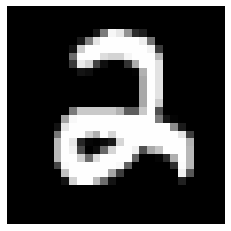
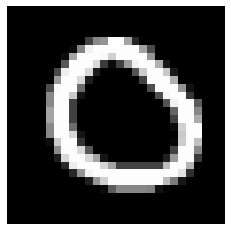
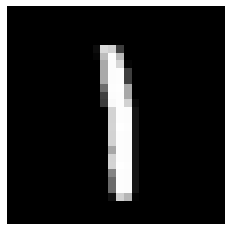
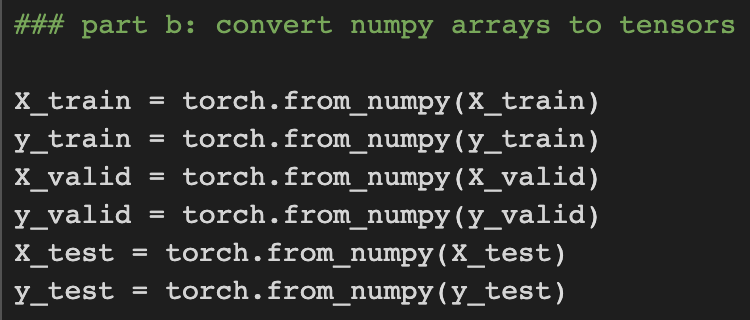
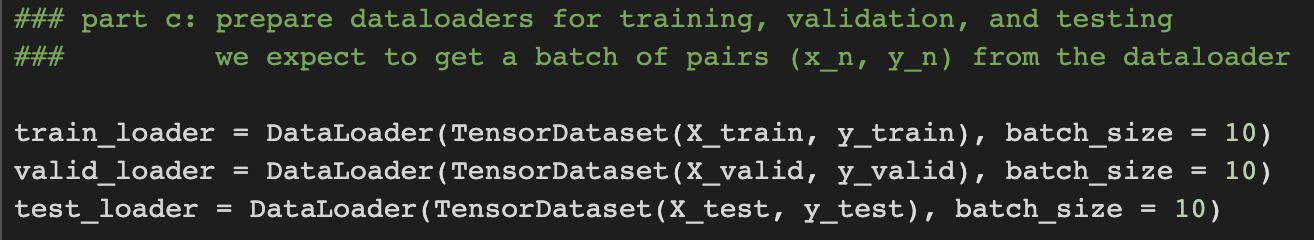
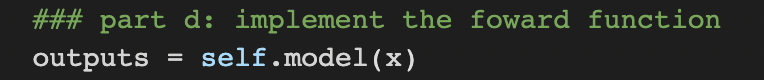
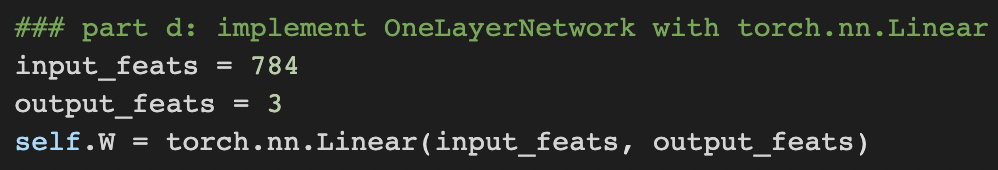
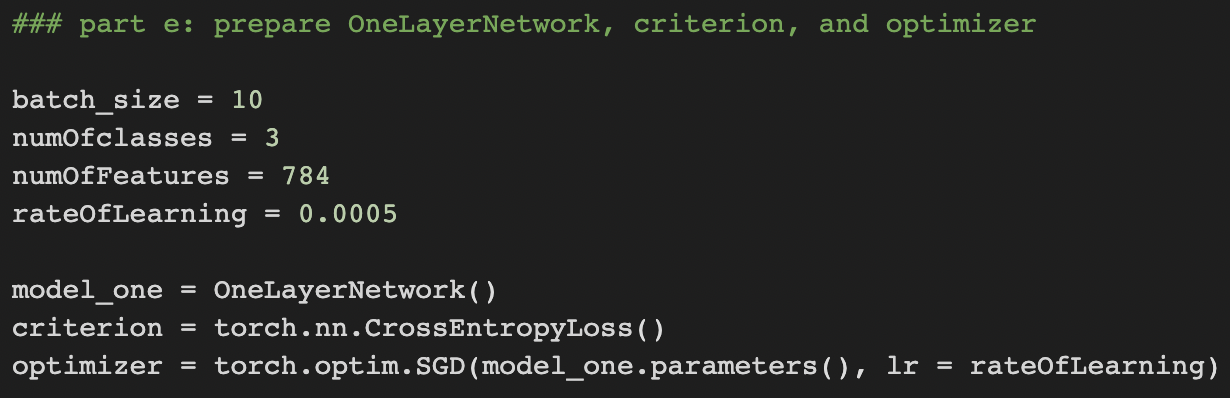
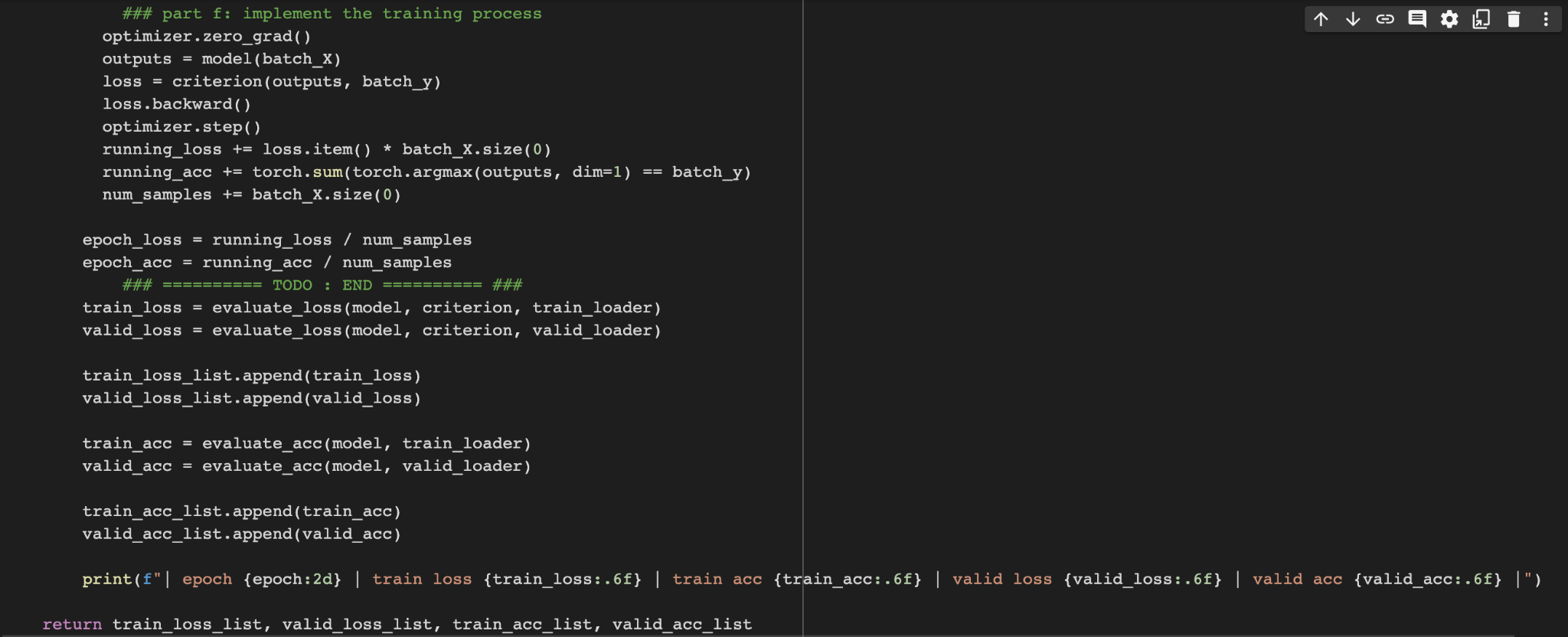
Preyasi Gaur

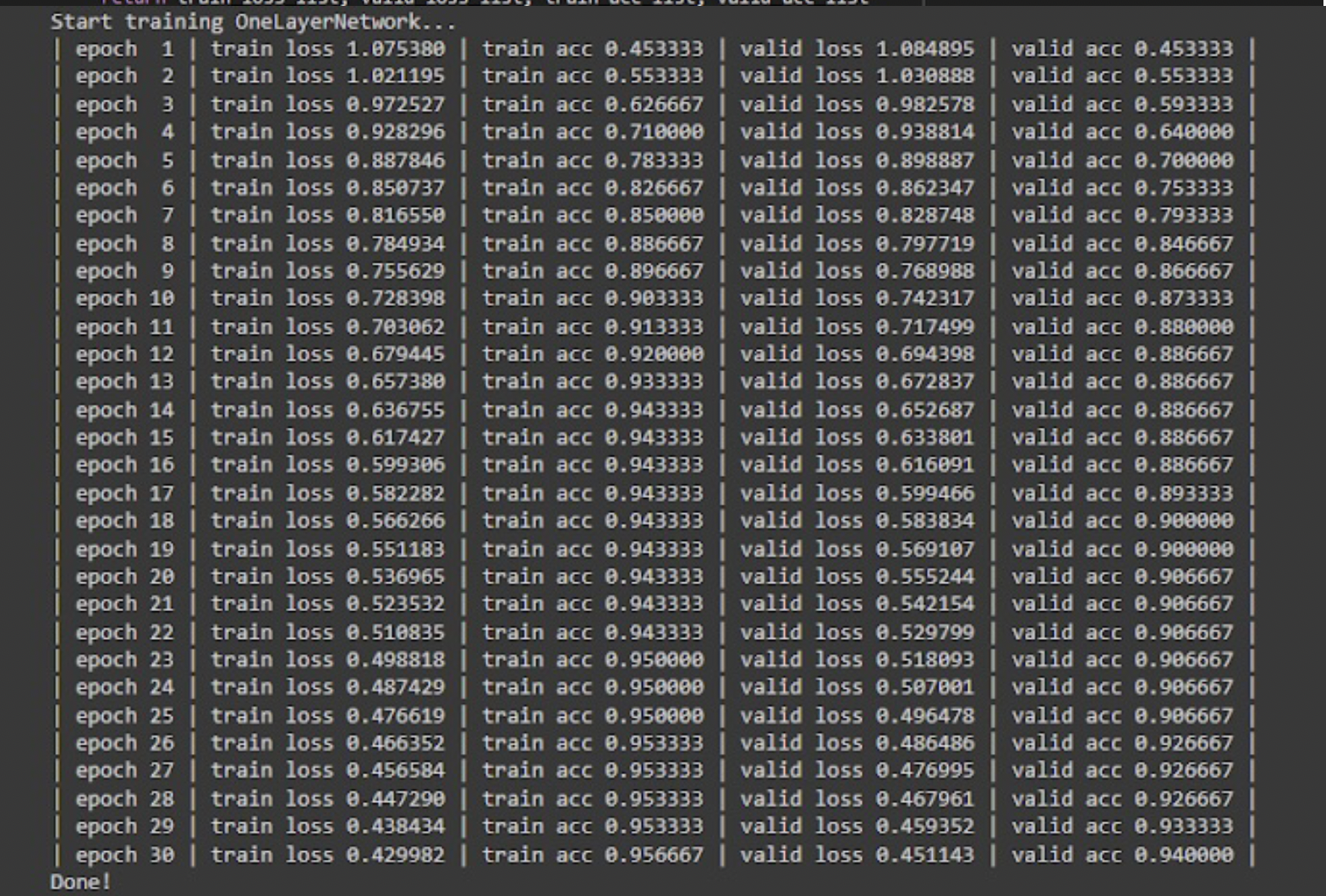
705704939

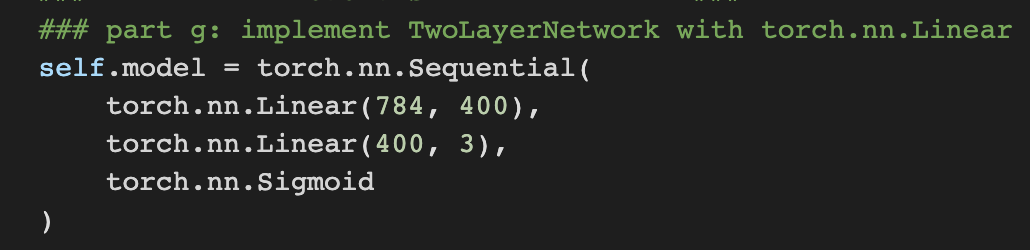
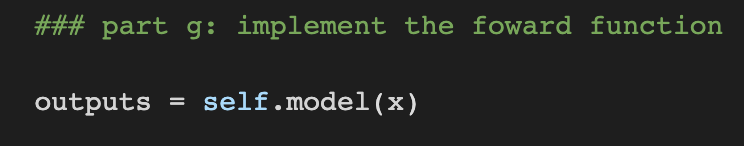
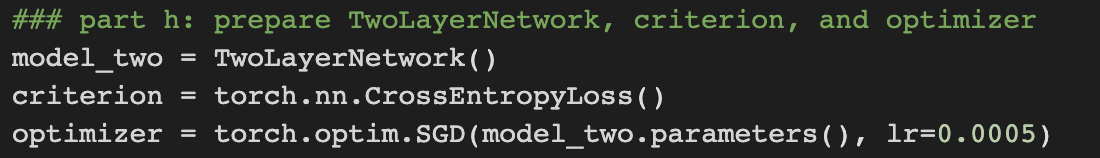
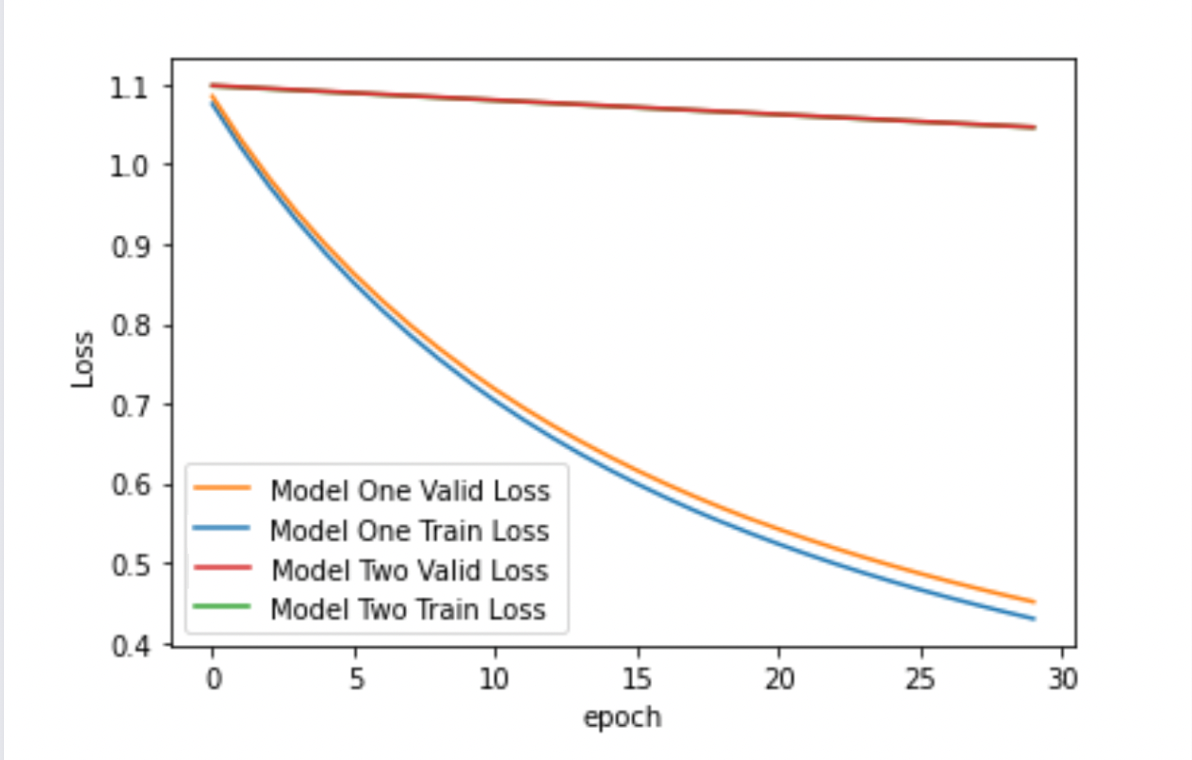
CS M146

Problem Set 3

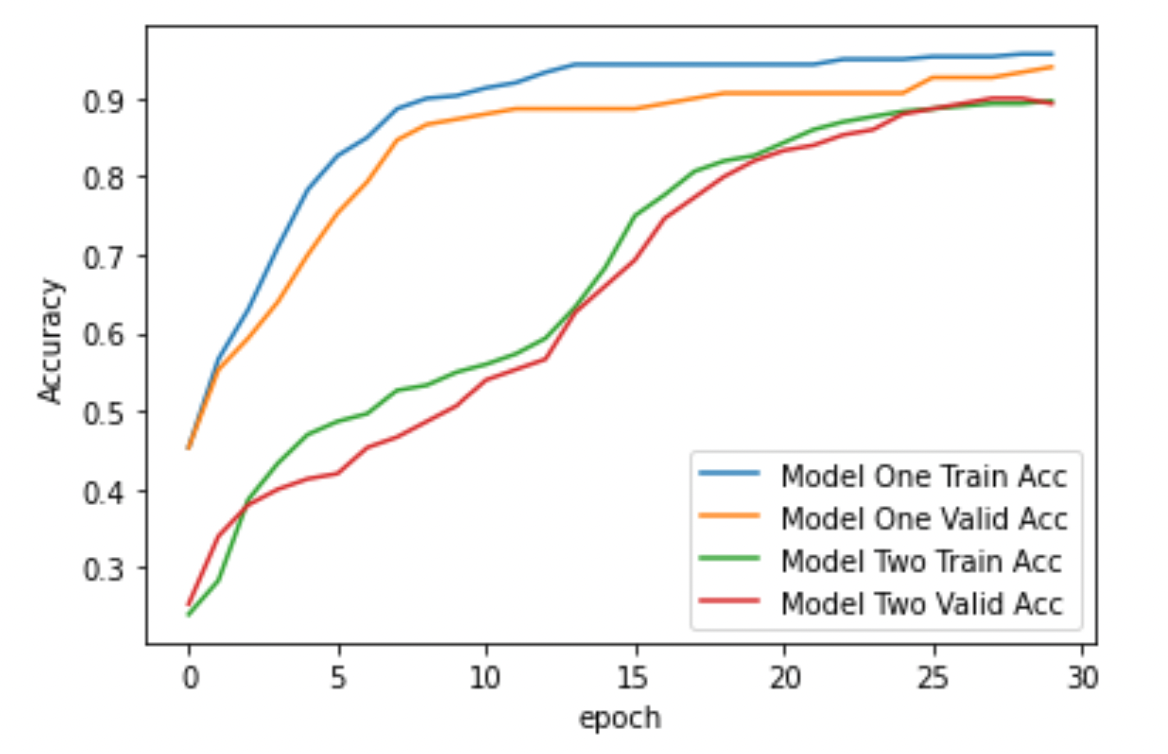
**4 Implementation: Digit Recognizer**

1. 
2. 
3. 
4. 
5. 
6. 



1.   
   
2. 
3. 

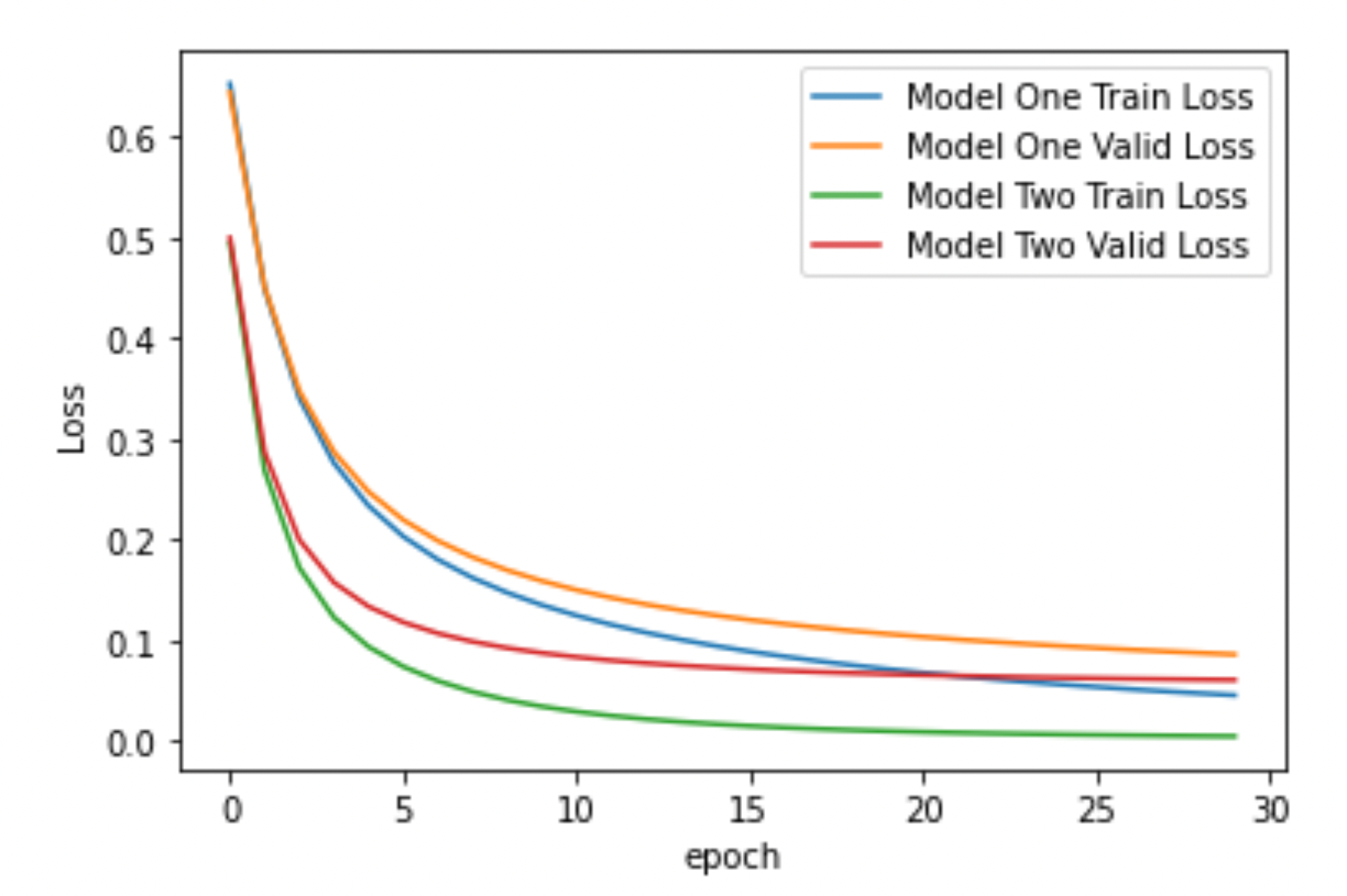
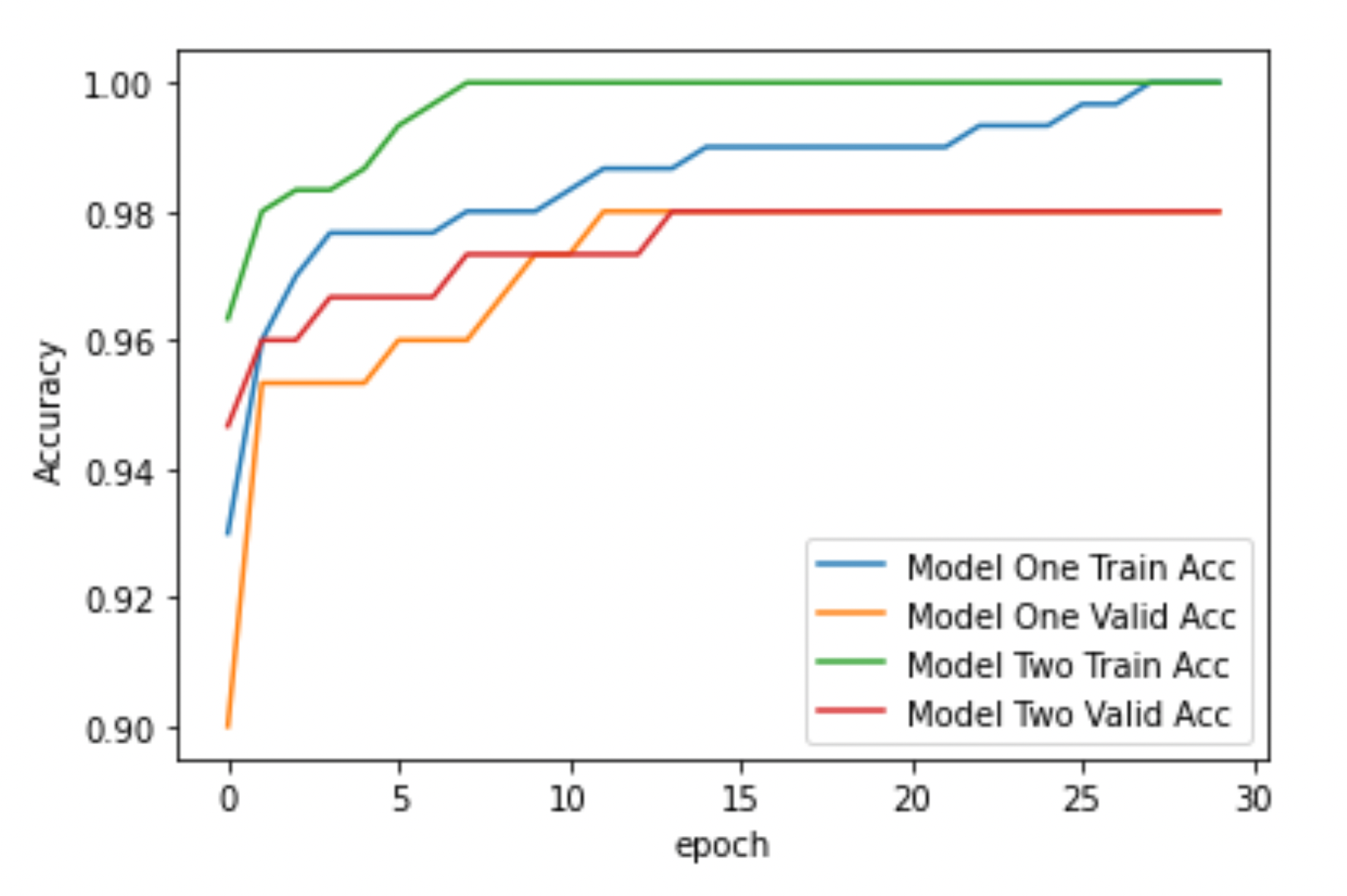
We can clearly see that the train loss and valid loss are close to each other for both model one and model two. We can also see that model two’s decrease in loss is much faster than model one’s relatively linear decrease in loss.

1. 

We can see that both the models train and valid accuracy stay closely together. In the beginning, model two’s accuracy grows at a relatively linear rate, and thus is slower than that of model two’s. However as the epoch grows sufficiently, both the models' accuracies reach a similar point.

1. Calculating the test accuracies for model one and two, we get:  
   Test Acc for Model One : 0.9599999785423279  
   Test Acc for Model Two: 0.8999999761581421  
   Notice, that here the test accuracy for the 2-layer model is less than the test accuracy for the 1-layer model. This wasn’t expected, as we expected the 2-layer model to be better fitting. This could likely be because of overfitting in the 2-layer model, which leads to better fitting the training data, and more error on the test data.

1. Upon replacing the SGD optimizer with the Adam optimizer, we get the following results:  
   Running the experiment for (i) again, we get,

  
Running the experiment for (j) again, we get,  


Recalculating the test accuracies for models one and two (part (k)), we get:  
Test Acc for Model One: 0.9733333587646484  
Test Acc for Model Two: 0.9666666388511658  
  
Here, using the Adam optimizer we can see that the models’ loss and accuracy graphs approach each other. Notice that the first graph (the loss graphs), model one’s and two’s losses decrease rapidly to reach an asymptote of ~0.05. Also notice that the second graph, model one’s and two’s accuracies increase rapidly to approach ~0.98 (for valid accuracies), and ~1.00 (for train accuracies). When we calculated the train accuracies, they have also improved for both the model but more significantly for model two. This could be because by using the Adam optimizer, we are adjusting the step size for each epoch which is leading to a better fitting estimation for the two-layer model.