Programming assignment 3: Recurrent Neural Network

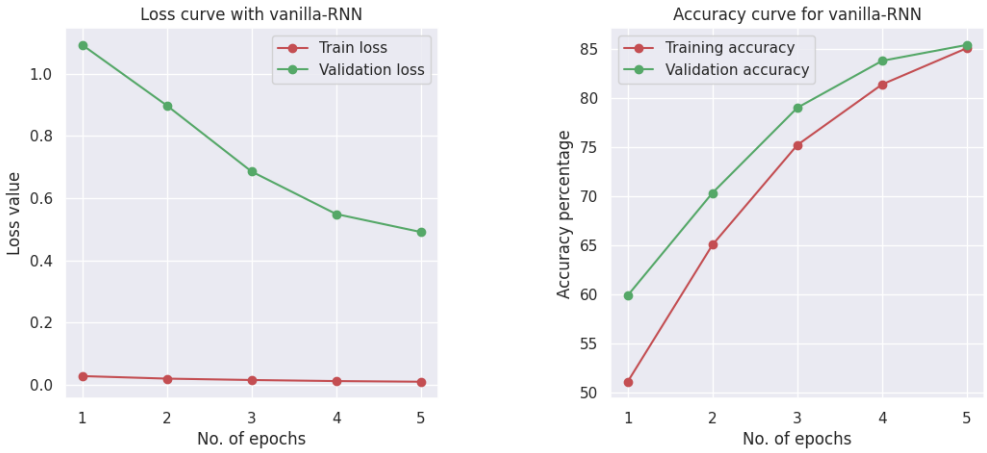
Arunima Sarkar

1 MNIST Classification using RNN.

The baseline models has :

1. input size= 28
2. no.of layers= 1 as we are not using stacked RNN
3. batch-size= 50
4. epochs= 5
5. Bidirectional flag = $False$
6. hidden layer size= 128

Vanilla-RNN

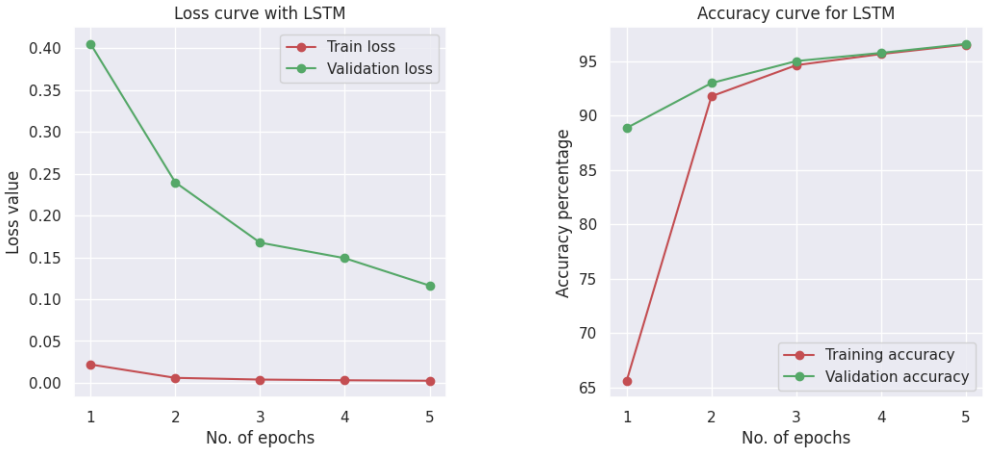


((a)) Loss curve ((b)) Accuracy curve

Figure 1: Plots for loss and accuracy for RNN

The average accuracy on the test set is 87.6%

LSTM

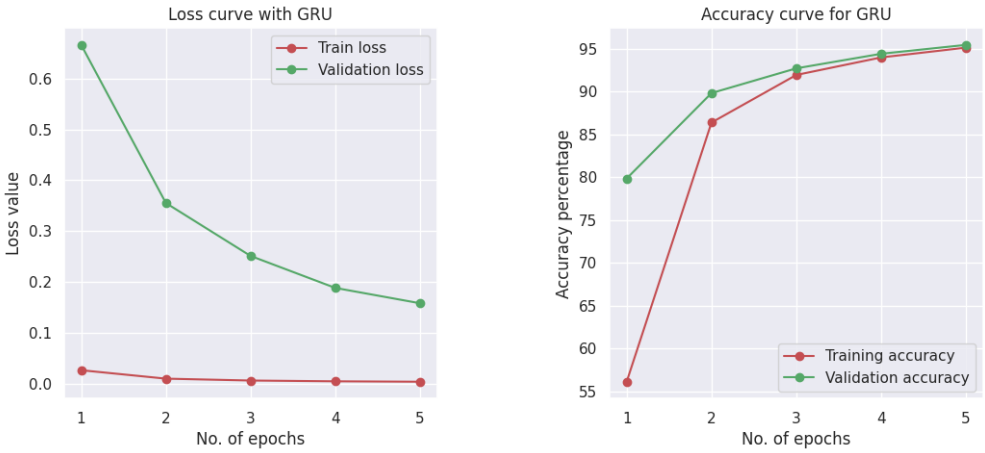


((a)) Loss curve ((b)) Accuracy curve

Figure 2: Plots for loss and accuracy for LSTM

The average accuracy on the test set is 96.83%

GRU



((a)) Loss curve ((b)) Accuracy curve

Figure 3: Plots for loss and accuracy for LSTM

The average accuracy on the test set is 95.7%

The accuracy of LSTM is best thus I chose it as my model to run it in randomly sampled test data.The results for 4 such iterations are shown below.

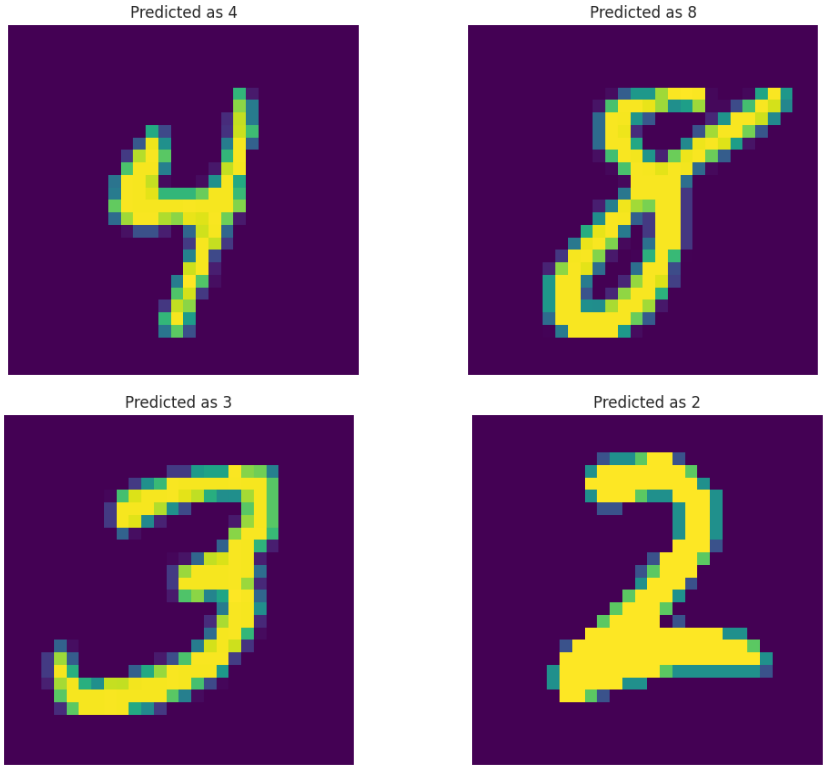
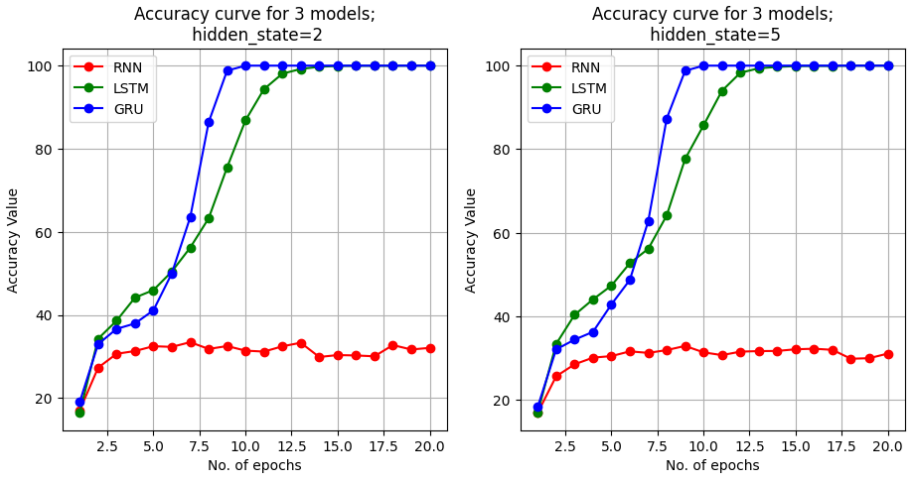


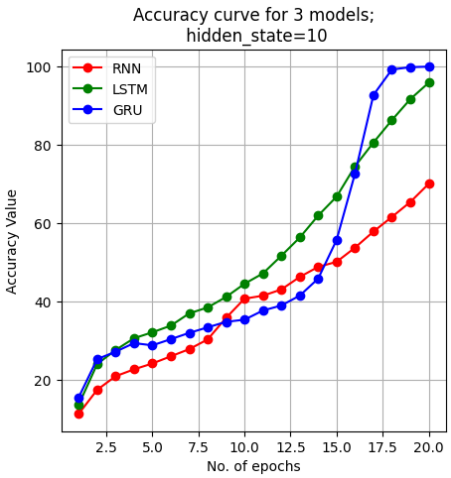
Figure 4: Images after random selection

2 Remembering the number at a particular index in a given sequence.

The plots below will show the loss and accuracy for hidden state of 2,5,10. Answer and execution of question 4(b) is shown in the notebook.

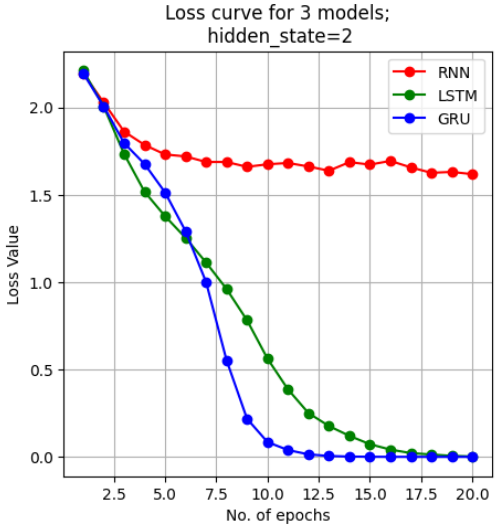


((a)) ((b))

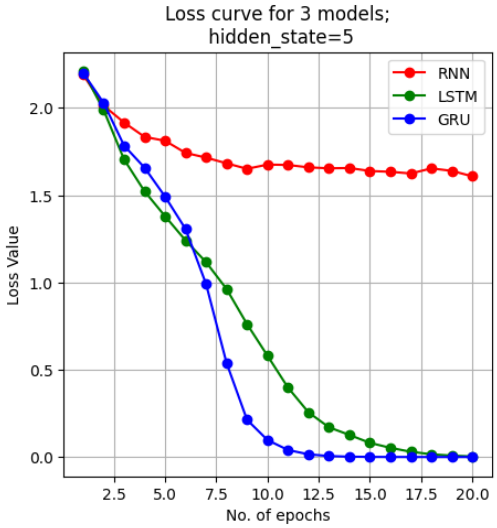


((c))

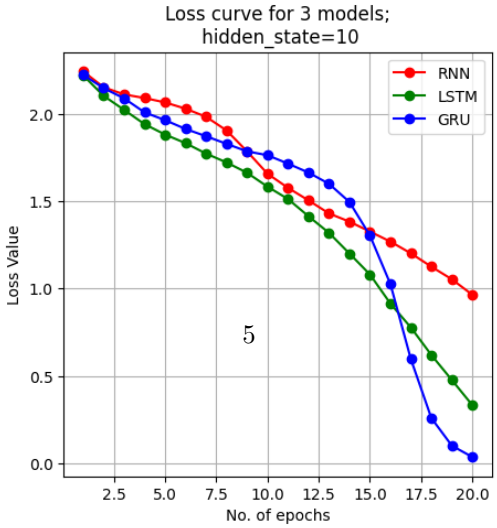
Figure 5: Accuracy curves


Loss curve for 3 models;
hidden_state=2
RNN
LSTM
GRU
Loss Value
No. of epochs
0.0
0.5
1.0
1.5
2.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0

((a))


Loss curve for 3 models;
hidden_state=5
RNN
LSTM
GRU
Loss Value
No. of epochs
0.0
0.5
1.0
1.5
2.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0

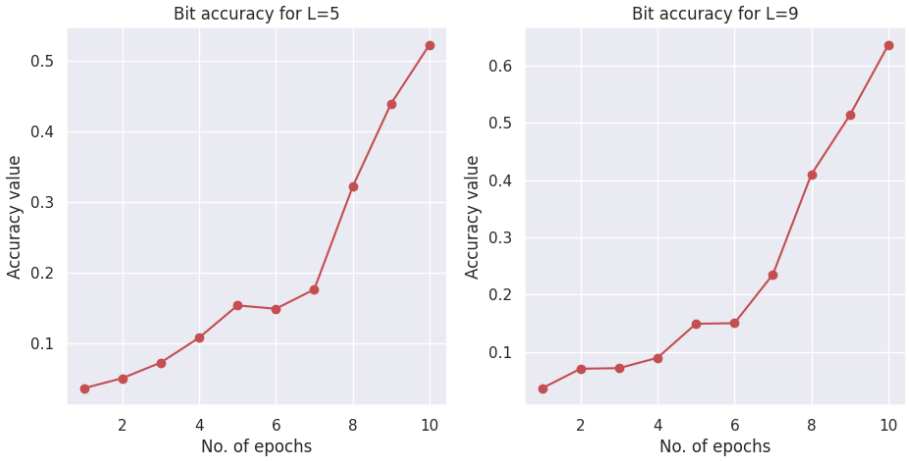
((b))


Loss curve for 3 models;
hidden_state=10
RNN
LSTM
GRU
Loss Value
No. of epochs
0.0
0.5
1.0
1.5
2.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0

((c))

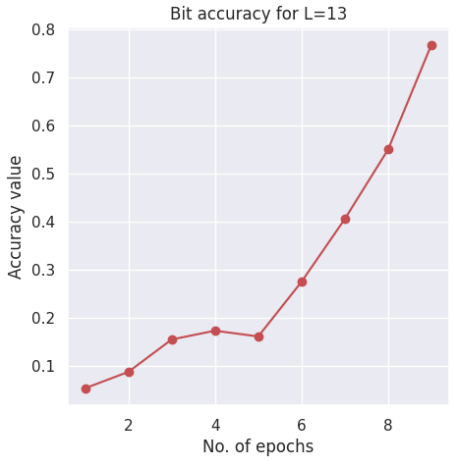
3 Adding two Binary strings

Upon varying the state vector size there is a increase in bit accuracy as evident from the graph below.



((a))

((b))



((c))

Figure 7: Accuracy curves

We can see that for L=13 the bit accuracy reaches 80%, whereas it is about 65% for L= 9 and 52% or 53 % for L = 5.

Upon comparing the accuracies for CE and MSE loss

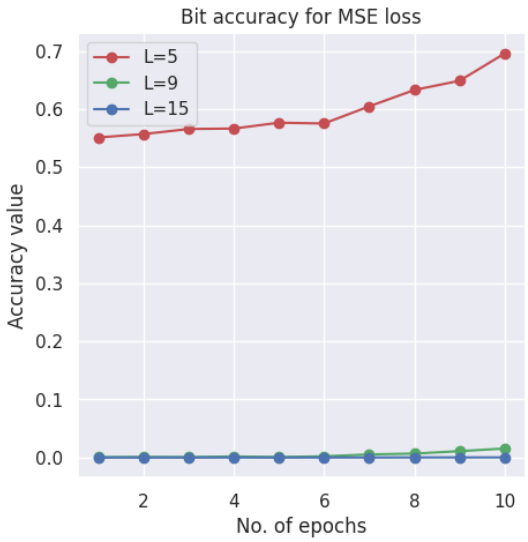


Figure 8

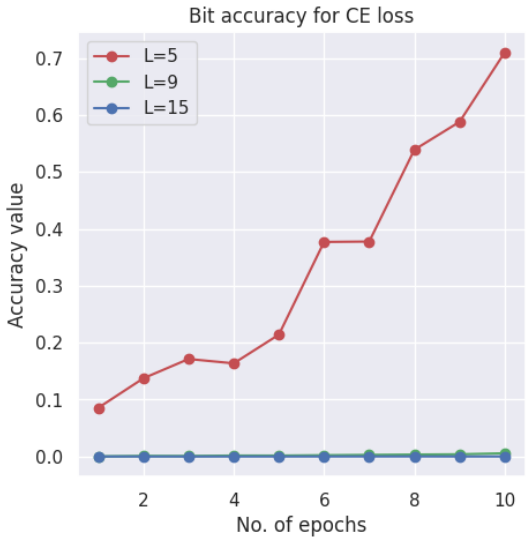
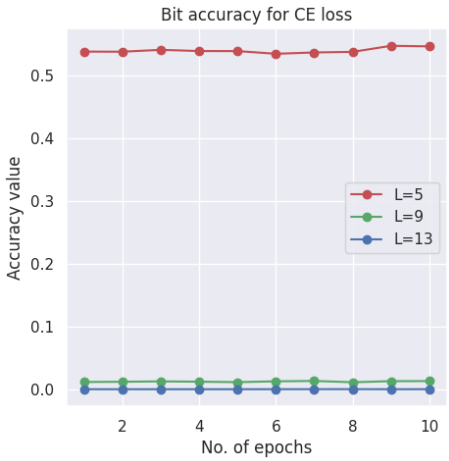


Figure 9

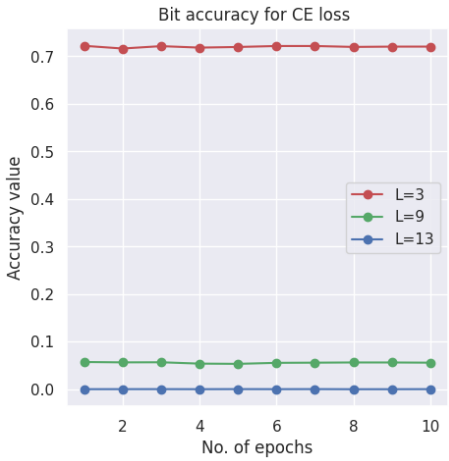
We can clearly see that CE loss is slightly better than MSE loss especially for lower L, however for higher L values both the losses perform equally.

Train on fixed length and test on varying

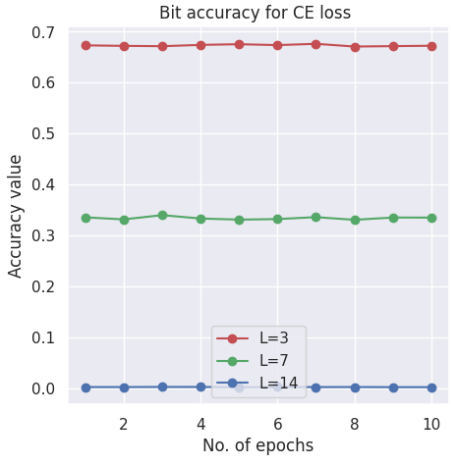
I purposefully chose CE loss as it is slightly better than MSE. We can see that when trained on L=3, the accuracy for 5 was high but for 9 and 13 are extremely low. Similarly when trained on L=5 the accuracy on 3 is high but very less low for 9 and quiet low for 13. And upon training on L=10, accuracy is high on 3 , moderately low on 7 and quiet low on 14.



((a)) Trained on L=3



((b)) Trained on L=5



((c)) Trained on L=10

Figure 10: Accuracy curves

3.4.3 Last part of the assignment is shown in notebook.