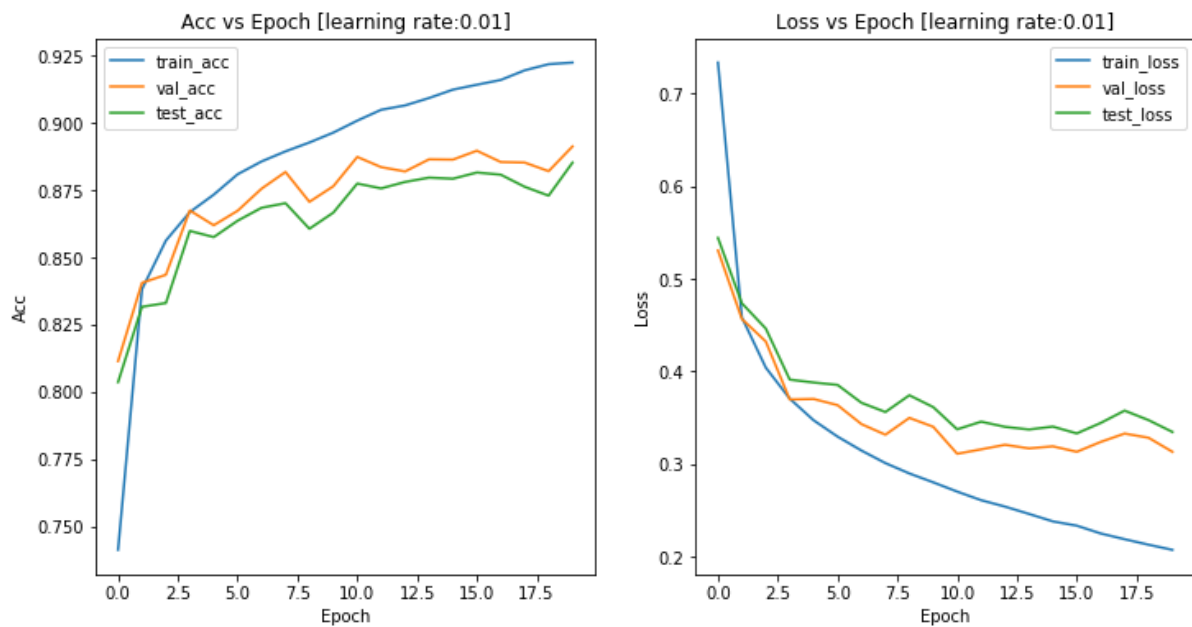
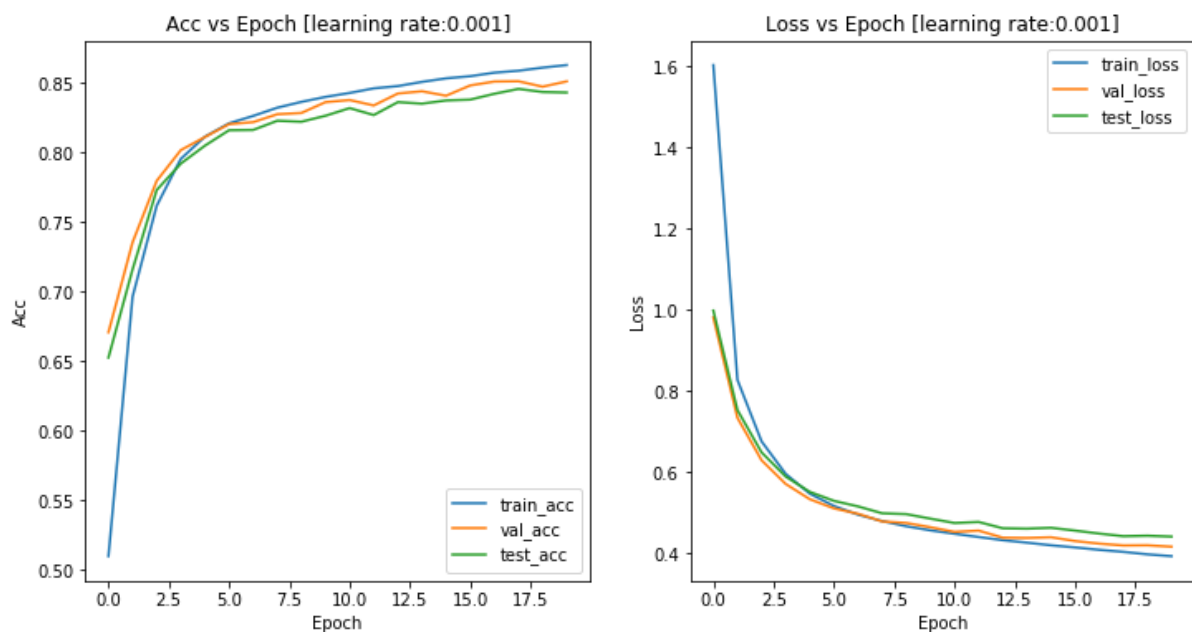


### 1. Graphs using Learning Rate = 0.01, Epoch = 20



### 2. Graphs using Learning Rate = 0.001, Epoch = 20



### 3. Question: Why you should not select the best model over all epochs on the test dataset?

The test dataset is used only to assess the performance of the trained model which should be done after the final model is chosen. After assessing the final model on the test dataset, we must not tune the model any further or else this would result in a biased error rate estimate of the final model on the test dataset. Selecting the best model should be done using the validation dataset to estimate how well the model has been trained.