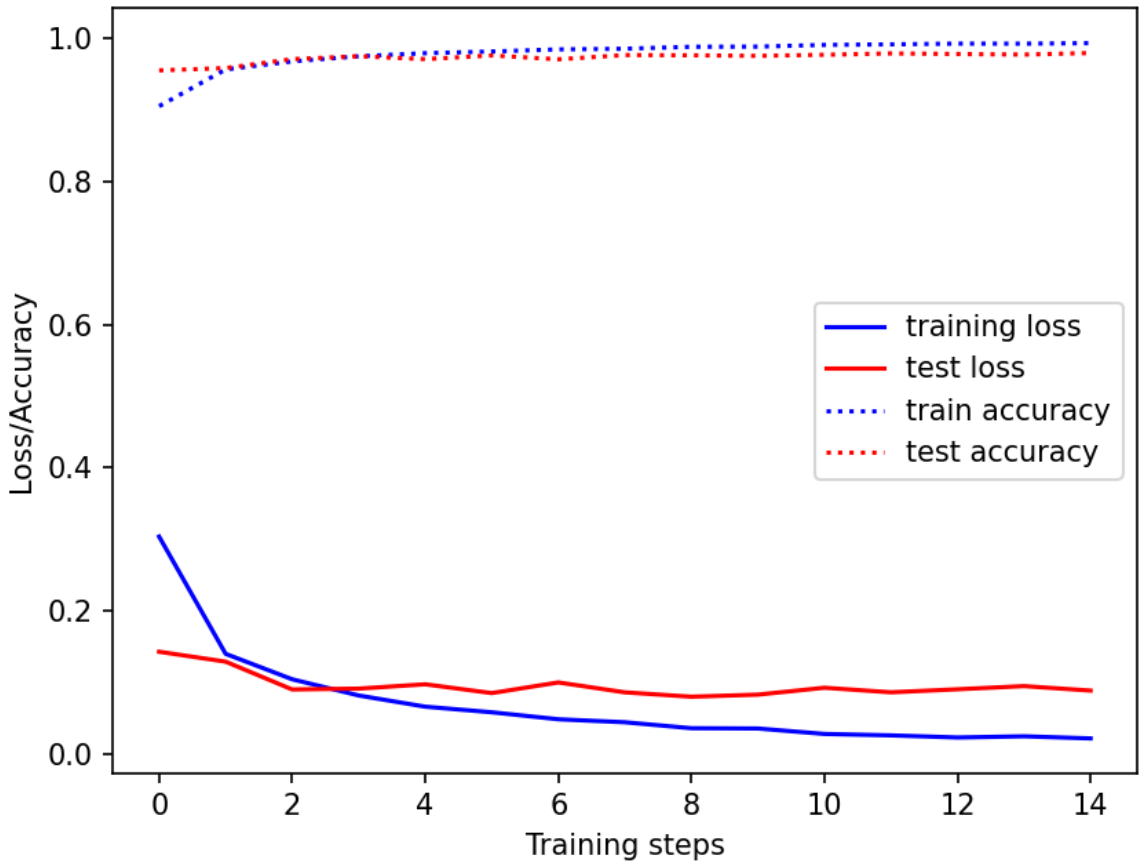
Small Model



Epochs: 6
Learning Rate: 0.5
Batchsize: 32
Number of Hidden Layers: 2
Size of Layers: 128, 128, 10
Optimizer: SGD with momentum 0.9

Comment:
Despite reducing multiple parameters, the model still performs decently again. It might be possible to cut down even more while still having a model that performs in a satisfactory manner, although some inconsistencies have snuck their way into the testing phases. Further reduction might also lead to underfitting.

Big Model



Epochs: 15
Learning Rate: 0.01
Batchsize: 32
Number of Hidden Layers: 4
Size of Layers: 256, 256, 256, 256, 10
Optimizer: SGD with momentum 0.9

Comment:
The end result of this longer model seems to be even more split between training and testing accuracies and losses. This implies that, despite having more time to learn, the model is simply overfitting on the training data even more, which is to be expected. The model has to retain some flexibility to correctly determine the digits in the test set, which it doesn't do if it overfits.