

This handout includes space for every question that requires a written response. Please feel free to use it to handwrite your solutions (legibly, please). If you choose to typeset your solutions, the `README.md` for this assignment includes instructions to regenerate this handout with your typeset L^AT_EX solutions.

3.a

By comparing the plots for $(N, K) = (2, 1), (3, 1)$ and $(4, 1)$ - increasing the number of classes N reduces the overall classification accuracy of the model on the meta-test query set. The learning rate (the rate at which the training and test loss reduces) is slower too.

As N increases, the model has to additionally learn new classes and unless it predicts these new classes with 100% accuracy, the overall accuracy will decrease.

3.b

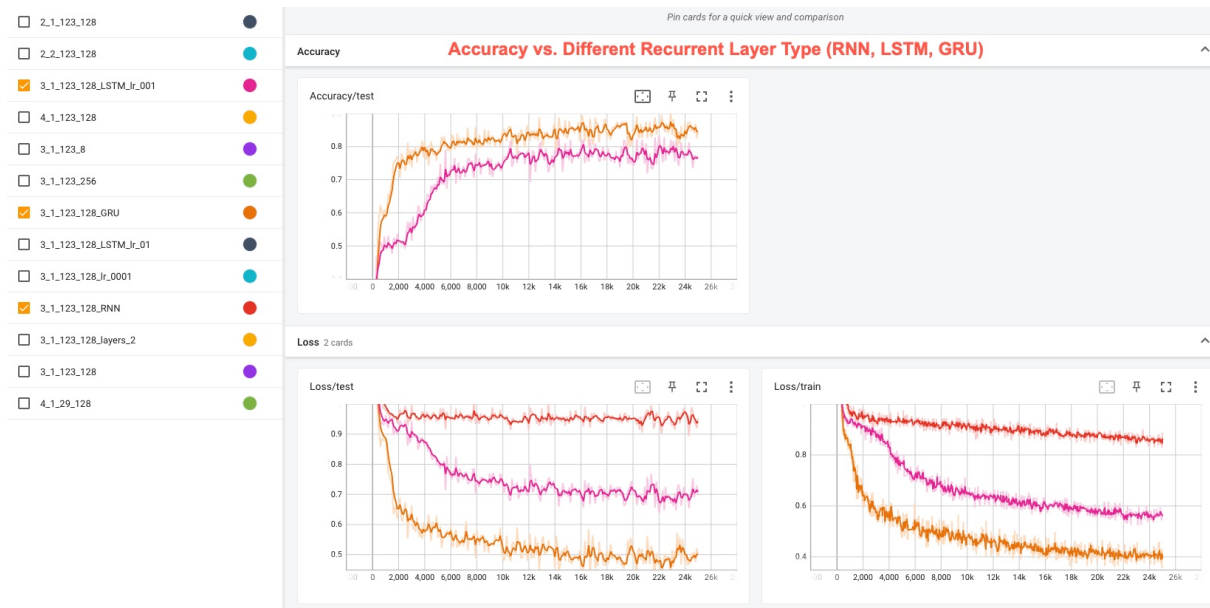
By comparing the plots for $(N, K) = (2, 1), (2, 2)$ - increasing the number of examples in the support set increases the accuracy of the model.

This is fairly intuitive as the model now has more examples to learn these new characters and hence does a better job in predicting these classes.

4.a

Performance vs. Model hyper-parameters

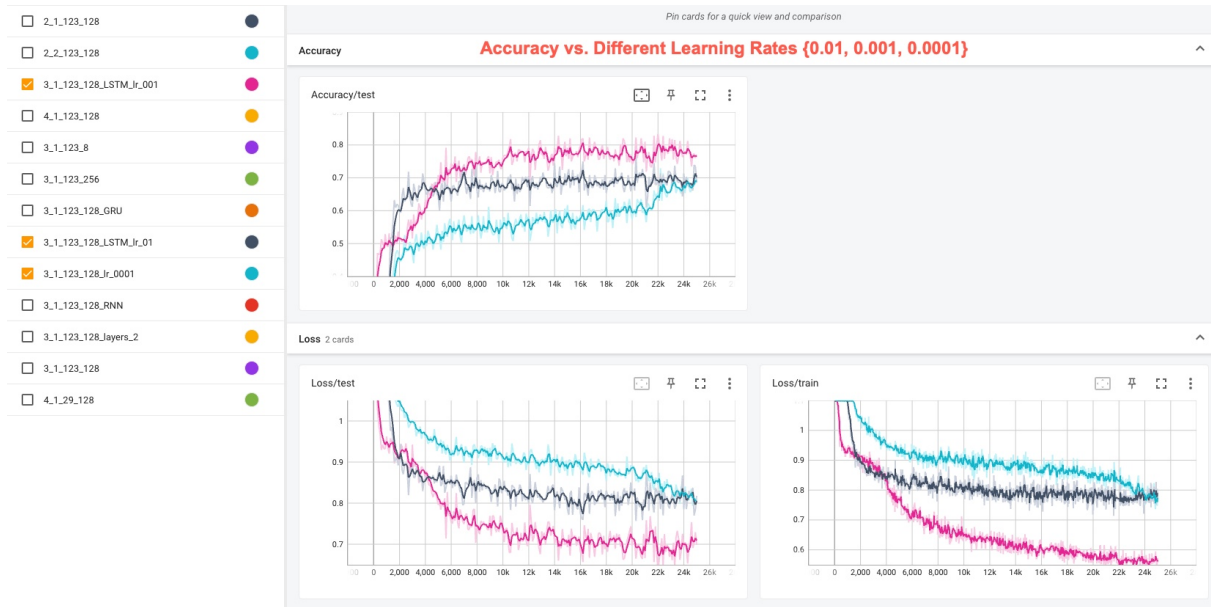
- Recurrent layer: Tried GRU and RNN as an alternate model architectures. GRU converged to a higher accuracy, and RNN converged to a lower accuracy.



- Size of hidden state: Increasing the size of hidden state would improve model performance. However, this would come with at the cost of a larger model size and higher memory requirements



- Learning rate: Increasing the learning rate would quicken the convergence to the optimum weights. However the variance once the model has converged is higher.



- Number of layers: A model with higher number of layers would generally yield higher accuracy.



4.b

Increasing the memory capacity improves the model performance (i.e., the accuracy on the meta-test query set) and likewise, decreasing the memory capacity deteriorates the model performance. The converged accuracy for $h = 8$ is less than the converged accuracy for $h = 128$ which in turn is less than the converged accuracy for $h = 256$.

h is a model hyper-parameter which corresponds to the number of hidden states. A model with higher number of hidden states can represent the information contained within the support set better and more accurately, which in turn is used to make predictions on the test dataset.

