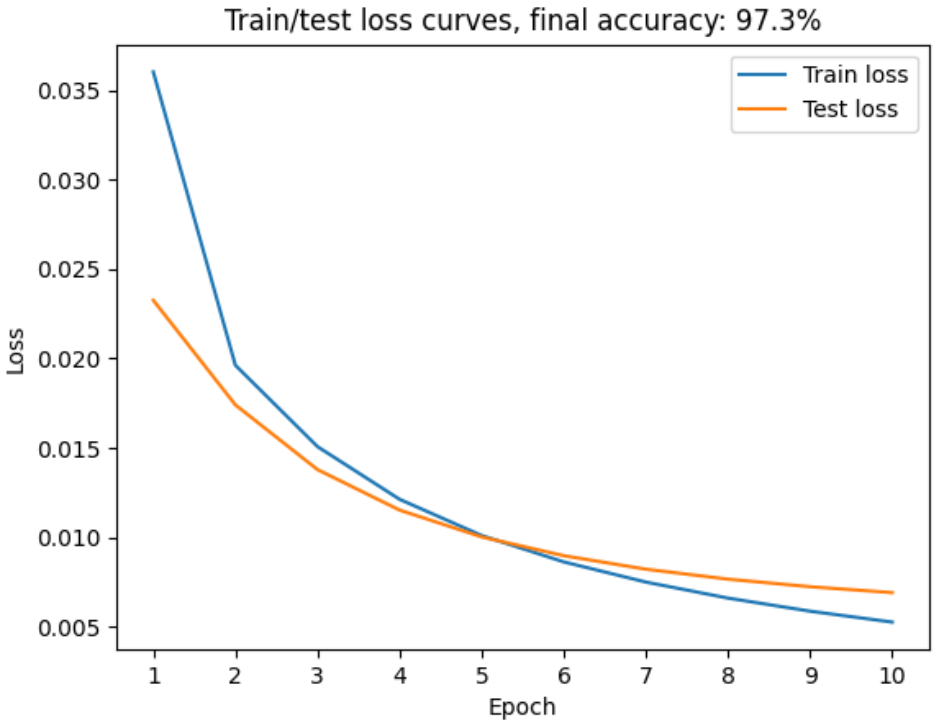
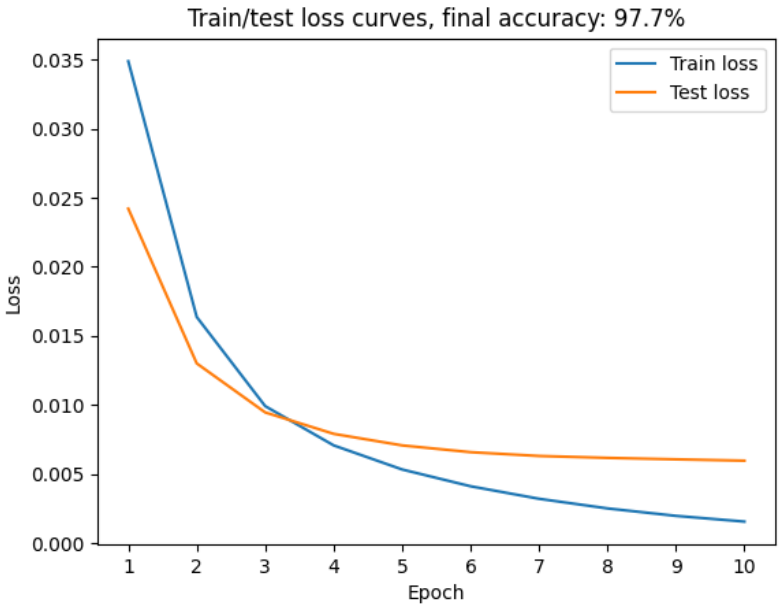
Report on the completion of the sixth task.

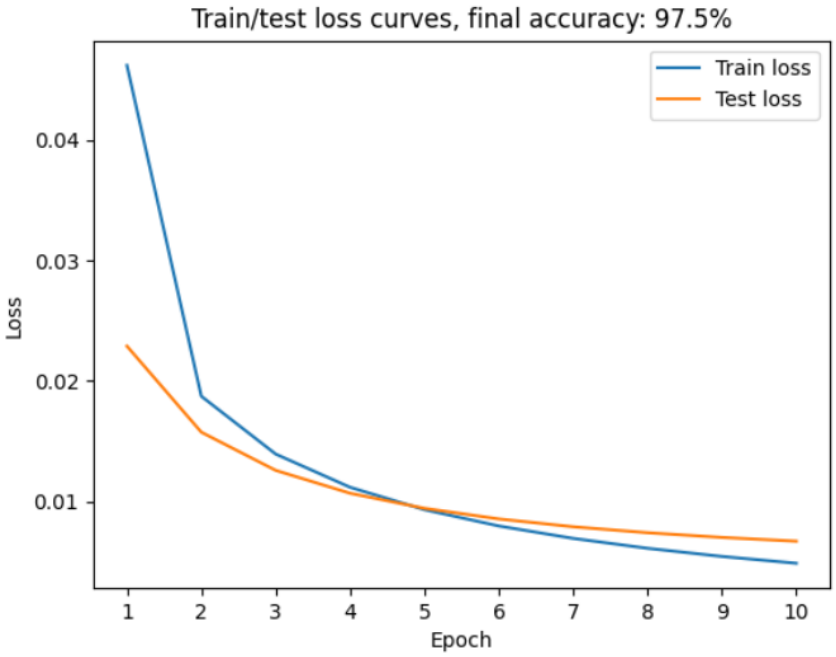
1. **The purpose of the task.**
   The main goal of this project is to investigate how do normalizations, dropouts, batch sizes and activation functions affect the NN performance.
2. **Experiments.**
   a) **Experiments with presence / absence of dropout and batch normalization.**
      i) **Experiment #1 (absence of dropout and batch normalization, sigmoid function, batch size = 12, layers = 2)**



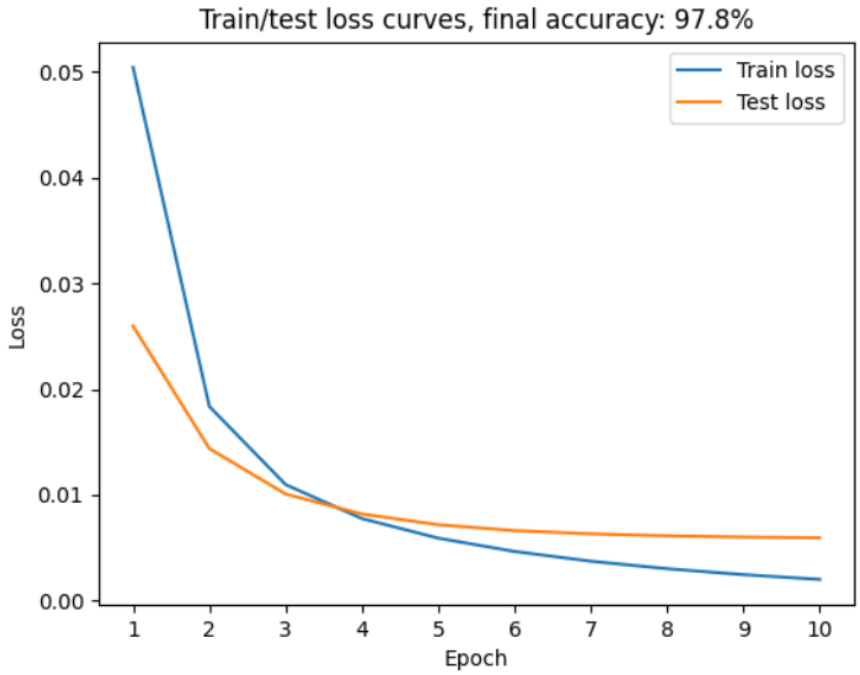
ii) Experiment #2 (absence of dropout and presence of batch normalization, sigmoid function, batch size = 12, layers = 2)



iii) Experiment #3 (presence of dropout and absence of batch normalization, sigmoid function, batch size = 12, layers = 2)

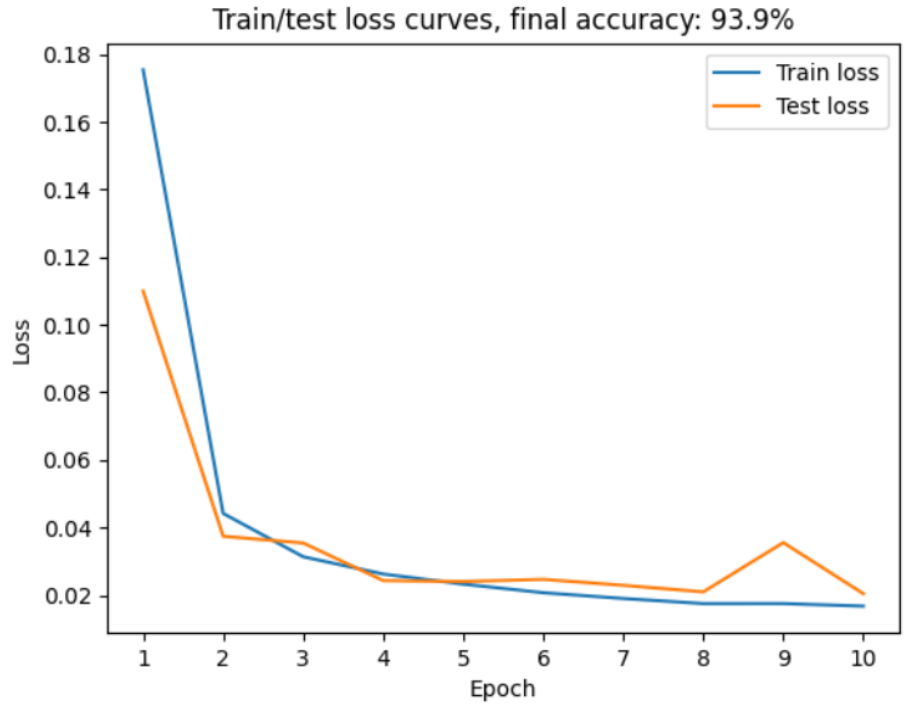


iv) **Experiment #4 (presence of dropout and batch normalization, sigmoid function, batch size = 12, layers = 2)**

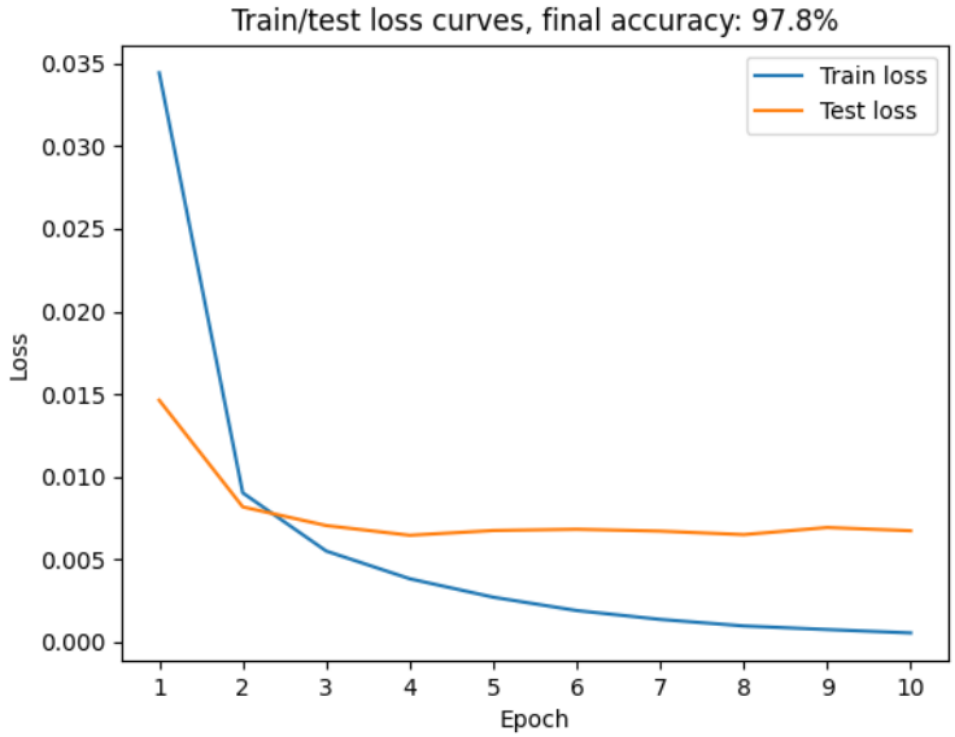


b) Experiments with the different activation functions.

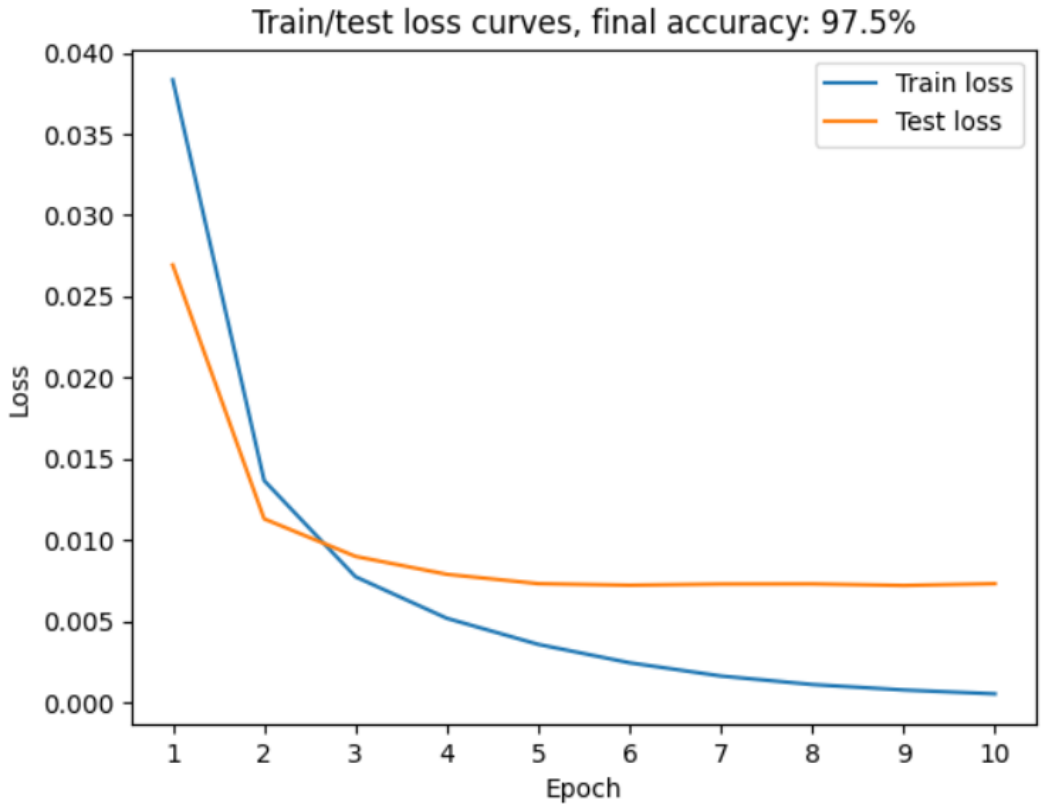
i) Experiment #1 (softmax, 2 layers, 12 batch size)



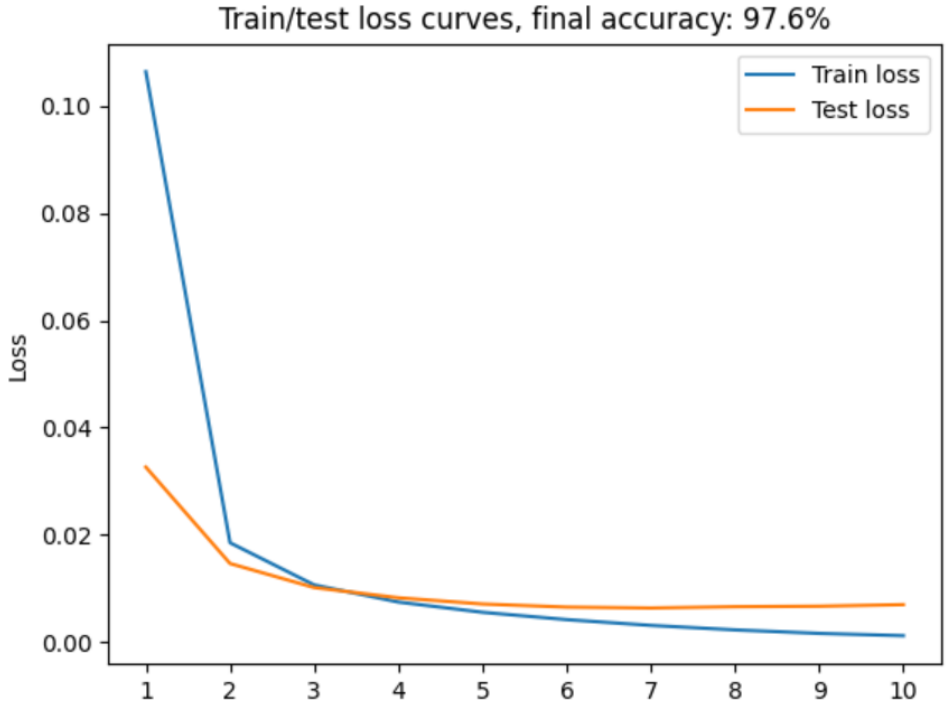
ii) Experiment #2 (ReLU, 2 layers, 12 batch size)



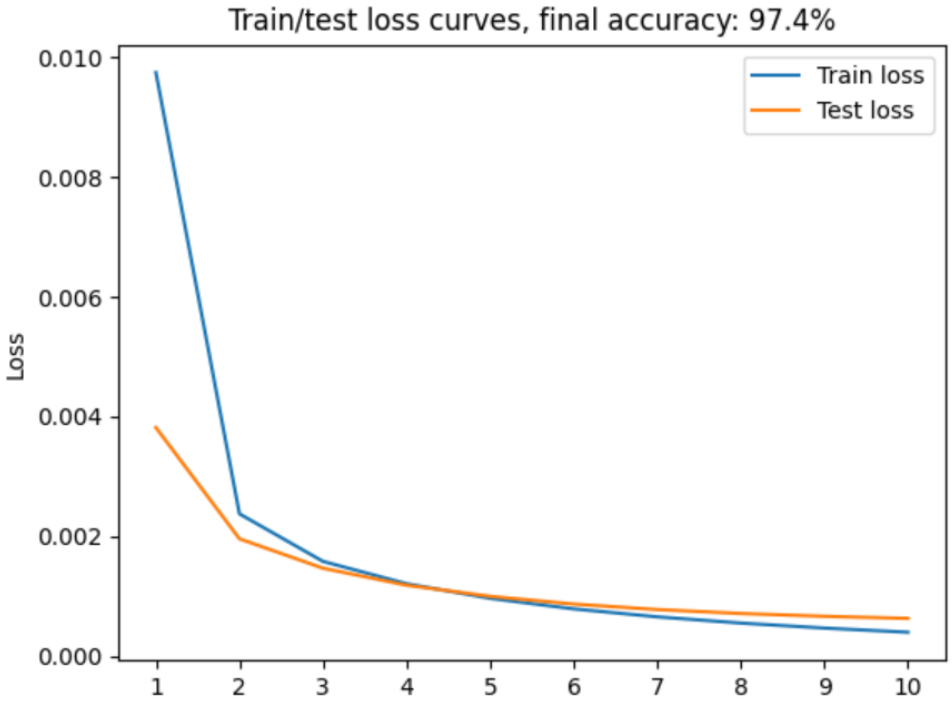
iii) Experiment #3 (Tanh, 2 layers, 12 batch size)



c) Experiment with 3 hidden layers (sigmoid function, 12 batch size)



d) Experiment with 128 batch size (sigmoid function, 3 layers)



3. **Conclusion.**

- As we can clearly see, the presence of batch normalization and dropout layer basically increases the accuracy of the model and gives very little loss on the train set.

-Softmax has the worst performance of all activation functions. ReLU and Tanh give slight loss on a training set, but the test set stays the same.
- Neither the increase in layers nor the size of batches give significant improvement to the network.