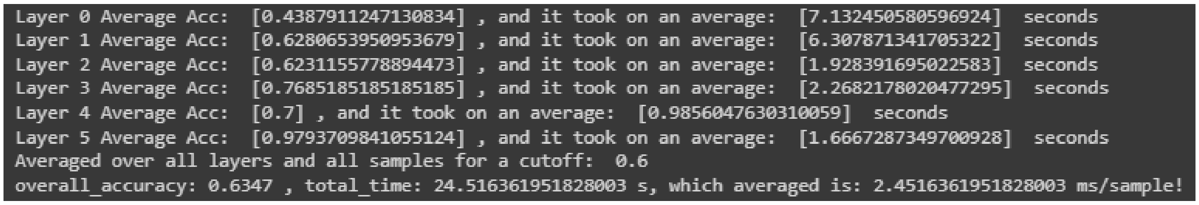
Assignment 3 - Dynamic Network Inference

Task 1

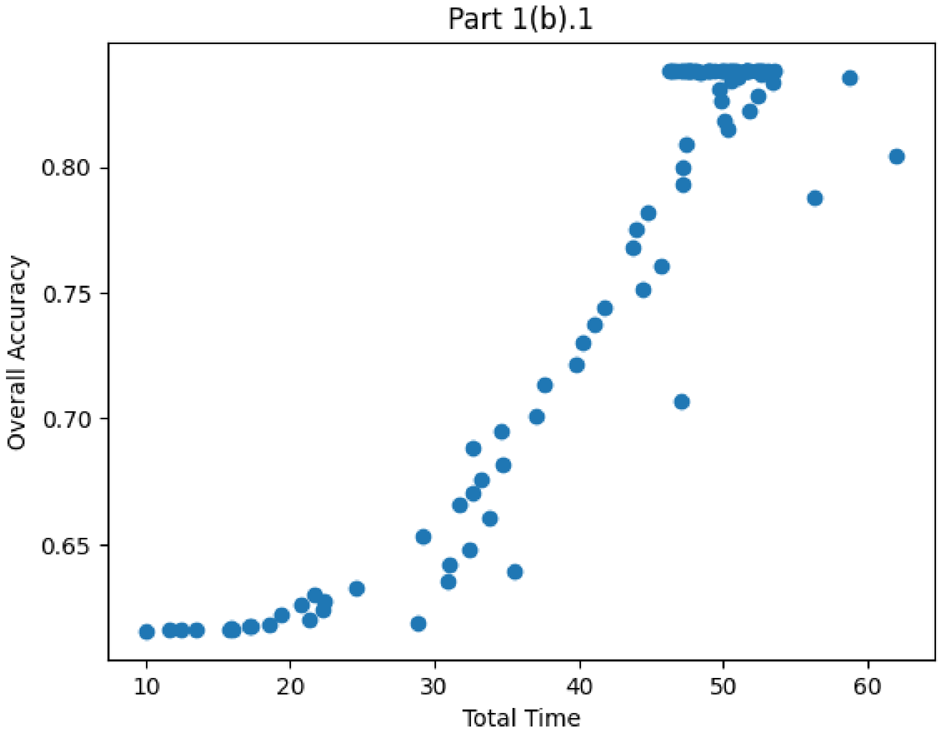
(a)

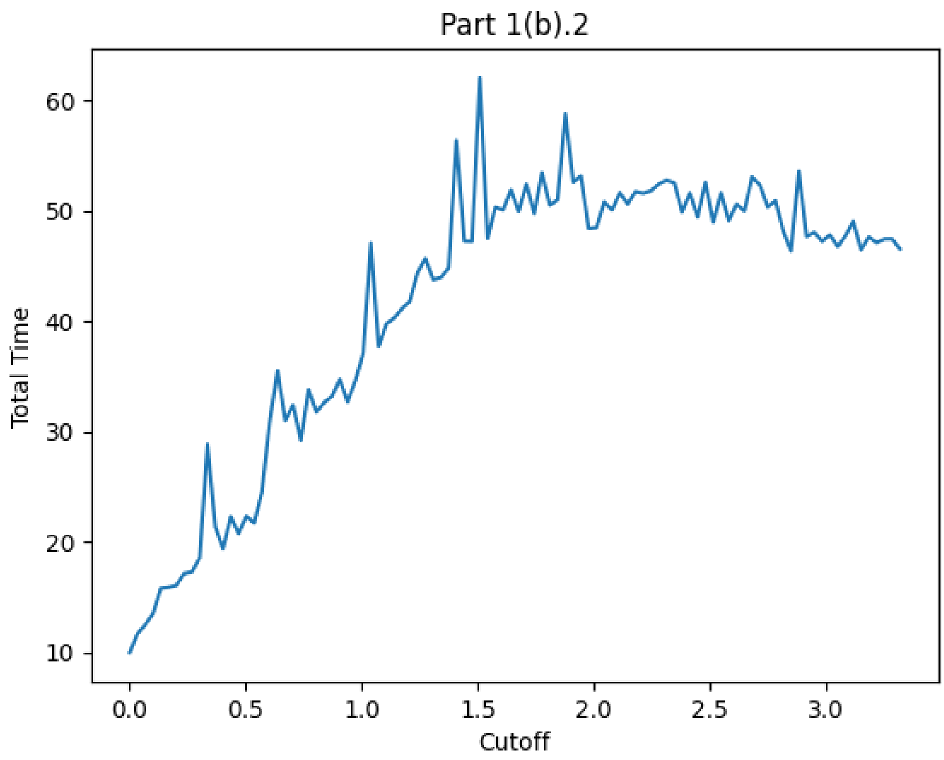


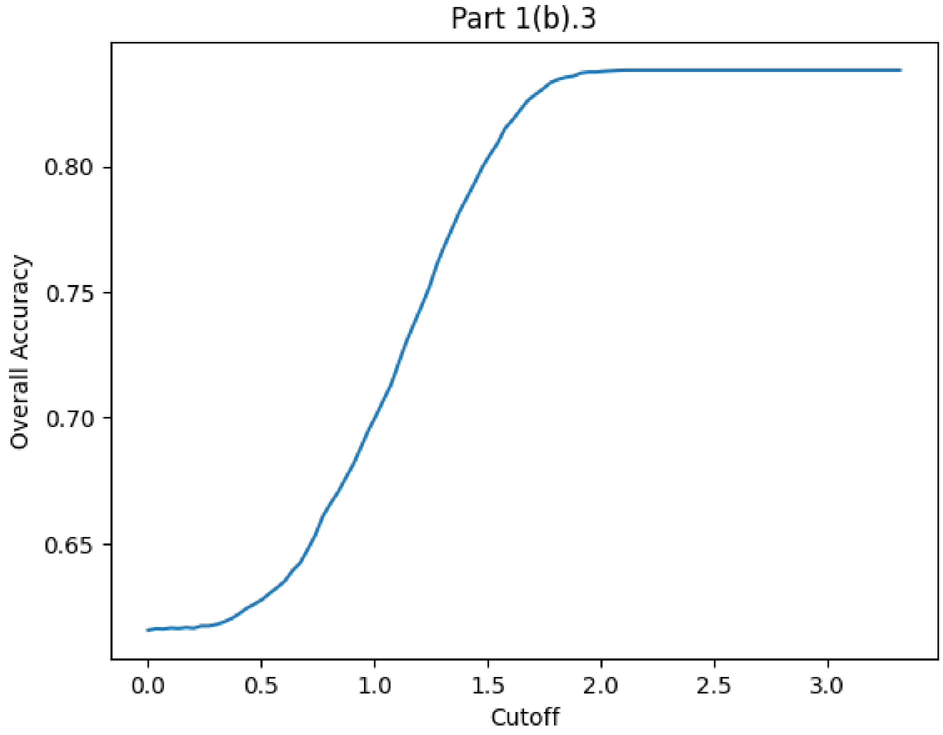
```
Layer 0 Average Acc:  [0.4387911247130834] , and it took on an average:  [7.132450580596924]  seconds
Layer 1 Average Acc:  [0.6280653950953679] , and it took on an average:  [6.307871341705322]  seconds
Layer 2 Average Acc:  [0.6231155778894473] , and it took on an average:  [1.928391695022583]  seconds
Layer 3 Average Acc:  [0.7685185185185185] , and it took on an average:  [2.2682178020477295]  seconds
Layer 4 Average Acc:  [0.7] , and it took on an average:  [0.9856047630310059]  seconds
Layer 5 Average Acc:  [0.9793709841055124] , and it took on an average:  [1.6667287349700928]  seconds
Averaged over all layers and all samples for a cutoff:  0.6
overall_accuracy: 0.6347 , total_time: 24.516361951828003 s, which averaged is: 2.4516361951828003 ms/sample!
```

Kindly note that the

(b)



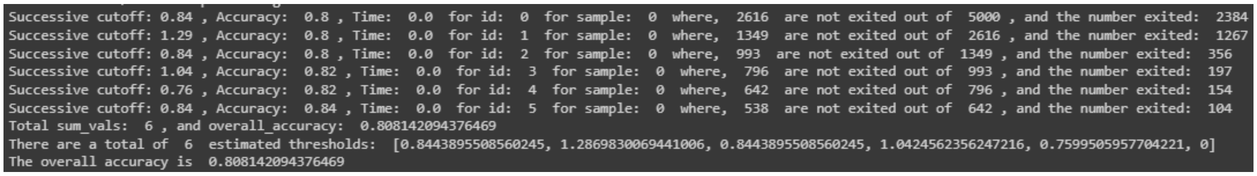




According to me, based on above, a threshold of ~2 works best for the least time and accuracy above 0.8 - so it’s a sweet spot.

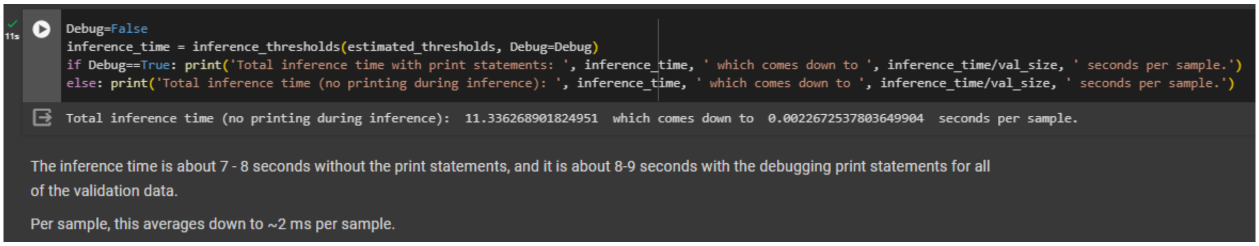
Task 2

(a)



```
Successive cutoff: 0.84 , Accuracy:  0.8 , Time:  0.0  for id:  0  for sample:  0  where,  2616  are not exited out of  5000 , and the number exited:  2384
Successive cutoff: 1.29 , Accuracy:  0.8 , Time:  0.0  for id:  1  for sample:  0  where,  1349  are not exited out of  2616 , and the number exited:  1267
Successive cutoff: 0.84 , Accuracy:  0.8 , Time:  0.0  for id:  2  for sample:  0  where,  993  are not exited out of  1349 , and the number exited:  356
Successive cutoff: 1.04 , Accuracy:  0.82 , Time:  0.0  for id:  3  for sample:  0  where,  796  are not exited out of  993 , and the number exited:  197
Successive cutoff: 0.76 , Accuracy:  0.82 , Time:  0.0  for id:  4  for sample:  0  where,  642  are not exited out of  796 , and the number exited:  154
Successive cutoff: 0.84 , Accuracy:  0.84 , Time:  0.0  for id:  5  for sample:  0  where,  538  are not exited out of  642 , and the number exited:  104
Total sum_vals:  6 , and overall_accuracy:  0.808142094376469
There are a total of  6  estimated thresholds:  [0.8443895508560245, 1.2869830069441006, 0.8443895508560245, 1.0424562356247216, 0.7599505957704221, 0]
The overall accuracy is  0.808142094376469
```

Please note that while a cutoff value is calculated for the final layer, a cutoff value of 0 is eventually used so that all the remaining samples can exit.



```
Debug=False
inference_time = inference_thresholds(estimated_thresholds, Debug=Debug)
if Debug==True: print('Total inference time with print statements: ', inference_time, ' which comes down to ', inference_time/val_size, ' seconds per sample.')
else: print('Total inference time (no printing during inference): ', inference_time, ' which comes down to ', inference_time/val_size, ' seconds per sample.')
```

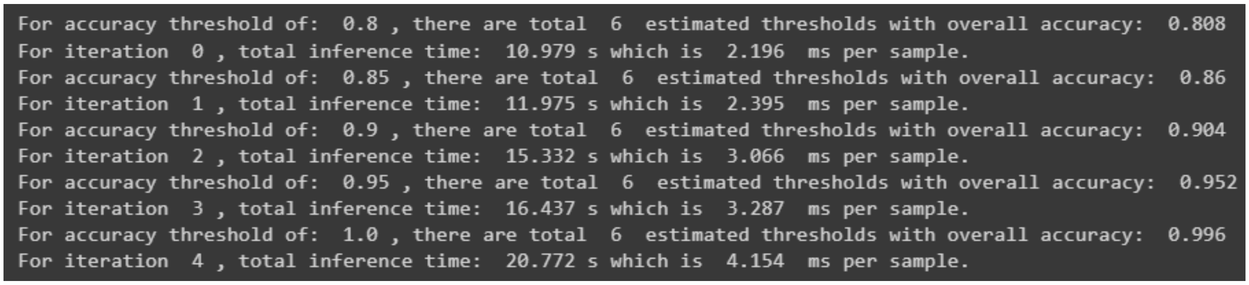
```
Total inference time (no printing during inference):  11.336268901824951  which comes down to  0.0022672537803649904  seconds per sample.
```

The inference time is about 7 - 8 seconds without the print statements, and it is about 8-9 seconds with the debugging print statements for all of the validation data.

Per sample, this averages down to ~2 ms per sample.

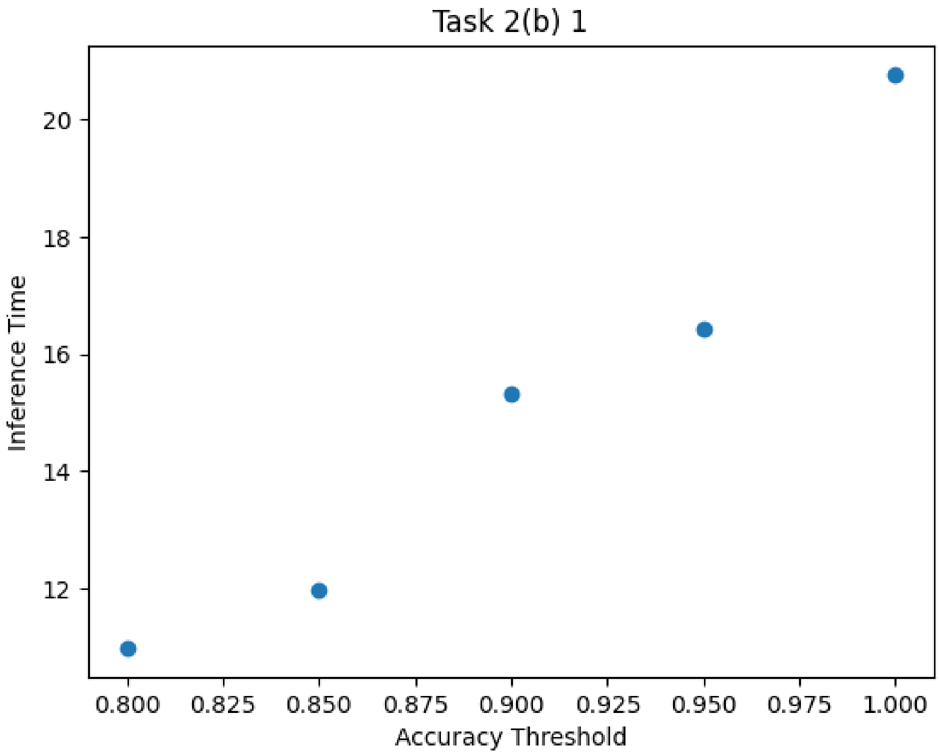
(b)

I choose 5 iterations between 0.8 and 1 to generate a plot of the same. The results for the different iterations are -

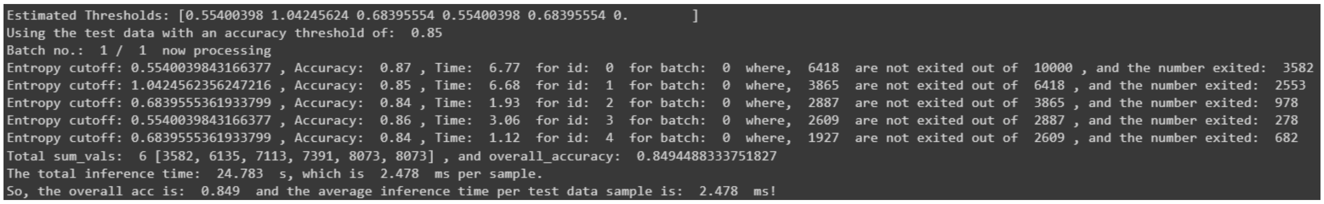


```
For accuracy threshold of:  0.8 , there are total  6  estimated thresholds with overall accuracy:  0.808
For iteration  0 , total inference time:  10.979 s which is  2.196  ms per sample.
For accuracy threshold of:  0.85 , there are total  6  estimated thresholds with overall accuracy:  0.86
For iteration  1 , total inference time:  11.975 s which is  2.395  ms per sample.
For accuracy threshold of:  0.9 , there are total  6  estimated thresholds with overall accuracy:  0.904
For iteration  2 , total inference time:  15.332 s which is  3.066  ms per sample.
For accuracy threshold of:  0.95 , there are total  6  estimated thresholds with overall accuracy:  0.952
For iteration  3 , total inference time:  16.437 s which is  3.287  ms per sample.
For accuracy threshold of:  1.0 , there are total  6  estimated thresholds with overall accuracy:  0.996
For iteration  4 , total inference time:  20.772 s which is  4.154  ms per sample.
```

Based on this, I choose a cutoff of 0.85 because while the values are increasing somewhat linearly, there is a slight drop for that value. And using the best threshold which I chose to be 0.85 accuracy, the accuracy and inference time on the test data is -



For this best accuracy, the overall accuracy is ~0.85, and the total inference time: 24.783 s, which is 2.478 ms per sample.



```
Estimated Thresholds: [0.55400398 1.04245624 0.68395554 0.55400398 0.68395554 0.        ]
Using the test data with an accuracy threshold of:  0.85
Batch no.:  1 /  1  now processing
Entropy cutoff: 0.5540039843166377 , Accuracy:  0.87 , Time:  6.77  for id:  0  for batch:  0  where,  6418  are not exited out of  10000 , and the number exited:  3582
Entropy cutoff: 1.0424562356247216 , Accuracy:  0.85 , Time:  6.68  for id:  1  for batch:  0  where,  3865  are not exited out of  6418 , and the number exited:  2553
Entropy cutoff: 0.6839555361933799 , Accuracy:  0.84 , Time:  1.93  for id:  2  for batch:  0  where,  2887  are not exited out of  3865 , and the number exited:  978
Entropy cutoff: 0.5540039843166377 , Accuracy:  0.86 , Time:  3.06  for id:  3  for batch:  0  where,  2609  are not exited out of  2887 , and the number exited:  278
Entropy cutoff: 0.6839555361933799 , Accuracy:  0.84 , Time:  1.12  for id:  4  for batch:  0  where,  1927  are not exited out of  2609 , and the number exited:  682
Total sum_vals:  6 [3582, 6135, 7113, 7391, 8073, 8073] , and overall_accuracy:  0.8494488333751827
The total inference time:  24.783  s, which is  2.478  ms per sample.
So, the overall acc is:  0.849  and the average inference time per test data sample is:  2.478  ms!
```