

Part 1:

a)

Batch Size	Trial 1 (seconds)	Trial 2 (seconds)	Trial 3 (seconds)	Trial 4 (seconds)	Trial 5 (seconds)	Mean	Std
16	177.29	162.58	163.90	166.27	162.93	166.59	5.50
32	81.75	82.98	82.63	84.04	84.67	83.21	1.03
64	64.00	66.19	68.61	70.61	64.45	66.77	2.51
128	45.86	56.82	45.12	58.34	45.09	50.25	6.015
160	44.47	46.51	56.95	44.66	46.35	47.79	4.65

Sample code result when 5 trials are ran and batch size is 128.

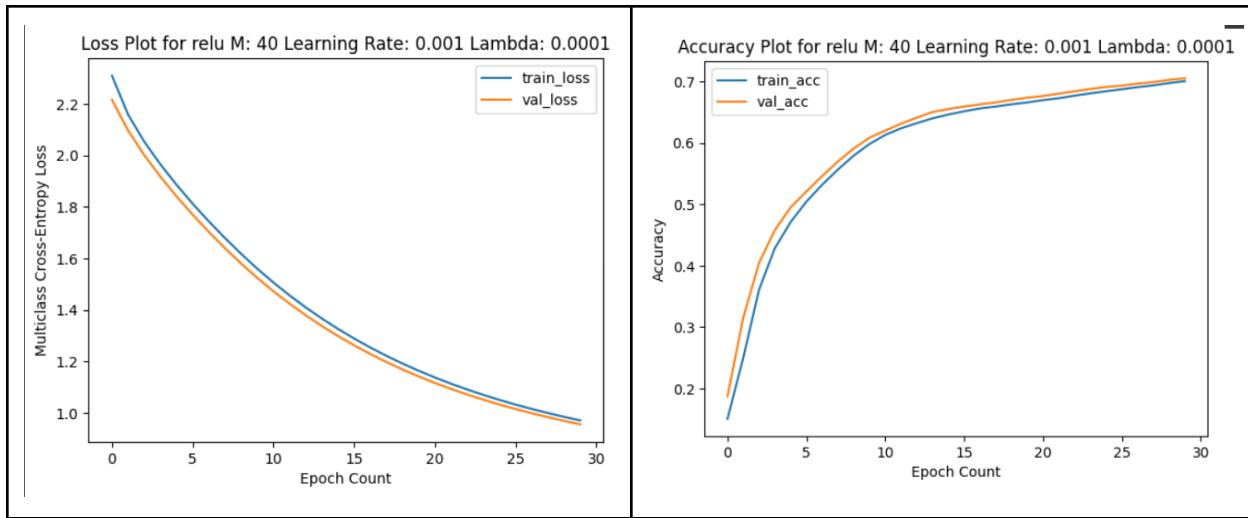
```
Batch Size: 128
Trail 0
375/375 [=====] - 11s 29ms/step - loss: 1.2139 - accuracy: 0.6376 - val_loss: 0.8363 - val_accuracy: 0.7386
375/375 [=====] - 13s 34ms/step - loss: 0.7560 - accuracy: 0.7586 - val_loss: 0.6880 - val_accuracy: 0.7770
375/375 [=====] - 11s 29ms/step - loss: 0.6539 - accuracy: 0.7880 - val_loss: 0.6218 - val_accuracy: 0.7950
375/375 [=====] - 11s 30ms/step - loss: 0.6007 - accuracy: 0.8030 - val_loss: 0.5836 - val_accuracy: 0.8033
Trail 1
375/375 [=====] - 11s 30ms/step - loss: 1.2199 - accuracy: 0.6201 - val_loss: 0.8500 - val_accuracy: 0.7201
375/375 [=====] - 11s 28ms/step - loss: 0.7720 - accuracy: 0.7424 - val_loss: 0.6987 - val_accuracy: 0.7669
375/375 [=====] - 12s 32ms/step - loss: 0.6692 - accuracy: 0.7769 - val_loss: 0.6352 - val_accuracy: 0.7860
375/375 [=====] - 12s 31ms/step - loss: 0.6138 - accuracy: 0.7956 - val_loss: 0.5926 - val_accuracy: 0.7994
375/375 [=====] - 11s 30ms/step - loss: 0.5779 - accuracy: 0.8080 - val_loss: 0.5647 - val_accuracy: 0.8080
Trail 2
375/375 [=====] - 11s 30ms/step - loss: 1.2376 - accuracy: 0.6199 - val_loss: 0.8406 - val_accuracy: 0.7320
375/375 [=====] - 11s 30ms/step - loss: 0.7582 - accuracy: 0.7513 - val_loss: 0.6872 - val_accuracy: 0.7742
375/375 [=====] - 11s 30ms/step - loss: 0.6543 - accuracy: 0.7848 - val_loss: 0.6208 - val_accuracy: 0.7952
375/375 [=====] - 11s 30ms/step - loss: 0.6004 - accuracy: 0.8021 - val_loss: 0.5823 - val_accuracy: 0.8062
Trail 3
375/375 [=====] - 12s 31ms/step - loss: 1.2962 - accuracy: 0.6260 - val_loss: 0.8748 - val_accuracy: 0.7197
375/375 [=====] - 12s 31ms/step - loss: 0.7811 - accuracy: 0.7497 - val_loss: 0.7052 - val_accuracy: 0.7691
375/375 [=====] - 12s 31ms/step - loss: 0.6678 - accuracy: 0.7848 - val_loss: 0.6317 - val_accuracy: 0.7893
375/375 [=====] - 11s 30ms/step - loss: 0.6107 - accuracy: 0.7995 - val_loss: 0.5909 - val_accuracy: 0.7987
375/375 [=====] - 12s 33ms/step - loss: 0.5743 - accuracy: 0.8098 - val_loss: 0.5620 - val_accuracy: 0.8096
Trail 4
375/375 [=====] - 11s 29ms/step - loss: 1.1719 - accuracy: 0.6506 - val_loss: 0.8320 - val_accuracy: 0.7415
375/375 [=====] - 12s 31ms/step - loss: 0.7564 - accuracy: 0.7596 - val_loss: 0.6855 - val_accuracy: 0.7772
375/375 [=====] - 11s 30ms/step - loss: 0.6581 - accuracy: 0.7869 - val_loss: 0.6227 - val_accuracy: 0.7924
375/375 [=====] - 11s 30ms/step - loss: 0.6058 - accuracy: 0.8007 - val_loss: 0.5867 - val_accuracy: 0.8017
```

The batch size with the smallest time is 160 with a mean of 47 seconds.

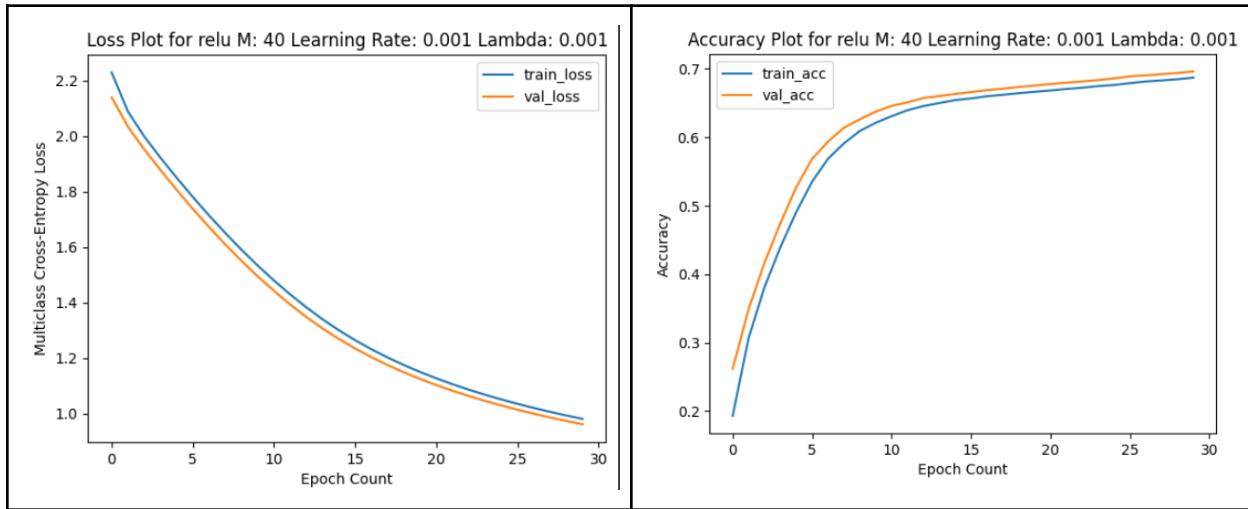
b) Perform a grid search over the following hyper-parameters:

Batch size for this is 160.

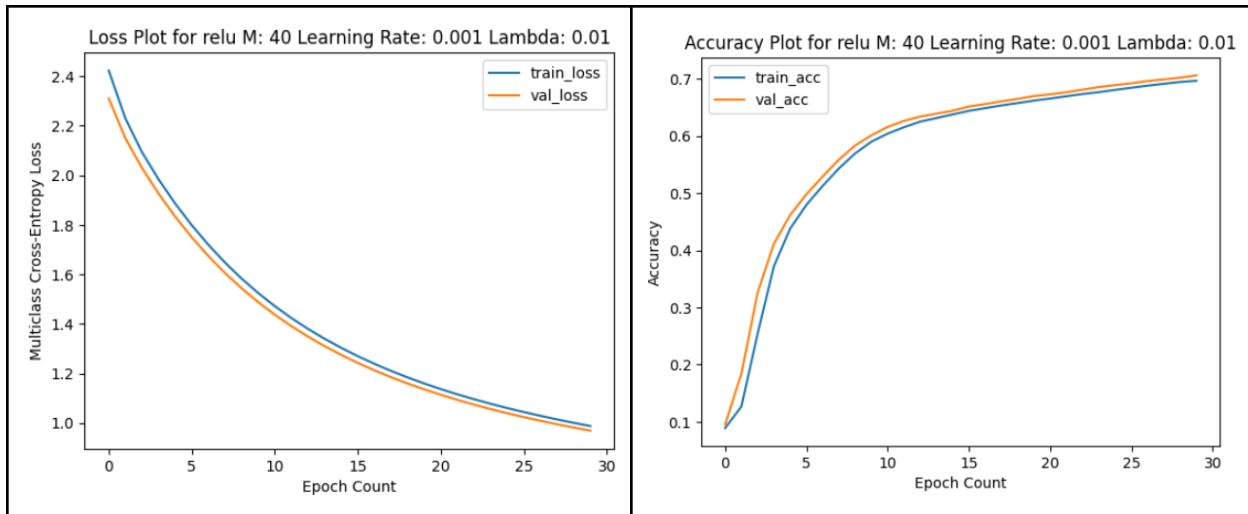
Creating and Traning Sequential model with M: 40 Learning Rate: 0.001 and Lambda: 0.0001



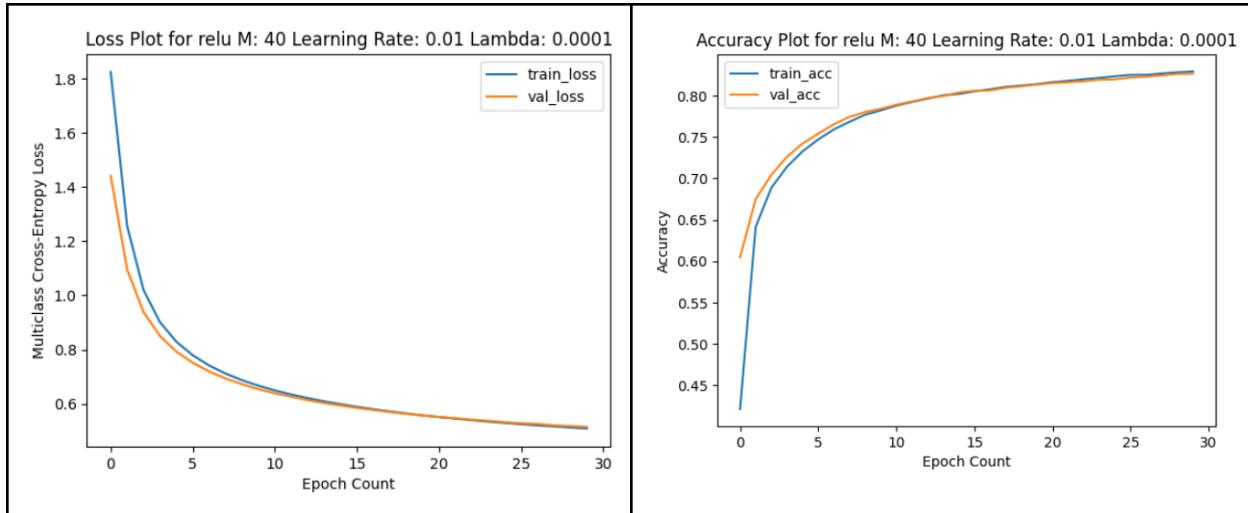
Creating and Traning Sequential model with M: 40 Learning Rate: 0.001 and Lambda: 0.001



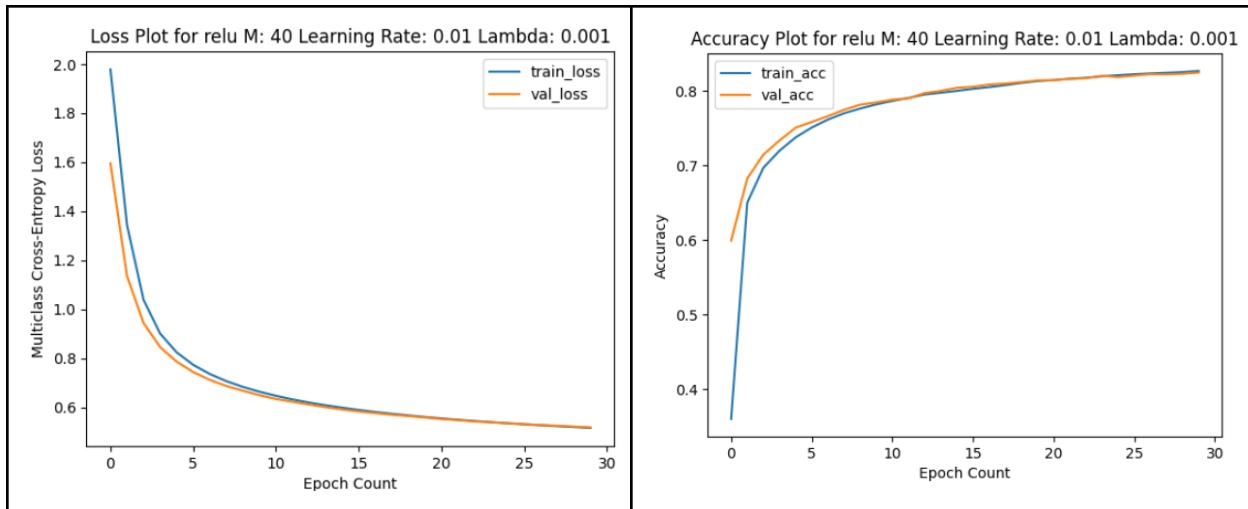
Creating and Traning Sequential model with M: 40 Learning Rate: 0.001 and Lambda: 0.01



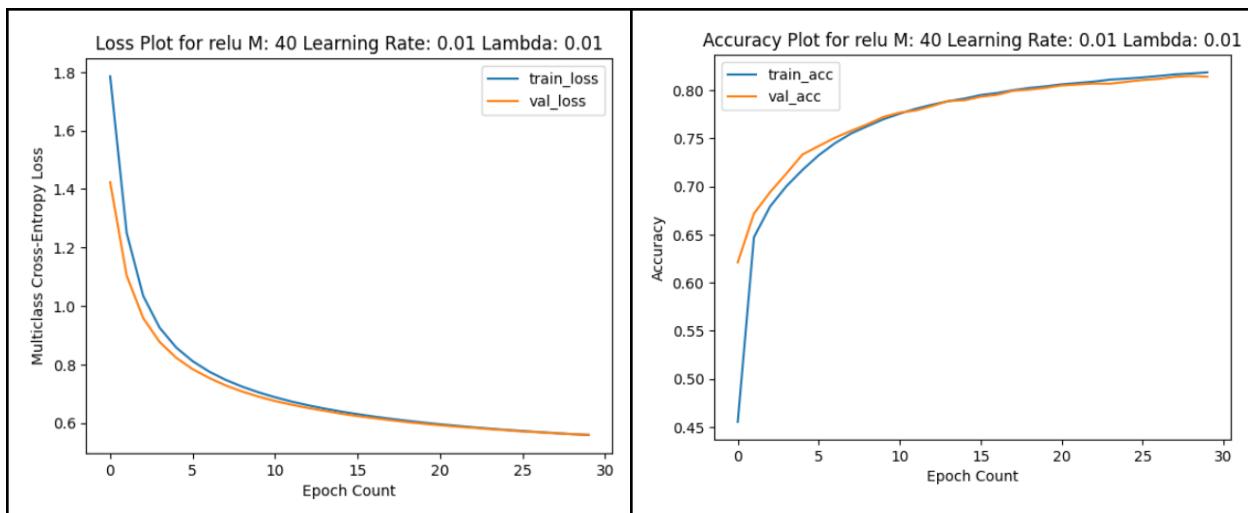
Creating and Traning Sequential model with M: 40 Learning Rate: 0.01 and Lambda: 0.0001



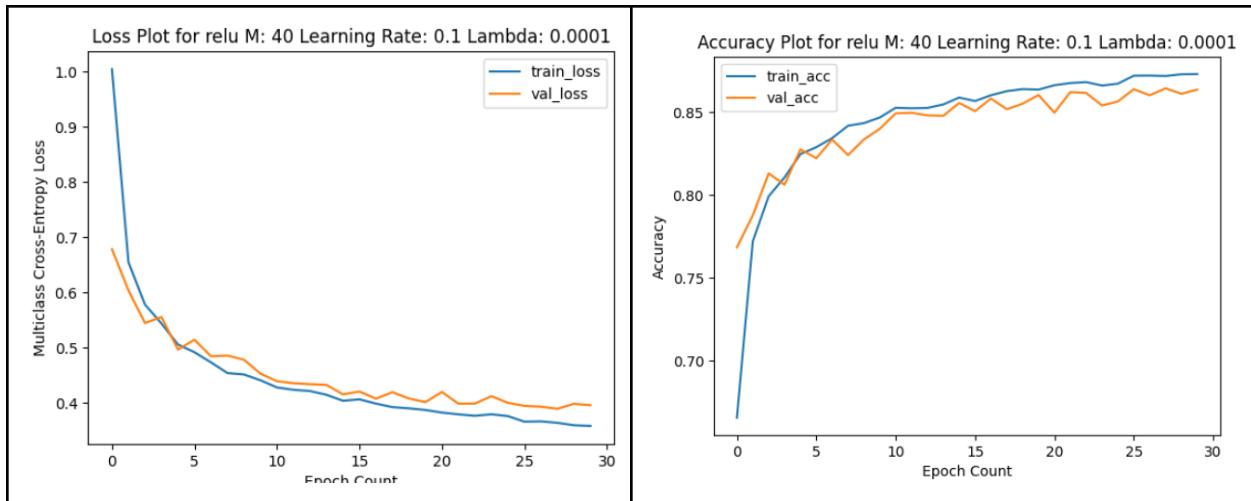
Creating and Traning Sequential model with M: 40 Learning Rate: 0.01 and Lambda: 0.001



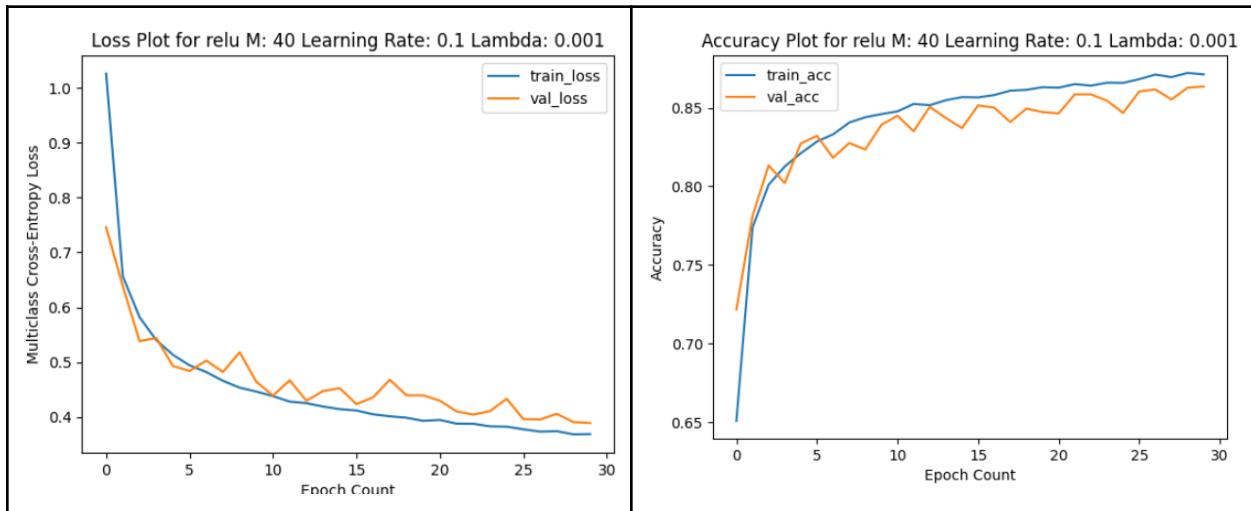
Creating and Traning Sequential model with M: 40 Learning Rate: 0.01 and Lambda: 0.01



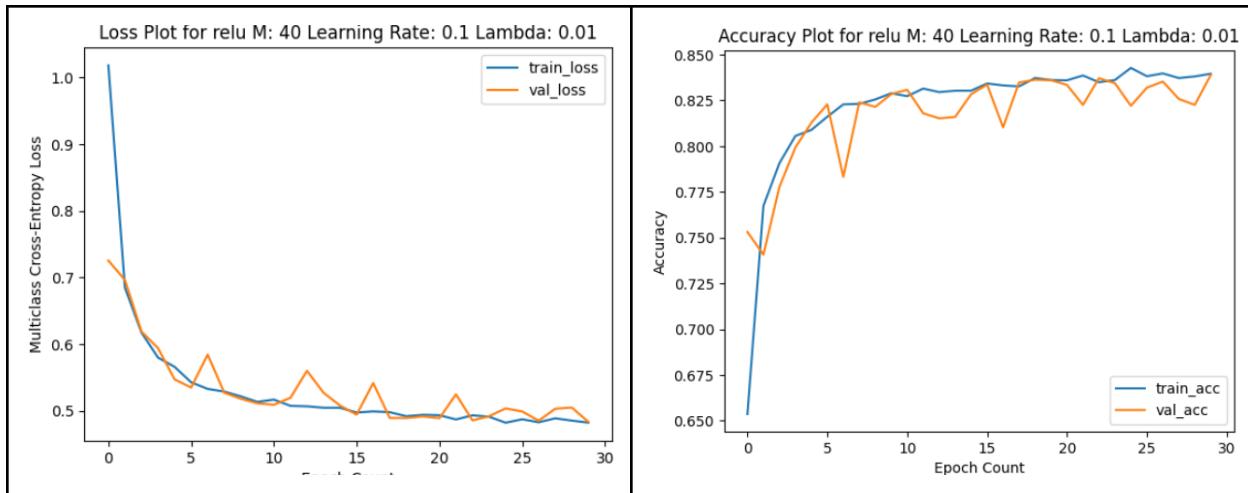
Creating and Traning Sequential model with M: 40 Learning Rate: 0.1 and Lambda: 0.0001



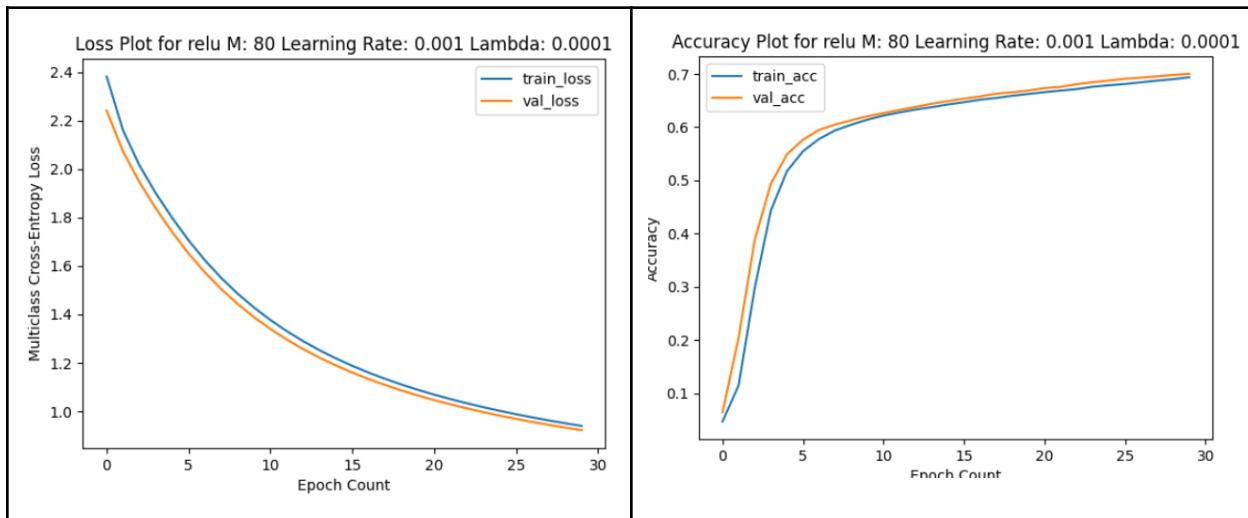
Creating and Traning Sequential model with M: 40 Learning Rate: 0.1 and Lambda: 0.001



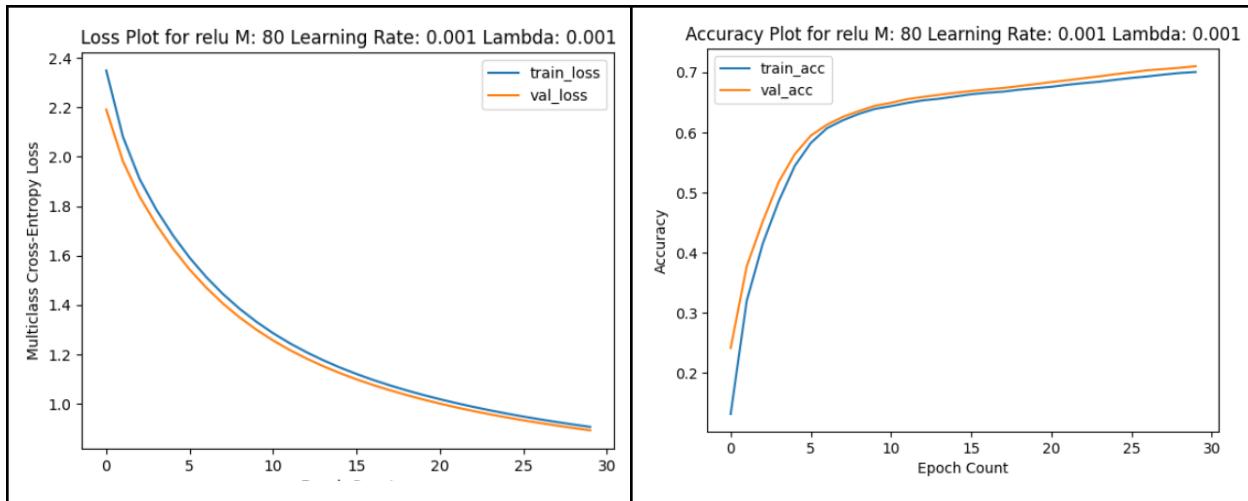
Creating and Traning Sequential model with M: 40 Learning Rate: 0.1 and Lambda: 0.01



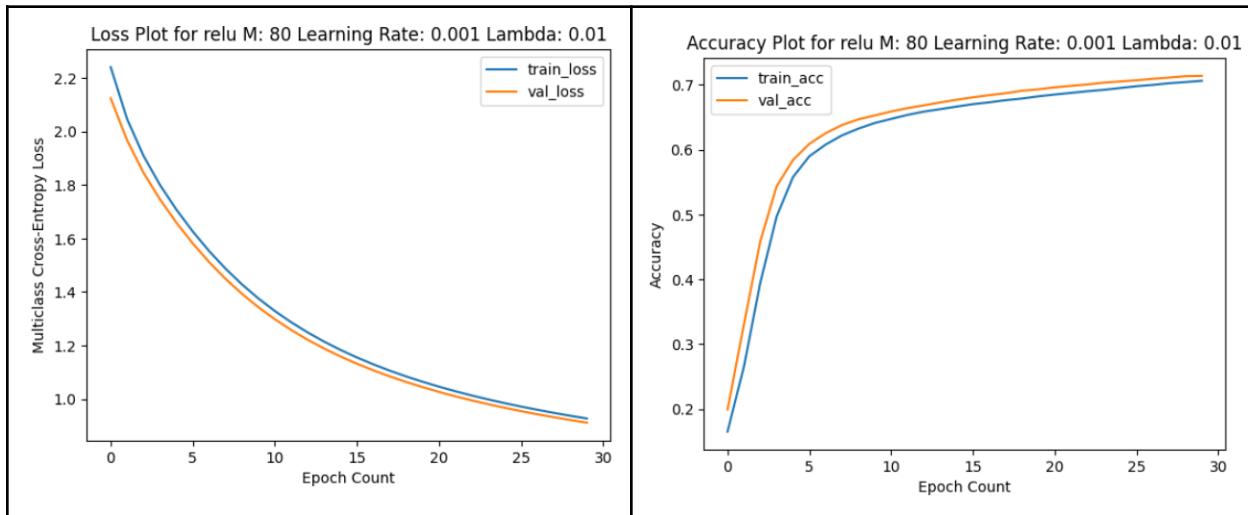
Creating and Traning Sequential model with M: 80 Learning Rate: 0.001 and Lambda: 0.0001



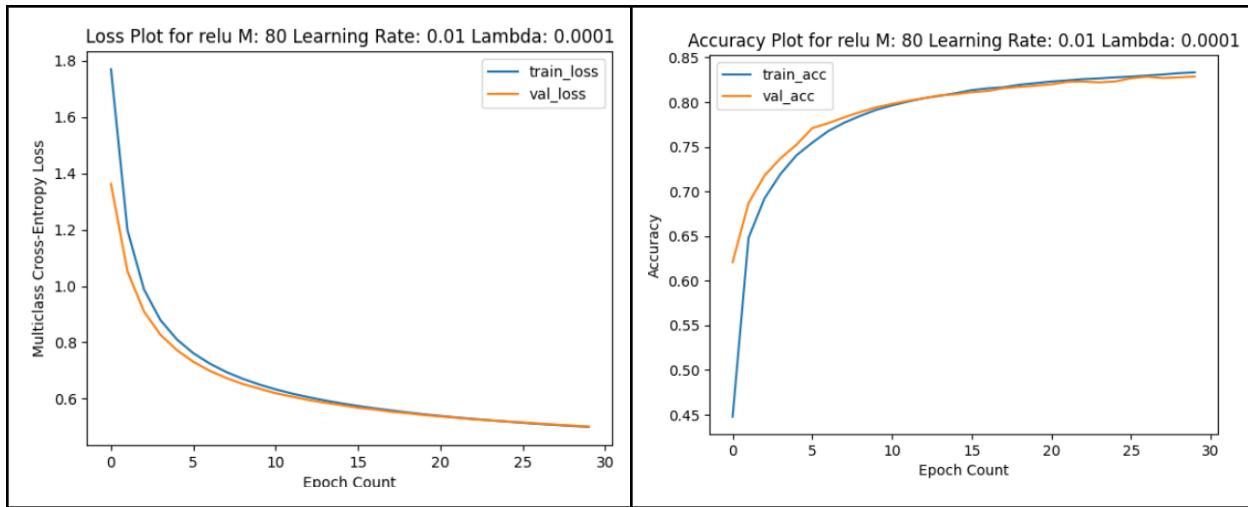
Creating and Traning Sequential model with M: 80 Learning Rate: 0.001 and Lambda: 0.001



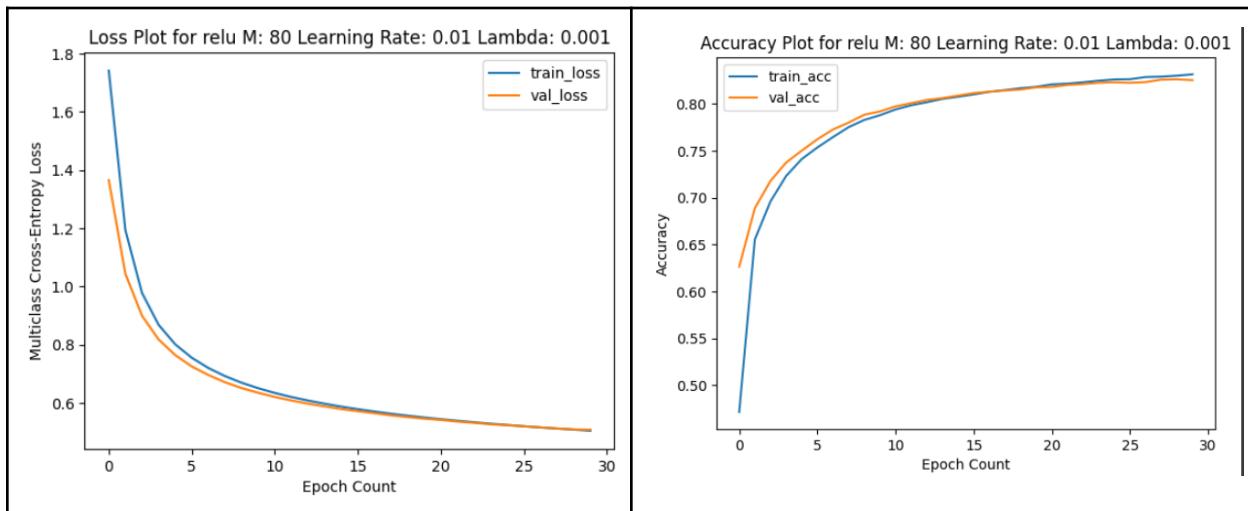
Creating and Traning Sequential model with M: 80 Learning Rate: 0.001 and Lambda: 0.01



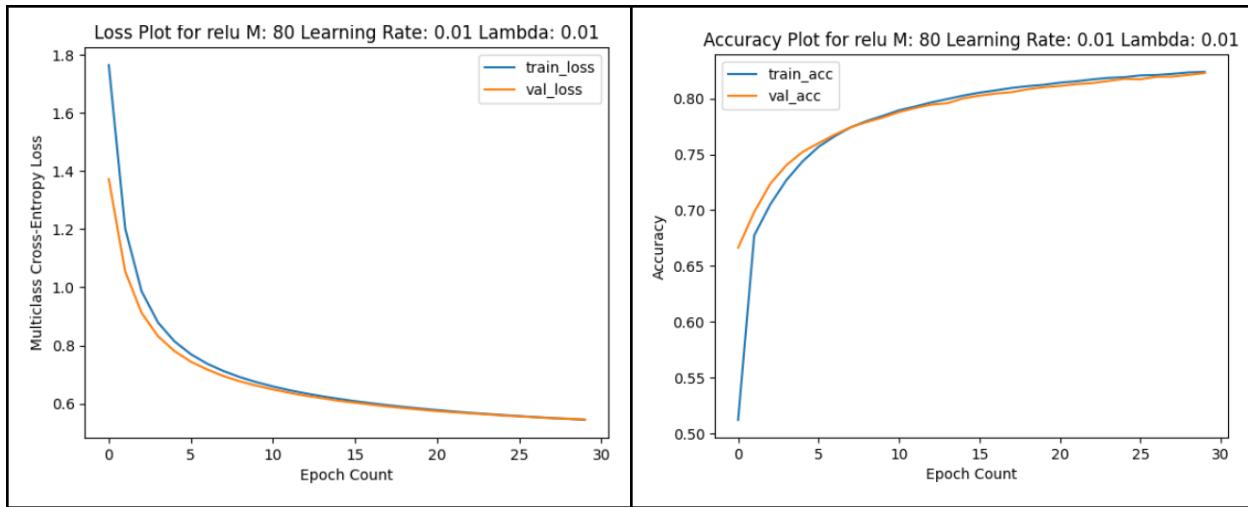
Creating and Traning Sequential model with M: 80 Learning Rate: 0.01 and Lambda: 0.0001



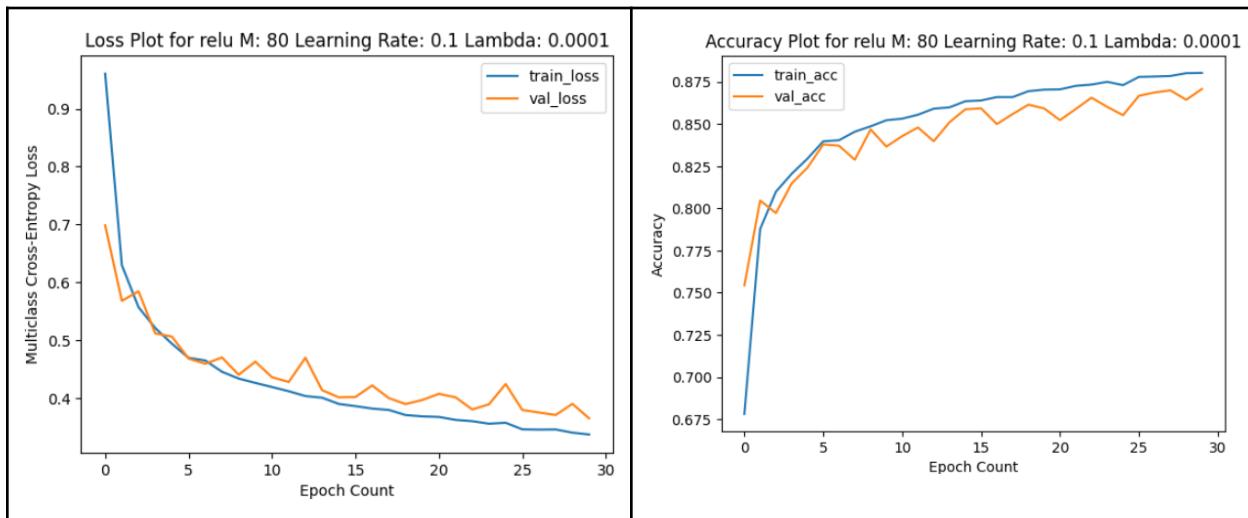
Creating and Traning Sequential model with M: 80 Learning Rate: 0.01 and Lambda: 0.001



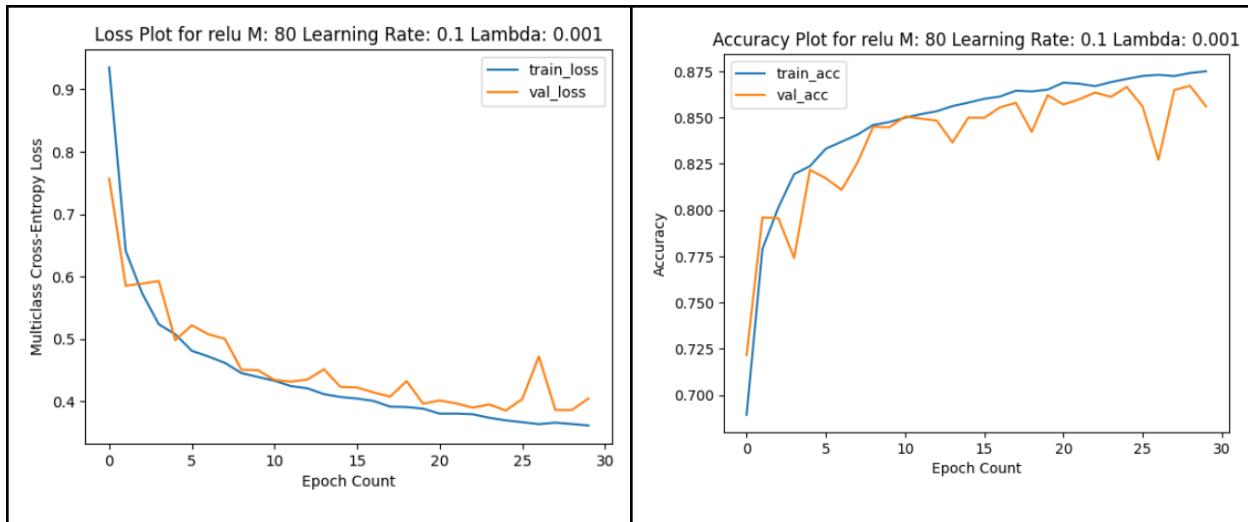
Creating and Traning Sequential model with M: 80 Learning Rate: 0.01 and Lambda: 0.01



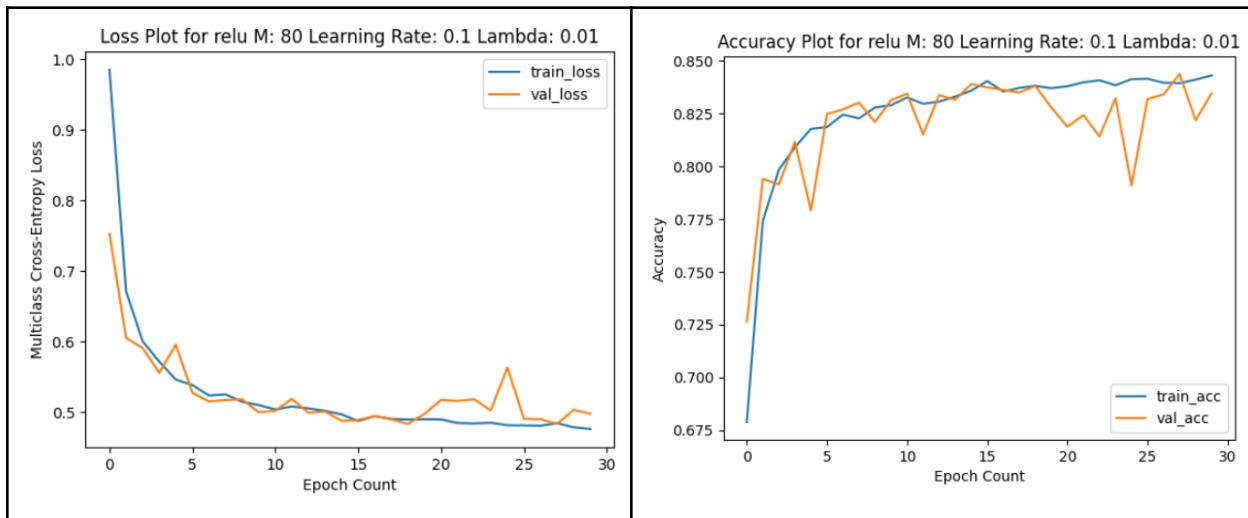
Creating and Traning Sequential model with M: 80 Learning Rate: 0.1 and Lambda: 0.0001



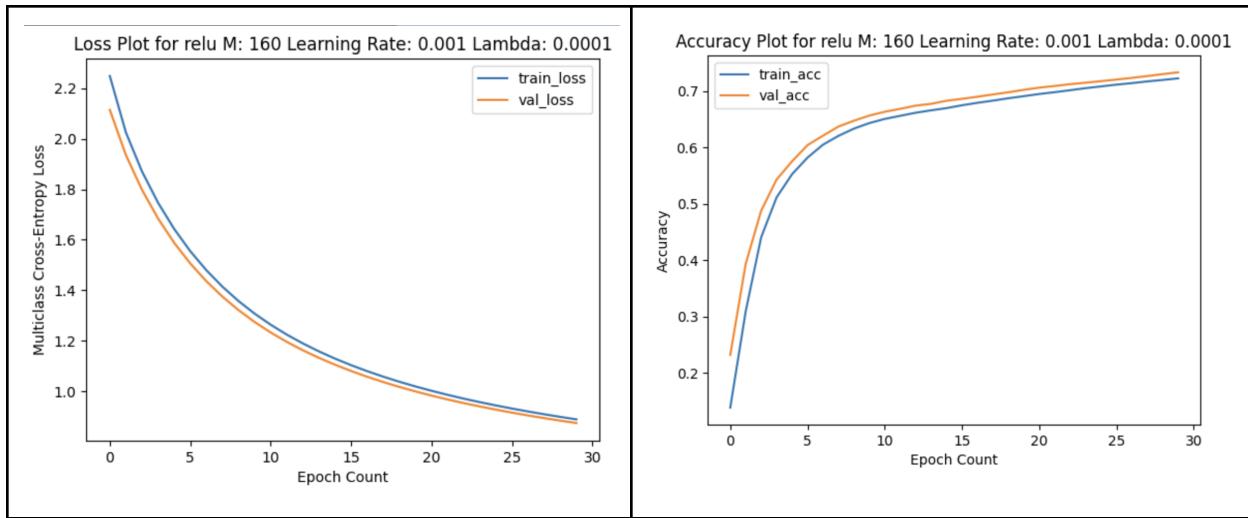
Creating and Traning Sequential model with M: 80 Learning Rate: 0.1 and Lambda: 0.001



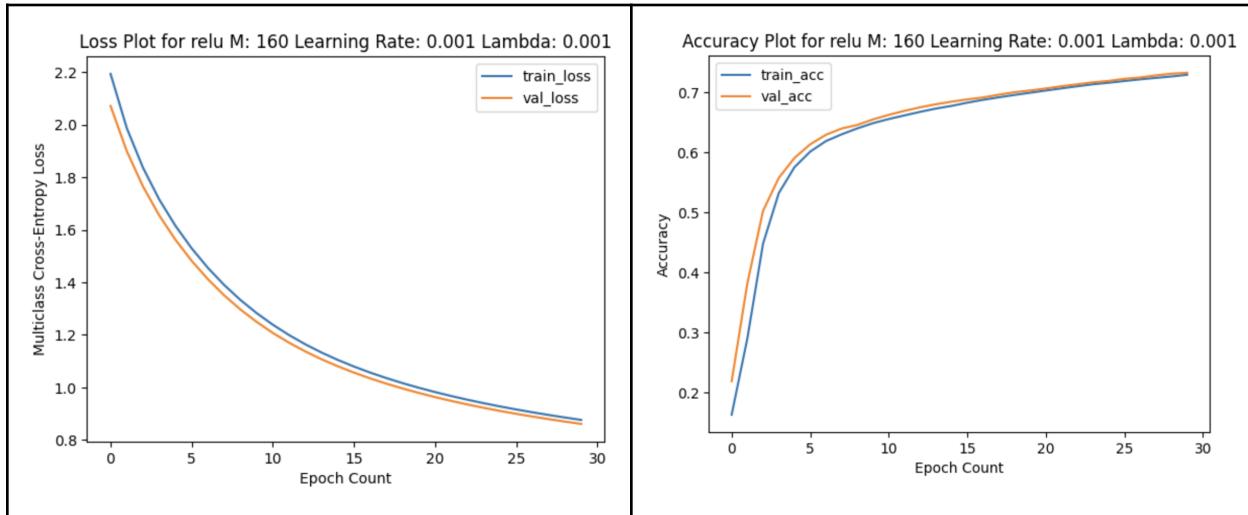
Creating and Traning Sequential model with M: 80 Learning Rate: 0.1 and Lambda: 0.01



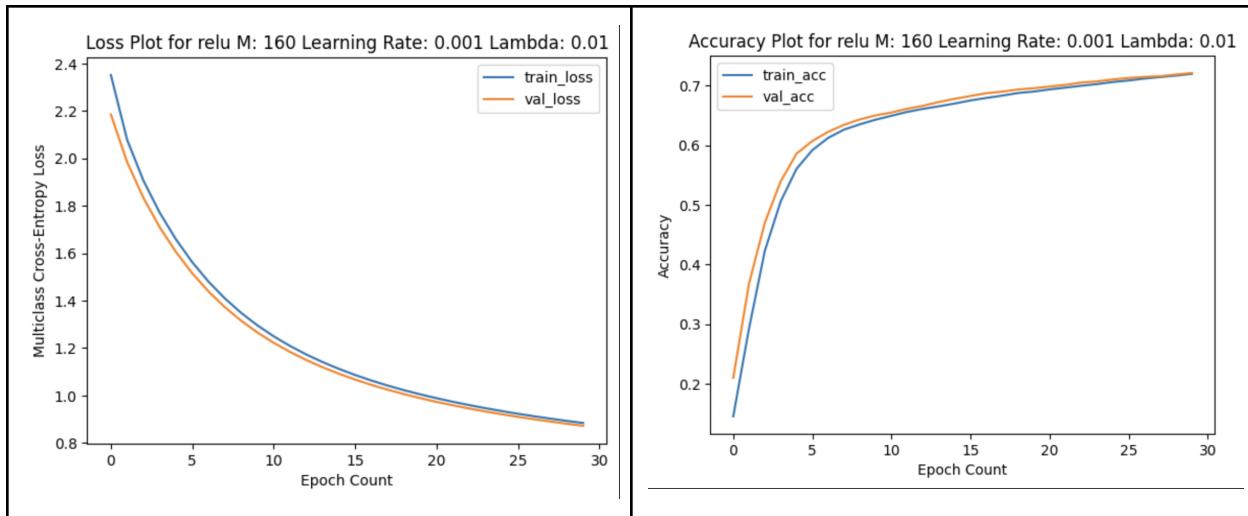
Creating and Traning Sequential model with M: 160 Learning Rate: 0.001 and Lambda: 0.0001



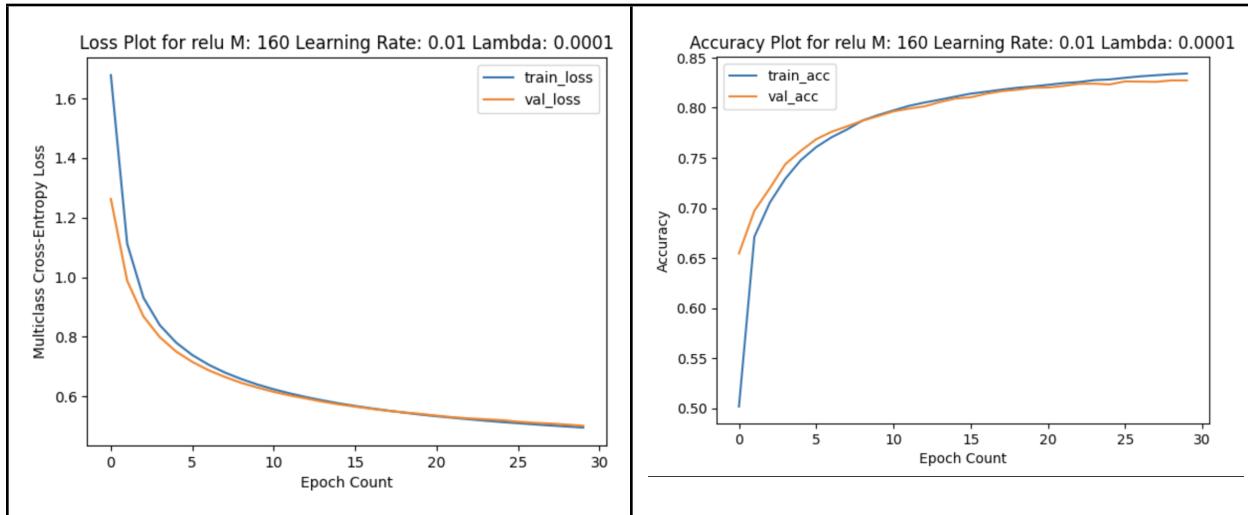
Creating and Traning Sequential model with M: 160 Learning Rate: 0.001 and Lambda: 0.001



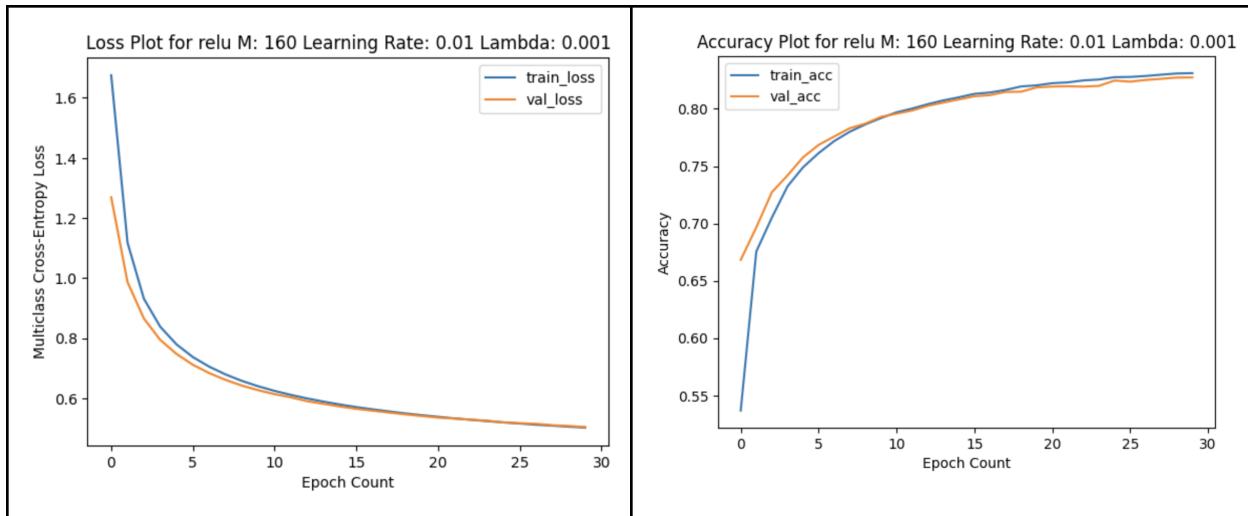
Creating and Traning Sequential model with M: 160 Learning Rate: 0.001 and Lambda: 0.01



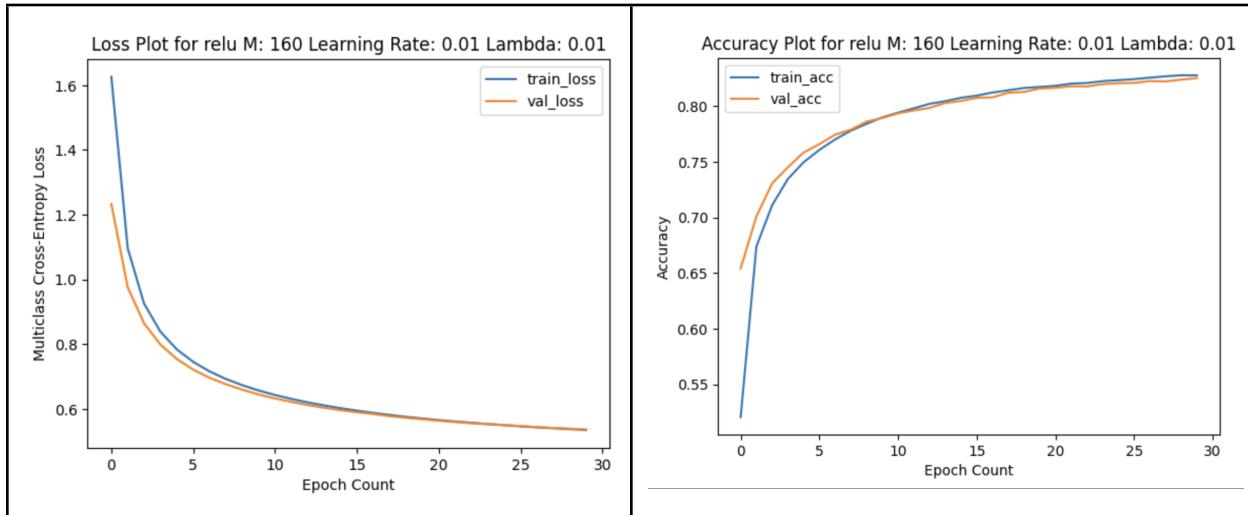
Creating and Traning Sequential model with M: 160 Learning Rate: 0.01 and Lambda: 0.0001



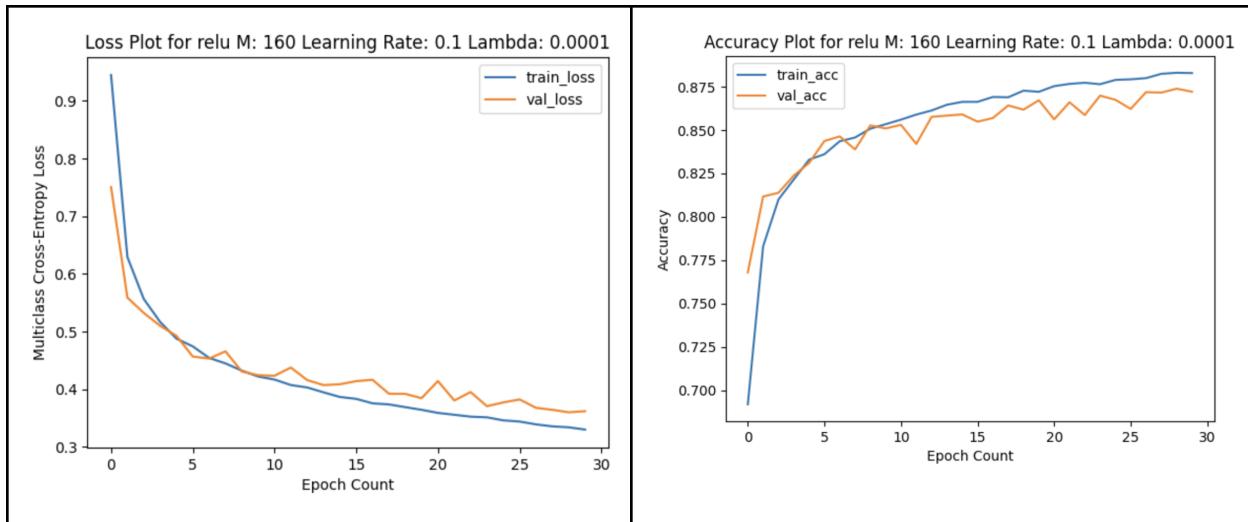
Creating and Traning Sequential model with M: 160 Learning Rate: 0.01 and Lambda: 0.001



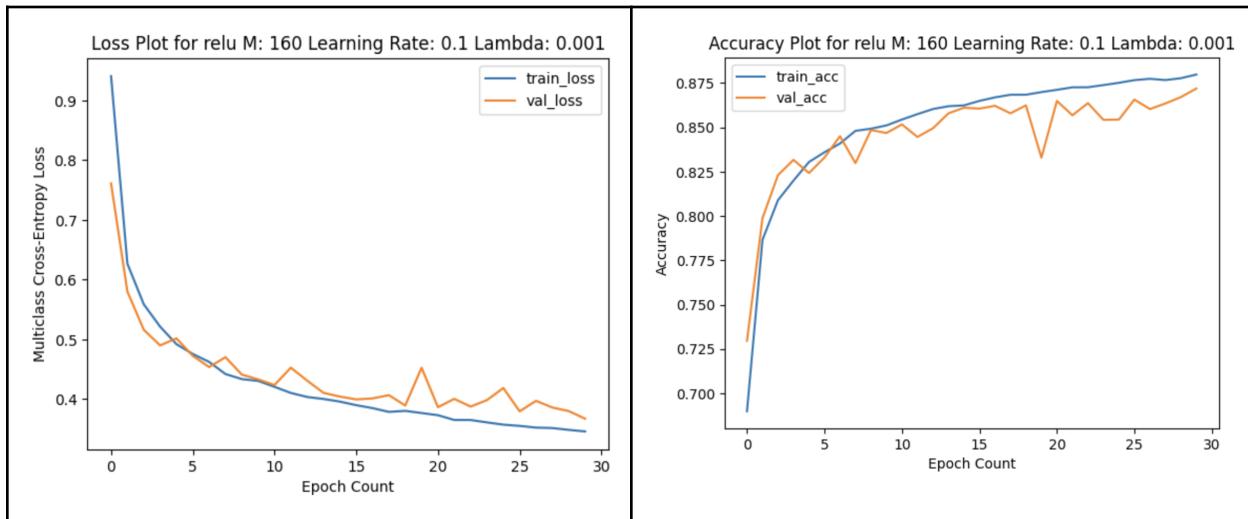
Creating and Traning Sequential model with M: 160 Learning Rate: 0.01 and Lambda: 0.01



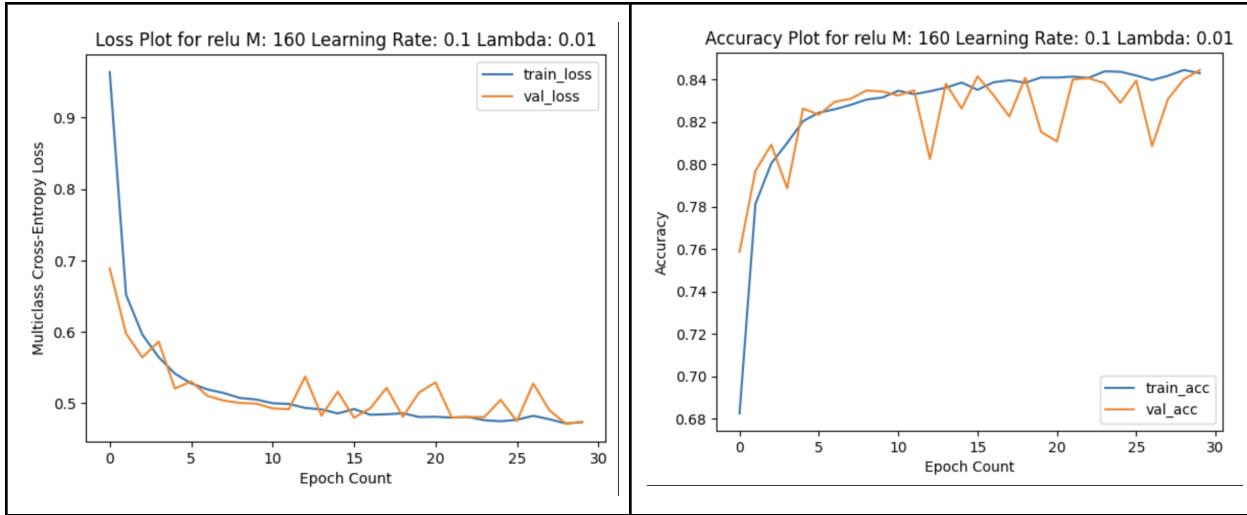
Creating and Traning Sequential model with M: 160 Learning Rate: 0.1 and Lambda: 0.0001



Creating and Traning Sequential model with M: 160 Learning Rate: 0.1 and Lambda: 0.001



Creating and Traning Sequential model with M: 160 Learning Rate: 0.1 and Lambda: 0.01



- c) For the best hyper-parameters found in part (b), run 5 training runs out to 100 epochs. Report the best accuracy (over epochs) on val for each run - this is 5 numbers. Compute, mean, max, and std deviation for these 5 values

d)

From best parameters, running model 5 times on 100 epochs

Mean: 0.895550012588501

Max: 0.8977500200271606

Standard Deviation: 0.001612972332337005

- e) Take best model from part c (highest val accuracy) and evaluate on test. Report the test accuracy. Report the number of trainable parameters and all hyper-parameters used to obtain this final best model

313/313 [=====] - 1s 2ms/step - loss: 0.3502 - accuracy: 0.8761
[0.3502149283885956, 0.8761000037193298]

Part 2:

- a) Calculate an approximate default choice γ

Sol.

$$\text{Alpha} = (4/M)^{1/2}$$

$\sigma =$