A0191508A Qiu Hai
CS5242 Neural Network and Deep Learning 2018/18 Sem 1
Assignment 3

Part 1

We can observe that the RMSE will initially decrease until length=5, then slightly going upwards and remain stable to the end. Because filter output depends only on last 5 input data points, so from length 1 to 5, the model got more and more relative input data and learned more relations between input and output, which leads to a better prediction result and less loss.

When the length comes to 6, unrelated noisy data starts to come in to the model, which comprised the normal learning process and deteriorate the training result. But the model can also find out that the data from 6 timestamps away are unrelated data and correct the output by either assign less weights to those data or let them cancel out each other (Figure 1-1,2). So the loss will increase just a little and basically remain stable from 6 to 10.

Part 2

We can observe that the loss of stateful RNN will decrease faster than stateless RNN from length 1 to 5. Because the stateful RNN will remember the state of last batch but stateless RNN won’t, and we have a batch size = 1, so the stateful RNN will get some related information from previous timestamps passed by the state of last batch, meanwhile stateless RNN can only predict the output based on current input which has a length smaller than 5, so some related information will be missed.

From length 6 to 10, the performance of both RNN model starts to worsen. Because we inverted the input sequence, making the most recent input comes first and the least recent input comes last, and valina RNN is not capable of learning “long-term dependencies”[1], so the output will be influenced by the least recent input most likely, which is independent from the real output. Besides, stateless RNN will outperform stateful RNN slightly from 6 to 10, because the stateful RNN will be biased by the rarely related state from last batch, but stateless RNN won’t

Part 3

The trends of losses from length 1 to 5 are very similar to RNN, but from 6 to 10 the performance didn’t deteriorate that much as RNN did. I think the reason is that, unlike vanilla RNN, LSTM model can learn the relationship between output and both long-term and short-term memory and decide which one is in dominance. So after length = 6, the LSTM model start to adjust the output in favor of depending more on long-term memory and less on short-term memory, preventing the performance from dropping too much.

Part 4 and other observation

The performance of each model is shown in the graph clearly.

We can also observe that the loss curve of stateful RNN and stateful LSTM start to drop drastically from length = 2, but the others from length = 3. My explanation is that data from t-2 have the most influence on the output, and the stateless model will not get the data from t-2 until length = 3 but stateful model will get it from the state inherited from last batch, so the stateful model can outperform stateless model when length = 2.

I assume the output is the weighted average of last 5 inputs, and make a model with only one dense layer of 5 inputs and one output. The RMSE on testing set is 4.4e-07, and weights for the last 5 inputs are [0.054, 0.244, 0.403, 0.244, 0.054], which supports my explanation.

[1] Understanding LSTM Networks http://colah.github.io/posts/2015-08-Understanding-LSTMs/

Appendix

Part 1

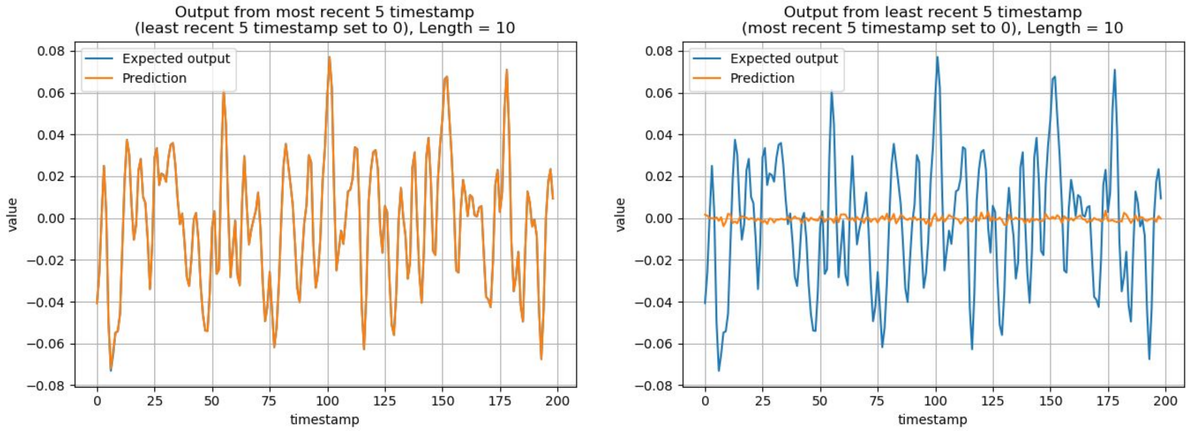
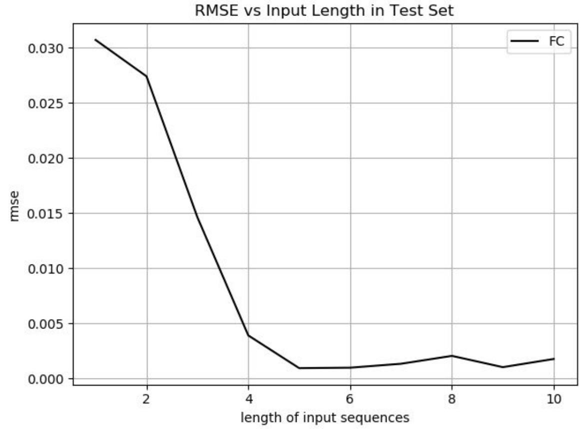
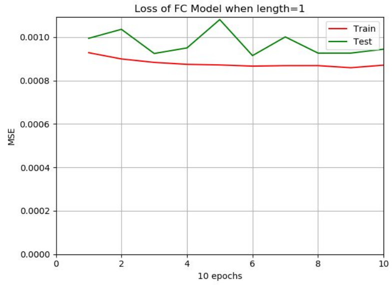
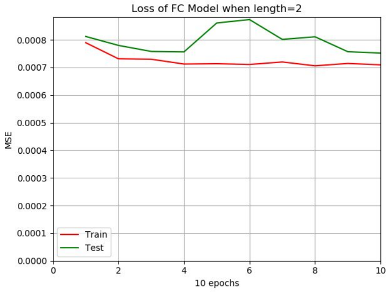
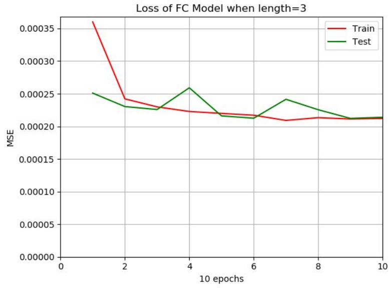


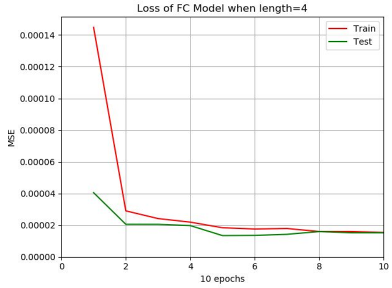
Figure 1-1,2

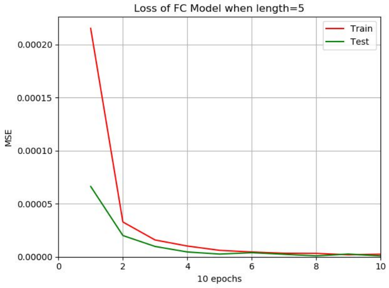


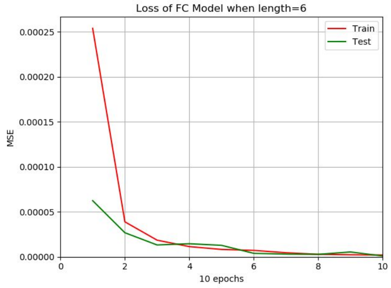


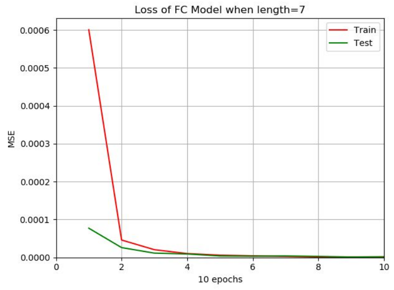


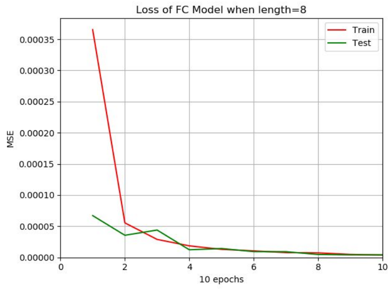


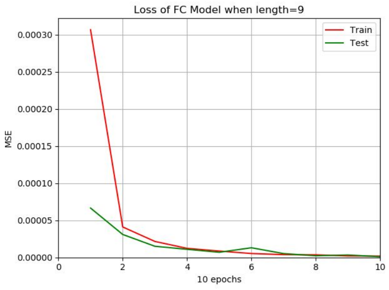


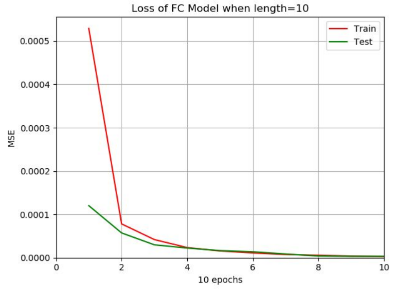




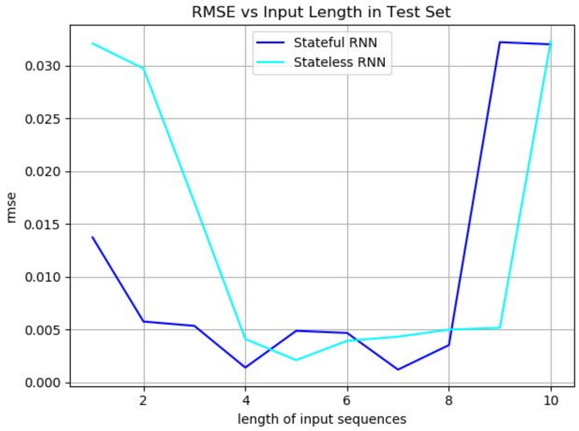

Loss of FC Model when length=7


Loss of FC Model when length=8

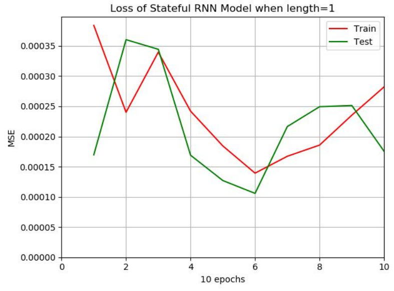

Loss of FC Model when length=9

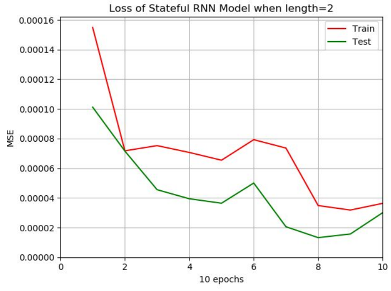

Loss of FC Model when length=10

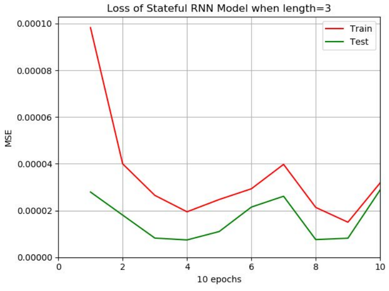
Part 2

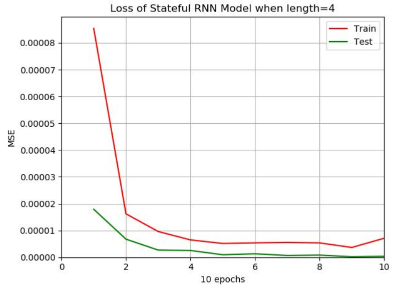

RMSE vs Input Length in Test Set

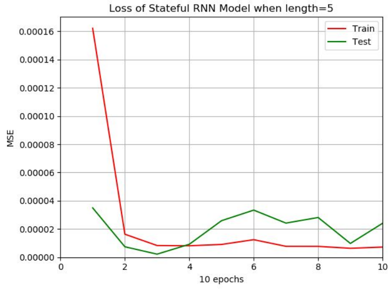
Stateful RNN:

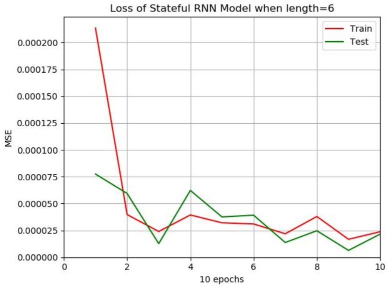

Loss of Stateful RNN Model when length=1

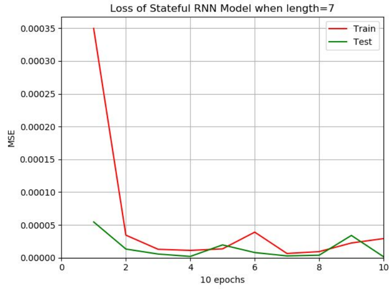

Loss of Stateful RNN Model when length=2


Loss of Stateful RNN Model when length=3

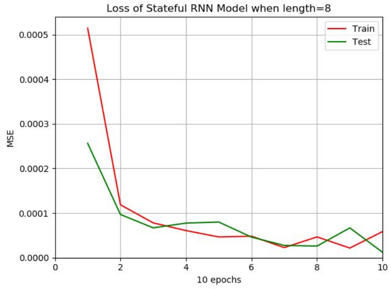

Loss of Stateful RNN Model when length=4


Loss of Stateful RNN Model when length=5

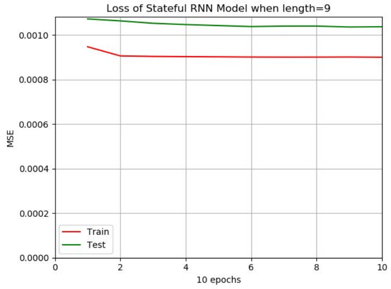

Loss of Stateful RNN Model when length=6



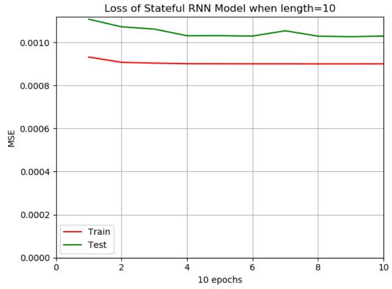
Loss of Stateful RNN Model when length=7



Loss of Stateful RNN Model when length=8

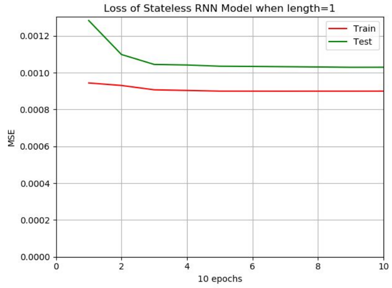


Loss of Stateful RNN Model when length=9

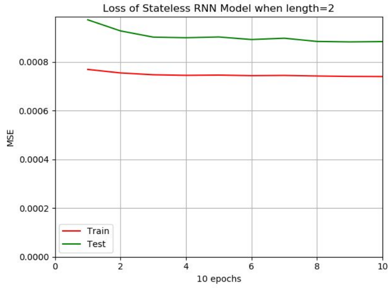


Loss of Stateful RNN Model when length=10

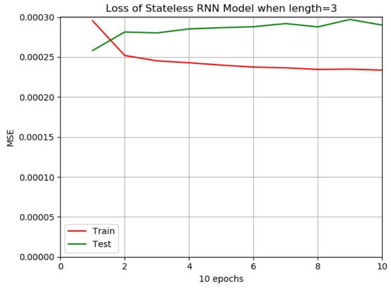
Stateless RNN:



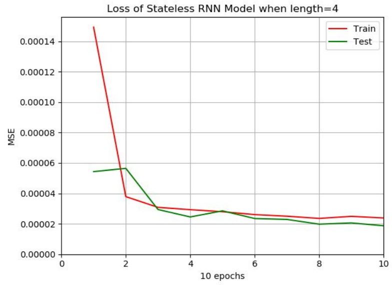
Loss of Stateless RNN Model when length=1



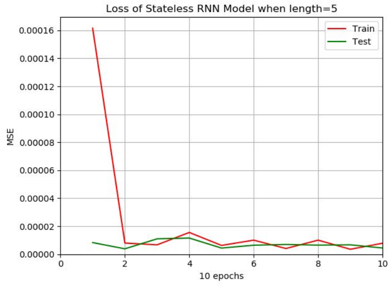
Loss of Stateless RNN Model when length=2



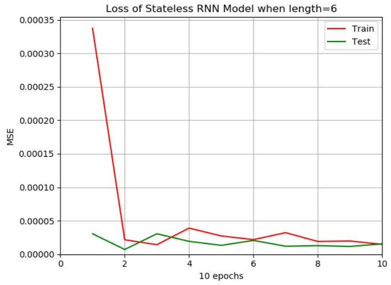
Loss of Stateless RNN Model when length=3



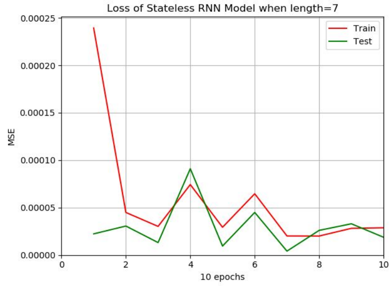
Loss of Stateless RNN Model when length=4



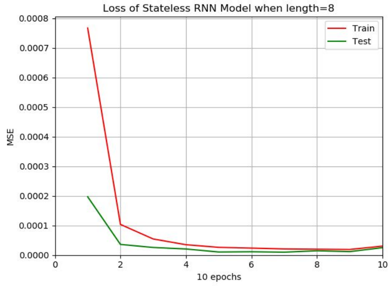
Loss of Stateless RNN Model when length=5



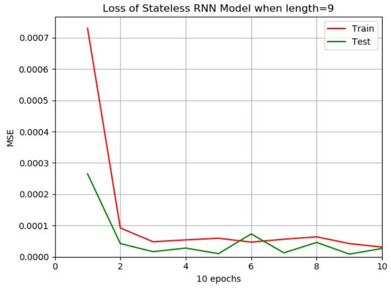
Loss of Stateless RNN Model when length=6



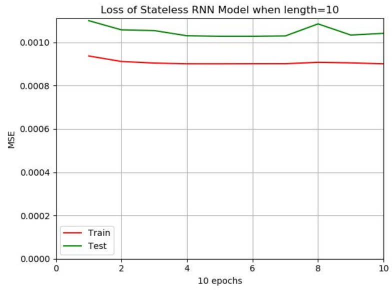
Loss of Stateless RNN Model when length=7



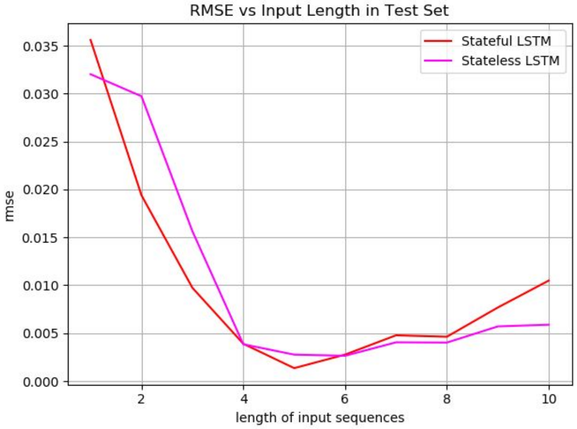
Loss of Stateless RNN Model when length=8



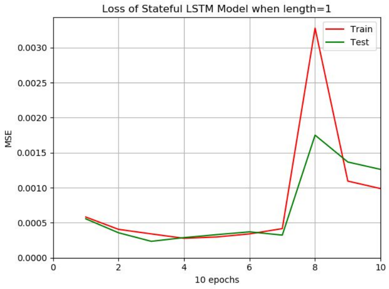
Loss of Stateless RNN Model when length=9

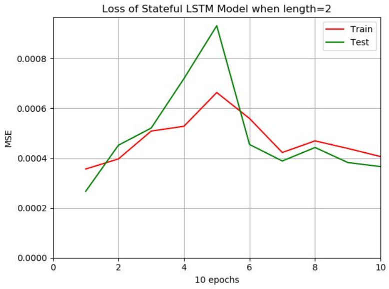

Loss of Stateless RNN Model when length=10

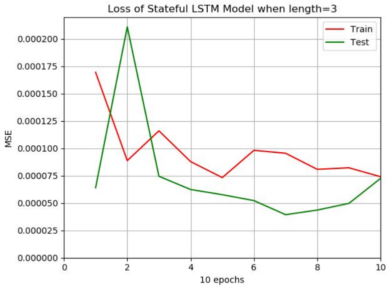
Part 3

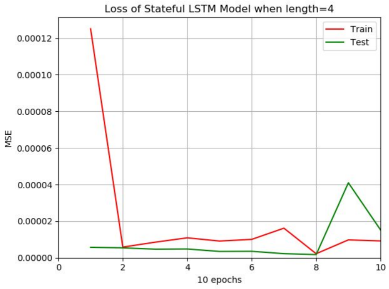

RMSE vs Input Length in Test Set

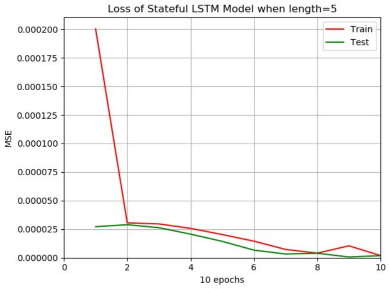
Stateful LSTM

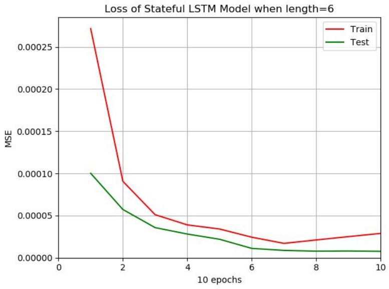

Loss of Stateful LSTM Model when length=1

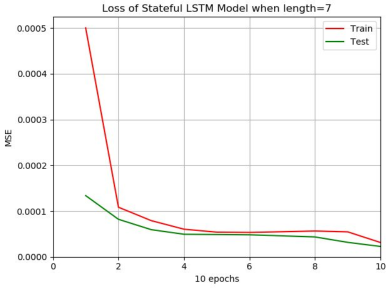

Loss of Stateful LSTM Model when length=2

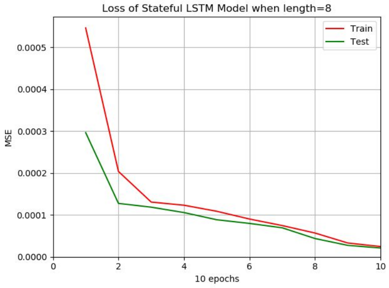

Loss of Stateful LSTM Model when length=3

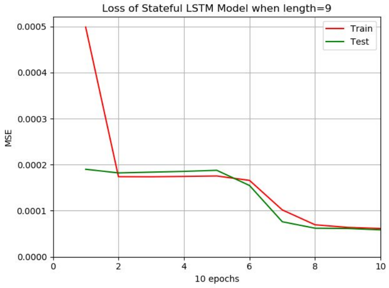

Loss of Stateful LSTM Model when length=4

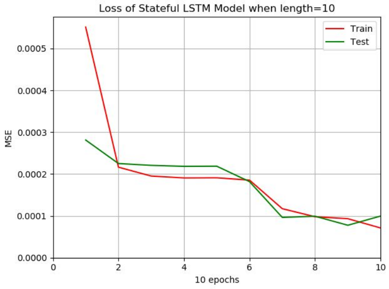

Loss of Stateful LSTM Model when length=5


Loss of Stateful LSTM Model when length=6

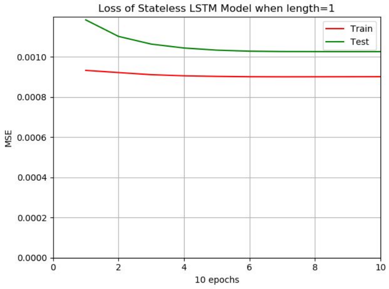

Loss of Stateful LSTM Model when length=7

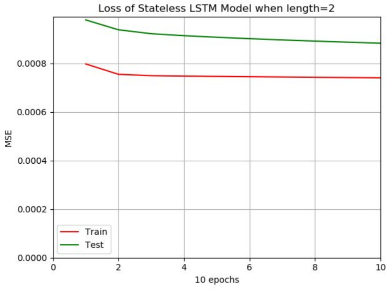

Loss of Stateful LSTM Model when length=8

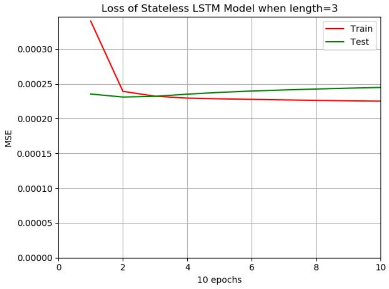

Loss of Stateful LSTM Model when length=9

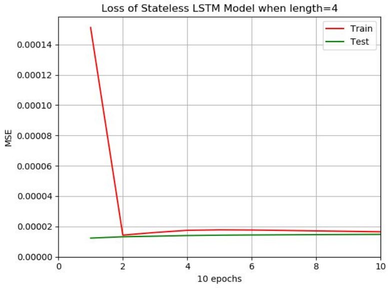


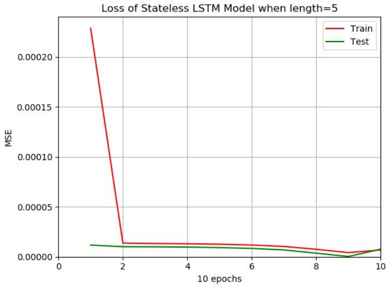
Stateless LSTM

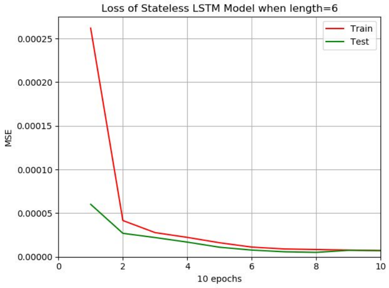


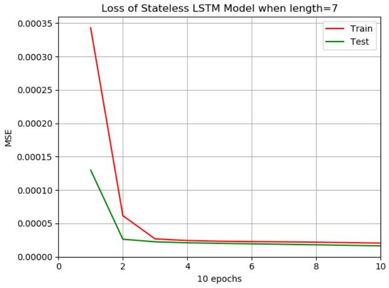


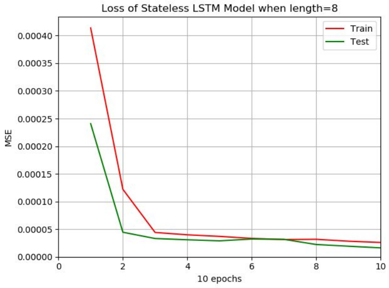


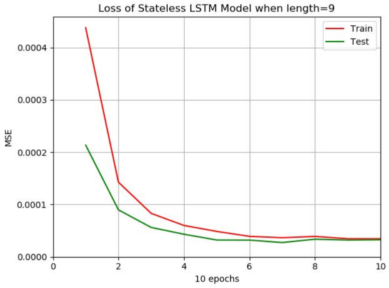


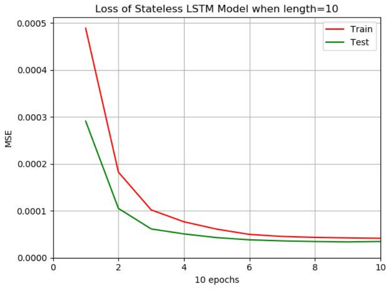




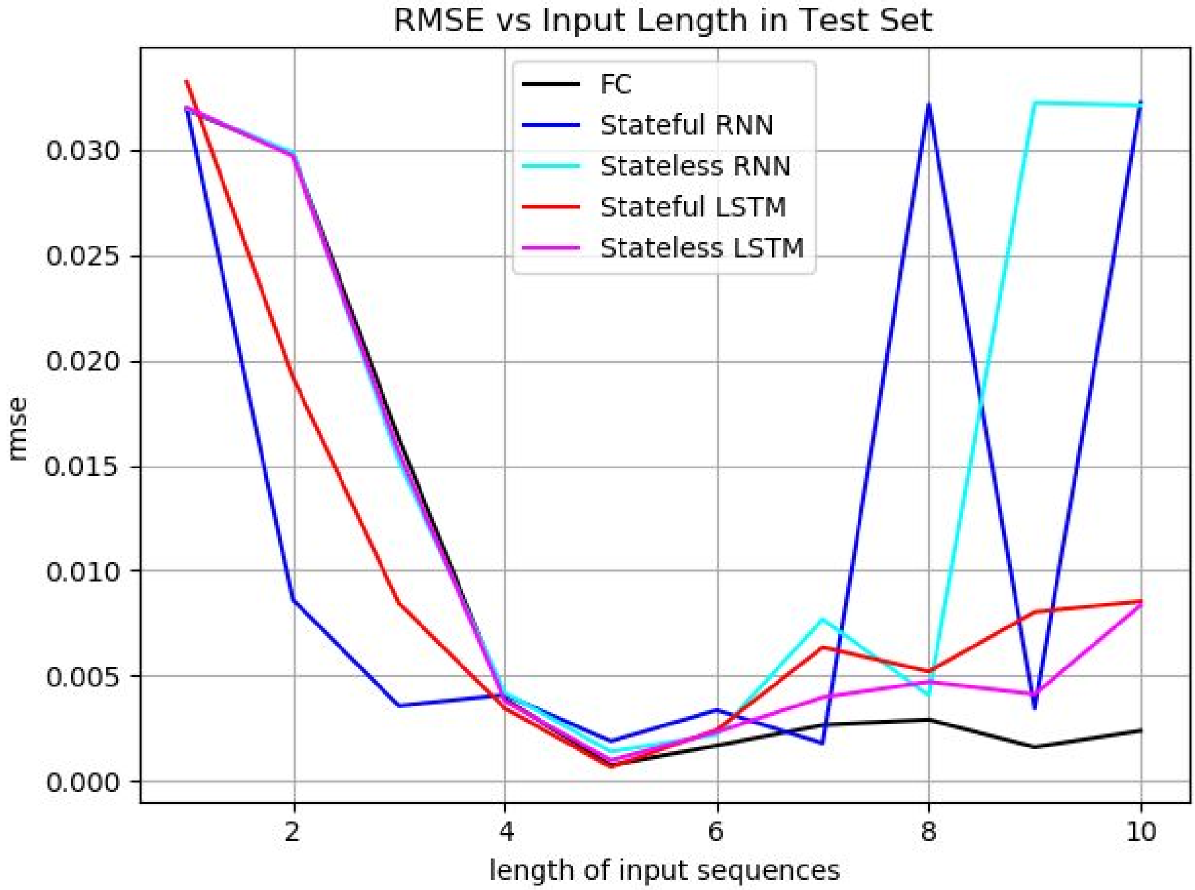








Part 4



RMSE vs Input Length in Test Set
FC
Stateful RNN
Stateless RNN
Stateful LSTM
Stateless LSTM
rmse
0.000
0.005
0.010
0.015
0.020
0.025
0.030
2
4
6
8
10
length of input sequences