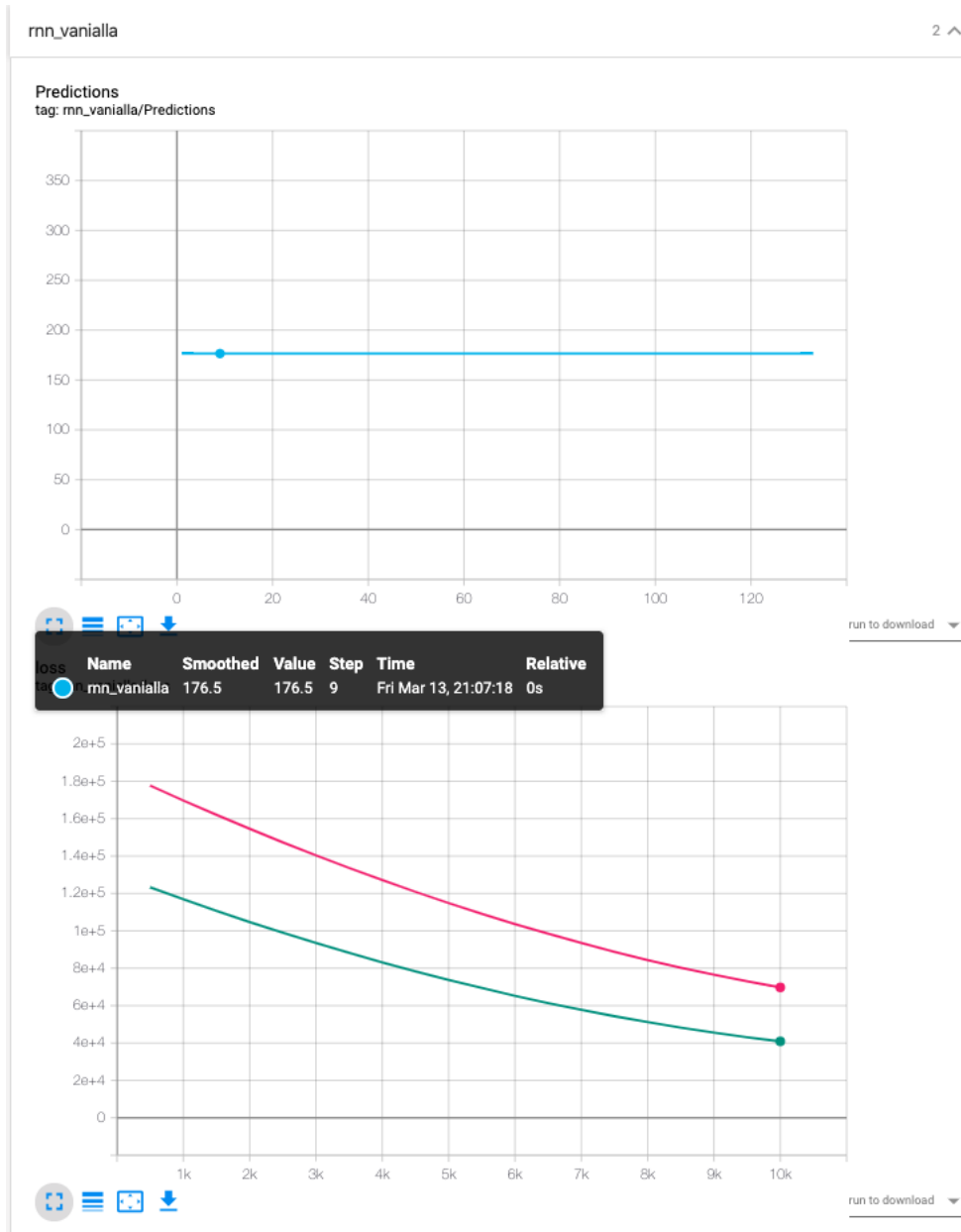


For all experiments, first 75% points are chosen as training set and the rest 25% points are chosen as testing. I couldn't colour code predictions to show this differentiation but you should be able to make out the difference by observing x-axis.

Four experiments are:

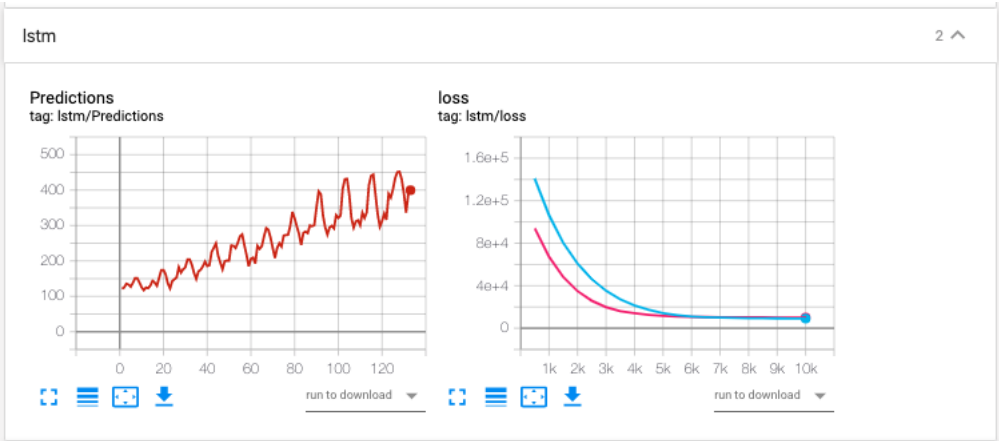
1. Vanilla RNN with just 2 hidden units (learns nothing)
2. RNN with 128 hidden units (learns the pattern)
3. LSTM with 256 hidden units (performs best)
4. GRU was made as a new predictor to answer 3rd question (performs better than RNN)

5.

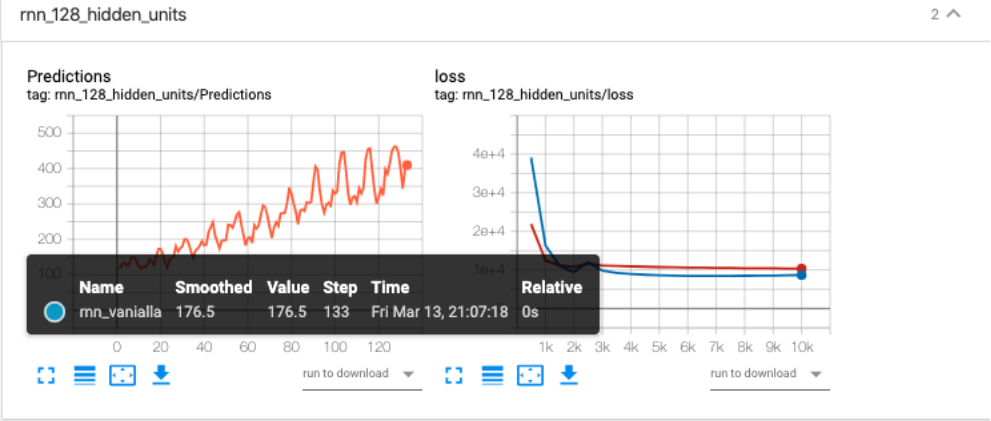


6.

7.



8.



9.

