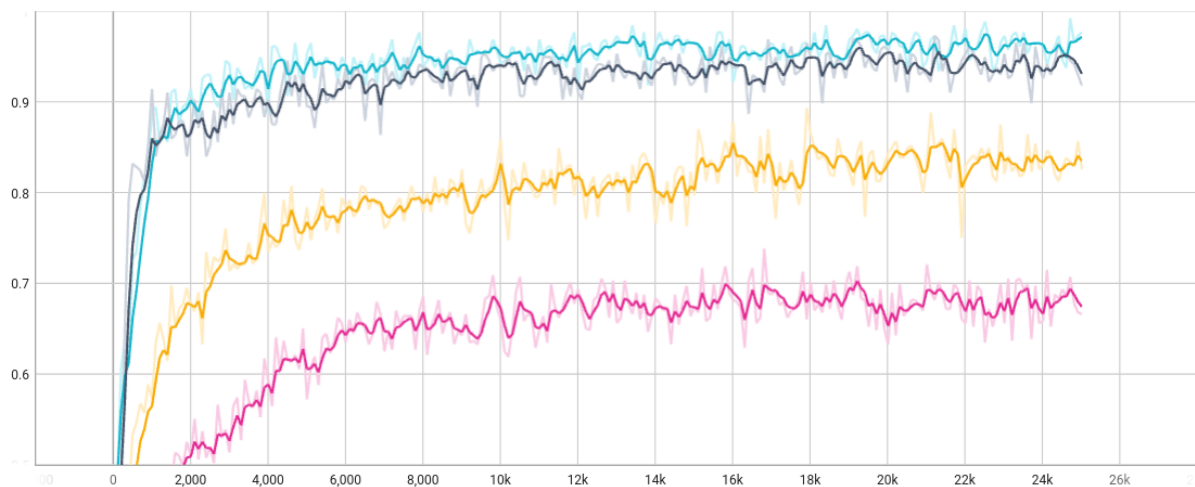


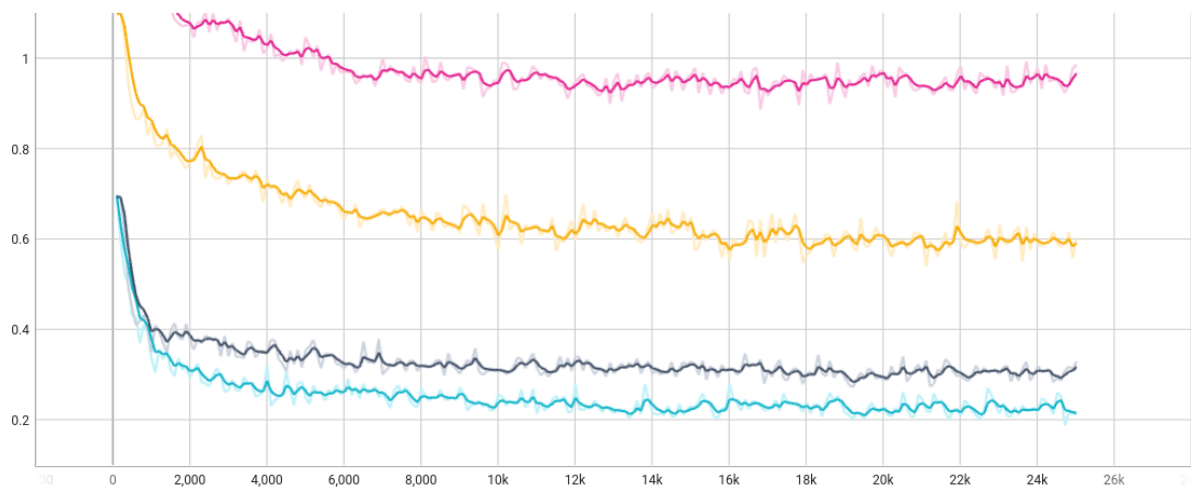
Following are the results for various configurations of K and N. As expected if we increase K, accuracy on the held out set also increases. If we increase N while keeping K fixed, accuracy cramps down.

In the figure below Blue is (K=2,N=2); Black is (K=1,N=2); Yellow is (K=1,N=3) and Red is (K=1,N=4)

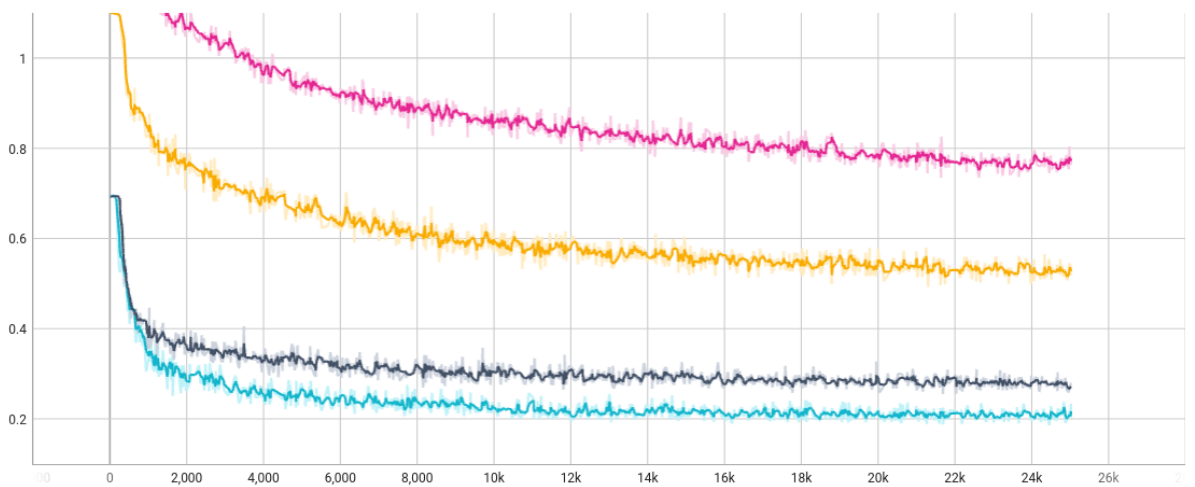
Accuracy/test



Loss/test



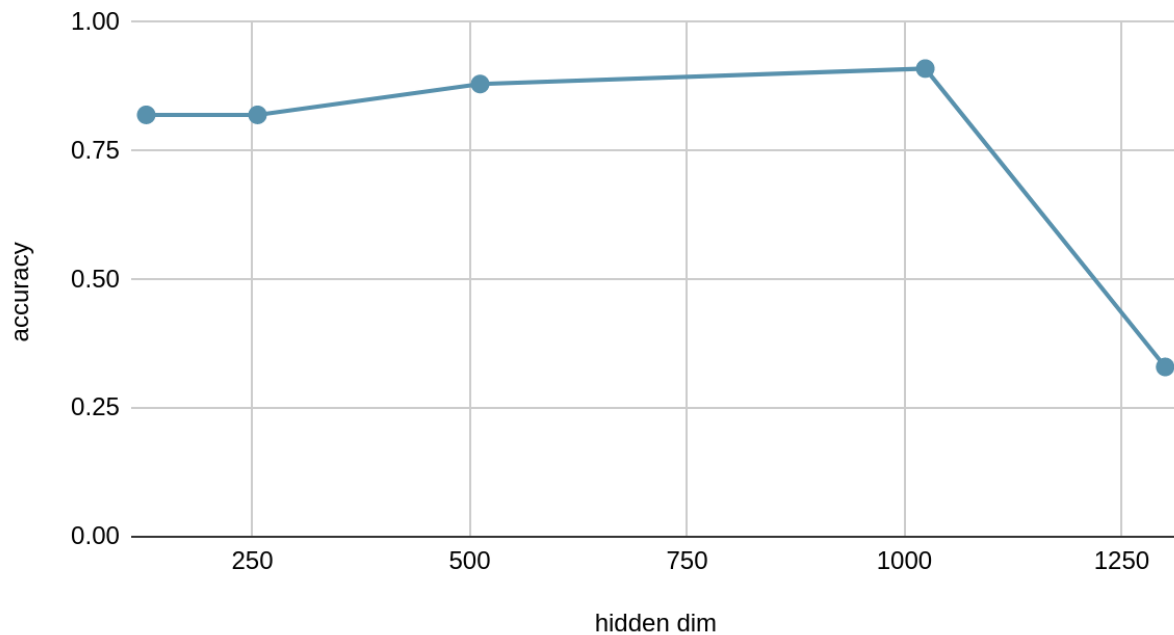
Loss/train



For the experimentation part, I tried sweeping across the hidden dimension of the LSTM from sizes 128->256->512->1024->1300 while keeping all other parameters fixed.

We see that the performance (measured as the query set accuracy) increases till the hidden dimension of 1024 but drops when increased further to 1300.

accuracy vs. hidden dim



At the largest hidden size, I noticed that the training loss remained relatively high. So it is possible that if we run for more epochs, the accuracy of hidden size model 1300 can increase. However increasing the number of epochs beyond a point leads to overfitting.