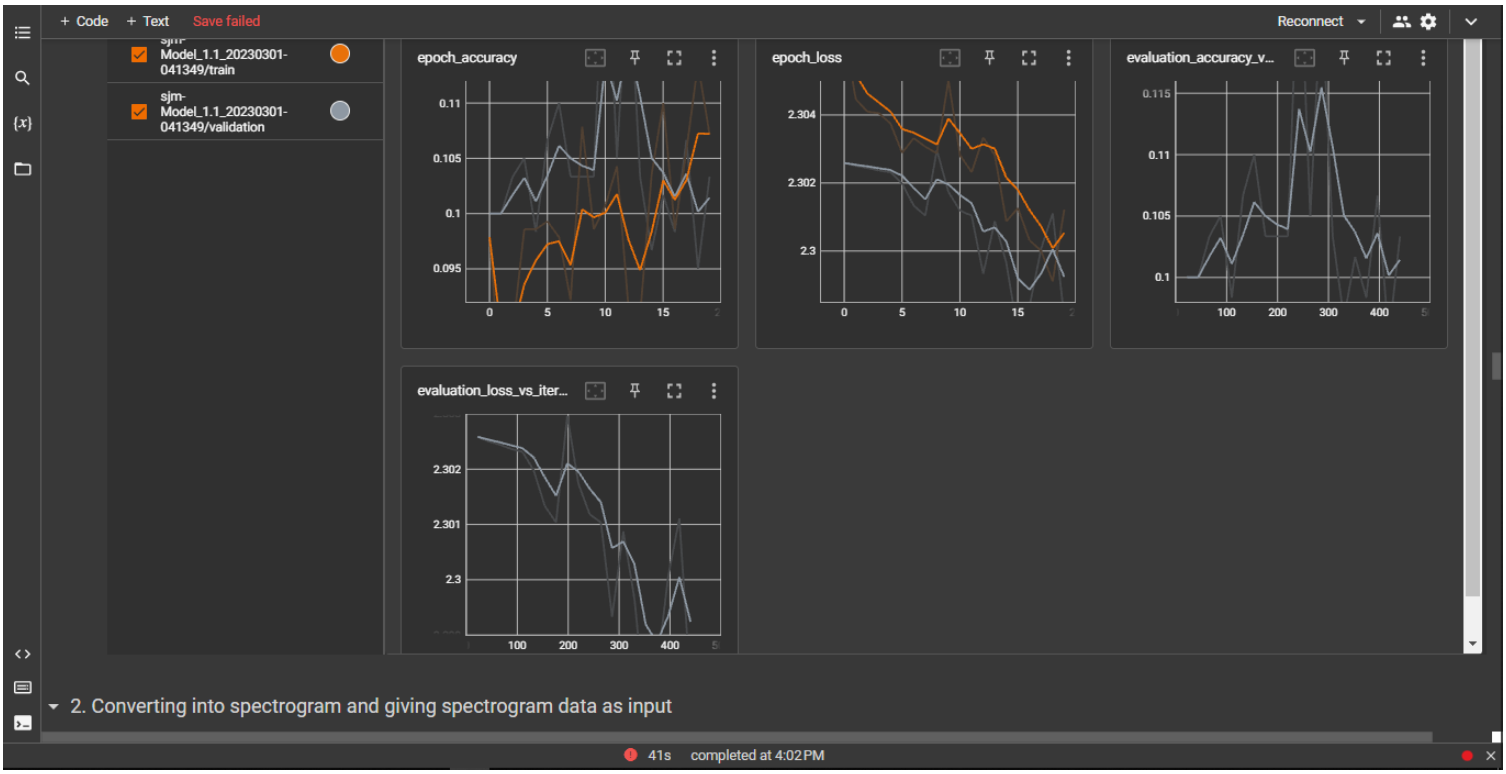
Spoken Digit Recognition Assignment

Model-1(Raw Data + LSTM)

Scalars

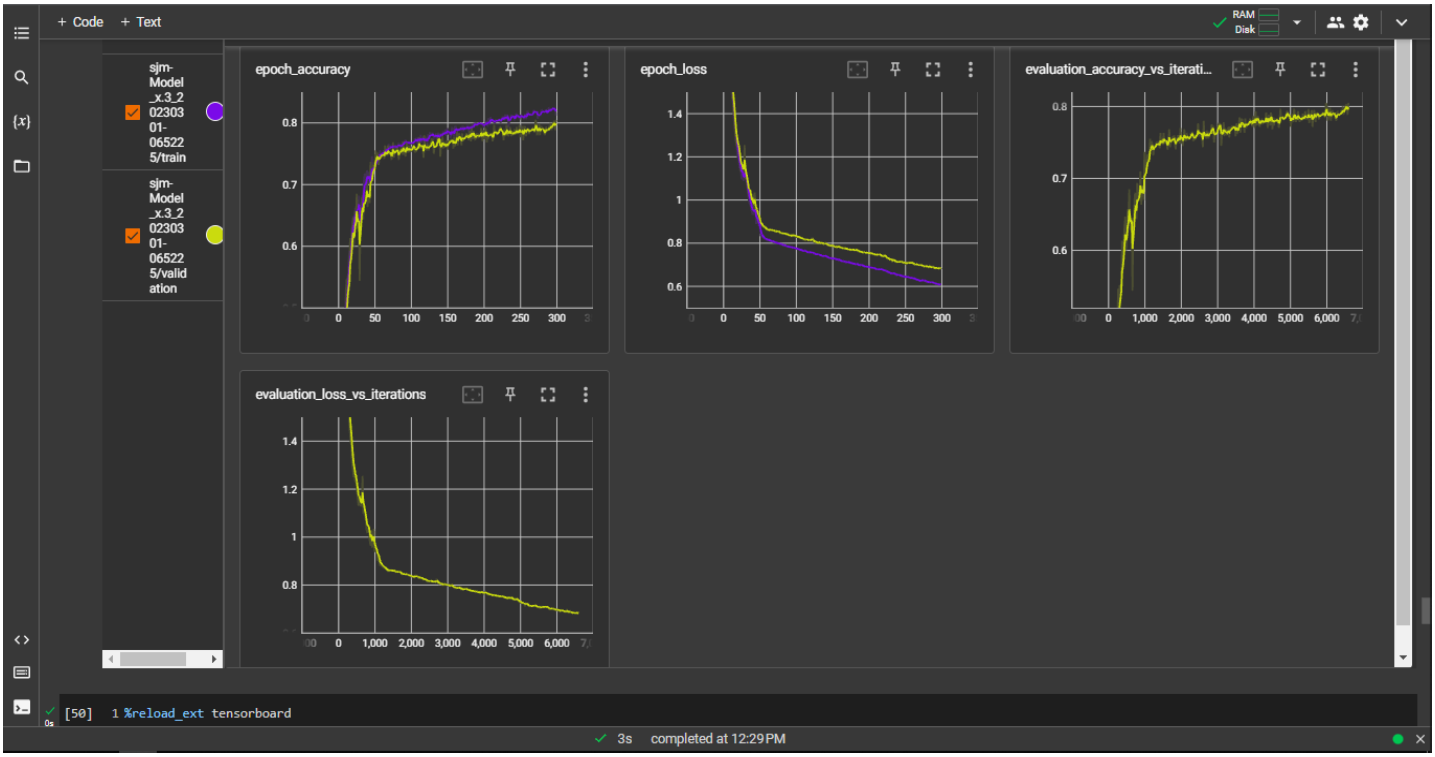


From the loss plots and the accuracy plots we can see that the overall trend of the models performance is increasing. The updates are very noisy as we have very little data to train Model-1(1400 data points).

Another reason for the model's low performance is the dimension of the data. In the train data each datapoint has 17640 time-steps with a single value at each step. This is a very long series of data, due to which the LSTM model is not able to learn from the data enough. Even though LSTMs are designed to handle long term-dependencies there is a threshold as to how long this “long-term” should be.

Model-2(Mel-Spectrogram + LSTM)

Scalars

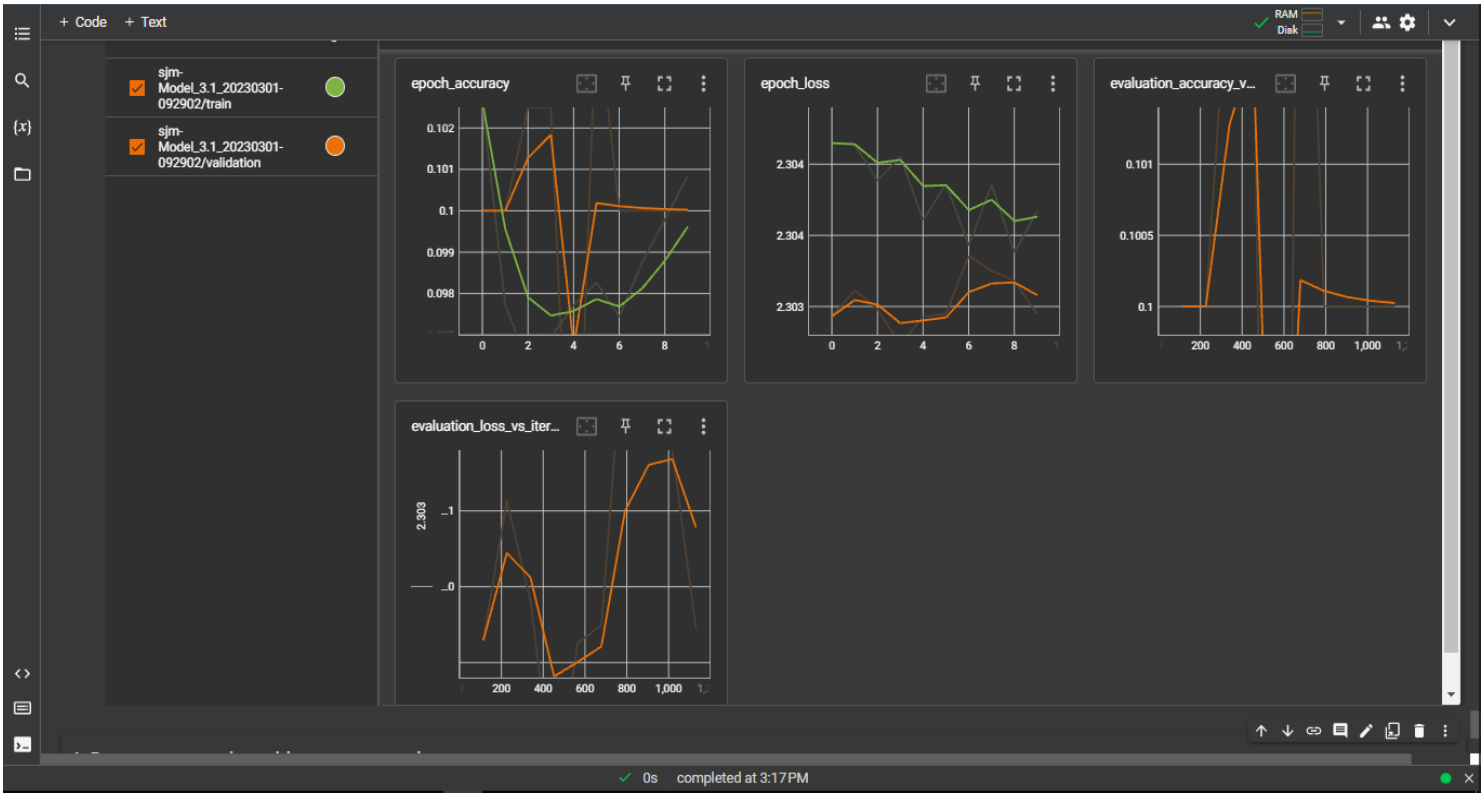


From the models loss and accuracy plots we can clearly observe that the performance is increasing with epochs. The updates in performance are very noisy in the sense that the accuracy/loss, overall, is increasing/decreasing but may decrease/increase a little after some epochs and increase/decrease again.

This model's performance is much better than model-1 due to its featurization(mel-spectrogram) and also lesser number of timesteps as compared to the first model.

Model-3(Raw - Augmented + LSTM)

Scalars



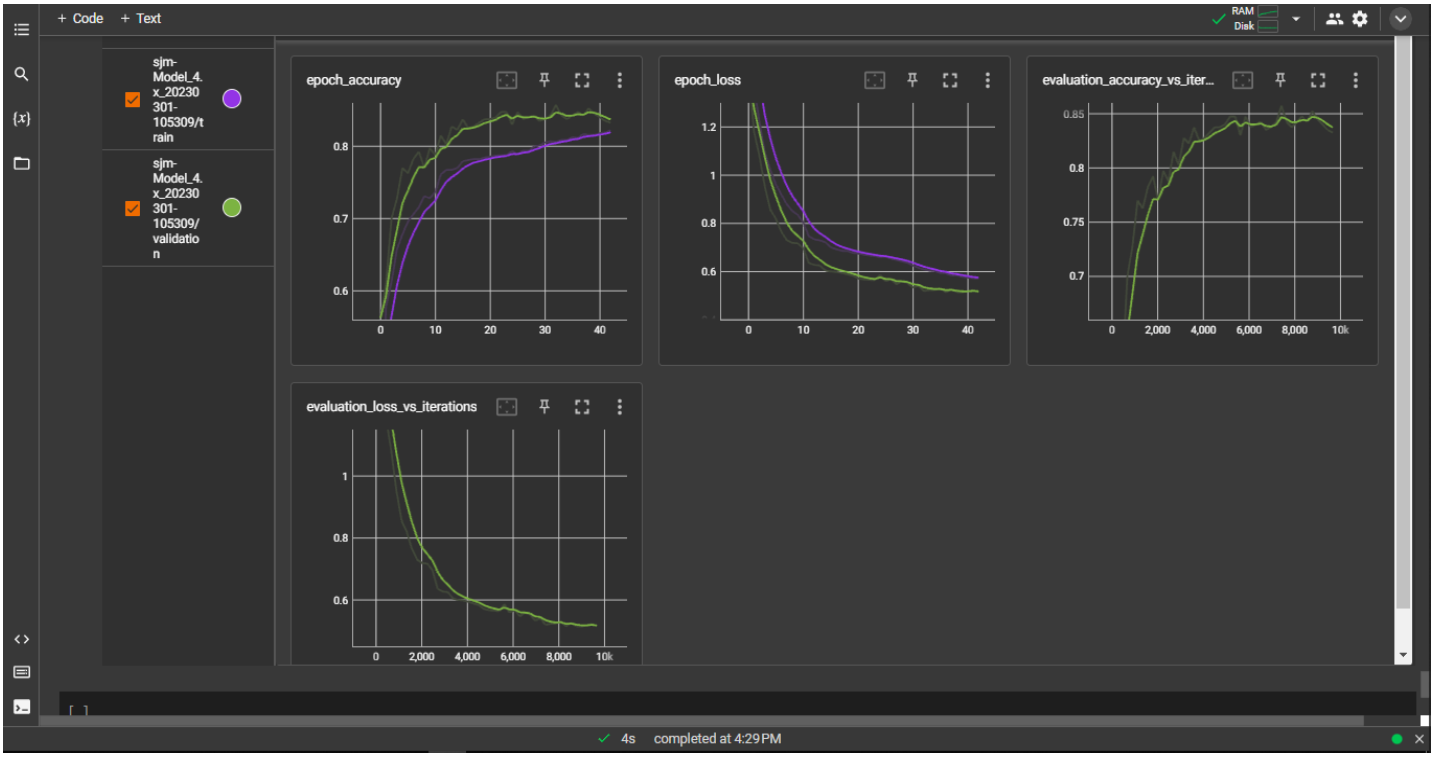
Epoch Accuracy:
The train accuracy first decreases and then starts to increase after some epochs. Whereas the test accuracy is somewhat fluctuating around but then becomes constant after the 5th epoch.

Epoch loss:
The loss for both the validation and train decreases overall. But on a closer look the train data loss keeps decreasing while the test data loss is somewhat on an increasing trend.

The model is not learning much from the data.

Model-4(Mel-Spectrogram of Augmented Data + LSTM)

Scalars



Epoch Accuray:
The accuracy over the epochs increases rapidly with increasing epochs and both the train and test accuracy start to show signs of saturation after the 40th epoch.

Epoch Loss:
This also follows the same trend as the accuracy, it decreases rapidly over the epochs and starts to saturate after the 40th epoch.

From these two plots alone we can say that the model is learning the patterns of the data quickly. This is due to the featurization of the train data as well as augmentation of the train data.