DL Assignment-1 Report

Task: Implement a neural network and utilize the CIFAR-10 dataset for the analysis

Objectives:

Question 1.

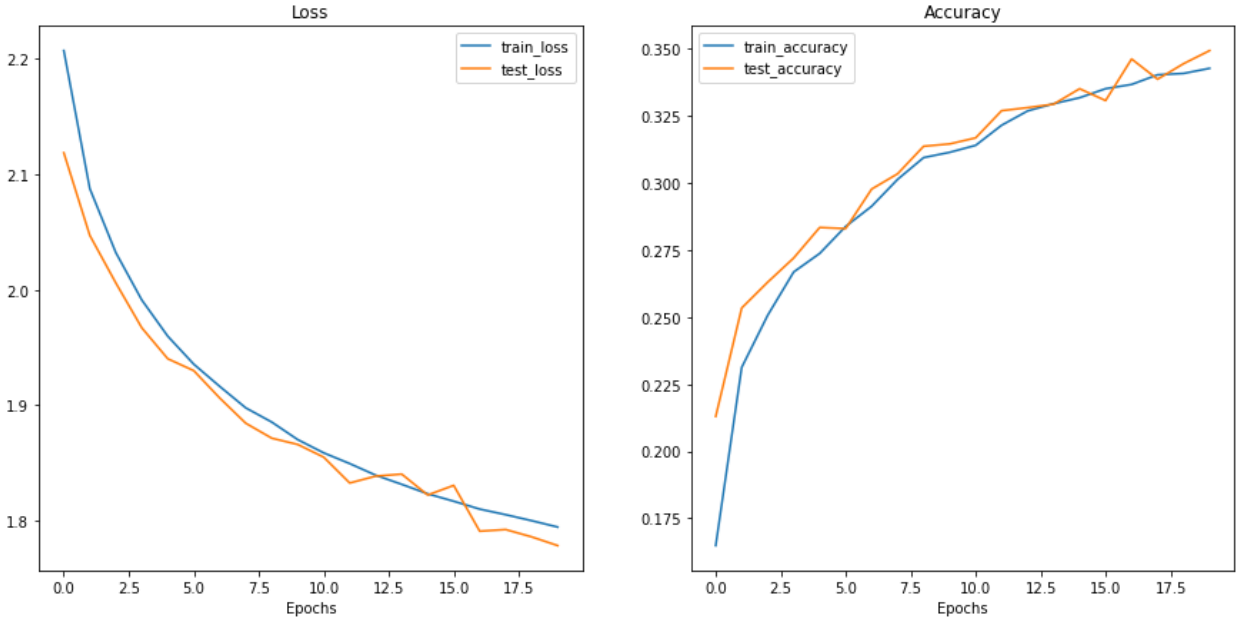
- Utilize various activation functions like sigmoid, tanh and critique the performance in each case.
- Increase the depth of the given network by adding more Fully-Connected layers until you encounter the vanishing gradient problem. With the help of the results, mention how to identify it.
- Suggest and implement methods to overcome the above problem.

Procedure:

- Import required packages. (PyTorch, NumPy, Matplotlib, torchmetrics …etc.)
- Making compose transform to convert image to grayscale and to tensors.
- Download the Data with torchvision.dataset.CIFAR10() function with required parameters and transform.
- Visualizing random image samples from the train data set.
- Creating the data loaders for train and test datasets respectively, with batch size 32.
- We are defining the PyTorch model class by inheriting nn.Module.
- Making model instance with appropriate arguments like numbers of layers and activation functions.
- Defining the training step, testing step & train function that will be used for training the neural network.
- We‘ve defined two models, one with sigmoid and the other with tanh for experimentation purposes.
- Training both models for 20 epochs, we get the results below.

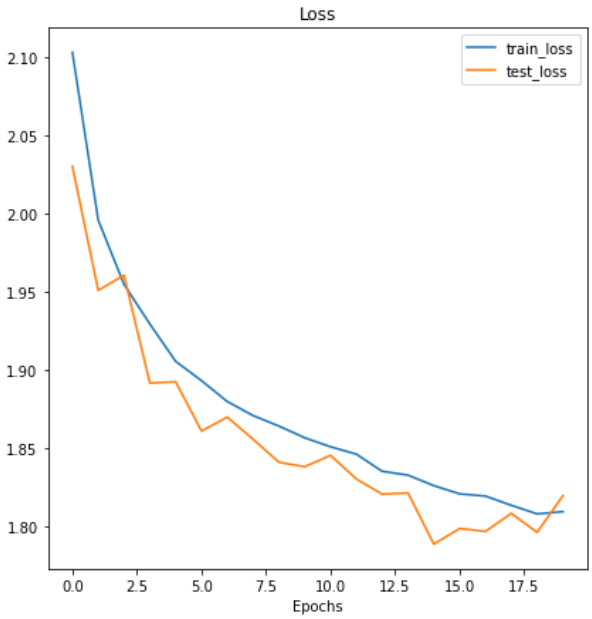
20/20 [06:53<00:00, 20.11s/it]
Epoch: 1 | train_loss: 2.2073 | train_acc: 0.1648 | test_loss: 2.1187 | test_acc: 0.2130
Epoch: 2 | train_loss: 2.0876 | train_acc: 0.2312 | test_loss: 2.0470 | test_acc: 0.2534
Epoch: 3 | train_loss: 2.0322 | train_acc: 0.2507 | test_loss: 2.0059 | test_acc: 0.2630
Epoch: 4 | train_loss: 1.9910 | train_acc: 0.2668 | test_loss: 1.9669 | test_acc: 0.2720
Epoch: 5 | train_loss: 1.9596 | train_acc: 0.2737 | test_loss: 1.9401 | test_acc: 0.2834
Epoch: 6 | train_loss: 1.9354 | train_acc: 0.2838 | test_loss: 1.9297 | test_acc: 0.2830
Epoch: 7 | train_loss: 1.9161 | train_acc: 0.2913 | test_loss: 1.9060 | test_acc: 0.2977
Epoch: 8 | train_loss: 1.8975 | train_acc: 0.3014 | test_loss: 1.8842 | test_acc: 0.3034
Epoch: 9 | train_loss: 1.8851 | train_acc: 0.3095 | test_loss: 1.8712 | test_acc: 0.3137
Epoch: 10 | train_loss: 1.8699 | train_acc: 0.3114 | test_loss: 1.8657 | test_acc: 0.3145
Epoch: 11 | train_loss: 1.8585 | train_acc: 0.3140 | test_loss: 1.8547 | test_acc: 0.3168
Epoch: 12 | train_loss: 1.8491 | train_acc: 0.3215 | test_loss: 1.8323 | test_acc: 0.3269
Epoch: 13 | train_loss: 1.8391 | train_acc: 0.3268 | test_loss: 1.8383 | test_acc: 0.3281
Epoch: 14 | train_loss: 1.8312 | train_acc: 0.3295 | test_loss: 1.8400 | test_acc: 0.3294

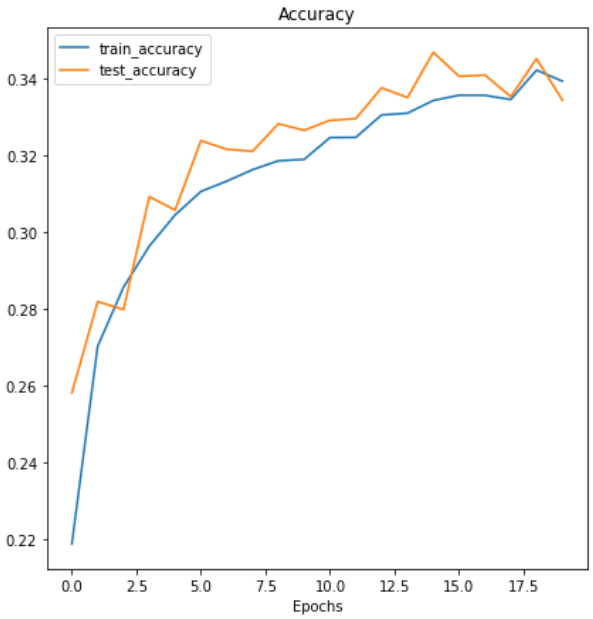
Epoch: 15 | train_loss: 1.8229 | train_acc: 0.3318 | test_loss: 1.8219 | test_acc: 0.3351
Epoch: 16 | train_loss: 1.8165 | train_acc: 0.3351 | test_loss: 1.8302 | test_acc: 0.3307
Epoch: 17 | train_loss: 1.8097 | train_acc: 0.3367 | test_loss: 1.7905 | test_acc: 0.3462
Epoch: 18 | train_loss: 1.8048 | train_acc: 0.3403 | test_loss: 1.7920 | test_acc: 0.3387
Epoch: 19 | train_loss: 1.7996 | train_acc: 0.3408 | test_loss: 1.7856 | test_acc: 0.3445
Epoch: 20 | train_loss: 1.7941 | train_acc: 0.3427 | test_loss: 1.7780 | test_acc: 0.3493
total training time: 413.382 sec.



20/20 [06:58<00:00, 20.23s/it]
Epoch: 1 | train_loss: 2.1030 | train_acc: 0.2187 | test_loss: 2.0301 | test_acc: 0.2581
Epoch: 2 | train_loss: 1.9959 | train_acc: 0.2703 | test_loss: 1.9509 | test_acc: 0.2819
Epoch: 3 | train_loss: 1.9547 | train_acc: 0.2857 | test_loss: 1.9605 | test_acc: 0.2799
Epoch: 4 | train_loss: 1.9294 | train_acc: 0.2964 | test_loss: 1.8917 | test_acc: 0.3093
Epoch: 5 | train_loss: 1.9056 | train_acc: 0.3045 | test_loss: 1.8925 | test_acc: 0.3058
Epoch: 6 | train_loss: 1.8932 | train_acc: 0.3107 | test_loss: 1.8611 | test_acc: 0.3239
Epoch: 7 | train_loss: 1.8799 | train_acc: 0.3133 | test_loss: 1.8699 | test_acc: 0.3216
Epoch: 8 | train_loss: 1.8709 | train_acc: 0.3163 | test_loss: 1.8558 | test_acc: 0.3211
Epoch: 9 | train_loss: 1.8642 | train_acc: 0.3186 | test_loss: 1.8411 | test_acc: 0.3283
Epoch: 10 | train_loss: 1.8568 | train_acc: 0.3190 | test_loss: 1.8383 | test_acc: 0.3266
Epoch: 11 | train_loss: 1.8510 | train_acc: 0.3247 | test_loss: 1.8456 | test_acc: 0.3292
Epoch: 12 | train_loss: 1.8463 | train_acc: 0.3248 | test_loss: 1.8304 | test_acc: 0.3296
Epoch: 13 | train_loss: 1.8353 | train_acc: 0.3306 | test_loss: 1.8208 | test_acc: 0.3377
Epoch: 14 | train_loss: 1.8329 | train_acc: 0.3310 | test_loss: 1.8215 | test_acc: 0.3351
Epoch: 15 | train_loss: 1.8262 | train_acc: 0.3344 | test_loss: 1.7889 | test_acc: 0.3469
Epoch: 16 | train_loss: 1.8209 | train_acc: 0.3357 | test_loss: 1.7988 | test_acc: 0.3406
Epoch: 17 | train_loss: 1.8195 | train_acc: 0.3357 | test_loss: 1.7970 | test_acc: 0.3410
Epoch: 18 | train_loss: 1.8136 | train_acc: 0.3346 | test_loss: 1.8085 | test_acc: 0.3354
Epoch: 19 | train_loss: 1.8081 | train_acc: 0.3423 | test_loss: 1.7964 | test_acc: 0.3453
Epoch: 20 | train_loss: 1.8096 | train_acc: 0.3394 | test_loss: 1.8197 | test_acc: 0.3345

total training time: 443.698 sec.

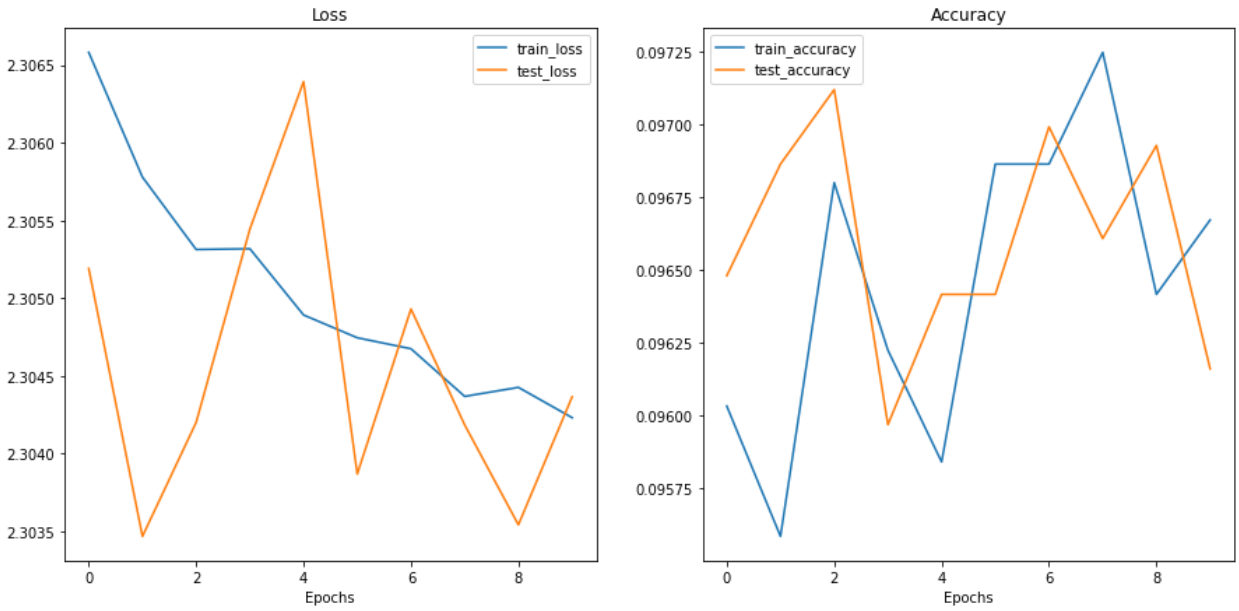




- Next, we start with Exploring the Vanishing Gradient problem.
- For this, we set up experiments as follows
    - Making models with layers between 10 to 20 e.g. 10,12,14... Layers, for both sigmoid and tanh.
    - We are doing this because we want to see up till with layer count the activation function is capable of resisting the problem of vanishing gradient problem.
    - The results of the experiments are shown below.
- Results For Sigmoid

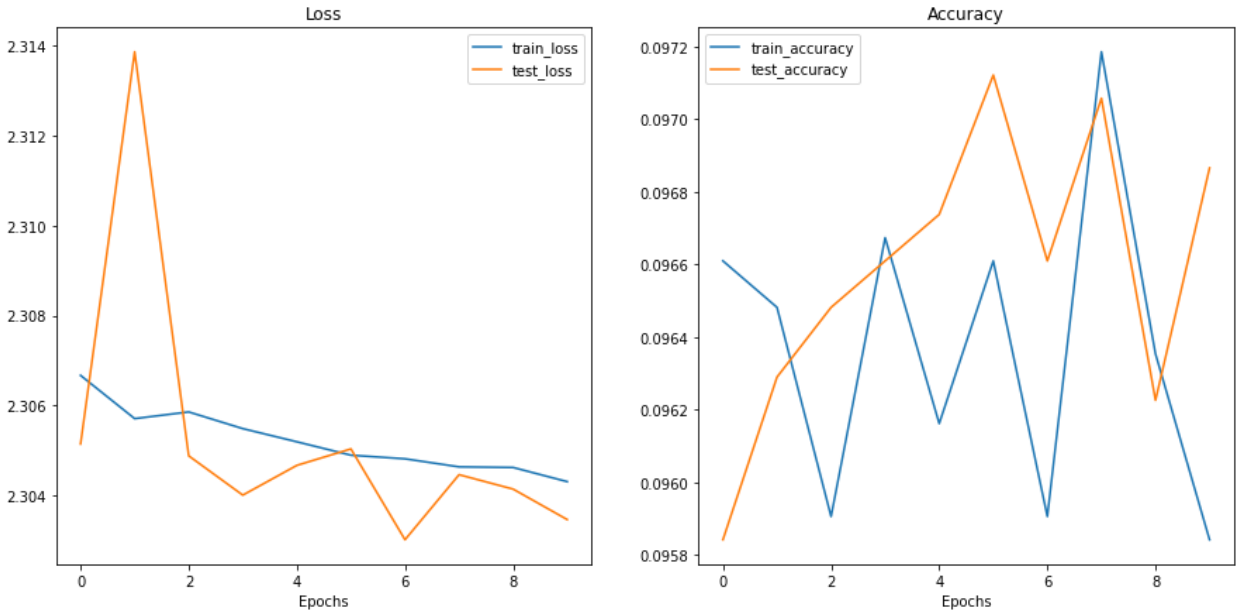
[INFO] training model_sigmode_10_layers

100%
10/10 [03:55<00:00, 22.30s/it]
Epoch: 1 | train_loss: 2.3066 | train_acc: 0.0960 | test_loss: 2.3052 | test_acc: 0.0965
Epoch: 2 | train_loss: 2.3058 | train_acc: 0.0956 | test_loss: 2.3035 | test_acc: 0.0969
Epoch: 3 | train_loss: 2.3053 | train_acc: 0.0968 | test_loss: 2.3042 | test_acc: 0.0971
Epoch: 4 | train_loss: 2.3053 | train_acc: 0.0962 | test_loss: 2.3054 | test_acc: 0.0960
Epoch: 5 | train_loss: 2.3049 | train_acc: 0.0958 | test_loss: 2.3064 | test_acc: 0.0964
Epoch: 6 | train_loss: 2.3047 | train_acc: 0.0969 | test_loss: 2.3039 | test_acc: 0.0964
Epoch: 7 | train_loss: 2.3047 | train_acc: 0.0969 | test_loss: 2.3049 | test_acc: 0.0970
Epoch: 8 | train_loss: 2.3044 | train_acc: 0.0972 | test_loss: 2.3042 | test_acc: 0.0966
Epoch: 9 | train_loss: 2.3044 | train_acc: 0.0964 | test_loss: 2.3035 | test_acc: 0.0969
Epoch: 10 | train_loss: 2.3042 | train_acc: 0.0967 | test_loss: 2.3044 | test_acc: 0.0962
total training time: 235.832 sec.



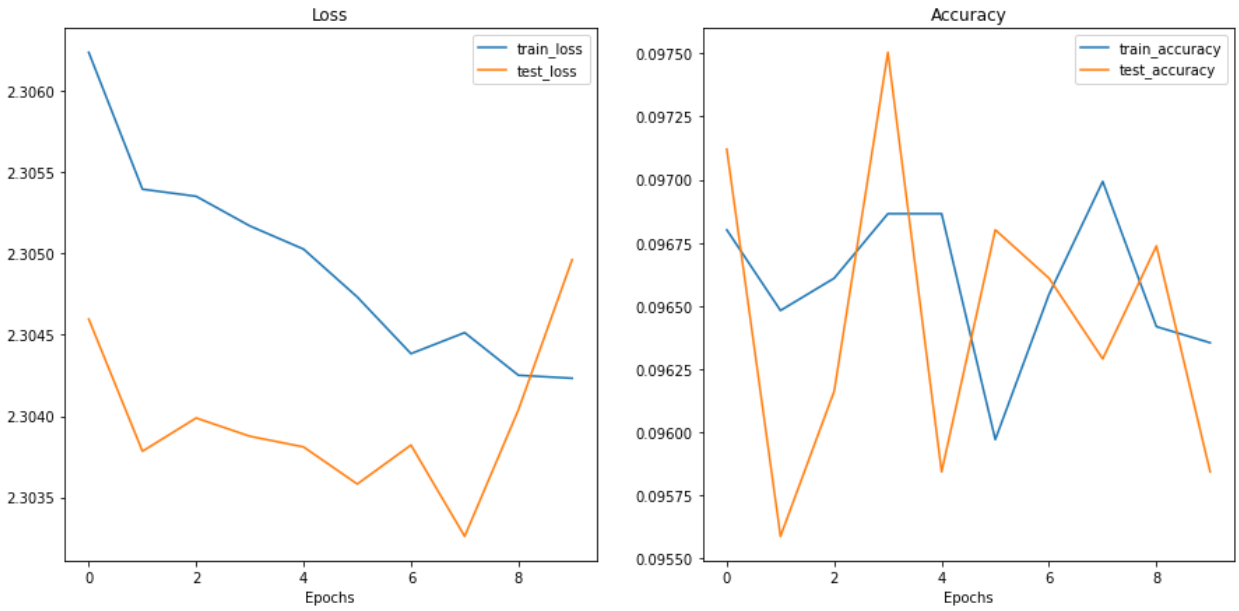
[INFO] training model_sigmode_12_layers

100%
10/10 [03:49<00:00, 22.84s/it]
Epoch: 1 | train_loss: 2.3067 | train_acc: 0.0966 | test_loss: 2.3051 | test_acc: 0.0958
Epoch: 2 | train_loss: 2.3057 | train_acc: 0.0965 | test_loss: 2.3139 | test_acc: 0.0963
Epoch: 3 | train_loss: 2.3059 | train_acc: 0.0959 | test_loss: 2.3049 | test_acc: 0.0965
Epoch: 4 | train_loss: 2.3055 | train_acc: 0.0967 | test_loss: 2.3040 | test_acc: 0.0966
Epoch: 5 | train_loss: 2.3052 | train_acc: 0.0962 | test_loss: 2.3047 | test_acc: 0.0967
Epoch: 6 | train_loss: 2.3049 | train_acc: 0.0966 | test_loss: 2.3050 | test_acc: 0.0971
Epoch: 7 | train_loss: 2.3048 | train_acc: 0.0959 | test_loss: 2.3030 | test_acc: 0.0966
Epoch: 8 | train_loss: 2.3046 | train_acc: 0.0972 | test_loss: 2.3045 | test_acc: 0.0971
Epoch: 9 | train_loss: 2.3046 | train_acc: 0.0964 | test_loss: 2.3041 | test_acc: 0.0962
Epoch: 10 | train_loss: 2.3043 | train_acc: 0.0958 | test_loss: 2.3035 | test_acc: 0.0969
total training time: 229.798 sec.



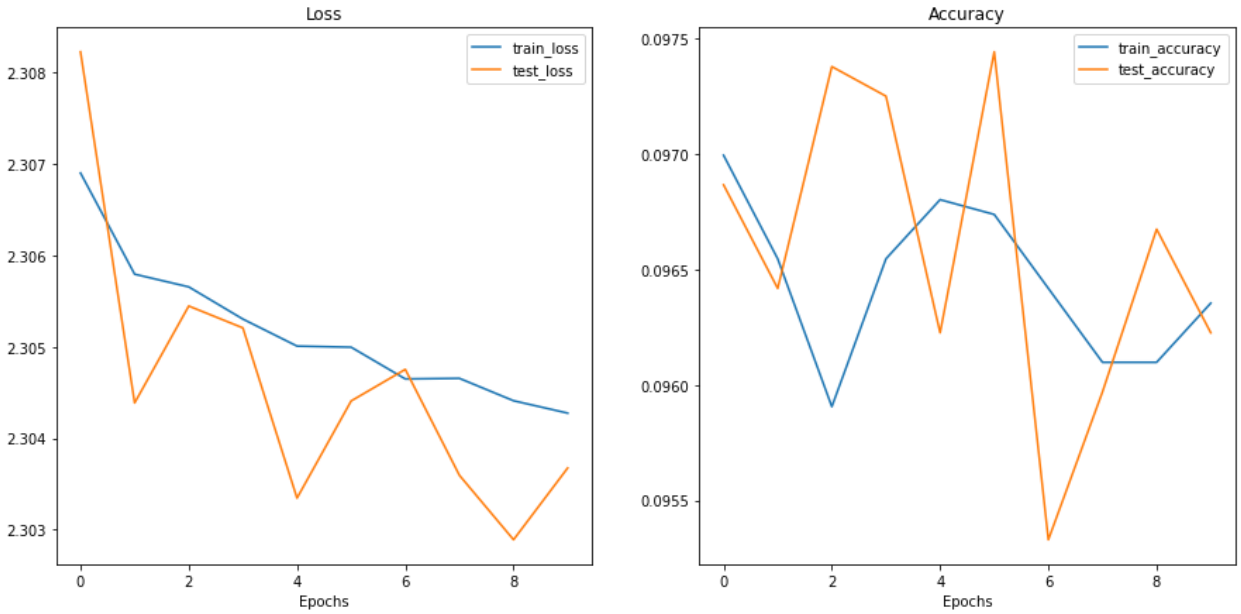
[INFO] training model_sigmode_14_layers

100%
10/10 [03:53<00:00, 23.32s/it]
Epoch: 1 | train_loss: 2.3062 | train_acc: 0.0968 | test_loss: 2.3046 | test_acc: 0.0971
Epoch: 2 | train_loss: 2.3054 | train_acc: 0.0965 | test_loss: 2.3038 | test_acc: 0.0956
Epoch: 3 | train_loss: 2.3054 | train_acc: 0.0966 | test_loss: 2.3040 | test_acc: 0.0962
Epoch: 4 | train_loss: 2.3052 | train_acc: 0.0969 | test_loss: 2.3039 | test_acc: 0.0975
Epoch: 5 | train_loss: 2.3050 | train_acc: 0.0969 | test_loss: 2.3038 | test_acc: 0.0958
Epoch: 6 | train_loss: 2.3047 | train_acc: 0.0960 | test_loss: 2.3036 | test_acc: 0.0968
Epoch: 7 | train_loss: 2.3044 | train_acc: 0.0965 | test_loss: 2.3038 | test_acc: 0.0966
Epoch: 8 | train_loss: 2.3045 | train_acc: 0.0970 | test_loss: 2.3033 | test_acc: 0.0963
Epoch: 9 | train_loss: 2.3043 | train_acc: 0.0964 | test_loss: 2.3040 | test_acc: 0.0967
Epoch: 10 | train_loss: 2.3042 | train_acc: 0.0964 | test_loss: 2.3050 | test_acc: 0.0958
total training time: 233.919 sec.



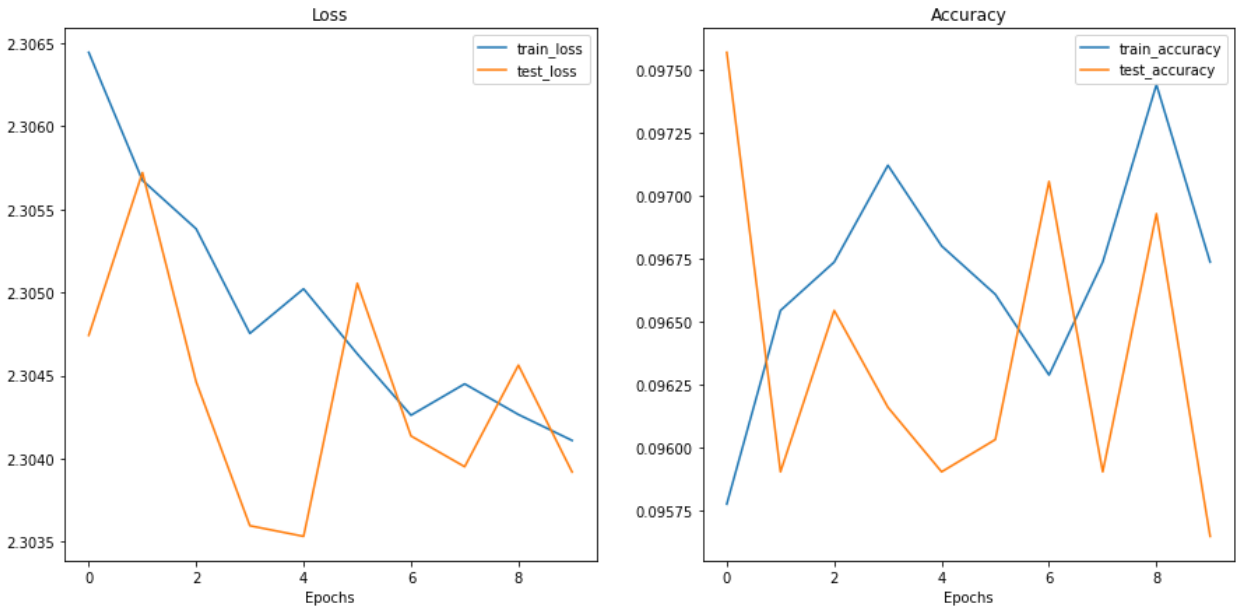
[INFO] training model_sigmode_16_layers

100%
10/10 [03:58<00:00, 23.83s/it]
Epoch: 1 | train_loss: 2.3069 | train_acc: 0.0970 | test_loss: 2.3082 | test_acc: 0.0969
Epoch: 2 | train_loss: 2.3058 | train_acc: 0.0965 | test_loss: 2.3044 | test_acc: 0.0964
Epoch: 3 | train_loss: 2.3057 | train_acc: 0.0959 | test_loss: 2.3054 | test_acc: 0.0974
Epoch: 4 | train_loss: 2.3053 | train_acc: 0.0965 | test_loss: 2.3052 | test_acc: 0.0972
Epoch: 5 | train_loss: 2.3050 | train_acc: 0.0968 | test_loss: 2.3033 | test_acc: 0.0962
Epoch: 6 | train_loss: 2.3050 | train_acc: 0.0967 | test_loss: 2.3044 | test_acc: 0.0974
Epoch: 7 | train_loss: 2.3046 | train_acc: 0.0964 | test_loss: 2.3048 | test_acc: 0.0953
Epoch: 8 | train_loss: 2.3047 | train_acc: 0.0961 | test_loss: 2.3036 | test_acc: 0.0960
Epoch: 9 | train_loss: 2.3044 | train_acc: 0.0961 | test_loss: 2.3029 | test_acc: 0.0967
Epoch: 10 | train_loss: 2.3043 | train_acc: 0.0964 | test_loss: 2.3037 | test_acc: 0.0962
total training time: 238.390 sec.



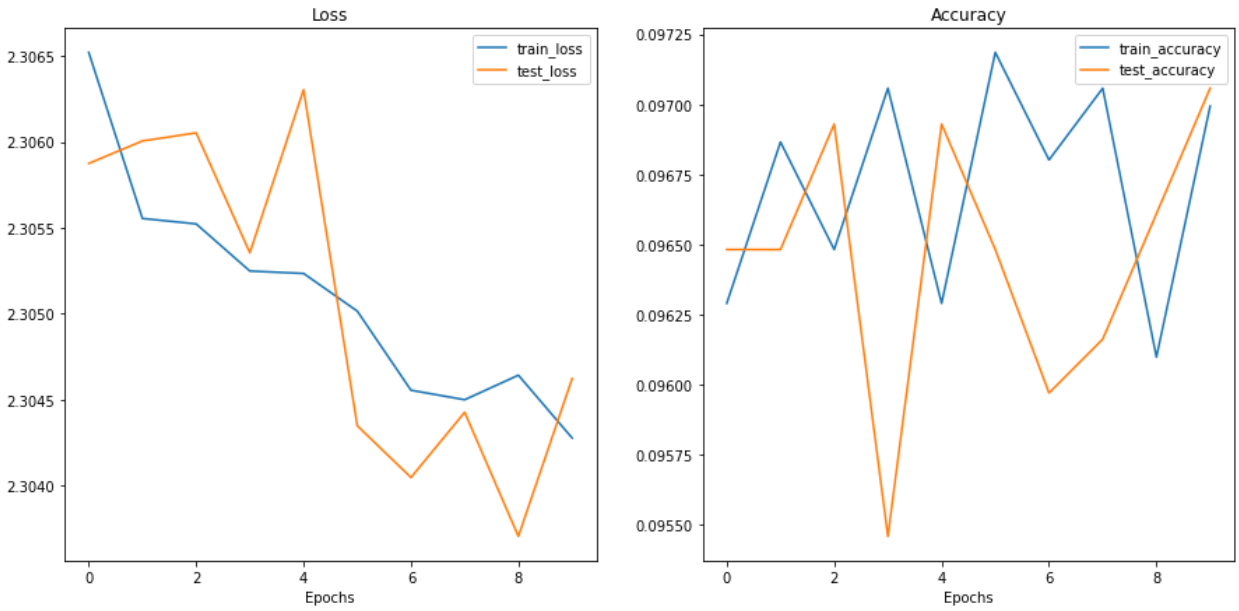
[INFO] training model_sigmode_18_layers

100%
10/10 [04:03<00:00, 24.36s/it]
Epoch: 1 | train_loss: 2.3064 | train_acc: 0.0958 | test_loss: 2.3047 | test_acc: 0.0976
Epoch: 2 | train_loss: 2.3057 | train_acc: 0.0965 | test_loss: 2.3057 | test_acc: 0.0959
Epoch: 3 | train_loss: 2.3054 | train_acc: 0.0967 | test_loss: 2.3045 | test_acc: 0.0965
Epoch: 4 | train_loss: 2.3048 | train_acc: 0.0971 | test_loss: 2.3036 | test_acc: 0.0962
Epoch: 5 | train_loss: 2.3050 | train_acc: 0.0968 | test_loss: 2.3035 | test_acc: 0.0959
Epoch: 6 | train_loss: 2.3046 | train_acc: 0.0966 | test_loss: 2.3051 | test_acc: 0.0960
Epoch: 7 | train_loss: 2.3043 | train_acc: 0.0963 | test_loss: 2.3041 | test_acc: 0.0971
Epoch: 8 | train_loss: 2.3044 | train_acc: 0.0967 | test_loss: 2.3040 | test_acc: 0.0959
Epoch: 9 | train_loss: 2.3043 | train_acc: 0.0974 | test_loss: 2.3046 | test_acc: 0.0969
Epoch: 10 | train_loss: 2.3041 | train_acc: 0.0967 | test_loss: 2.3039 | test_acc: 0.0956
total training time: 243.344 sec.



[INFO] training model_sigmode_20_layers

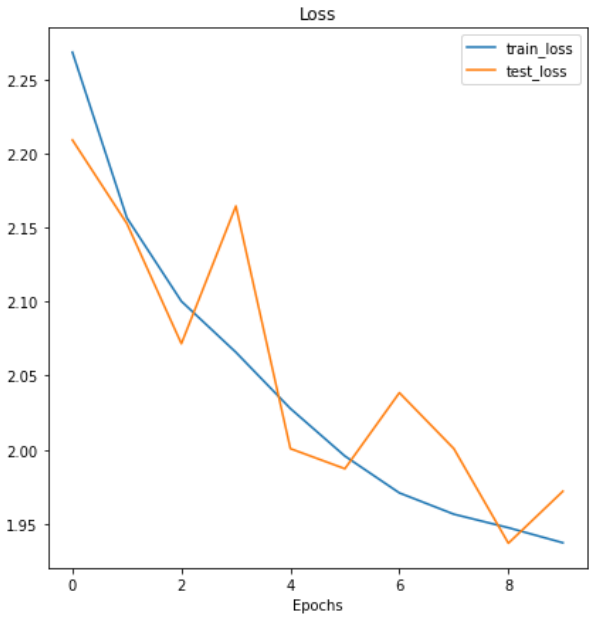
100%
10/10 [04:11<00:00, 25.21s/it]
Epoch: 1 | train_loss: 2.3065 | train_acc: 0.0963 | test_loss: 2.3059 | test_acc: 0.0965
Epoch: 2 | train_loss: 2.3056 | train_acc: 0.0969 | test_loss: 2.3060 | test_acc: 0.0965
Epoch: 3 | train_loss: 2.3055 | train_acc: 0.0965 | test_loss: 2.3061 | test_acc: 0.0969
Epoch: 4 | train_loss: 2.3052 | train_acc: 0.0971 | test_loss: 2.3054 | test_acc: 0.0955
Epoch: 5 | train_loss: 2.3052 | train_acc: 0.0963 | test_loss: 2.3063 | test_acc: 0.0969
Epoch: 6 | train_loss: 2.3050 | train_acc: 0.0972 | test_loss: 2.3043 | test_acc: 0.0965
Epoch: 7 | train_loss: 2.3046 | train_acc: 0.0968 | test_loss: 2.3040 | test_acc: 0.0960
Epoch: 8 | train_loss: 2.3045 | train_acc: 0.0971 | test_loss: 2.3044 | test_acc: 0.0962
Epoch: 9 | train_loss: 2.3046 | train_acc: 0.0961 | test_loss: 2.3037 | test_acc: 0.0966
Epoch: 10 | train_loss: 2.3043 | train_acc: 0.0970 | test_loss: 2.3046 | test_acc: 0.0971
total training time: 251.416 sec.

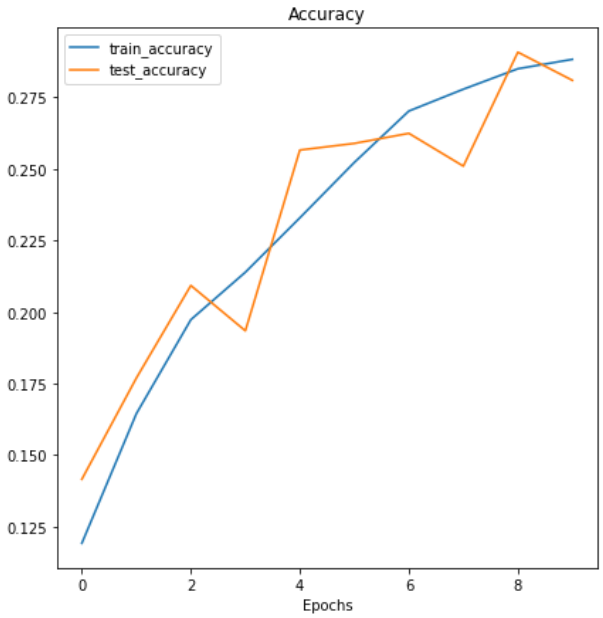


- Results For Tanh

[INFO] training model_tanh_10_layers

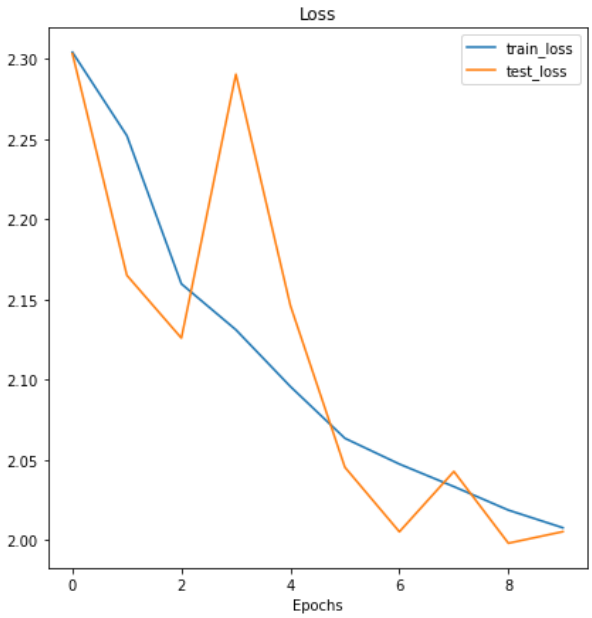
100%
10/10 [03:42<00:00, 22.09s/it]
Epoch: 1 | train_loss: 2.2684 | train_acc: 0.1192 | test_loss: 2.2091 | test_acc: 0.1416
Epoch: 2 | train_loss: 2.1565 | train_acc: 0.1645 | test_loss: 2.1527 | test_acc: 0.1769
Epoch: 3 | train_loss: 2.1002 | train_acc: 0.1973 | test_loss: 2.0716 | test_acc: 0.2093
Epoch: 4 | train_loss: 2.0658 | train_acc: 0.2139 | test_loss: 2.1645 | test_acc: 0.1934
Epoch: 5 | train_loss: 2.0278 | train_acc: 0.2329 | test_loss: 2.0007 | test_acc: 0.2566
Epoch: 6 | train_loss: 1.9957 | train_acc: 0.2523 | test_loss: 1.9871 | test_acc: 0.2589
Epoch: 7 | train_loss: 1.9708 | train_acc: 0.2702 | test_loss: 2.0384 | test_acc: 0.2624
Epoch: 8 | train_loss: 1.9564 | train_acc: 0.2778 | test_loss: 2.0006 | test_acc: 0.2510
Epoch: 9 | train_loss: 1.9473 | train_acc: 0.2850 | test_loss: 1.9368 | test_acc: 0.2908
Epoch: 10 | train_loss: 1.9371 | train_acc: 0.2883 | test_loss: 1.9719 | test_acc: 0.2809
total training time: 222.701 sec.

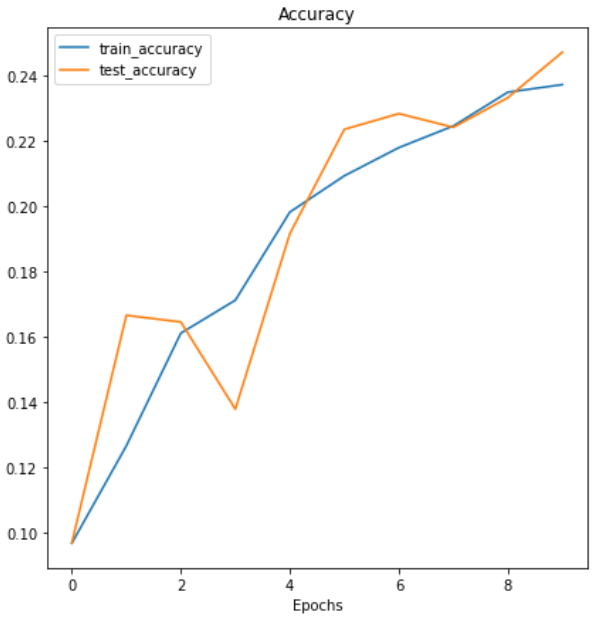




[INFO] training model_tanh_12_layers

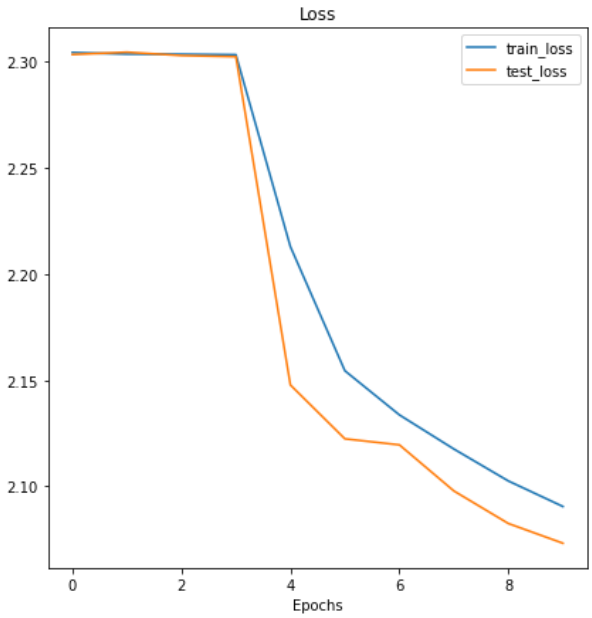
100%
10/10 [03:46<00:00, 22.59s/it]
Epoch: 1 | train_loss: 2.3042 | train_acc: 0.0969 | test_loss: 2.3033 | test_acc: 0.0971
Epoch: 2 | train_loss: 2.2522 | train_acc: 0.1267 | test_loss: 2.1651 | test_acc: 0.1666
Epoch: 3 | train_loss: 2.1598 | train_acc: 0.1611 | test_loss: 2.1260 | test_acc: 0.1646
Epoch: 4 | train_loss: 2.1312 | train_acc: 0.1712 | test_loss: 2.2905 | test_acc: 0.1379
Epoch: 5 | train_loss: 2.0958 | train_acc: 0.1982 | test_loss: 2.1463 | test_acc: 0.1916
Epoch: 6 | train_loss: 2.0636 | train_acc: 0.2093 | test_loss: 2.0455 | test_acc: 0.2235
Epoch: 7 | train_loss: 2.0475 | train_acc: 0.2179 | test_loss: 2.0053 | test_acc: 0.2283
Epoch: 8 | train_loss: 2.0335 | train_acc: 0.2245 | test_loss: 2.0429 | test_acc: 0.2241
Epoch: 9 | train_loss: 2.0188 | train_acc: 0.2348 | test_loss: 1.9982 | test_acc: 0.2332
Epoch: 10 | train_loss: 2.0078 | train_acc: 0.2371 | test_loss: 2.0054 | test_acc: 0.2471
total training time: 226.303 sec.

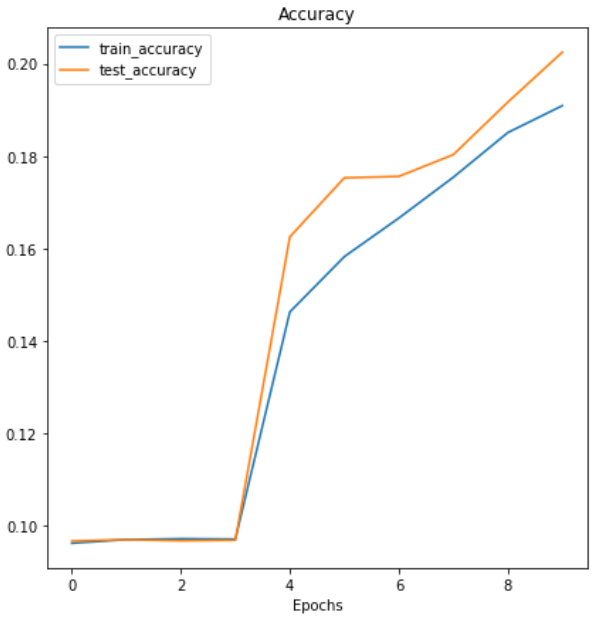




[INFO] training model_tanh_14_layers

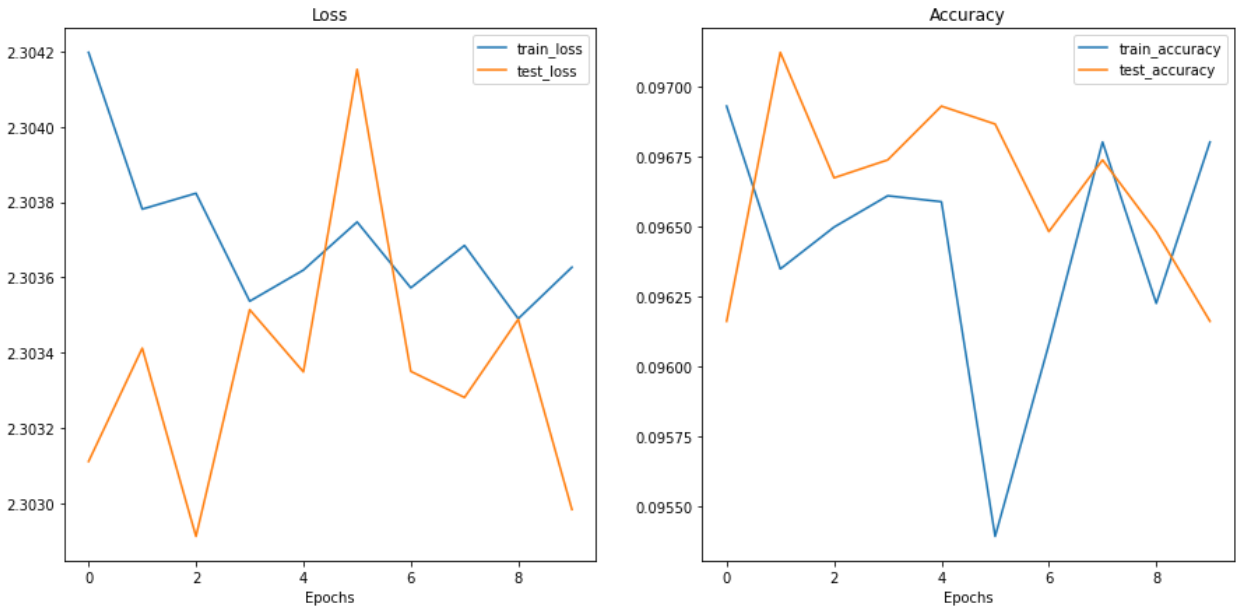
100%
10/10 [03:52<00:00, 23.22s/it]
Epoch: 1 | train_loss: 2.3044 | train_acc: 0.0963 | test_loss: 2.3035 | test_acc: 0.0967
Epoch: 2 | train_loss: 2.3037 | train_acc: 0.0971 | test_loss: 2.3046 | test_acc: 0.0971
Epoch: 3 | train_loss: 2.3036 | train_acc: 0.0972 | test_loss: 2.3030 | test_acc: 0.0968
Epoch: 4 | train_loss: 2.3034 | train_acc: 0.0972 | test_loss: 2.3025 | test_acc: 0.0969
Epoch: 5 | train_loss: 2.2130 | train_acc: 0.1463 | test_loss: 2.1478 | test_acc: 0.1625
Epoch: 6 | train_loss: 2.1545 | train_acc: 0.1582 | test_loss: 2.1224 | test_acc: 0.1753
Epoch: 7 | train_loss: 2.1337 | train_acc: 0.1666 | test_loss: 2.1195 | test_acc: 0.1756
Epoch: 8 | train_loss: 2.1176 | train_acc: 0.1754 | test_loss: 2.0978 | test_acc: 0.1803
Epoch: 9 | train_loss: 2.1025 | train_acc: 0.1851 | test_loss: 2.0825 | test_acc: 0.1916
Epoch: 10 | train_loss: 2.0905 | train_acc: 0.1909 | test_loss: 2.0732 | test_acc: 0.2024
total training time: 232.271 sec.





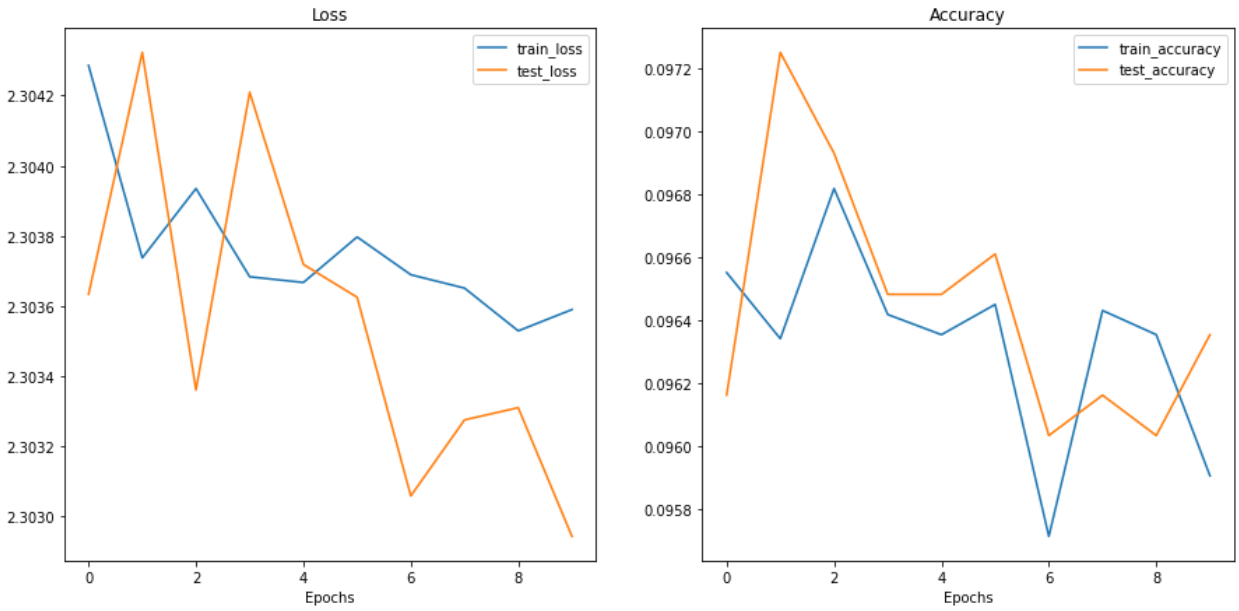
[INFO] training model_tanh_16_layers

100%
10/10 [03:59<00:00, 24.05s/it]
Epoch: 1 | train_loss: 2.3042 | train_acc: 0.0969 | test_loss: 2.3031 | test_acc: 0.0962
Epoch: 2 | train_loss: 2.3038 | train_acc: 0.0963 | test_loss: 2.3034 | test_acc: 0.0971
Epoch: 3 | train_loss: 2.3038 | train_acc: 0.0965 | test_loss: 2.3029 | test_acc: 0.0967
Epoch: 4 | train_loss: 2.3035 | train_acc: 0.0966 | test_loss: 2.3035 | test_acc: 0.0967
Epoch: 5 | train_loss: 2.3036 | train_acc: 0.0966 | test_loss: 2.3033 | test_acc: 0.0969
Epoch: 6 | train_loss: 2.3037 | train_acc: 0.0954 | test_loss: 2.3042 | test_acc: 0.0969
Epoch: 7 | train_loss: 2.3036 | train_acc: 0.0961 | test_loss: 2.3033 | test_acc: 0.0965
Epoch: 8 | train_loss: 2.3037 | train_acc: 0.0968 | test_loss: 2.3033 | test_acc: 0.0967
Epoch: 9 | train_loss: 2.3035 | train_acc: 0.0962 | test_loss: 2.3035 | test_acc: 0.0965
Epoch: 10 | train_loss: 2.3036 | train_acc: 0.0968 | test_loss: 2.3030 | test_acc: 0.0962
total training time: 239.509 sec.



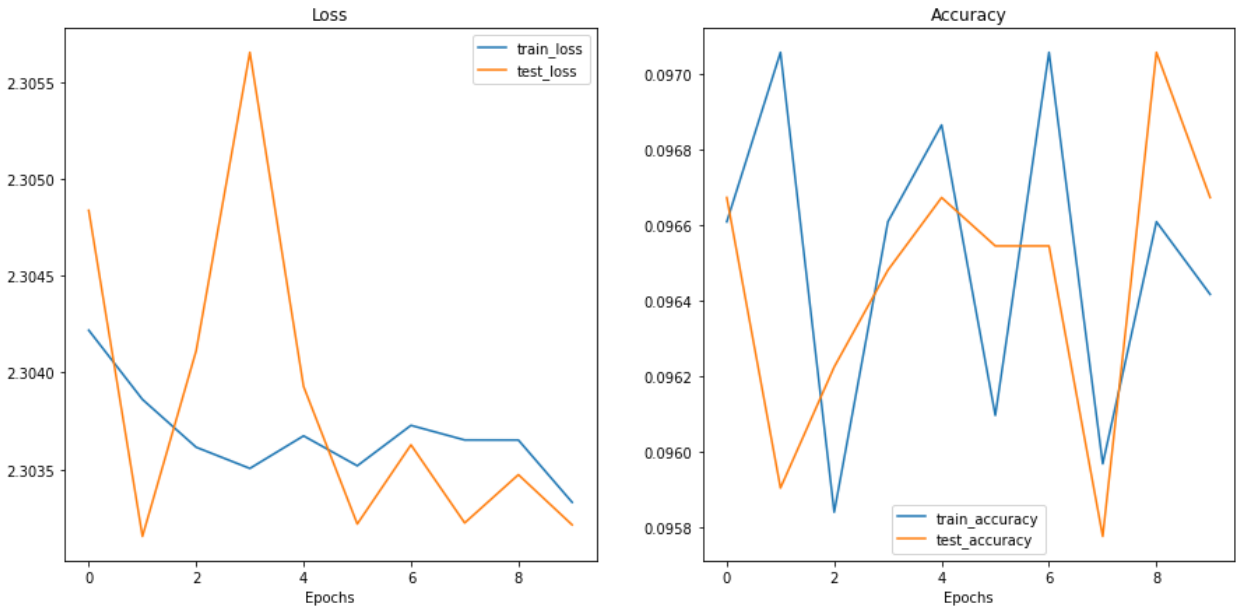
[INFO] training model_tanh_18_layers

100%
10/10 [04:06<00:00, 24.68s/it]
Epoch: 1 | train_loss: 2.3043 | train_acc: 0.0966 | test_loss: 2.3036 | test_acc: 0.0962
Epoch: 2 | train_loss: 2.3037 | train_acc: 0.0963 | test_loss: 2.3043 | test_acc: 0.0972
Epoch: 3 | train_loss: 2.3039 | train_acc: 0.0968 | test_loss: 2.3034 | test_acc: 0.0969
Epoch: 4 | train_loss: 2.3037 | train_acc: 0.0964 | test_loss: 2.3042 | test_acc: 0.0965
Epoch: 5 | train_loss: 2.3037 | train_acc: 0.0964 | test_loss: 2.3037 | test_acc: 0.0965
Epoch: 6 | train_loss: 2.3038 | train_acc: 0.0964 | test_loss: 2.3036 | test_acc: 0.0966
Epoch: 7 | train_loss: 2.3037 | train_acc: 0.0957 | test_loss: 2.3031 | test_acc: 0.0960
Epoch: 8 | train_loss: 2.3037 | train_acc: 0.0964 | test_loss: 2.3033 | test_acc: 0.0962
Epoch: 9 | train_loss: 2.3035 | train_acc: 0.0964 | test_loss: 2.3033 | test_acc: 0.0960
Epoch: 10 | train_loss: 2.3036 | train_acc: 0.0959 | test_loss: 2.3029 | test_acc: 0.0964
total training time: 246.773 sec.



[INFO] training model_tanh_20_layers

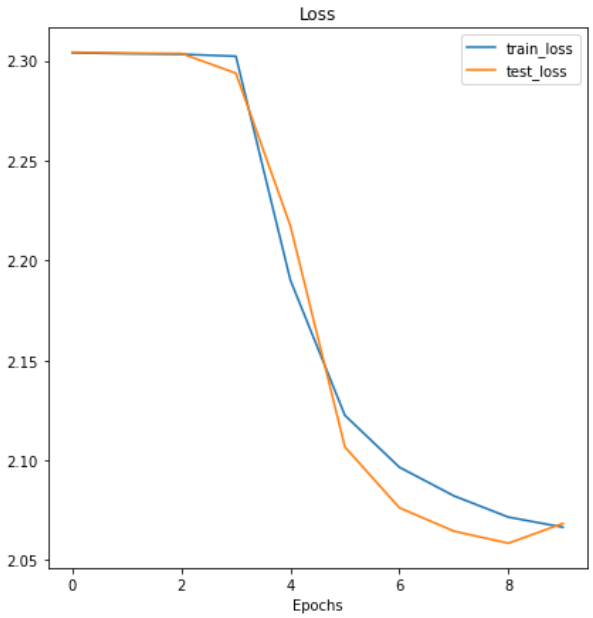
100%
10/10 [04:12<00:00, 25.27s/it]
Epoch: 1 | train_loss: 2.3042 | train_acc: 0.0966 | test_loss: 2.3048 | test_acc: 0.0967
Epoch: 2 | train_loss: 2.3039 | train_acc: 0.0971 | test_loss: 2.3032 | test_acc: 0.0959
Epoch: 3 | train_loss: 2.3036 | train_acc: 0.0958 | test_loss: 2.3041 | test_acc: 0.0962
Epoch: 4 | train_loss: 2.3035 | train_acc: 0.0966 | test_loss: 2.3057 | test_acc: 0.0965
Epoch: 5 | train_loss: 2.3037 | train_acc: 0.0969 | test_loss: 2.3039 | test_acc: 0.0967
Epoch: 6 | train_loss: 2.3035 | train_acc: 0.0961 | test_loss: 2.3032 | test_acc: 0.0965
Epoch: 7 | train_loss: 2.3037 | train_acc: 0.0971 | test_loss: 2.3036 | test_acc: 0.0965
Epoch: 8 | train_loss: 2.3037 | train_acc: 0.0960 | test_loss: 2.3032 | test_acc: 0.0958
Epoch: 9 | train_loss: 2.3037 | train_acc: 0.0966 | test_loss: 2.3035 | test_acc: 0.0971
Epoch: 10 | train_loss: 2.3033 | train_acc: 0.0964 | test_loss: 2.3032 | test_acc: 0.0967
total training time: 252.762 sec.

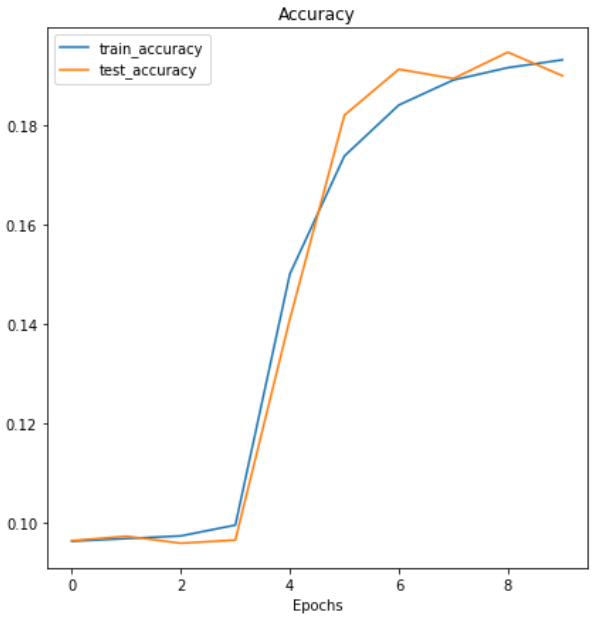


- Solution for vanishing gradient problem using non-linear activation function (ReLU)

[INFO] training model_relu_10_layers

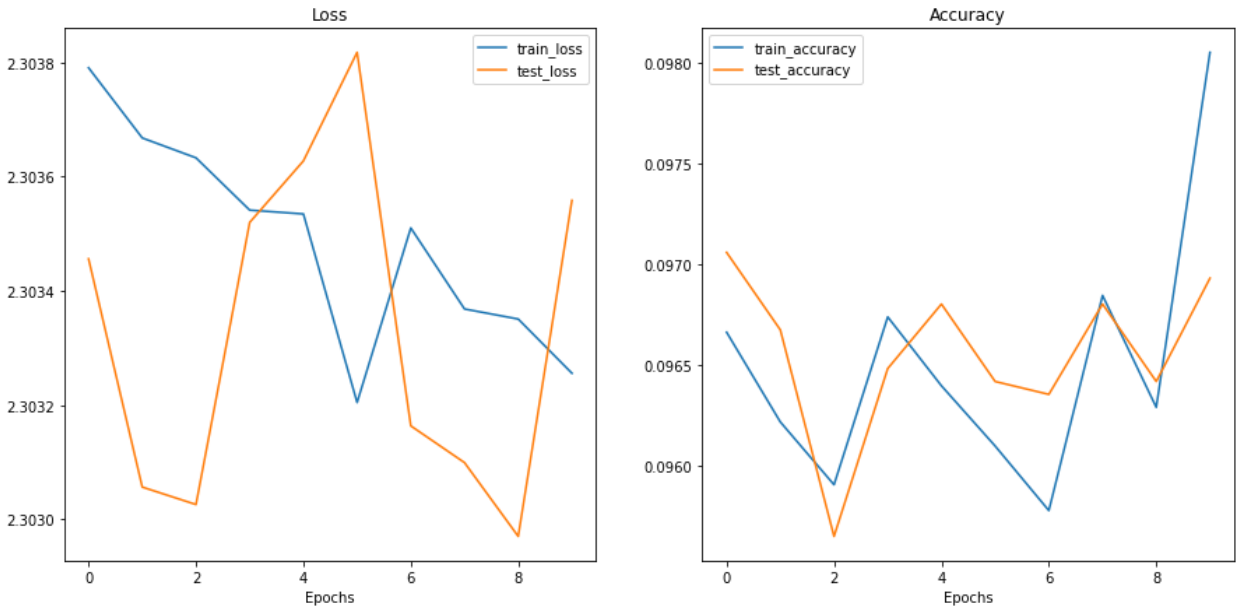
100%
10/10 [03:46<00:00, 22.48s/it]
Epoch: 1 | train_loss: 2.3039 | train_acc: 0.0963 | test_loss: 2.3042 | test_acc: 0.0964
Epoch: 2 | train_loss: 2.3035 | train_acc: 0.0968 | test_loss: 2.3037 | test_acc: 0.0973
Epoch: 3 | train_loss: 2.3032 | train_acc: 0.0974 | test_loss: 2.3036 | test_acc: 0.0959
Epoch: 4 | train_loss: 2.3022 | train_acc: 0.0995 | test_loss: 2.2937 | test_acc: 0.0965
Epoch: 5 | train_loss: 2.1901 | train_acc: 0.1501 | test_loss: 2.2173 | test_acc: 0.1410
Epoch: 6 | train_loss: 2.1224 | train_acc: 0.1738 | test_loss: 2.1066 | test_acc: 0.1820
Epoch: 7 | train_loss: 2.0964 | train_acc: 0.1841 | test_loss: 2.0761 | test_acc: 0.1913
Epoch: 8 | train_loss: 2.0821 | train_acc: 0.1891 | test_loss: 2.0644 | test_acc: 0.1894
Epoch: 9 | train_loss: 2.0714 | train_acc: 0.1916 | test_loss: 2.0584 | test_acc: 0.1947
Epoch: 10 | train_loss: 2.0664 | train_acc: 0.1932 | test_loss: 2.0682 | test_acc: 0.1900
total training time: 226.109 sec.





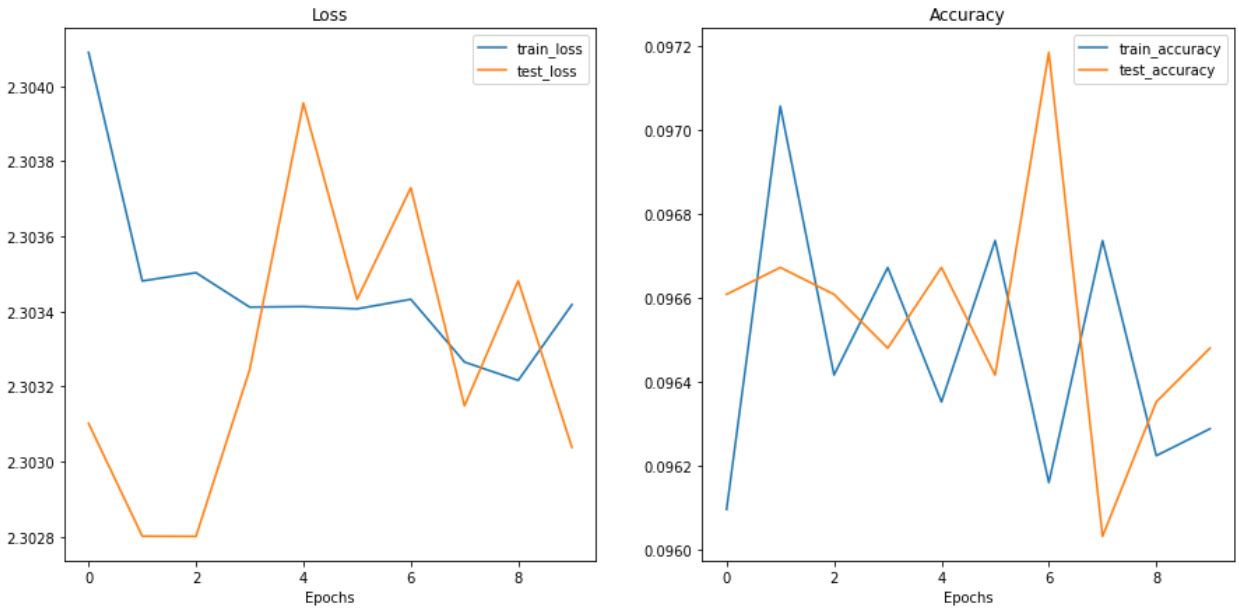
[INFO] training model_relu_12_layers

100%
10/10 [03:49<00:00, 22.86s/it]
Epoch: 1 | train_loss: 2.3038 | train_acc: 0.0967 | test_loss: 2.3035 | test_acc: 0.0971
Epoch: 2 | train_loss: 2.3037 | train_acc: 0.0962 | test_loss: 2.3031 | test_acc: 0.0967
Epoch: 3 | train_loss: 2.3036 | train_acc: 0.0959 | test_loss: 2.3030 | test_acc: 0.0956
Epoch: 4 | train_loss: 2.3035 | train_acc: 0.0967 | test_loss: 2.3035 | test_acc: 0.0965
Epoch: 5 | train_loss: 2.3035 | train_acc: 0.0964 | test_loss: 2.3036 | test_acc: 0.0968
Epoch: 6 | train_loss: 2.3032 | train_acc: 0.0961 | test_loss: 2.3038 | test_acc: 0.0964
Epoch: 7 | train_loss: 2.3035 | train_acc: 0.0958 | test_loss: 2.3032 | test_acc: 0.0964
Epoch: 8 | train_loss: 2.3034 | train_acc: 0.0968 | test_loss: 2.3031 | test_acc: 0.0968
Epoch: 9 | train_loss: 2.3034 | train_acc: 0.0963 | test_loss: 2.3030 | test_acc: 0.0964
Epoch: 10 | train_loss: 2.3033 | train_acc: 0.0980 | test_loss: 2.3036 | test_acc: 0.0969
total training time: 229.629 sec.



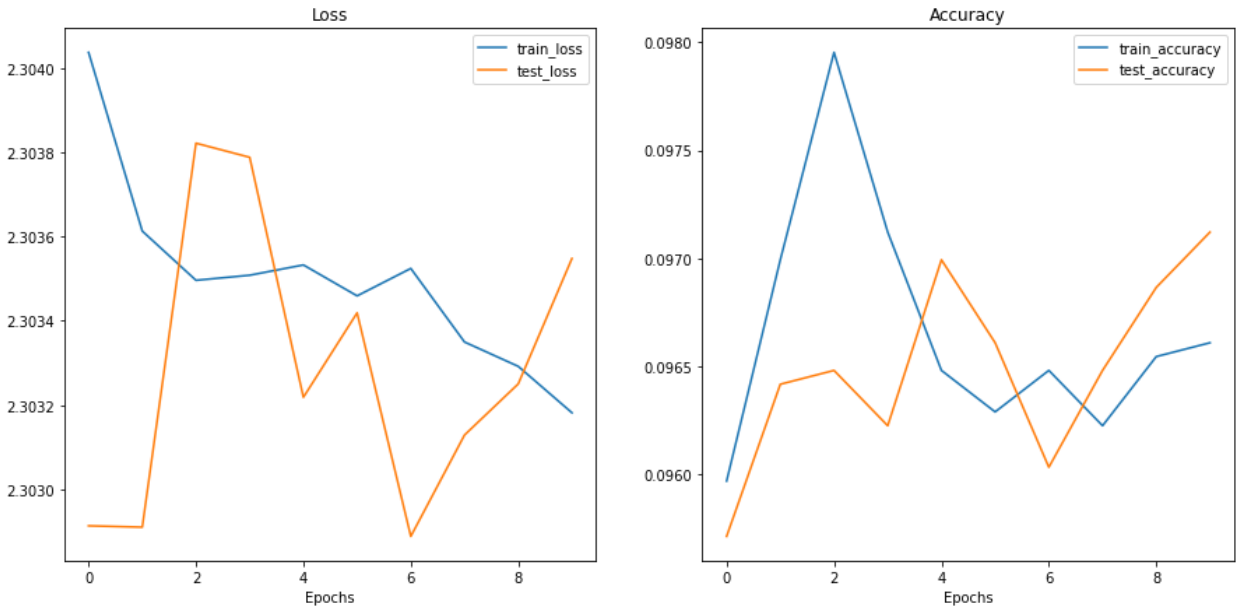
[INFO] training model_relu_14_layers

100%
10/10 [03:54<00:00, 23.49s/it]
Epoch: 1 | train_loss: 2.3041 | train_acc: 0.0961 | test_loss: 2.3031 | test_acc: 0.0966
Epoch: 2 | train_loss: 2.3035 | train_acc: 0.0971 | test_loss: 2.3028 | test_acc: 0.0967
Epoch: 3 | train_loss: 2.3035 | train_acc: 0.0964 | test_loss: 2.3028 | test_acc: 0.0966
Epoch: 4 | train_loss: 2.3034 | train_acc: 0.0967 | test_loss: 2.3032 | test_acc: 0.0965
Epoch: 5 | train_loss: 2.3034 | train_acc: 0.0964 | test_loss: 2.3040 | test_acc: 0.0967
Epoch: 6 | train_loss: 2.3034 | train_acc: 0.0967 | test_loss: 2.3034 | test_acc: 0.0964
Epoch: 7 | train_loss: 2.3034 | train_acc: 0.0962 | test_loss: 2.3037 | test_acc: 0.0972
Epoch: 8 | train_loss: 2.3033 | train_acc: 0.0967 | test_loss: 2.3031 | test_acc: 0.0960
Epoch: 9 | train_loss: 2.3032 | train_acc: 0.0962 | test_loss: 2.3035 | test_acc: 0.0964
Epoch: 10 | train_loss: 2.3034 | train_acc: 0.0963 | test_loss: 2.3030 | test_acc: 0.0965
total training time: 234.786 sec.



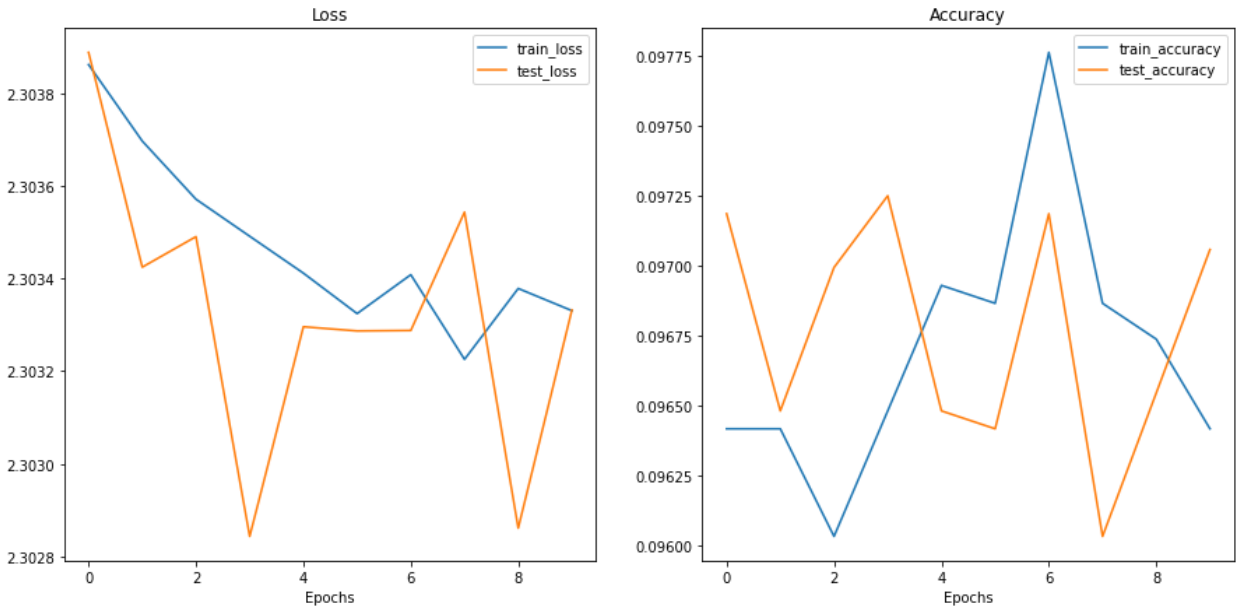
[INFO] training model_relu_16_layers

100%
10/10 [04:00<00:00, 24.07s/it]
Epoch: 1 | train_loss: 2.3040 | train_acc: 0.0960 | test_loss: 2.3029 | test_acc: 0.0957
Epoch: 2 | train_loss: 2.3036 | train_acc: 0.0970 | test_loss: 2.3029 | test_acc: 0.0964
Epoch: 3 | train_loss: 2.3035 | train_acc: 0.0980 | test_loss: 2.3038 | test_acc: 0.0965
Epoch: 4 | train_loss: 2.3035 | train_acc: 0.0971 | test_loss: 2.3038 | test_acc: 0.0962
Epoch: 5 | train_loss: 2.3035 | train_acc: 0.0965 | test_loss: 2.3032 | test_acc: 0.0970
Epoch: 6 | train_loss: 2.3035 | train_acc: 0.0963 | test_loss: 2.3034 | test_acc: 0.0966
Epoch: 7 | train_loss: 2.3035 | train_acc: 0.0965 | test_loss: 2.3029 | test_acc: 0.0960
Epoch: 8 | train_loss: 2.3033 | train_acc: 0.0962 | test_loss: 2.3031 | test_acc: 0.0965
Epoch: 9 | train_loss: 2.3033 | train_acc: 0.0965 | test_loss: 2.3033 | test_acc: 0.0969
Epoch: 10 | train_loss: 2.3032 | train_acc: 0.0966 | test_loss: 2.3035 | test_acc: 0.0971
total training time: 240.799 sec.



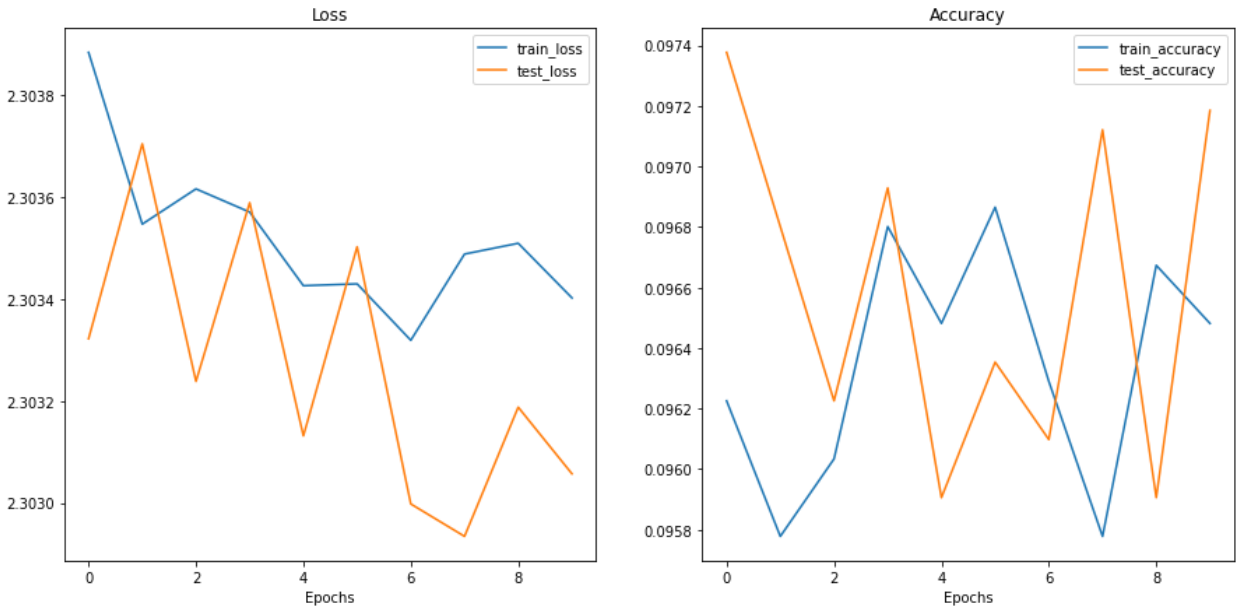
[INFO] training model_relu_18_layers

100%
10/10 [04:07<00:00, 24.60s/it]
Epoch: 1 | train_loss: 2.3039 | train_acc: 0.0964 | test_loss: 2.3039 | test_acc: 0.0972
Epoch: 2 | train_loss: 2.3037 | train_acc: 0.0964 | test_loss: 2.3034 | test_acc: 0.0965
Epoch: 3 | train_loss: 2.3036 | train_acc: 0.0960 | test_loss: 2.3035 | test_acc: 0.0970
Epoch: 4 | train_loss: 2.3035 | train_acc: 0.0965 | test_loss: 2.3028 | test_acc: 0.0972
Epoch: 5 | train_loss: 2.3034 | train_acc: 0.0969 | test_loss: 2.3033 | test_acc: 0.0965
Epoch: 6 | train_loss: 2.3033 | train_acc: 0.0969 | test_loss: 2.3033 | test_acc: 0.0964
Epoch: 7 | train_loss: 2.3034 | train_acc: 0.0978 | test_loss: 2.3033 | test_acc: 0.0972
Epoch: 8 | train_loss: 2.3032 | train_acc: 0.0969 | test_loss: 2.3035 | test_acc: 0.0960
Epoch: 9 | train_loss: 2.3034 | train_acc: 0.0967 | test_loss: 2.3029 | test_acc: 0.0965
Epoch: 10 | train_loss: 2.3033 | train_acc: 0.0964 | test_loss: 2.3033 | test_acc: 0.0971
total training time: 247.086 sec.



[INFO] training model_relu_20_layers

100%
10/10 [04:15<00:00, 25.61s/it]
Epoch: 1 | train_loss: 2.3039 | train_acc: 0.0962 | test_loss: 2.3033 | test_acc: 0.0974
Epoch: 2 | train_loss: 2.3035 | train_acc: 0.0958 | test_loss: 2.3037 | test_acc: 0.0968
Epoch: 3 | train_loss: 2.3036 | train_acc: 0.0960 | test_loss: 2.3032 | test_acc: 0.0962
Epoch: 4 | train_loss: 2.3036 | train_acc: 0.0968 | test_loss: 2.3036 | test_acc: 0.0969
Epoch: 5 | train_loss: 2.3034 | train_acc: 0.0965 | test_loss: 2.3031 | test_acc: 0.0959
Epoch: 6 | train_loss: 2.3034 | train_acc: 0.0969 | test_loss: 2.3035 | test_acc: 0.0964
Epoch: 7 | train_loss: 2.3033 | train_acc: 0.0963 | test_loss: 2.3030 | test_acc: 0.0961
Epoch: 8 | train_loss: 2.3035 | train_acc: 0.0958 | test_loss: 2.3029 | test_acc: 0.0971
Epoch: 9 | train_loss: 2.3035 | train_acc: 0.0967 | test_loss: 2.3032 | test_acc: 0.0959
Epoch: 10 | train_loss: 2.3034 | train_acc: 0.0965 | test_loss: 2.3031 | test_acc: 0.0972
total training time: 255.403 sec.



NOTE 1:- The Graphs of all the experiments might not look too different from each other because we are doing only ten epochs per model; if we were going up to 100-500 epochs, we would see a clear difference b/w all of them.

NOTE 2:- We can also see this with the updates in weights matrices, but that would become very hard to interpret as we humans are not very good at analyzing a large set of numerical data.

Task: Implement a neural network and utilize the CIFAR-10 dataset for the analysis

Objectives:

Question 2.

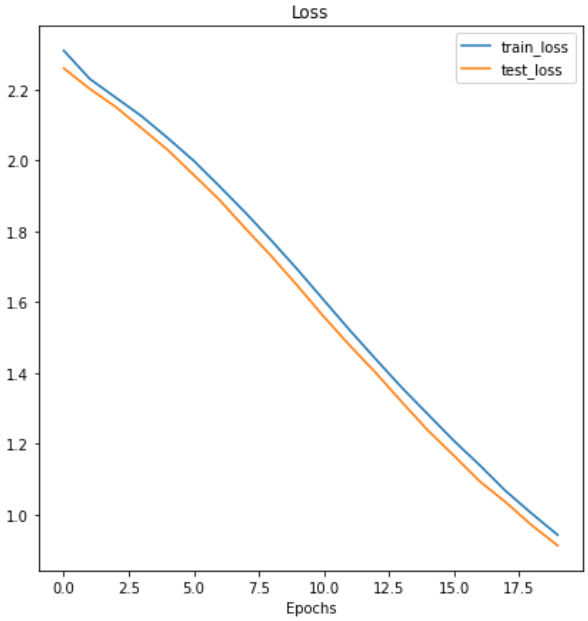
- Implement a neural network on the Gurmukhi dataset and implement the following regularization techniques from scratch.
- L-1 regularization
- L-2 regularization
- Dropouts
- Implement gradient checking (from scratch) to verify the values of gradients during backpropagation.

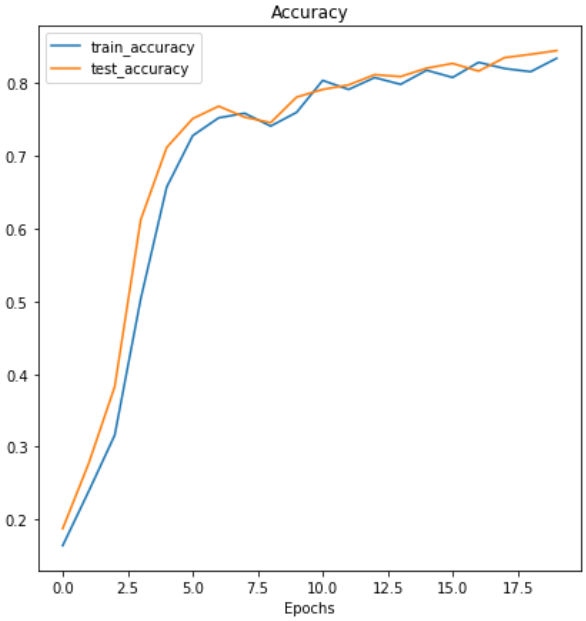
Procedure:

- Import required packages. (PyTorch, NumPy, Matplotlib, torchmetrics …etc.)
- Making compose transform to convert image to grayscale and to tensors.
- Even though the images from the dataset looked like grayscale images, they were not. Hence we use transform.
- Writing Script to download the dataset with the help of the requests library.
- Loading the Data with torchvision.dataset.ImageFolder() function with required parameters and transform.
- Visualization of random data samples from the dataset.
- Creating the data loaders for train and test datasets respectively, with batch size 32.
- We are defining the PyTorch model class by inheriting nn.Module.
- Making model instances with appropriate arguments like numbers of layers and activation functions.
- Defining the training step, testing step & train function that will be used for training the neural network.
- Defining function for Penalty for L1 regularization
- Defining function for Penalty for L2 regularization
- Defining plot graph function.
- We‘ve defined two models, one with sigmoid and the other with tanh, for experimentation purposes.
- Experiments done are:-
    - training model_sigmode_3_layers without regularization
    - training model_tanh_3_layers without regularization
    - training model_sigmode_3_layers with l1 regularization
    - training model_tanh_3_layers with l1 regularization
    - training model_sigmode_3_layers with l2 regularization
    - training model_tanh_3_layers with l2 regularization
    - training model_sigmode_3_layers with dropouts with p=0.25
    - training model_tanh_3_layers with dropouts with p=0.25
    - gradient check
- Results:-

- training model_sigmode_3_layers without regularization

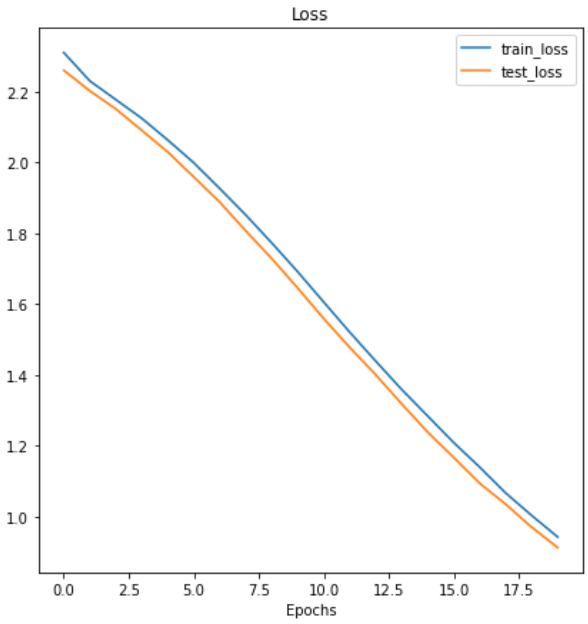
20/20 [00:18<00:00, 1.08it/s]
Epoch: 1 | train_loss: 2.3096 | train_acc: 0.1641 | test_loss: 2.2596 | test_acc: 0.1875
Epoch: 2 | train_loss: 2.2297 | train_acc: 0.2388 | test_loss: 2.2017 | test_acc: 0.2772
Epoch: 3 | train_loss: 2.1768 | train_acc: 0.3159 | test_loss: 2.1505 | test_acc: 0.3827
Epoch: 4 | train_loss: 2.1240 | train_acc: 0.5032 | test_loss: 2.0901 | test_acc: 0.6121
Epoch: 5 | train_loss: 2.0627 | train_acc: 0.6572 | test_loss: 2.0289 | test_acc: 0.7116
Epoch: 6 | train_loss: 1.9985 | train_acc: 0.7279 | test_loss: 1.9585 | test_acc: 0.7513
Epoch: 7 | train_loss: 1.9256 | train_acc: 0.7524 | test_loss: 1.8872 | test_acc: 0.7684
Epoch: 8 | train_loss: 1.8511 | train_acc: 0.7587 | test_loss: 1.8063 | test_acc: 0.7533
Epoch: 9 | train_loss: 1.7721 | train_acc: 0.7410 | test_loss: 1.7280 | test_acc: 0.7458
Epoch: 10 | train_loss: 1.6906 | train_acc: 0.7600 | test_loss: 1.6446 | test_acc: 0.7809
Epoch: 11 | train_loss: 1.6051 | train_acc: 0.8038 | test_loss: 1.5583 | test_acc: 0.7914
Epoch: 12 | train_loss: 1.5200 | train_acc: 0.7915 | test_loss: 1.4769 | test_acc: 0.7977
Epoch: 13 | train_loss: 1.4383 | train_acc: 0.8078 | test_loss: 1.3997 | test_acc: 0.8118
Epoch: 14 | train_loss: 1.3581 | train_acc: 0.7985 | test_loss: 1.3175 | test_acc: 0.8091
Epoch: 15 | train_loss: 1.2832 | train_acc: 0.8178 | test_loss: 1.2376 | test_acc: 0.8207
Epoch: 16 | train_loss: 1.2082 | train_acc: 0.8079 | test_loss: 1.1662 | test_acc: 0.8272
Epoch: 17 | train_loss: 1.1388 | train_acc: 0.8287 | test_loss: 1.0932 | test_acc: 0.8167
Epoch: 18 | train_loss: 1.0659 | train_acc: 0.8202 | test_loss: 1.0341 | test_acc: 0.8352
Epoch: 19 | train_loss: 1.0023 | train_acc: 0.8158 | test_loss: 0.9692 | test_acc: 0.8397
Epoch: 20 | train_loss: 0.9417 | train_acc: 0.8342 | test_loss: 0.9117 | test_acc: 0.8448
total training time: 18.555 sec.

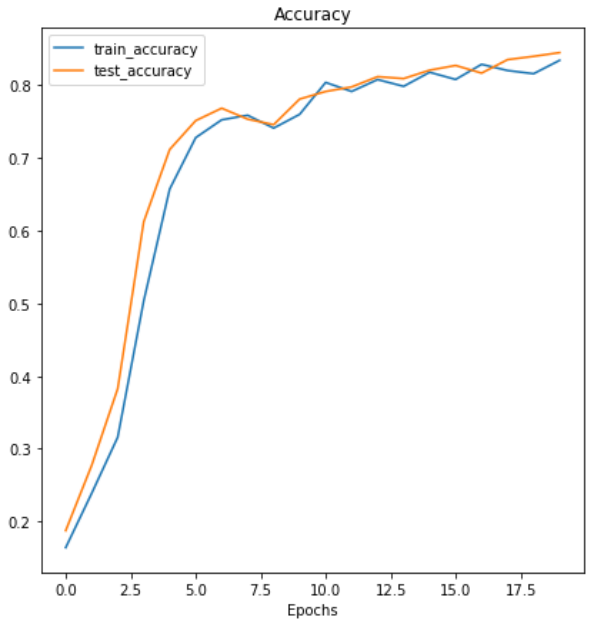




- training model_tanh_3_layers without regularization

20/20 [00:18<00:00, 1.03it/s]
Epoch: 1 | train_loss: 1.9821 | train_acc: 0.4414 | test_loss: 1.7151 | test_acc: 0.5905
Epoch: 2 | train_loss: 1.5498 | train_acc: 0.7051 | test_loss: 1.3751 | test_acc: 0.7741
Epoch: 3 | train_loss: 1.2483 | train_acc: 0.7977 | test_loss: 1.1013 | test_acc: 0.8424
Epoch: 4 | train_loss: 0.9997 | train_acc: 0.8650 | test_loss: 0.8892 | test_acc: 0.8655
Epoch: 5 | train_loss: 0.8178 | train_acc: 0.8457 | test_loss: 0.7227 | test_acc: 0.8935
Epoch: 6 | train_loss: 0.6569 | train_acc: 0.8799 | test_loss: 0.5865 | test_acc: 0.8899
Epoch: 7 | train_loss: 0.5339 | train_acc: 0.9134 | test_loss: 0.4676 | test_acc: 0.9200
Epoch: 8 | train_loss: 0.4436 | train_acc: 0.9174 | test_loss: 0.3983 | test_acc: 0.9245
Epoch: 9 | train_loss: 0.3680 | train_acc: 0.9064 | test_loss: 0.3487 | test_acc: 0.9347
Epoch: 10 | train_loss: 0.3156 | train_acc: 0.9176 | test_loss: 0.2873 | test_acc: 0.9429
Epoch: 11 | train_loss: 0.2634 | train_acc: 0.9165 | test_loss: 0.2354 | test_acc: 0.9319
Epoch: 12 | train_loss: 0.2198 | train_acc: 0.9520 | test_loss: 0.2006 | test_acc: 0.9505
Epoch: 13 | train_loss: 0.1872 | train_acc: 0.9508 | test_loss: 0.1661 | test_acc: 0.9598
Epoch: 14 | train_loss: 0.1666 | train_acc: 0.9418 | test_loss: 0.1471 | test_acc: 0.9277
Epoch: 15 | train_loss: 0.1493 | train_acc: 0.9465 | test_loss: 0.1303 | test_acc: 0.9537
Epoch: 16 | train_loss: 0.1250 | train_acc: 0.9452 | test_loss: 0.1194 | test_acc: 0.9612
Epoch: 17 | train_loss: 0.1090 | train_acc: 0.9457 | test_loss: 0.0994 | test_acc: 0.9552
Epoch: 18 | train_loss: 0.0953 | train_acc: 0.9607 | test_loss: 0.0861 | test_acc: 0.9556
Epoch: 19 | train_loss: 0.0812 | train_acc: 0.9453 | test_loss: 0.0750 | test_acc: 0.9557
Epoch: 20 | train_loss: 0.0732 | train_acc: 0.9682 | test_loss: 0.0681 | test_acc: 0.9552
total training time: 18.493 sec.

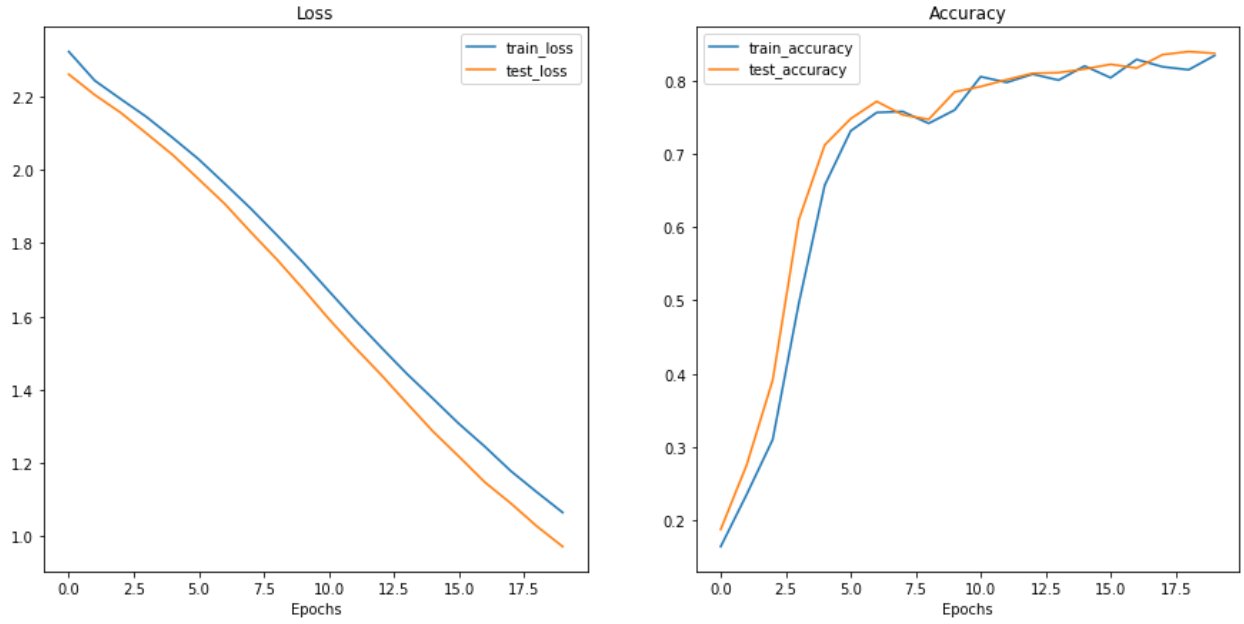




- training model_sigmode_3_layers with l1 regularization

20/20 [00:19<00:00, 1.04it/s]
Epoch: 1 | train_loss: 2.3229 | train_acc: 0.1641 | test_loss: 2.2609 | test_acc: 0.1875
Epoch: 2 | train_loss: 2.2437 | train_acc: 0.2353 | test_loss: 2.2047 | test_acc: 0.2757
Epoch: 3 | train_loss: 2.1934 | train_acc: 0.3101 | test_loss: 2.1560 | test_acc: 0.3913

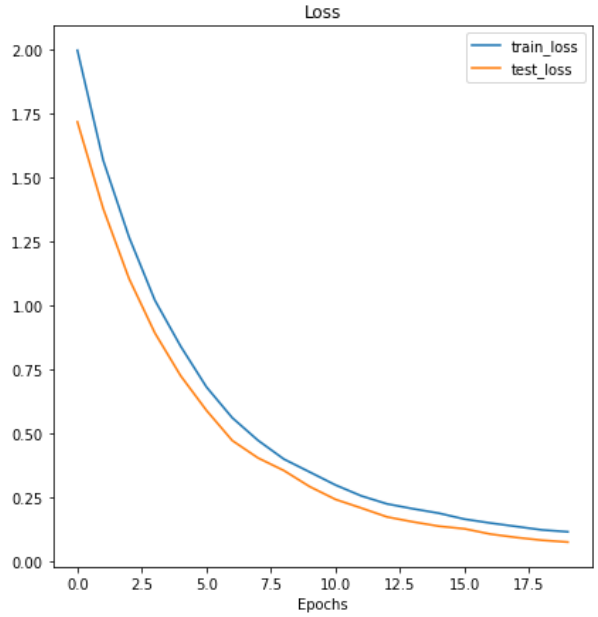
Epoch: 4 | train_loss: 2.1441 | train_acc: 0.4952 | test_loss: 2.0988 | test_acc: 0.6095
Epoch: 5 | train_loss: 2.0874 | train_acc: 0.6569 | test_loss: 2.0407 | test_acc: 0.7119
Epoch: 6 | train_loss: 2.0288 | train_acc: 0.7311 | test_loss: 1.9745 | test_acc: 0.7477
Epoch: 7 | train_loss: 1.9623 | train_acc: 0.7562 | test_loss: 1.9071 | test_acc: 0.7714
Epoch: 8 | train_loss: 1.8945 | train_acc: 0.7576 | test_loss: 1.8302 | test_acc: 0.7530
Epoch: 9 | train_loss: 1.8223 | train_acc: 0.7414 | test_loss: 1.7558 | test_acc: 0.7469
Epoch: 10 | train_loss: 1.7480 | train_acc: 0.7596 | test_loss: 1.6760 | test_acc: 0.7843
Epoch: 11 | train_loss: 1.6694 | train_acc: 0.8051 | test_loss: 1.5933 | test_acc: 0.7917
Epoch: 12 | train_loss: 1.5914 | train_acc: 0.7972 | test_loss: 1.5156 | test_acc: 0.8010
Epoch: 13 | train_loss: 1.5166 | train_acc: 0.8086 | test_loss: 1.4417 | test_acc: 0.8097
Epoch: 14 | train_loss: 1.4436 | train_acc: 0.8004 | test_loss: 1.3635 | test_acc: 0.8106
Epoch: 15 | train_loss: 1.3759 | train_acc: 0.8197 | test_loss: 1.2866 | test_acc: 0.8155
Epoch: 16 | train_loss: 1.3072 | train_acc: 0.8037 | test_loss: 1.2177 | test_acc: 0.8220
Epoch: 17 | train_loss: 1.2448 | train_acc: 0.8286 | test_loss: 1.1473 | test_acc: 0.8168
Epoch: 18 | train_loss: 1.1780 | train_acc: 0.8186 | test_loss: 1.0897 | test_acc: 0.8352
Epoch: 19 | train_loss: 1.1202 | train_acc: 0.8145 | test_loss: 1.0272 | test_acc: 0.8393
Epoch: 20 | train_loss: 1.0644 | train_acc: 0.8340 | test_loss: 0.9713 | test_acc: 0.8370
total training time: 19.236 sec.

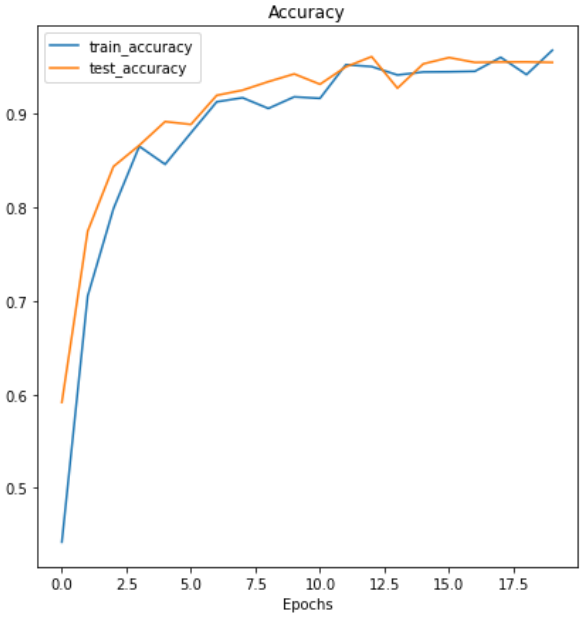


- training model_tanh_3_layers with l1 regularization

20/20 [00:20<00:00, 1.04it/s]
Epoch: 1 | train_loss: 1.9964 | train_acc: 0.4422 | test_loss: 1.7164 | test_acc: 0.5917
Epoch: 2 | train_loss: 1.5660 | train_acc: 0.7053 | test_loss: 1.3771 | test_acc: 0.7747
Epoch: 3 | train_loss: 1.2661 | train_acc: 0.7989 | test_loss: 1.1037 | test_acc: 0.8439
Epoch: 4 | train_loss: 1.0191 | train_acc: 0.8655 | test_loss: 0.8912 | test_acc: 0.8666
Epoch: 5 | train_loss: 0.8390 | train_acc: 0.8462 | test_loss: 0.7236 | test_acc: 0.8919
Epoch: 6 | train_loss: 0.6796 | train_acc: 0.8799 | test_loss: 0.5881 | test_acc: 0.8889
Epoch: 7 | train_loss: 0.5592 | train_acc: 0.9131 | test_loss: 0.4701 | test_acc: 0.9200

Epoch: 8 | train_loss: 0.4707 | train_acc: 0.9174 | test_loss: 0.4020 | test_acc: 0.9255
Epoch: 9 | train_loss: 0.3976 | train_acc: 0.9060 | test_loss: 0.3529 | test_acc: 0.9347
Epoch: 10 | train_loss: 0.3471 | train_acc: 0.9184 | test_loss: 0.2901 | test_acc: 0.9429
Epoch: 11 | train_loss: 0.2963 | train_acc: 0.9168 | test_loss: 0.2400 | test_acc: 0.9319
Epoch: 12 | train_loss: 0.2537 | train_acc: 0.9528 | test_loss: 0.2063 | test_acc: 0.9505
Epoch: 13 | train_loss: 0.2224 | train_acc: 0.9508 | test_loss: 0.1715 | test_acc: 0.9614
Epoch: 14 | train_loss: 0.2035 | train_acc: 0.9418 | test_loss: 0.1523 | test_acc: 0.9277
Epoch: 15 | train_loss: 0.1862 | train_acc: 0.9450 | test_loss: 0.1354 | test_acc: 0.9537
Epoch: 16 | train_loss: 0.1629 | train_acc: 0.9452 | test_loss: 0.1251 | test_acc: 0.9604
Epoch: 17 | train_loss: 0.1477 | train_acc: 0.9457 | test_loss: 0.1045 | test_acc: 0.9552
Epoch: 18 | train_loss: 0.1342 | train_acc: 0.9607 | test_loss: 0.0913 | test_acc: 0.9556
Epoch: 19 | train_loss: 0.1207 | train_acc: 0.9422 | test_loss: 0.0804 | test_acc: 0.9557
Epoch: 20 | train_loss: 0.1134 | train_acc: 0.9682 | test_loss: 0.0732 | test_acc: 0.9552
total training time: 20.045 sec.

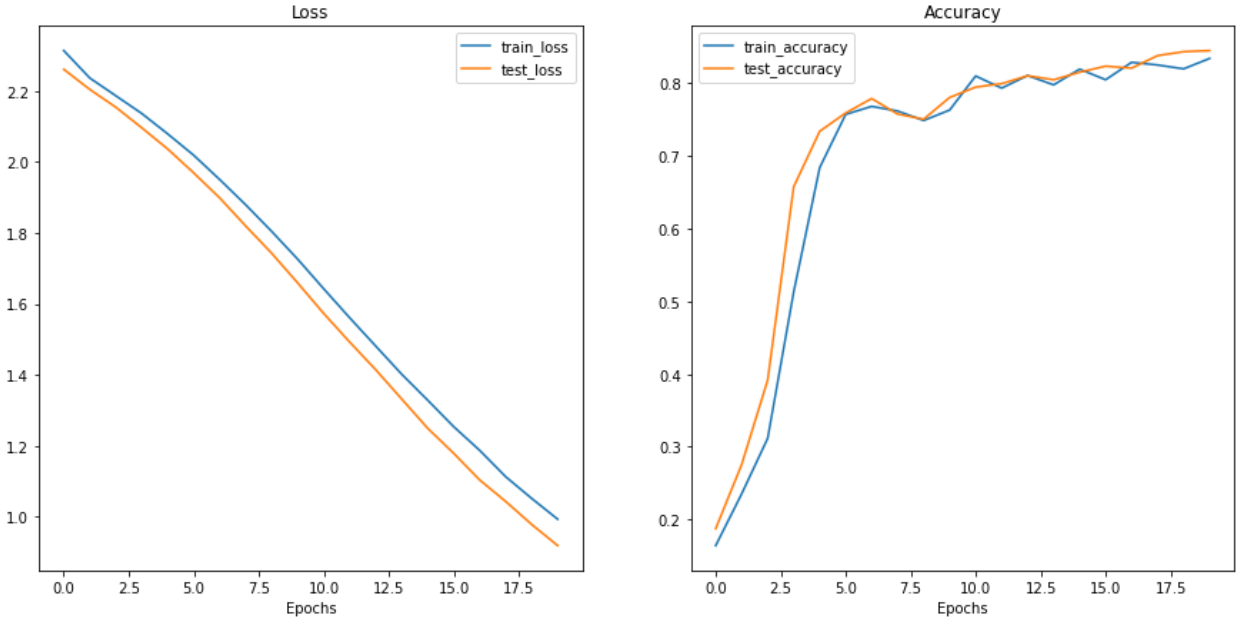




- training model_sigmode_3_layers with l2 regularization

20/20 [00:20<00:00, 1.01it/s]
Epoch: 1 | train_loss: 2.3131 | train_acc: 0.1641 | test_loss: 2.2604 | test_acc: 0.1875
Epoch: 2 | train_loss: 2.2359 | train_acc: 0.2353 | test_loss: 2.2038 | test_acc: 0.2757
Epoch: 3 | train_loss: 2.1860 | train_acc: 0.3117 | test_loss: 2.1538 | test_acc: 0.3918
Epoch: 4 | train_loss: 2.1360 | train_acc: 0.5129 | test_loss: 2.0958 | test_acc: 0.6574
Epoch: 5 | train_loss: 2.0784 | train_acc: 0.6837 | test_loss: 2.0359 | test_acc: 0.7337
Epoch: 6 | train_loss: 2.0184 | train_acc: 0.7570 | test_loss: 1.9686 | test_acc: 0.7589
Epoch: 7 | train_loss: 1.9499 | train_acc: 0.7680 | test_loss: 1.8982 | test_acc: 0.7787
Epoch: 8 | train_loss: 1.8788 | train_acc: 0.7616 | test_loss: 1.8191 | test_acc: 0.7574
Epoch: 9 | train_loss: 1.8039 | train_acc: 0.7486 | test_loss: 1.7423 | test_acc: 0.7504
Epoch: 10 | train_loss: 1.7257 | train_acc: 0.7629 | test_loss: 1.6588 | test_acc: 0.7802
Epoch: 11 | train_loss: 1.6424 | train_acc: 0.8098 | test_loss: 1.5729 | test_acc: 0.7945
Epoch: 12 | train_loss: 1.5606 | train_acc: 0.7931 | test_loss: 1.4923 | test_acc: 0.7992

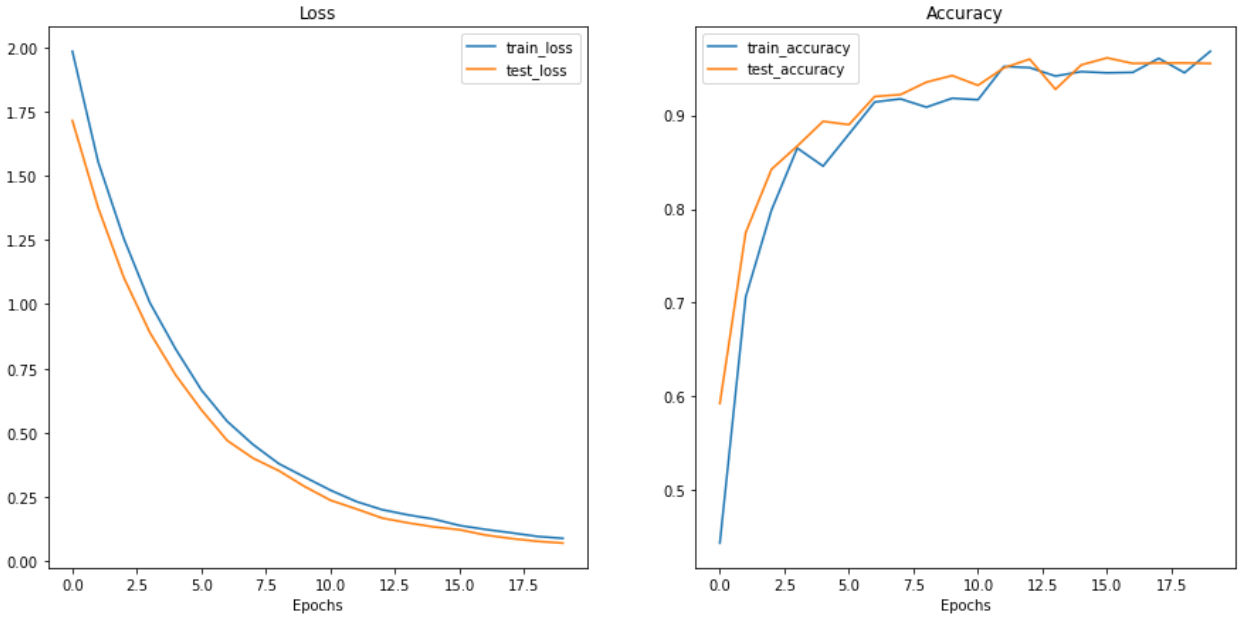
Epoch: 13 | train_loss: 1.4812 | train_acc: 0.8104 | test_loss: 1.4146 | test_acc: 0.8101
Epoch: 14 | train_loss: 1.4018 | train_acc: 0.7976 | test_loss: 1.3320 | test_acc: 0.8046
Epoch: 15 | train_loss: 1.3287 | train_acc: 0.8192 | test_loss: 1.2497 | test_acc: 0.8150
Epoch: 16 | train_loss: 1.2542 | train_acc: 0.8046 | test_loss: 1.1791 | test_acc: 0.8233
Epoch: 17 | train_loss: 1.1875 | train_acc: 0.8286 | test_loss: 1.1040 | test_acc: 0.8205
Epoch: 18 | train_loss: 1.1134 | train_acc: 0.8249 | test_loss: 1.0437 | test_acc: 0.8378
Epoch: 19 | train_loss: 1.0522 | train_acc: 0.8196 | test_loss: 0.9789 | test_acc: 0.8432
Epoch: 20 | train_loss: 0.9937 | train_acc: 0.8339 | test_loss: 0.9194 | test_acc: 0.8447
total training time: 20.195 sec.



- training model_tanh_3_layers with l2 regularization

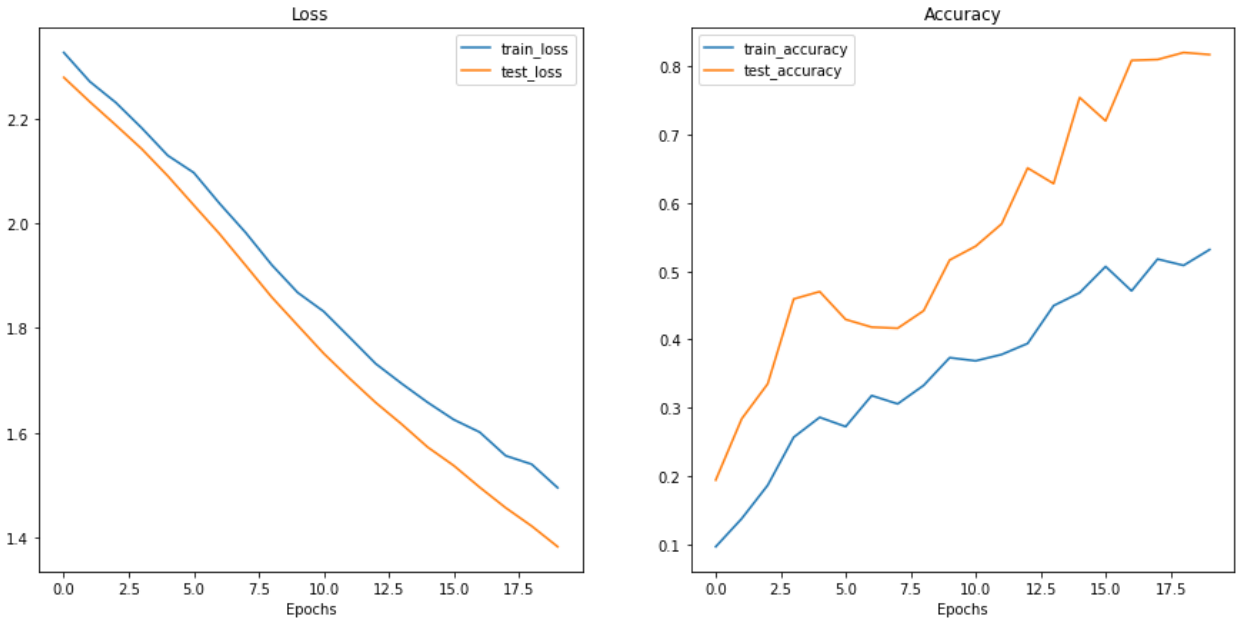
20/20 [00:21<00:00, 1.14s/it]
Epoch: 1 | train_loss: 1.9855 | train_acc: 0.4437 | test_loss: 1.7153 | test_acc: 0.5926
Epoch: 2 | train_loss: 1.5542 | train_acc: 0.7061 | test_loss: 1.3756 | test_acc: 0.7747
Epoch: 3 | train_loss: 1.2537 | train_acc: 0.7988 | test_loss: 1.1020 | test_acc: 0.8424
Epoch: 4 | train_loss: 1.0062 | train_acc: 0.8650 | test_loss: 0.8899 | test_acc: 0.8672
Epoch: 5 | train_loss: 0.8253 | train_acc: 0.8457 | test_loss: 0.7241 | test_acc: 0.8935
Epoch: 6 | train_loss: 0.6655 | train_acc: 0.8799 | test_loss: 0.5881 | test_acc: 0.8899
Epoch: 7 | train_loss: 0.5433 | train_acc: 0.9141 | test_loss: 0.4687 | test_acc: 0.9200
Epoch: 8 | train_loss: 0.4532 | train_acc: 0.9174 | test_loss: 0.4000 | test_acc: 0.9220
Epoch: 9 | train_loss: 0.3784 | train_acc: 0.9086 | test_loss: 0.3512 | test_acc: 0.9353
Epoch: 10 | train_loss: 0.3265 | train_acc: 0.9181 | test_loss: 0.2897 | test_acc: 0.9423
Epoch: 11 | train_loss: 0.2752 | train_acc: 0.9165 | test_loss: 0.2361 | test_acc: 0.9319
Epoch: 12 | train_loss: 0.2312 | train_acc: 0.9520 | test_loss: 0.2024 | test_acc: 0.9505
Epoch: 13 | train_loss: 0.1994 | train_acc: 0.9508 | test_loss: 0.1668 | test_acc: 0.9598
Epoch: 14 | train_loss: 0.1792 | train_acc: 0.9418 | test_loss: 0.1481 | test_acc: 0.9277
Epoch: 15 | train_loss: 0.1627 | train_acc: 0.9465 | test_loss: 0.1324 | test_acc: 0.9537
Epoch: 16 | train_loss: 0.1378 | train_acc: 0.9452 | test_loss: 0.1213 | test_acc: 0.9612
Epoch: 17 | train_loss: 0.1226 | train_acc: 0.9457 | test_loss: 0.1007 | test_acc: 0.9552

Epoch: 18 | train_loss: 0.1094 | train_acc: 0.9607 | test_loss: 0.0871 | test_acc: 0.9556
Epoch: 19 | train_loss: 0.0955 | train_acc: 0.9453 | test_loss: 0.0763 | test_acc: 0.9557
Epoch: 20 | train_loss: 0.0878 | train_acc: 0.9682 | test_loss: 0.0693 | test_acc: 0.9552
total training time: 21.102 sec.



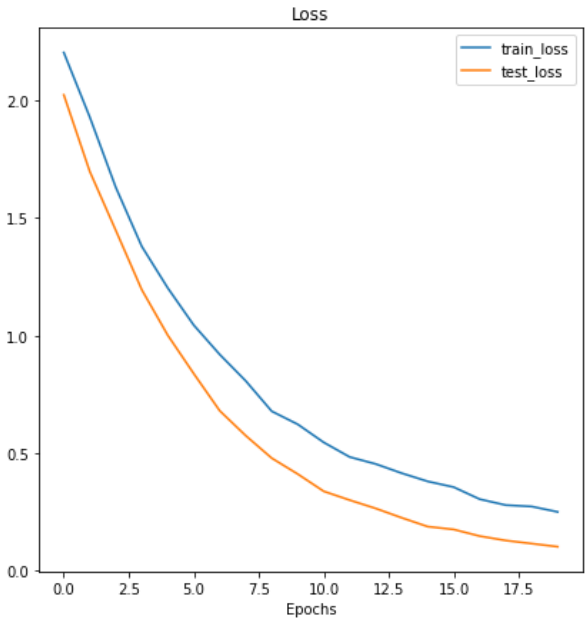
- training model_sigmode_3_layers with dropouts with p=0.25

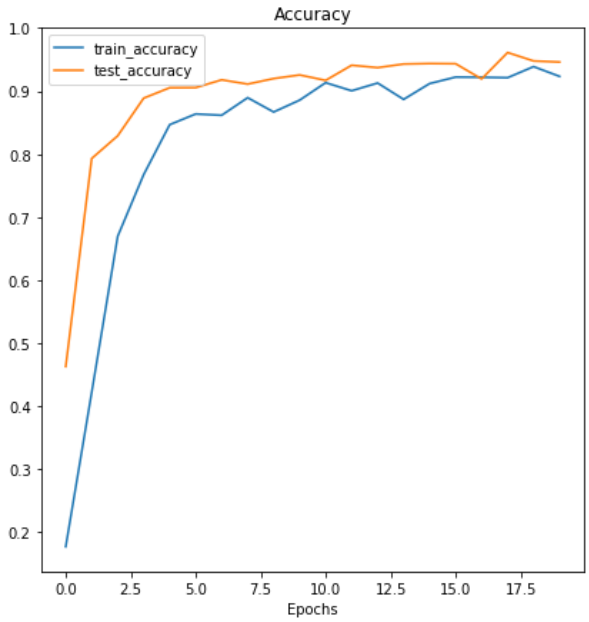
20/20 [00:18<00:00, 1.05it/s]
Epoch: 1 | train_loss: 2.3256 | train_acc: 0.0960 | test_loss: 2.2781 | test_acc: 0.1938
Epoch: 2 | train_loss: 2.2702 | train_acc: 0.1372 | test_loss: 2.2316 | test_acc: 0.2835
Epoch: 3 | train_loss: 2.2302 | train_acc: 0.1861 | test_loss: 2.1871 | test_acc: 0.3348
Epoch: 4 | train_loss: 2.1815 | train_acc: 0.2565 | test_loss: 2.1416 | test_acc: 0.4592
Epoch: 5 | train_loss: 2.1291 | train_acc: 0.2857 | test_loss: 2.0901 | test_acc: 0.4701
Epoch: 6 | train_loss: 2.0963 | train_acc: 0.2721 | test_loss: 2.0343 | test_acc: 0.4292
Epoch: 7 | train_loss: 2.0371 | train_acc: 0.3176 | test_loss: 1.9790 | test_acc: 0.4179
Epoch: 8 | train_loss: 1.9817 | train_acc: 0.3054 | test_loss: 1.9191 | test_acc: 0.4163
Epoch: 9 | train_loss: 1.9203 | train_acc: 0.3326 | test_loss: 1.8588 | test_acc: 0.4420
Epoch: 10 | train_loss: 1.8674 | train_acc: 0.3732 | test_loss: 1.8049 | test_acc: 0.5163
Epoch: 11 | train_loss: 1.8314 | train_acc: 0.3685 | test_loss: 1.7512 | test_acc: 0.5366
Epoch: 12 | train_loss: 1.7818 | train_acc: 0.3778 | test_loss: 1.7034 | test_acc: 0.5693
Epoch: 13 | train_loss: 1.7317 | train_acc: 0.3940 | test_loss: 1.6575 | test_acc: 0.6513
Epoch: 14 | train_loss: 1.6937 | train_acc: 0.4494 | test_loss: 1.6163 | test_acc: 0.6284
Epoch: 15 | train_loss: 1.6578 | train_acc: 0.4683 | test_loss: 1.5724 | test_acc: 0.7547
Epoch: 16 | train_loss: 1.6250 | train_acc: 0.5069 | test_loss: 1.5374 | test_acc: 0.7204
Epoch: 17 | train_loss: 1.6011 | train_acc: 0.4711 | test_loss: 1.4959 | test_acc: 0.8091
Epoch: 18 | train_loss: 1.5561 | train_acc: 0.5177 | test_loss: 1.4567 | test_acc: 0.8103
Epoch: 19 | train_loss: 1.5399 | train_acc: 0.5085 | test_loss: 1.4218 | test_acc: 0.8206
Epoch: 20 | train_loss: 1.4949 | train_acc: 0.5316 | test_loss: 1.3824 | test_acc: 0.8176
total training time: 18.989 sec.



- training model_tanh_3_layers with dropouts with p=0.25

20/20 [00:18<00:00, 1.06it/s]
Epoch: 1 | train_loss: 2.2041 | train_acc: 0.1766 | test_loss: 2.0240 | test_acc: 0.4626
Epoch: 2 | train_loss: 1.9293 | train_acc: 0.4217 | test_loss: 1.6979 | test_acc: 0.7929
Epoch: 3 | train_loss: 1.6306 | train_acc: 0.6694 | test_loss: 1.4475 | test_acc: 0.8288
Epoch: 4 | train_loss: 1.3794 | train_acc: 0.7674 | test_loss: 1.1951 | test_acc: 0.8886
Epoch: 5 | train_loss: 1.2025 | train_acc: 0.8465 | test_loss: 1.0015 | test_acc: 0.9052
Epoch: 6 | train_loss: 1.0445 | train_acc: 0.8635 | test_loss: 0.8381 | test_acc: 0.9054
Epoch: 7 | train_loss: 0.9196 | train_acc: 0.8615 | test_loss: 0.6801 | test_acc: 0.9177
Epoch: 8 | train_loss: 0.8072 | train_acc: 0.8894 | test_loss: 0.5744 | test_acc: 0.9110
Epoch: 9 | train_loss: 0.6783 | train_acc: 0.8666 | test_loss: 0.4788 | test_acc: 0.9197
Epoch: 10 | train_loss: 0.6227 | train_acc: 0.8855 | test_loss: 0.4112 | test_acc: 0.9255
Epoch: 11 | train_loss: 0.5458 | train_acc: 0.9132 | test_loss: 0.3378 | test_acc: 0.9169
Epoch: 12 | train_loss: 0.4837 | train_acc: 0.9004 | test_loss: 0.3003 | test_acc: 0.9409
Epoch: 13 | train_loss: 0.4540 | train_acc: 0.9127 | test_loss: 0.2649 | test_acc: 0.9371
Epoch: 14 | train_loss: 0.4149 | train_acc: 0.8867 | test_loss: 0.2252 | test_acc: 0.9428
Epoch: 15 | train_loss: 0.3800 | train_acc: 0.9118 | test_loss: 0.1877 | test_acc: 0.9436
Epoch: 16 | train_loss: 0.3556 | train_acc: 0.9221 | test_loss: 0.1756 | test_acc: 0.9433
Epoch: 17 | train_loss: 0.3043 | train_acc: 0.9218 | test_loss: 0.1473 | test_acc: 0.9192
Epoch: 18 | train_loss: 0.2791 | train_acc: 0.9211 | test_loss: 0.1286 | test_acc: 0.9611
Epoch: 19 | train_loss: 0.2732 | train_acc: 0.9387 | test_loss: 0.1154 | test_acc: 0.9477
Epoch: 20 | train_loss: 0.2502 | train_acc: 0.9232 | test_loss: 0.1021 | test_acc: 0.9460
total training time: 19.000 sec.





- gradient check

the execution will throw an error if grad. was not calculated corrected !!!
100%
2/2 [00:02<00:00, 1.01s/it]
<ipython-input-341-7f6d21f4fb37>:45: VisibleDeprecationWarning: Creating an ndarray from ragged nested sequences (which is a list-or-tuple of lists-or-tuples-or ndarrays with different lengths or shapes) is deprecated. If you meant to do this, you must specify 'dtype=object' when creating the ndarray.
theta = np.array([i.grad.detach().numpy() for i in model.parameters()])
Epoch: 1 | train_loss: 2.3169 | train_acc: 0.1095 | test_loss: 2.2882 | test_acc: 0.1623
<ipython-input-341-7f6d21f4fb37>:89: VisibleDeprecationWarning: Creating an ndarray from ragged nested sequences (which is a list-or-tuple of lists-or-tuples-or ndarrays with different lengths or shapes) is deprecated. If you meant to do this, you must specify 'dtype=object' when creating the ndarray.
actual_grad = np.array([i.grad.detach().numpy() for i in model.parameters()])
<ipython-input-341-7f6d21f4fb37>:92: DeprecationWarning: elementwise comparison failed; this will raise an error in the future.
dm = (np.average([np.linalg.norm(i) in actual_grad]) + np.average([np.linalg.norm(i) for i in grad_approx]))/2
Epoch: 2 | train_loss: 2.2834 | train_acc: 0.1488 | test_loss: 2.2426 | test_acc: 0.2233
total training time: 2.019 sec.

NOTE 1:- The Gradient Check will through exception if the gradient was not calculated correctly.
NOTE 2:- Numpy warnings are present because of the use of deprecated functions.

References:

https://pytorch.org/docs/stable/index.html
https://numpy.org/doc/stable/reference/index.html
https://torchmetrics.readthedocs.io/en/stable/
https://stackoverflow.com/
https://www.youtube.com/watch?v=Z_ikDlimN6A&t=10644s
Class ppts.

Links:

Question 1:-
https://colab.research.google.com/drive/1Xu90hjxQifbjEyGDxyWB_P4ZoOon3Zh2?usp=sharing

Question 2:-
https://colab.research.google.com/drive/1LQoj5RWZzDFvEp44ziv34mDjKWjGYa_Y?usp=sharing