COL774
Assignment 3

Vaibhav Seth

October 2023

Contents

1 Decision Trees

1.1 Decision Tree Construction and Testing

Upon experimenting with Trees of Maximum depth from the the set 5,10,15,20,25, the following results were obtained:

The following accuracy vs max depth plot was obtained:

	precision	recall	f1-score	support
0	0.914127	0.848984	0.880352	3887.000000
1	0.860802	0.921320	0.890033	3940.000000
accuracy	0.885397	0.885397	0.885397	0.885397
macro avg	0.887464	0.885152	0.885193	7827.000000
weighted avg	0.887284	0.885397	0.885225	7827.000000

(a) Metrics on training data for tree of depth 5

	precision	recall	f1-score	support
0	0.591611	0.550308	0.570213	487.000000
1	0.573930	0.614583	0.593561	480.000000
accuracy	0.582213	0.582213	0.582213	0.582213
macro avg	0.582771	0.582446	0.581887	967.000000
weighted avg	0.582835	0.582213	0.581803	967.000000

(b) Metrics on testing data for tree of depth 5

Figure 1: Testing and Training metrics for various max depth = 5

	precision	recall	f1-score	support
0	1.000000	0.993054	0.996515	3887.00000
1	0.993194	1.000000	0.996585	3940.00000
accuracy	0.996550	0.996550	0.996550	0.99655
macro avg	0.996597	0.996527	0.996550	7827.00000
weighted avg	0.996574	0.996550	0.996550	7827.00000

(a) Metrics on training data for tree of depth 10

	precision	recall	f1-score	support
0	0.603774	0.591376	0.597510	487.000000
1	0.593878	0.606250	0.600000	480.000000
accuracy	0.598759	0.598759	0.598759	0.598759
macro avg	0.598826	0.598813	0.598755	967.000000
weighted avg	0.598861	0.598759	0.598746	967.000000

(b) Metrics on testing data for tree of depth 10

Figure 2: Testing and Training metrics for various max depth = 10

	precision	recall	f1-score	support
0	1.000000	0.994340	0.997162	3887.000000
1	0.994447	1.000000	0.997216	3940.000000
accuracy	0.997189	0.997189	0.997189	0.997189
macro avg	0.997224	0.997170	0.997189	7827.000000
weighted avg	0.997205	0.997189	0.997189	7827.000000

(a) Metrics on training data for tree of depth 15

	precision	recall	f1-score	support
0	0.602510	0.591376	0.596891	487.000000
1	0.593047	0.604167	0.598555	480.000000
accuracy	0.597725	0.597725	0.597725	0.597725
macro avg	0.597779	0.597771	0.597723	967.000000
weighted avg	0.597813	0.597725	0.597717	967.000000

(b) Metrics on testing data for tree of depth 15

Figure 3: Testing and Training metrics for various max depth = 15

	precision	recall	f1-score	support
0	1.000000	0.994340	0.997162	3887.000000
1	0.994447	1.000000	0.997216	3940.000000
accuracy	0.997189	0.997189	0.997189	0.997189
macro avg	0.997224	0.997170	0.997189	7827.000000
weighted avg	0.997205	0.997189	0.997189	7827.000000

(a) Metrics on training data for tree of depth 20

	precision	recall	f1-score	support
0	0.602510	0.591376	0.596891	487.000000
1	0.593047	0.604167	0.598555	480.000000
accuracy	0.597725	0.597725	0.597725	0.597725
macro avg	0.597779	0.597771	0.597723	967.000000
weighted avg	0.597813	0.597725	0.597717	967.000000

(b) Metrics on testing data for tree of depth 20

Figure 4: Testing and Training metrics for various max depth = 20

	precision	recall	f1-score	support
0	0.914127	0.848984	0.880352	3887.000000
1	0.860802	0.921320	0.890033	3940.000000
accuracy	0.885397	0.885397	0.885397	0.885397
macro avg	0.887464	0.885152	0.885193	7827.000000
weighted avg	0.887284	0.885397	0.885225	7827.000000

(a) Metrics on training data for tree of depth 25

	precision	recall	f1-score	support
0	0.602510	0.591376	0.596891	487.000000
1	0.593047	0.604167	0.598555	480.000000
accuracy	0.597725	0.597725	0.597725	0.597725
macro avg	0.597779	0.597771	0.597723	967.000000
weighted avg	0.597813	0.597725	0.597717	967.000000

(b) Metrics on testing data for tree of depth 25

Figure 5: Testing and Training metrics for various max depth = 25

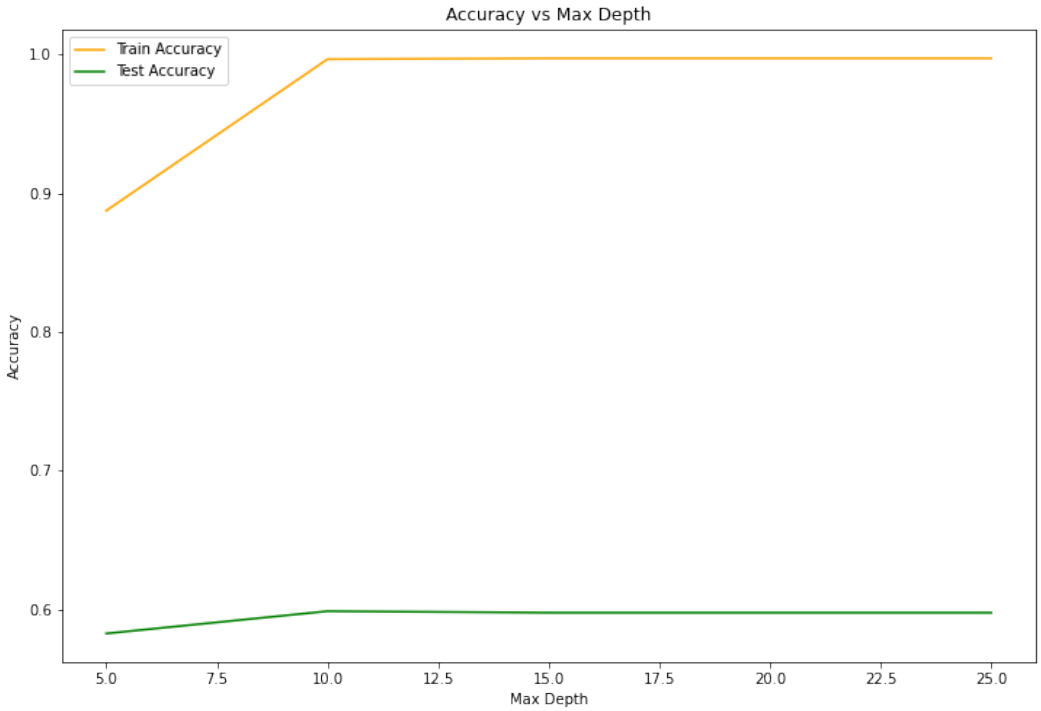


Figure 6: Plot of test and train accuracy vs maximum depth

We observe for large maximum depth the training and test accuracies become almost constant. This due to overfitting. It's a well known fact that decision trees are prone to overfitting and it can be seen here. As the max depth increases the model tries to fit parameters that are very specific to the dataset and as a result there is a massive difference between test and training accuracies. The accuracies plateau because the tree overfits sufficiently for smaller depths and as a result at larger depths there is not much new information left for the tree to work on.

```
tree.node_level_dict[12]
✓ 0.0s

[Depth = 12 | IS_LEAF : False and value : 1 | SPLIT ON : 0 and GAIN : 0.0,
 Depth = 12 | IS_LEAF : True and value : 1 | SPLIT ON : None and GAIN : None,
 Depth = 12 | IS_LEAF : True and value : 0 | SPLIT ON : None and GAIN : None,
 Depth = 12 | IS_LEAF : False and value : 1 | SPLIT ON : 0 and GAIN : 0.0,
 Depth = 12 | IS_LEAF : False and value : 1 | SPLIT ON : 0 and GAIN : 0.0,
 Depth = 12 | IS_LEAF : False and value : 1 | SPLIT ON : 0 and GAIN : 0.0,
 Depth = 12 | IS_LEAF : False and value : 1 | SPLIT ON : 0 and GAIN : 0.0,
 Depth = 12 | IS_LEAF : False and value : 1 | SPLIT ON : 0 and GAIN : 0.0,
 Depth = 12 | IS_LEAF : False and value : 1 | SPLIT ON : 0 and GAIN : 0.0,
 Depth = 12 | IS_LEAF : False and value : 1 | SPLIT ON : 0 and GAIN : 0.0,
 Depth = 12 | IS_LEAF : False and value : 1 | SPLIT ON : 0 and GAIN : 0.0,
 Depth = 12 | IS_LEAF : False and value : 1 | SPLIT ON : 0 and GAIN : 0.0,
 Depth = 12 | IS_LEAF : False and value : 1 | SPLIT ON : 0 and GAIN : 0.0]
```

Figure 7: Information Gains for different nodes at depth = 12 for a tree with max depth = 25

```
tree.node_level_dict[11]
✓ 0.0s

[Depth = 11 | IS_LEAF : True and value : 0 | SPLIT ON : None and GAIN : None,
 Depth = 11 | IS_LEAF : True and value : 1 | SPLIT ON : None and GAIN : None,
 Depth = 11 | IS_LEAF : True and value : 0 | SPLIT ON : None and GAIN : None,
 Depth = 11 | IS_LEAF : True and value : 1 | SPLIT ON : None and GAIN : None,
 Depth = 11 | IS_LEAF : True and value : 1 | SPLIT ON : None and GAIN : None,
 Depth = 11 | IS_LEAF : True and value : 0 | SPLIT ON : None and GAIN : None,
 Depth = 11 | IS_LEAF : False and value : 1 | SPLIT ON : 0 and GAIN : 0.0,
 Depth = 11 | IS_LEAF : False and value : 1 | SPLIT ON : 5 and GAIN : 1.0,
 Depth = 11 | IS_LEAF : True and value : 1 | SPLIT ON : None and GAIN : None,
 Depth = 11 | IS_LEAF : False and value : 1 | SPLIT ON : 0 and GAIN : 0.0,
 Depth = 11 | IS_LEAF : True and value : 0 | SPLIT ON : None and GAIN : None,
 Depth = 11 | IS_LEAF : True and value : 1 | SPLIT ON : None and GAIN : None,
 Depth = 11 | IS_LEAF : False and value : 1 | SPLIT ON : 0 and GAIN : 0.0,
 Depth = 11 | IS_LEAF : False and value : 1 | SPLIT ON : 0 and GAIN : 0.0,
 Depth = 11 | IS_LEAF : False and value : 1 | SPLIT ON : 0 and GAIN : 0.0,
 Depth = 11 | IS_LEAF : False and value : 1 | SPLIT ON : 0 and GAIN : 0.0,
 Depth = 11 | IS_LEAF : False and value : 1 | SPLIT ON : 0 and GAIN : 0.0,
 Depth = 11 | IS_LEAF : False and value : 1 | SPLIT ON : 0 and GAIN : 0.0,
 Depth = 11 | IS_LEAF : False and value : 1 | SPLIT ON : 0 and GAIN : 0.0,
 Depth = 11 | IS_LEAF : False and value : 1 | SPLIT ON : 0 and GAIN : 0.0,
 Depth = 11 | IS_LEAF : False and value : 1 | SPLIT ON : 0 and GAIN : 0.0]
```

Figure 8: Information Gains for different nodes at depth = 11 for a tree with max depth = 25

In the above figures we see that there is no information gain left at depth 12 , and at depth 11 there is just one node that offers any information gain. The same follows for all trees with depths

more than 11.

Only win and only loss accuracies :

Max depth	Only Win	Only Loss
5	57.39299610894941%	59.16114790286976%
10	59.38775510204082%	60.37735849056604%
15	59.30470347648262%	60.25104602510461%
20	59.30470347648262%	60.25104602510461%
25	59.30470347648262%	60.25104602510461%

1.2 Decision Tree One-Hot encoding

Upon experimenting with Trees of Maximum depth from the the set {15,25,35,45]}, the following results were obtained:

	precision	recall	f1-score	support
0	0.828488	0.513249	0.633836	3887.000000
1	0.650858	0.895178	0.753713	3940.000000
accuracy	0.705507	0.705507	0.705507	0.705507
macro avg	0.739673	0.704213	0.693775	7827.000000
weighted avg	0.739072	0.705507	0.694181	7827.000000

(a) Metrics on training data for tree of depth 15

	precision	recall	f1-score	support
0	0.619048	0.347023	0.444737	487.000000
1	0.541787	0.783333	0.640545	480.000000
accuracy	0.563599	0.563599	0.563599	0.563599
macro avg	0.580417	0.565178	0.542641	967.000000
weighted avg	0.580697	0.563599	0.541932	967.000000

(b) Metrics on testing data for tree of depth 15

Figure 9: Testing and Training metrics for various max depth = 15

	precision	recall	f1-score	support
0	0.901308	0.780036	0.836298	3887.000000
1	0.808425	0.915736	0.858741	3940.000000
accuracy	0.848345	0.848345	0.848345	0.848345
macro avg	0.854866	0.847886	0.847520	7827.000000
weighted avg	0.854552	0.848345	0.847596	7827.000000

(a) Metrics on training data for tree of depth 25

	precision	recall	f1-score	support
0	0.648379	0.533881	0.585586	487.000000
1	0.598940	0.706250	0.648184	480.000000
accuracy	0.619442	0.619442	0.619442	0.619442
macro avg	0.623659	0.620065	0.616885	967.000000
weighted avg	0.623838	0.619442	0.616658	967.000000

(b) Metrics on testing data for tree of depth 25

Figure 10: Testing and Training metrics for various max depth = 25

	precision	recall	f1-score	support
0	0.964749	0.880113	0.920490	3887.000000
1	0.891147	0.968274	0.928111	3940.000000
accuracy	0.924492	0.924492	0.924492	0.924492
macro avg	0.927948	0.924194	0.924300	7827.000000
weighted avg	0.927699	0.924492	0.924326	7827.000000

(a) Metrics on training data for tree of depth 35

	precision	recall	f1-score	support
0	0.634434	0.552361	0.590560	487.000000
1	0.598527	0.677083	0.635386	480.000000
accuracy	0.614271	0.614271	0.614271	0.614271
macro avg	0.616480	0.614722	0.612973	967.000000
weighted avg	0.616610	0.614271	0.612811	967.000000

(b) Metrics on testing data for tree of depth 35

Figure 11: Testing and Training metrics for various max depth = 35

	precision	recall	f1-score	support
0	0.994034	0.985850	0.989925	3887.000000
1	0.986153	0.994162	0.990142	3940.000000
accuracy	0.990034	0.990034	0.990034	0.990034
macro avg	0.990093	0.990006	0.990033	7827.000000
weighted avg	0.990067	0.990034	0.990034	7827.000000

(a) Metrics on training data for tree of depth 45

	precision	recall	f1-score	support
0	0.625000	0.585216	0.604454	487.000000
1	0.604697	0.643750	0.623613	480.000000
accuracy	0.614271	0.614271	0.614271	0.614271
macro avg	0.614848	0.614483	0.614033	967.000000
weighted avg	0.614922	0.614271	0.613964	967.000000

(b) Metrics on testing data for tree of depth 45

Figure 12: Testing and Training metrics for various max depth = 45

	precision	recall	f1-score	support
0	0.999484	0.996398	0.997939	3887.000000
1	0.996457	0.999492	0.997973	3940.000000
accuracy	0.997956	0.997956	0.997956	0.997956
macro avg	0.997971	0.997945	0.997956	7827.000000
weighted avg	0.997960	0.997956	0.997956	7827.000000

(a) Metrics on training data for tree of depth 55

	precision	recall	f1-score	support
0	0.629386	0.589322	0.608696	487.000000
1	0.608611	0.647917	0.627649	480.000000
accuracy	0.618407	0.618407	0.618407	0.618407
macro avg	0.618998	0.618620	0.618172	967.000000
weighted avg	0.619073	0.618407	0.618104	967.000000

(b) Metrics on testing data for tree of depth 55

Figure 13: Testing and Training metrics for various max depth = 55

	precision	recall	f1-score	support
0	0.999484	0.996398	0.997939	3887.000000
1	0.996457	0.999492	0.997973	3940.000000
accuracy	0.997956	0.997956	0.997956	0.997956
macro avg	0.997971	0.997945	0.997956	7827.000000
weighted avg	0.997960	0.997956	0.997956	7827.000000

(a) Metrics on training data for tree of depth 75

	precision	recall	f1-score	support
0	0.629386	0.589322	0.608696	487.000000
1	0.608611	0.647917	0.627649	480.000000
accuracy	0.618407	0.618407	0.618407	0.618407
macro avg	0.618998	0.618620	0.618172	967.000000
weighted avg	0.619073	0.618407	0.618104	967.000000

(b) Metrics on testing data for tree of depth 75

Figure 14: Testing and Training metrics for various max depth = 75

	precision	recall	f1-score	support
0	0.999484	0.996398	0.997939	3887.000000
1	0.996457	0.999492	0.997973	3940.000000
accuracy	0.997956	0.997956	0.997956	0.997956
macro avg	0.997971	0.997945	0.997956	7827.000000
weighted avg	0.997960	0.997956	0.997956	7827.000000

(a) Metrics on training data for tree of depth 100

	precision	recall	f1-score	support
0	0.629386	0.589322	0.608696	487.000000
1	0.608611	0.647917	0.627649	480.000000
accuracy	0.618407	0.618407	0.618407	0.618407
macro avg	0.618998	0.618620	0.618172	967.000000
weighted avg	0.619073	0.618407	0.618104	967.000000

(b) Metrics on testing data for tree of depth 100

Figure 15: Testing and Training metrics for various max depth = 100

The following accuracy vs max depth plot was obtained:

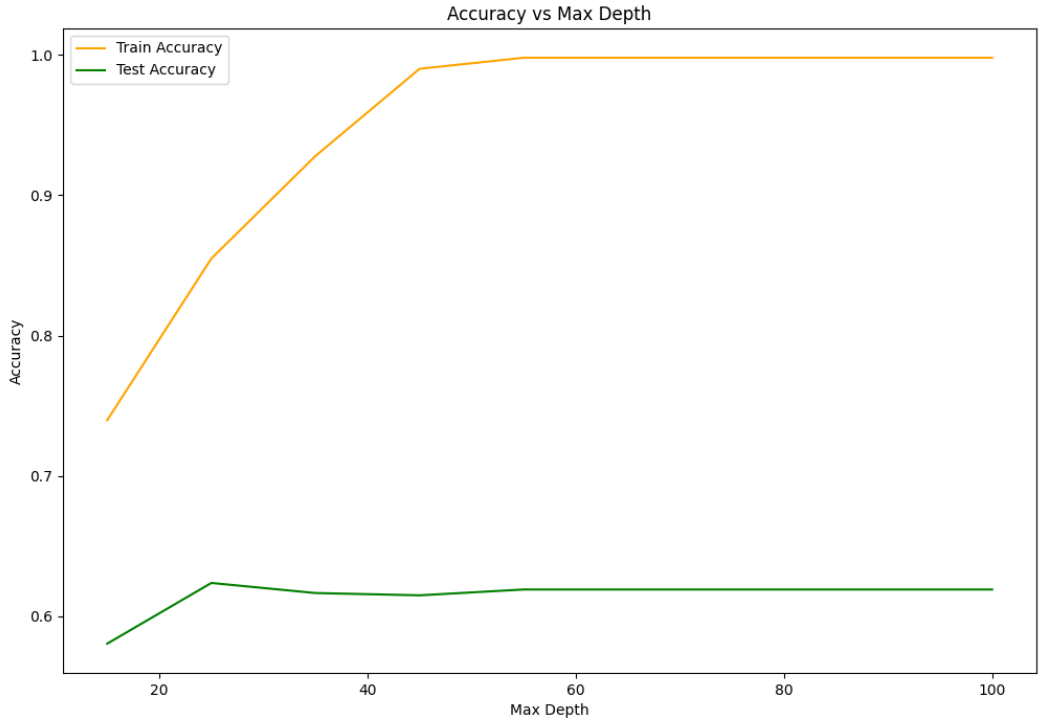


Figure 16: Plot of test and train accuracy vs maximum depth

The training accuracies decrease when compared to part a, but the testing accuracies increase. This could be due to introduction of new features, the model has more dimensions to learn, but becuase the max depth is still quite restricted the model prevents overfitting (as can be seen from upward sloping accuracies). As a result, the model performs nicely on testing data.

In a similar fashion to the previous part we see that the train and test accuracy plateau. The reason is same.

1.3 Decision Tree Post Pruning

The pruning algorithm has been implemented in a slightly different manner. It is still a greedy algorithm, but instead of calculating accuracy for each node removal, we follow a layer-wise bottom-up approach. The first Nodes to be pruned are the ones in the level above the lowest level(which has only leaves). We remove all nodes that result in an increased accuracy over validation set in the level and then we move one level up and do the same. We keep doing this process till we reach the root node. The approach might prove to be ineffective if there's a limit on number of nodes that can be removed, but if considering removing all nodes that can be removed, then this approach works.

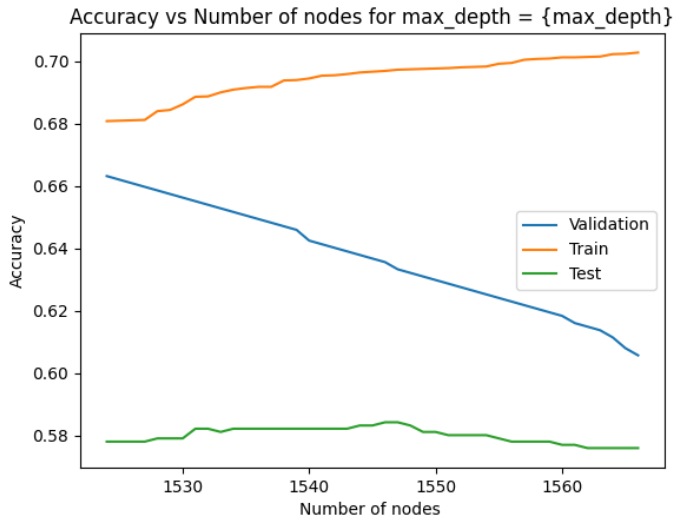


Figure 17: Plot of accuracies vs number of nodes in the tree for depth = 15

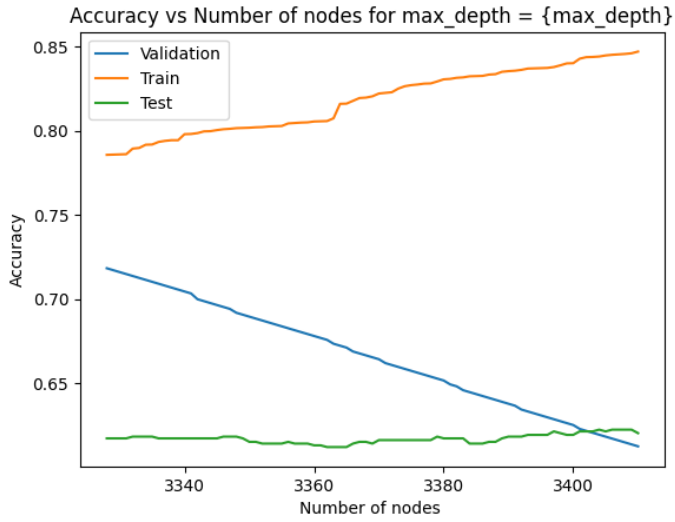


Figure 18: Plot of accuracies vs number of nodes in the tree for depth = 25

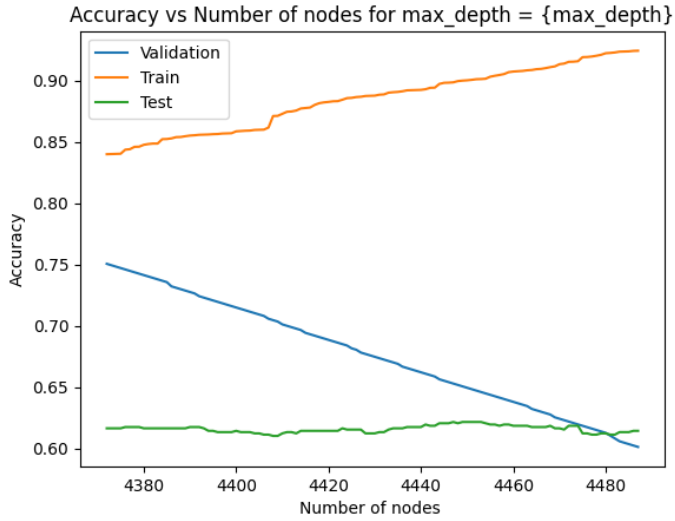


Figure 19: Plot of accuracies vs number of nodes in the tree for depth = 35

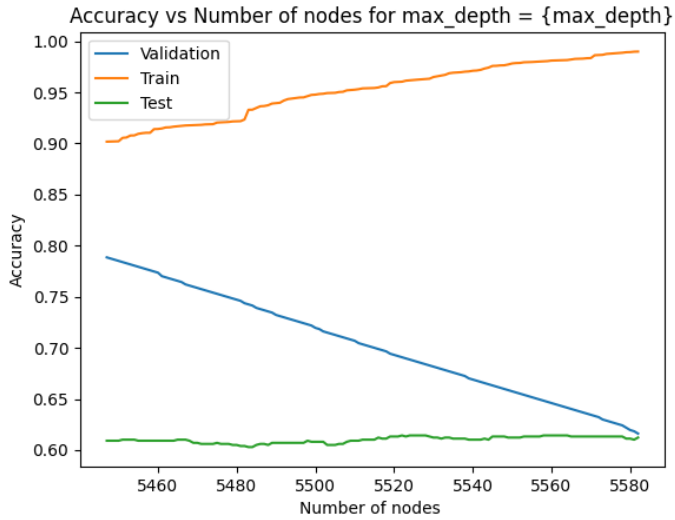


Figure 20: Plot of accuracies vs number of nodes in the tree for depth = 45

We can see as we remove more nodes the overfitting reduces and validation accuracy increases, but the test accuracy isn't improved. A possible reason could be that because we are doing pruning till accuracies are increasing, we might be somewhat overfitting on the validation set. Hence an eraly stopping condition could improve test accuracies.

1.4 Decision Tree sci-kit learn

Accuracies and plots :

	precision	recall	f1-score	support
0	0.815597	0.546180	0.654237	3887.0000
1	0.662328	0.878173	0.755129	3940.0000
accuracy	0.713300	0.713300	0.713300	0.7133
macro avg	0.738963	0.712176	0.704683	7827.0000
weighted avg	0.738444	0.713300	0.705025	7827.0000

(a) Metrics on training data for tree of depth 15

	precision	recall	f1-score	support
0	0.661585	0.445585	0.532515	487.000000
1	0.577465	0.768750	0.659517	480.000000
accuracy	0.605998	0.605998	0.605998	0.605998
macro avg	0.619525	0.607168	0.596016	967.000000
weighted avg	0.619830	0.605998	0.595557	967.000000

(b) Metrics on testing data for tree of depth 15

Figure 21: Testing and Training metrics for various max depth = 15

	precision	recall	f1-score	support
0	0.916465	0.779007	0.842164	3887.000000
1	0.810082	0.929949	0.865887	3940.000000
accuracy	0.854989	0.854989	0.854989	0.854989
macro avg	0.863273	0.854478	0.854025	7827.000000
weighted avg	0.862913	0.854989	0.854106	7827.000000

(a) Metrics on training data for tree of depth 25

	precision	recall	f1-score	support
0	0.679790	0.531828	0.596774	487.000000
1	0.610922	0.745833	0.671670	480.000000
accuracy	0.638056	0.638056	0.638056	0.638056
macro avg	0.645356	0.638830	0.634222	967.000000
weighted avg	0.645605	0.638056	0.633951	967.000000

(b) Metrics on testing data for tree of depth 25

Figure 22: Testing and Training metrics for various max depth = 25

	precision	recall	f1-score	support
0	0.961004	0.925650	0.942996	3887.000000
1	0.929219	0.962944	0.945781	3940.000000
accuracy	0.944423	0.944423	0.944423	0.944423
macro avg	0.945111	0.944297	0.944388	7827.000000
weighted avg	0.945004	0.944423	0.944398	7827.000000

(a) Metrics on training data for tree of depth 35

	precision	recall	f1-score	support
0	0.665939	0.626283	0.645503	487.000000
1	0.642436	0.681250	0.661274	480.000000
accuracy	0.653568	0.653568	0.653568	0.653568
macro avg	0.654188	0.653767	0.653388	967.000000
weighted avg	0.654273	0.653568	0.653331	967.000000

(b) Metrics on testing data for tree of depth 35

Figure 23: Testing and Training metrics for various max depth = 35

	precision	recall	f1-score	support
0	0.992331	0.998714	0.995512	3887.000000
1	0.998723	0.992386	0.995544	3940.000000
accuracy	0.995528	0.995528	0.995528	0.995528
macro avg	0.995527	0.995550	0.995528	7827.000000
weighted avg	0.995549	0.995528	0.995528	7827.000000

(a) Metrics on training data for tree of depth 45

	precision	recall	f1-score	support
0	0.644491	0.636550	0.640496	487.000000
1	0.635802	0.643750	0.639752	480.000000
accuracy	0.640124	0.640124	0.640124	0.640124
macro avg	0.640147	0.640150	0.640124	967.000000
weighted avg	0.640178	0.640124	0.640126	967.000000

(b) Metrics on testing data for tree of depth 45

Figure 24: Testing and Training metrics for various max depth = 45

The following accuracy vs max depth plot was obtained:

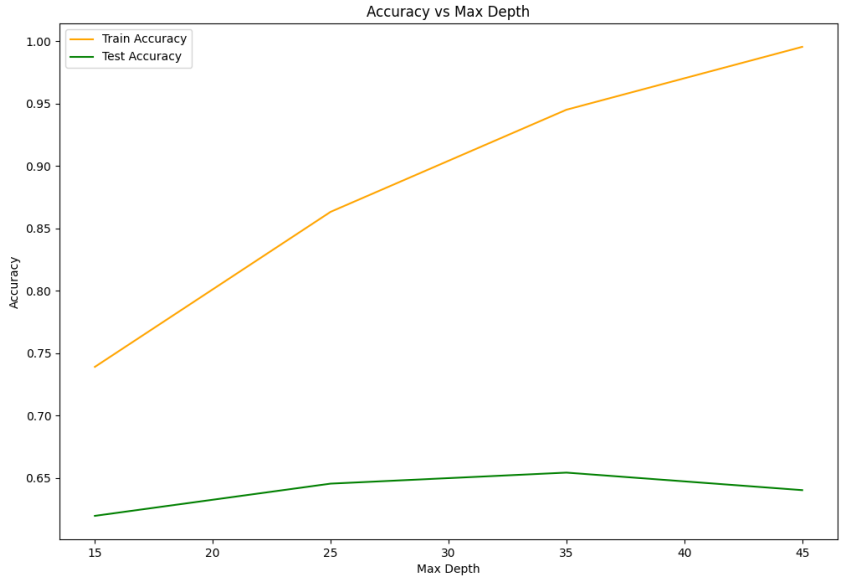


Figure 25: Plot of test and train accuracy vs maximum depth

The following accuracy vs max depth plot was obtained over validation data:

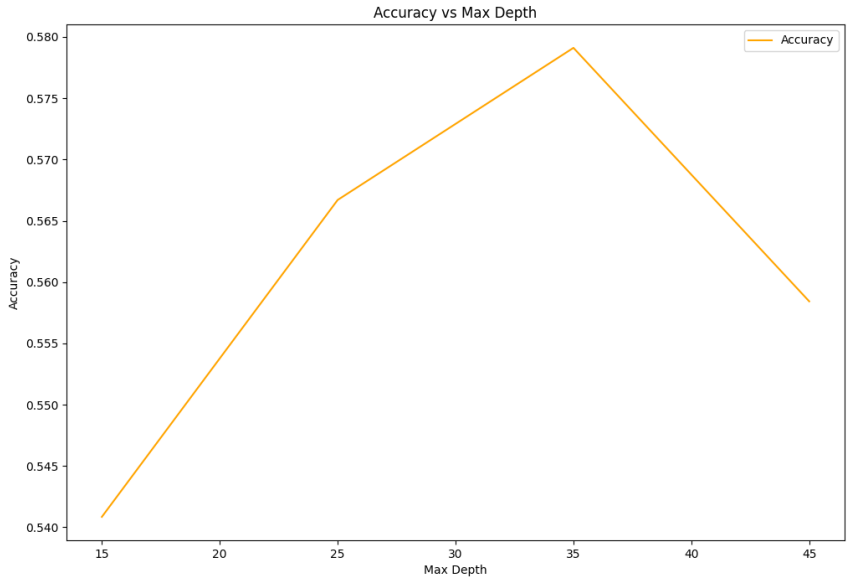


Figure 26: Plot of test and train accuracy vs maximum depth over val

We can clearly see that we get the best accuracy at max depth 35. The accuracy is around 0.579

The following accuracy vs ccp alpha plot was obtained over validation data:

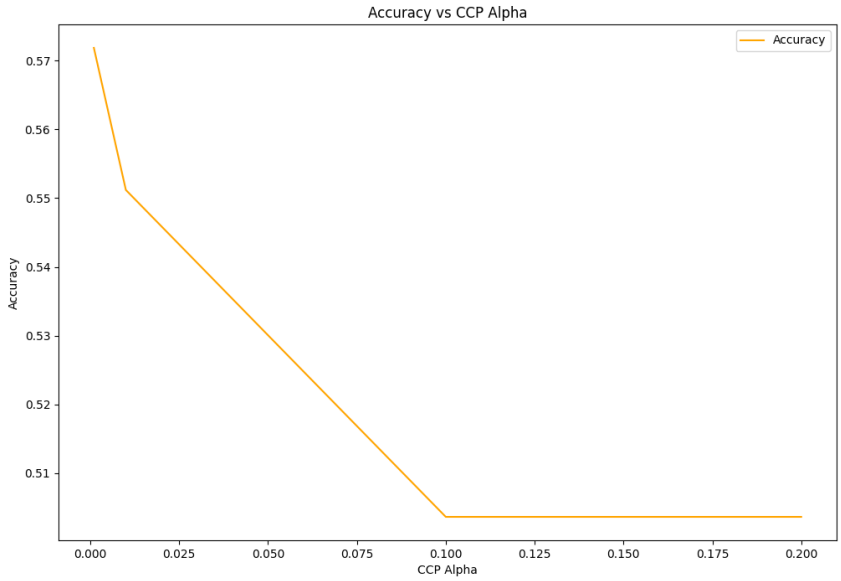


Figure 27: Plot of test and train accuracy vs ccp alpha over val

We can clearly see that we get the best accuracy at ccp alpha = 0.001. The accuracy is around 0.572

When comparing the two max depth models, we see that we get similar performance at max depth = 35. This could be because the model does not over fit or under fit too much.

When comparing ccp alpha and pruning models we find that our pruning model slightly outperformed the sk learn model over validation set. This could be because our pruning is unrestricted while sklearn has an upper bound decided by ccp alpha.

1.5 Random Forests

Best Parameters obtained after grid search with OOB score :

$$\begin{aligned} \text{max features} &= 0.9 \\ \text{min samples split} &= 8 \\ \text{n estimators} &= 250 \end{aligned} \tag{1}$$

The following results were obtained on the best model :

$$\begin{aligned} \text{Training Accuracy} &= 0.9793024147182828 \\ \text{Test Accuracy} &= 0.7280248190279214 \\ \text{Validation Accuracy} &= 0.7091954022988506 \\ \text{OOB Score} &= 0.721732464545803 \end{aligned} \tag{2}$$

The test accuracy is higher than all the decision tree models.

2 Neural Networks

2.1 Single Hidden Layer with sigmoid activation

The following stopping criteria were used:

$$\begin{aligned} \text{MAX\_EPOCH} &= 1000 \\ \text{Constant\_Threshold} &= 1e-7 \end{aligned} \tag{3}$$

If the number of epochs surpasses MAX_EPOCH, the training is terminated. If we detect 5 successive iterations where loss doesn't improve(Constant_Threshold) then we stop training. Also if there are 5 consecutive epoch for which loss increase we stop training and return the best model (least loss). A limit of thousand was set on epochs due to computational limitations. Training for further epoch could have led to improved accuracies over test and train.

Metrics for hidden layers : {1, 5, 10, 50, 100}

	Unnamed: 0	precision	recall	f1-score	support
0	0	0.903930	0.932432	0.917960	222.000
1	1	0.621212	0.754601	0.681440	163.000
2	2	0.447236	0.566879	0.500000	157.000
3	3	0.315508	0.412587	0.357576	143.000
4	4	0.967914	0.574603	0.721116	315.000
5	accuracy	0.659000	0.659000	0.659000	0.659
6	macro avg	0.651160	0.648221	0.635618	1000.000
7	weighted avg	0.722157	0.659000	0.671647	1000.000

(a) Metrics for layer dim 1 on test data

	precision	recall	f1-score	support
0	0.904110	0.910112	0.907101	1958.000
1	0.649141	0.778654	0.708023	1649.000
2	0.489242	0.600629	0.539243	1590.000
3	0.320219	0.449650	0.374055	1430.000
4	0.949785	0.588793	0.726940	3373.000
accuracy	0.665000	0.665000	0.665000	0.665
macro avg	0.662499	0.665568	0.651072	10000.000
weighted avg	0.728011	0.665000	0.678790	10000.000

(b) Metrics for layer dim 1 on train data

Figure 28: Metrics for hidden layer : {1}

	Unnamed: 0	precision	recall	f1-score	support
0	0	0.921397	0.909483	0.915401	232.000
1	1	0.752525	0.693023	0.721550	215.000
2	2	0.703518	0.507246	0.589474	276.000
3	3	0.390374	0.496599	0.437126	147.000
4	4	0.550802	0.792308	0.649842	130.000
5	accuracy	0.676000	0.676000	0.676000	0.676
6	macro avg	0.663723	0.679732	0.662679	1000.000
7	weighted avg	0.698717	0.676000	0.678938	1000.000

(a) Metrics for layer dim {5} on test data

	precision	recall	f1-score	support
0	0.927955	0.897008	0.912219	2039.0000
1	0.798281	0.721994	0.758223	2187.0000
2	0.722336	0.529081	0.610786	2665.0000
3	0.433765	0.524699	0.474918	1660.0000
4	0.586801	0.846791	0.693220	1449.0000
accuracy	0.691600	0.691600	0.691600	0.6916
macro avg	0.693828	0.703914	0.689874	10000.0000
weighted avg	0.713329	0.691600	0.693884	10000.0000

(b) Metrics for layer dim {5} on train data

Figure 29: Metrics for hidden layer : {5}

	Unnamed: 0	precision	recall	f1-score	support
0	0	0.943231	0.939130	0.941176	230.000
1	1	0.712121	0.801136	0.754011	176.000
2	2	0.577889	0.649718	0.611702	177.000
3	3	0.609626	0.532710	0.568579	214.000
4	4	0.775401	0.714286	0.743590	203.000
5	accuracy	0.731000	0.731000	0.731000	0.731
6	macro avg	0.723654	0.727396	0.723812	1000.000
7	weighted avg	0.732429	0.731000	0.730072	1000.000

(a) Metrics for layer dim {10} on test data

	precision	recall	f1-score	support
0	0.958904	0.927835	0.943114	2037.0000
1	0.757836	0.834168	0.794172	1797.0000
2	0.614242	0.699942	0.654297	1713.0000
3	0.606574	0.599115	0.602821	2033.0000
4	0.858919	0.742149	0.796276	2420.0000
accuracy	0.760200	0.760200	0.760200	0.7602
macro avg	0.759295	0.760642	0.758136	10000.0000
weighted avg	0.767906	0.760200	0.762158	10000.0000

(b) Metrics for layer dim {10} on train data

Figure 30: Metrics for hidden layer : {10}

	Unnamed: 0	precision	recall	f1-score	support
0	0	0.973799	0.961207	0.967462	232.000
1	1	0.813131	0.800995	0.807018	201.000
2	2	0.648241	0.712707	0.678947	181.000
3	3	0.582888	0.615819	0.598901	177.000
4	4	0.812834	0.727273	0.767677	209.000
5	accuracy	0.774000	0.774000	0.774000	0.774
6	macro avg	0.766179	0.763600	0.764001	1000.000
7	weighted avg	0.779746	0.774000	0.776001	1000.000

(a) Metrics for layer dim {50} on test data

	precision	recall	f1-score	support
0	0.982243	0.937984	0.959603	2064.0000
1	0.837209	0.835520	0.836364	1982.0000
2	0.689549	0.725216	0.706933	1856.0000
3	0.558267	0.672869	0.610234	1666.0000
4	0.879005	0.755757	0.812735	2432.0000
accuracy	0.789700	0.789700	0.789700	0.7897
macro avg	0.789255	0.785469	0.785174	10000.0000
weighted avg	0.803431	0.789700	0.794358	10000.0000

(b) Metrics for layer dim {50} on train data

Figure 31: Metrics for hidden layer : {50}

	Unnamed: 0	precision	recall	f1-score	support
0	0	0.938865	0.968468	0.953437	222.000
1	1	0.818182	0.786408	0.801980	206.000
2	2	0.628141	0.714286	0.668449	175.000
3	3	0.663102	0.599034	0.629442	207.000
4	4	0.775401	0.763158	0.769231	190.000
5	accuracy	0.771000	0.771000	0.771000	0.771
6	macro avg	0.764738	0.766271	0.764508	1000.000
7	weighted avg	0.771486	0.771000	0.770298	1000.000

(a) Metrics for layer dim {100} on test data

	precision	recall	f1-score	support
0	0.948757	0.965909	0.957256	1936.0000
1	0.852376	0.830133	0.841108	2031.0000
2	0.674693	0.732073	0.702213	1799.0000
3	0.635956	0.639780	0.637862	1996.0000
4	0.837877	0.782842	0.809425	2238.0000
accuracy	0.790200	0.790200	0.790200	0.7902
macro avg	0.789932	0.790147	0.789573	10000.0000
weighted avg	0.792628	0.790200	0.790948	10000.0000

(b) Metrics for layer dim {100} on train data

Figure 32: Metrics for hidden layer : {100}

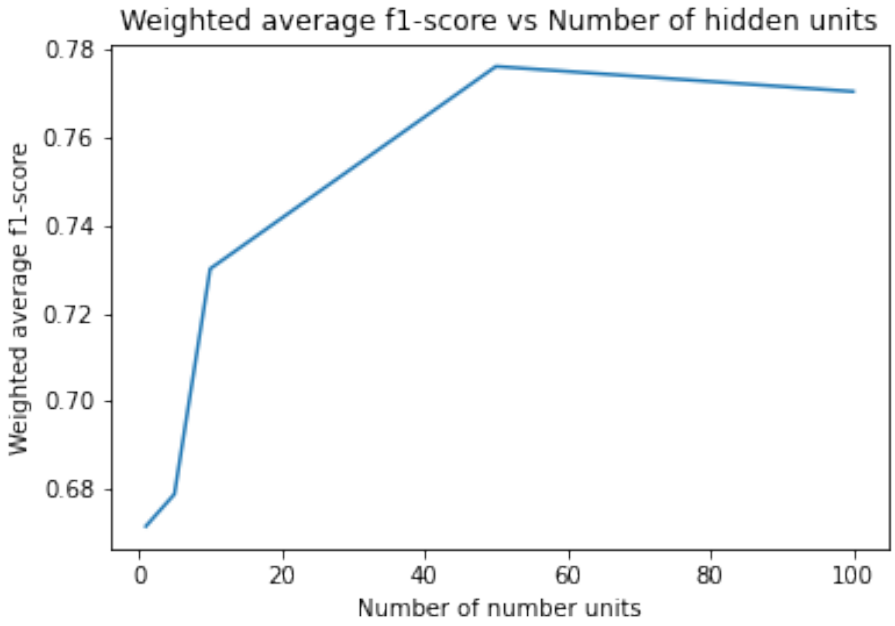


Figure 33: Plot of weighted average test F1 score vs number units

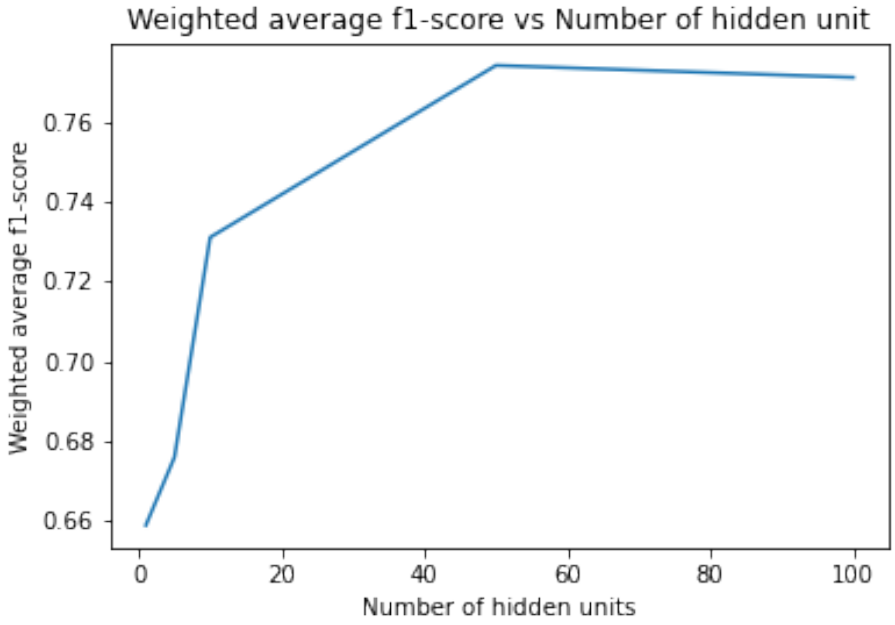


Figure 34: Plot of average test accuracy score vs number units

2.2 Multiple Hidden Layer Architecture

Metrics for hidden layers : {{512}, {512, 256}, {512, 256, 128}, {512, 256, 128, 64}}.

	Unnamed: 0	precision	recall	f1-score	support
0	0	0.947598	0.960177	0.953846	226.000
1	1	0.732323	0.805556	0.767196	180.000
2	2	0.673367	0.635071	0.653659	211.000
3	3	0.620321	0.563107	0.590331	206.000
4	4	0.721925	0.762712	0.741758	177.000
5	accuracy	0.747000	0.747000	0.747000	0.747
6	macro avg	0.739107	0.745324	0.741358	1000.000
7	weighted avg	0.743623	0.747000	0.744486	1000.000

(a) Metrics on test data

	precision	recall	f1-score	support
0	0.955860	0.944361	0.950076	1995.000
1	0.771486	0.839846	0.804216	1817.000
2	0.727459	0.671078	0.698132	2116.000
3	0.611554	0.628776	0.620045	1953.000
4	0.799617	0.789051	0.794299	2119.000
accuracy	0.773000	0.773000	0.773000	0.773
macro avg	0.773195	0.774622	0.773354	10000.000
weighted avg	0.773679	0.773000	0.772798	10000.000

(b) Metrics on train data

Figure 35: Metrics for hidden layer : {512}

	Unnamed: 0	precision	recall	f1-score	support
0	0	0.991266	0.978448	0.984816	232.000
1	1	0.838384	0.864583	0.851282	192.000
2	2	0.693467	0.770950	0.730159	179.000
3	3	0.786096	0.625532	0.696682	235.000
4	4	0.727273	0.839506	0.779370	162.000
5	accuracy	0.814000	0.814000	0.814000	0.814
6	macro avg	0.807297	0.815804	0.808462	1000.000
7	weighted avg	0.817625	0.814000	0.812600	1000.000

(a) Metrics on test data

	precision	recall	f1-score	support
0	0.991882	0.979459	0.985631	1996.0000
1	0.901416	0.916238	0.908767	1946.0000
2	0.746414	0.804528	0.774382	1811.0000
3	0.750498	0.638559	0.690018	2360.0000
4	0.758489	0.840488	0.797386	1887.0000
accuracy	0.828800	0.828800	0.828800	0.8288
macro avg	0.829740	0.835854	0.831237	10000.0000
weighted avg	0.828815	0.828800	0.827130	10000.0000

(b) Metrics on train data

Figure 36: Metrics for hidden layer : {512, 256}

	Unnamed: 0	precision	recall	f1-score	support
0	0	0.991266	0.970085	0.980562	234.000
1	1	0.909091	0.849057	0.878049	212.000
2	2	0.718593	0.764706	0.740933	187.000
3	3	0.631016	0.648352	0.639566	182.000
4	4	0.796791	0.805405	0.801075	185.000
5	accuracy	0.817000	0.817000	0.817000	0.817
6	macro avg	0.809352	0.807521	0.808037	1000.000
7	weighted avg	0.821312	0.817000	0.818752	1000.000

(a) Metrics on test data

	Unnamed: 0	precision	recall	f1-score	support
0	0	0.991266	0.970085	0.980562	234.000
1	1	0.909091	0.849057	0.878049	212.000
2	2	0.718593	0.764706	0.740933	187.000
3	3	0.631016	0.648352	0.639566	182.000
4	4	0.796791	0.805405	0.801075	185.000
5	accuracy	0.817000	0.817000	0.817000	0.817
6	macro avg	0.809352	0.807521	0.808037	1000.000
7	weighted avg	0.821312	0.817000	0.818752	1000.000

(b) Metrics on train data

Figure 37: Metrics for hidden layer : {512, 256, 128}

	Unnamed: 0	precision	recall	f1-score	support
0	0	1.000000	0.970339	0.984946	236.000
1	1	0.904040	0.895000	0.899497	200.000
2	2	0.753769	0.797872	0.775194	188.000
3	3	0.663102	0.666667	0.664879	186.000
4	4	0.818182	0.805263	0.811671	190.000
5	accuracy	0.835000	0.835000	0.835000	0.835
6	macro avg	0.827819	0.827028	0.827238	1000.000
7	weighted avg	0.837308	0.835000	0.835968	1000.000

(a) Metrics on test data

	precision	recall	f1-score	support
0	0.997463	0.965619	0.981283	2036.0000
1	0.919616	0.914990	0.917297	1988.0000
2	0.797131	0.810417	0.803719	1920.0000
3	0.642928	0.694086	0.667528	1860.0000
4	0.838833	0.798725	0.818288	2196.0000
accuracy	0.838600	0.838600	0.838600	0.8386
macro avg	0.839194	0.836767	0.837623	10000.0000
weighted avg	0.842745	0.838600	0.840318	10000.0000

(b) Metrics on train data

Figure 38: Metrics for hidden layer : {512, 256, 128, 64}

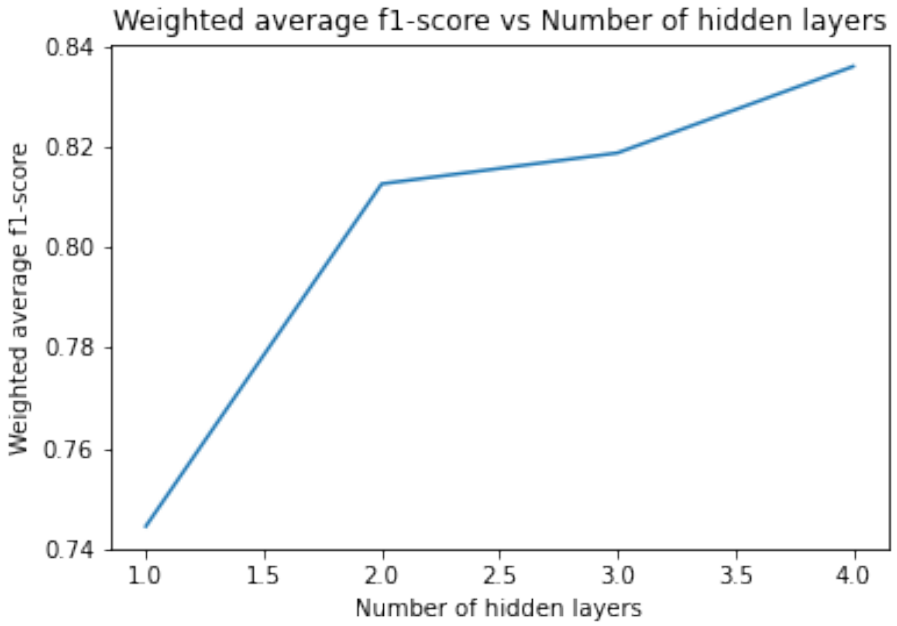


Figure 39: Plot of weighted average test F1 score vs network depth

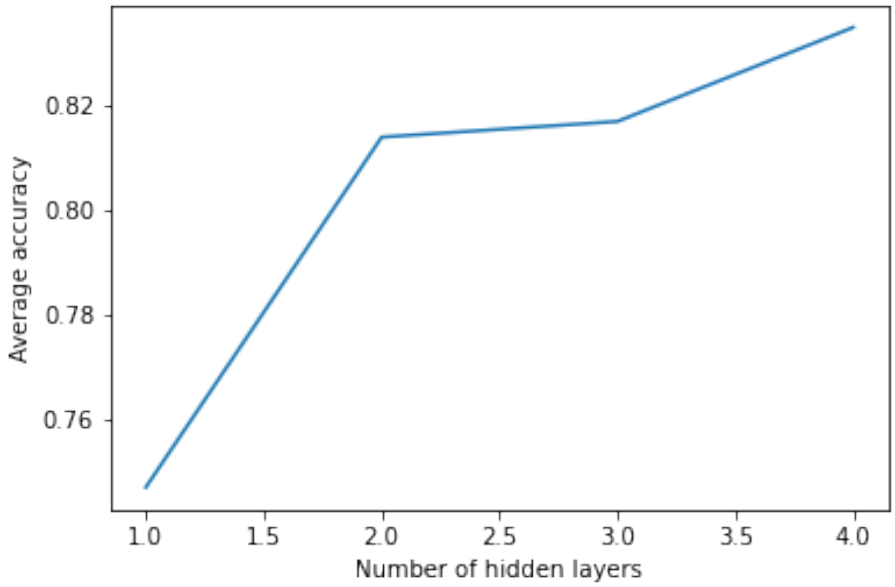


Figure 40: Plot of average test accuracy score vs network depth

2.3 Adaptive Learning Rate

Metrics for hidden layers : {{512}, {512, 256}, {512, 256, 128}, {512, 256, 128, 64}}.

	precision	recall	f1-score	support
0	0.903930	0.900000	0.901961	230.000
1	0.676768	0.666667	0.671679	201.000
2	0.497487	0.568966	0.530831	174.000
3	0.534759	0.478469	0.505051	209.000
4	0.689840	0.693548	0.691689	186.000
accuracy	0.669000	0.669000	0.669000	0.669
macro avg	0.660557	0.661530	0.660242	1000.000
weighted avg	0.670572	0.669000	0.669033	1000.000

(a) Metrics on test data

	precision	recall	f1-score	support
0	0.900051	0.867906	0.883686	2044.0000
1	0.694641	0.686314	0.690452	2002.0000
2	0.543033	0.559662	0.551222	1894.0000
3	0.470120	0.512486	0.490390	1842.0000
4	0.748924	0.706041	0.726851	2218.0000
accuracy	0.671800	0.671800	0.671800	0.6718
macro avg	0.671354	0.666482	0.668520	10000.0000
weighted avg	0.678595	0.671800	0.674801	10000.0000

(b) Metrics on train data

Figure 41: Metrics for hidden layer : {512}

	precision	recall	f1-score	support
0	0.912664	0.896996	0.904762	233.000
1	0.661616	0.678756	0.670077	193.000
2	0.507538	0.577143	0.540107	175.000
3	0.540107	0.469767	0.502488	215.000
4	0.668449	0.679348	0.673854	184.000
accuracy	0.667000	0.667000	0.667000	0.667
macro avg	0.658075	0.660402	0.658258	1000.000
weighted avg	0.668279	0.667000	0.666677	1000.000

(a) Metrics on test data

	precision	recall	f1-score	support
0	0.901065	0.867188	0.883802	2048.0000
1	0.688069	0.690512	0.689288	1971.0000
2	0.536885	0.556559	0.546545	1883.0000
3	0.460657	0.494652	0.477050	1870.0000
4	0.740316	0.694794	0.716833	2228.0000
accuracy	0.665800	0.665800	0.665800	0.6658
macro avg	0.665398	0.660741	0.662704	10000.0000
weighted avg	0.672337	0.665800	0.668694	10000.0000

(b) Metrics on train data

Figure 42: Metrics for hidden layer : {512, 256}

	precision	recall	f1-score	support
0	0.903930	0.869748	0.886510	238.000
1	0.611111	0.654054	0.631854	185.000
2	0.472362	0.556213	0.510870	169.000
3	0.486631	0.464286	0.475196	196.000
4	0.737968	0.650943	0.691729	212.000
accuracy	0.651000	0.651000	0.651000	0.651
macro avg	0.642400	0.639049	0.639232	1000.000
weighted avg	0.659849	0.651000	0.654004	1000.000

(a) Metrics on test data

	precision	recall	f1-score	support
0	0.903930	0.869748	0.886510	238.000
1	0.611111	0.654054	0.631854	185.000
2	0.472362	0.556213	0.510870	169.000
3	0.486631	0.464286	0.475196	196.000
4	0.737968	0.650943	0.691729	212.000
accuracy	0.651000	0.651000	0.651000	0.651
macro avg	0.642400	0.639049	0.639232	1000.000
weighted avg	0.659849	0.651000	0.654004	1000.000

(b) Metrics on train data

Figure 43: Metrics for hidden layer : {512, 256, 128}

	precision	recall	f1-score	support
0	0.000	0.0000	0.000000	0.000
1	0.000	0.0000	0.000000	0.000
2	0.000	0.0000	0.000000	0.000
3	0.000	0.0000	0.000000	0.000
4	1.000	0.1870	0.315080	1000.000
accuracy	0.187	0.1870	0.187000	0.187
macro avg	0.200	0.0374	0.063016	1000.000
weighted avg	1.000	0.1870	0.315080	1000.000

(a) Metrics on test data

	precision	recall	f1-score	support
0	0.0000	0.00000	0.000000	0.0000
1	0.0000	0.00000	0.000000	0.0000
2	0.0000	0.00000	0.000000	0.0000
3	0.0000	0.00000	0.000000	0.0000
4	1.0000	0.20910	0.345877	10000.0000
accuracy	0.2091	0.20910	0.209100	0.2091
macro avg	0.2000	0.04182	0.069175	10000.0000
weighted avg	1.0000	0.20910	0.345877	10000.0000

(b) Metrics on train data

Figure 44: Metrics for hidden layer : {512, 256, 128, 64}

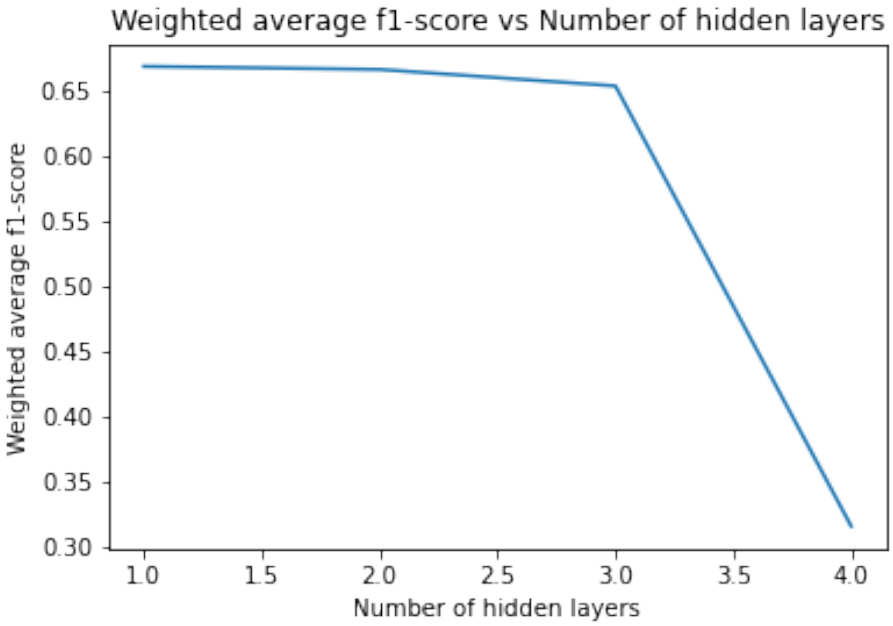


Figure 45: Plot of weighted average test F1 score vs network depth

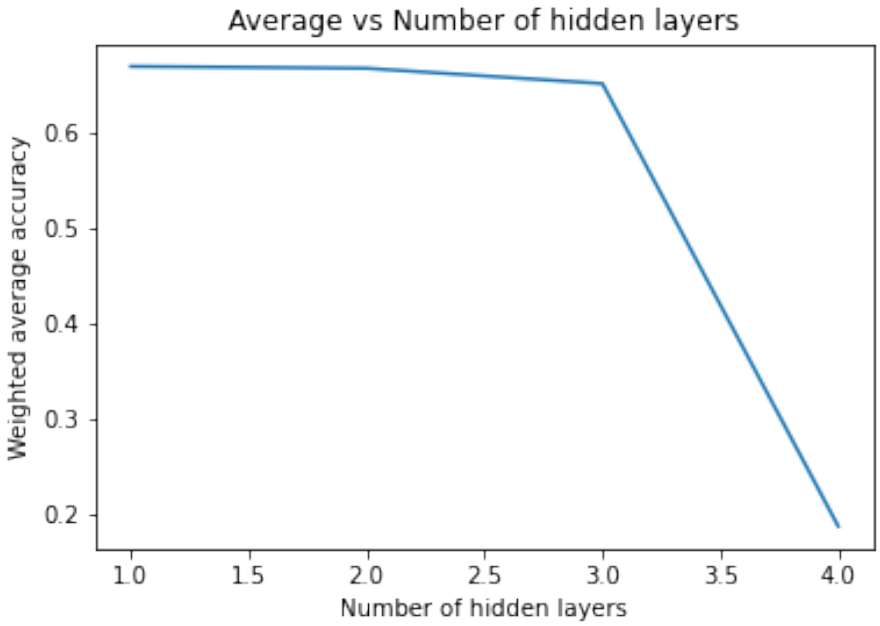


Figure 46: Plot of average test accuracy score vs network depth

We observe a decrease in F1-score here. This could be due to the fact that for more hidden layers it takes much more steps to reach the optimum value. Over here we are also decreasing the learning rate, as a result the step size decreases and the larger network is unable to reach optimum in given number of epochs. For other layer sizes as well, the performance is nowhere near the

previous sub-part because the gradient update steps become too small.

There isn't much boost to the training time as convergence in the given order was not achieved, but if the threshold was minimized, an early convergence would have surely occurred.

2.4 Tarining with ReLU Activation

Metrics for hidden layers : {{512}, {512, 256}, {512, 256, 128}, {512, 256, 128, 64}}.

	precision	recall	f1-score	support
0	0.908297	0.920354	0.914286	226.000
1	0.696970	0.722513	0.709512	191.000
2	0.582915	0.604167	0.593350	192.000
3	0.577540	0.521739	0.548223	207.000
4	0.711230	0.722826	0.716981	184.000
accuracy	0.703000	0.703000	0.703000	0.703
macro avg	0.695390	0.698320	0.696470	1000.000
weighted avg	0.700733	0.703000	0.701475	1000.000

(a) Metrics on test data

	Unnamed: 0	precision	recall	f1-score	support
0	0	0.900051	0.895055	0.897546	1982.0000
1	1	0.720425	0.731144	0.725745	1949.0000
2	2	0.630635	0.608202	0.619215	2024.0000
3	3	0.533865	0.583878	0.557752	1836.0000
4	4	0.796270	0.753735	0.774419	2209.0000
5	accuracy	0.716700	0.716700	0.716700	0.7167
6	macro avg	0.716249	0.714403	0.714935	10000.0000
7	weighted avg	0.720355	0.716700	0.718143	10000.0000

(b) Metrics on train data

Figure 47: Metrics for hidden layer : {512}

	precision	recall	f1-score	support
0	0.995633	0.970213	0.982759	235.000
1	0.868687	0.843137	0.855721	204.000
2	0.728643	0.714286	0.721393	203.000
3	0.620321	0.610526	0.615385	190.000
4	0.711230	0.791667	0.749296	168.000
accuracy	0.794000	0.794000	0.794000	0.794
macro avg	0.784903	0.785966	0.784911	1000.000
weighted avg	0.796448	0.794000	0.794763	1000.000

(a) Metrics on test data

	precision	recall	f1-score	support
0	0.989346	0.955414	0.972084	2041.0000
1	0.881699	0.867662	0.874624	2010.0000
2	0.779201	0.728100	0.752784	2089.0000
3	0.628984	0.666139	0.647029	1896.0000
4	0.783357	0.834012	0.807891	1964.0000
accuracy	0.811600	0.811600	0.811600	0.8116
macro avg	0.812517	0.810265	0.810882	10000.0000
weighted avg	0.815029	0.811600	0.812805	10000.0000

(b) Metrics on train data

Figure 48: Metrics for hidden layer : {512, 256}

	precision	recall	f1-score	support
0	1.000000	0.978632	0.989201	234.000
1	0.949495	0.903846	0.926108	208.000
2	0.788945	0.867403	0.826316	181.000
3	0.796791	0.659292	0.721550	226.000
4	0.705882	0.874172	0.781065	151.000
accuracy	0.855000	0.855000	0.855000	0.855
macro avg	0.848223	0.856669	0.848848	1000.000
weighted avg	0.860957	0.855000	0.854678	1000.000

(a) Metrics on test data

	precision	recall	f1-score	support
0	1.000000	0.978632	0.989201	234.000
1	0.949495	0.903846	0.926108	208.000
2	0.788945	0.867403	0.826316	181.000
3	0.796791	0.659292	0.721550	226.000
4	0.705882	0.874172	0.781065	151.000
accuracy	0.855000	0.855000	0.855000	0.855
macro avg	0.848223	0.856669	0.848848	1000.000
weighted avg	0.860957	0.855000	0.854678	1000.000

(b) Metrics on train data

Figure 49: Metrics for hidden layer : {512, 256, 128}

	precision	recall	f1-score	support
0	1.000000	0.966245	0.982833	237.000
1	0.939394	0.873239	0.905109	213.000
2	0.798995	0.823834	0.811224	193.000
3	0.764706	0.680952	0.720403	210.000
4	0.705882	0.897959	0.790419	147.000
accuracy	0.849000	0.849000	0.849000	0.849
macro avg	0.841795	0.848446	0.841998	1000.000
weighted avg	0.855650	0.849000	0.849762	1000.000

(a) Metrics on test data

	precision	recall	f1-score	support
0	1.000000	0.983042	0.991449	2005.0000
1	0.981800	0.954300	0.967854	2035.0000
2	0.950307	0.927964	0.939003	1999.0000
3	0.920319	0.903667	0.911917	2045.0000
4	0.907700	0.990605	0.947342	1916.0000
accuracy	0.951400	0.951400	0.951400	0.9514
macro avg	0.952025	0.951916	0.951513	10000.0000
weighted avg	0.952383	0.951400	0.951448	10000.0000

(b) Metrics on train data

Figure 50: Metrics for hidden layer : {512, 256, 128, 64}

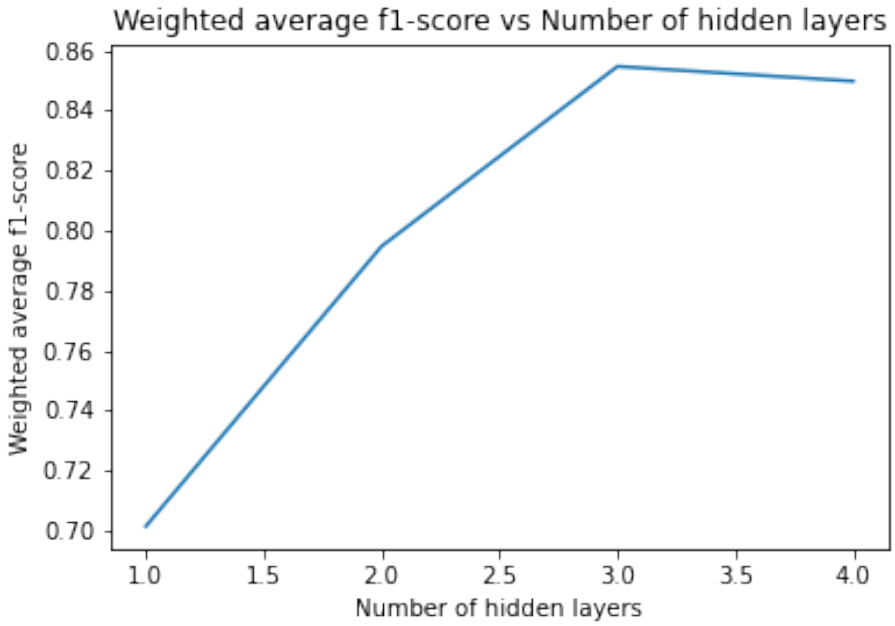


Figure 51: Plot of weighted average test F1 score vs network depth

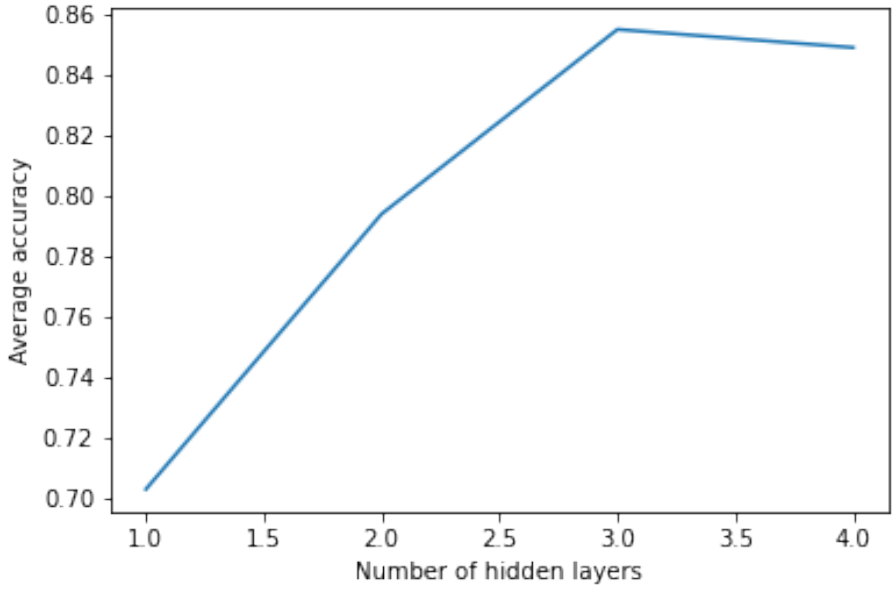
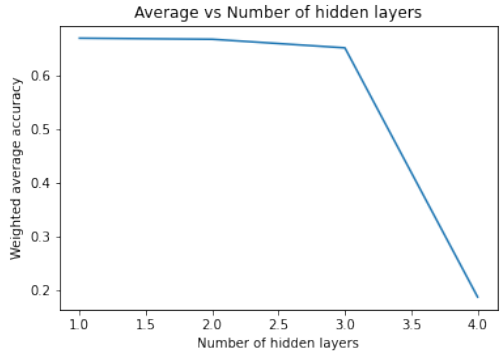
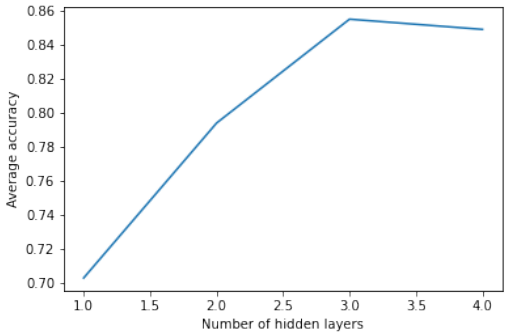


Figure 52: Plot of average test accuracy score vs network depth

The following plots are for comparing the accuracies of the two models : sigmoid vs relu with variable learning rate



(a) Accuracy for part d



(b) Accuracy for part e

Figure 53: Comparison of accuracies

We can clearly see that ReLU model performs much better than sigmoid. The gradients in ReLU are simple and are either 0 or 1. This means a faster training in less epochs. Hence when we train a network using ReLU we get better results for the same epoch even with variable learning rate.

2.5 MLPClassifier

Stopping Criteria : Max Iteration = 1000 and defualt tolerance

Metrics for hidden layers : {{512}, {512, 256}, {512, 256, 128}, {512, 256, 128, 64}}.

	precision	recall	f1-score	support
0	0.890830	0.607143	0.722124	336.00
1	0.227273	0.401786	0.290323	112.00
2	0.150754	0.410959	0.220588	73.00
3	0.213904	0.333333	0.260586	120.00
4	0.860963	0.448468	0.589744	359.00
accuracy	0.480000	0.480000	0.480000	0.48
macro avg	0.468745	0.440338	0.416673	1000.00
weighted avg	0.670532	0.480000	0.534241	1000.00

(a) Metrics on test data

	precision	recall	f1-score	support
0	0.908676	0.533036	0.671919	3360.0000
1	0.199191	0.372049	0.259467	1059.0000
2	0.131660	0.392966	0.197237	654.0000
3	0.192729	0.345536	0.247442	1120.0000
4	0.855093	0.469661	0.606307	3807.0000
accuracy	0.461700	0.461700	0.461700	0.4617
macro avg	0.457470	0.422650	0.396474	10000.0000
weighted avg	0.682140	0.461700	0.524676	10000.0000

(b) Metrics on train data

Figure 54: Metrics for hidden layer : {512}

	precision	recall	f1-score	support
0	0.938865	0.571809	0.710744	376.000
1	0.136364	0.375000	0.200000	72.000
2	0.080402	0.333333	0.129555	48.000
3	0.122995	0.252747	0.165468	91.000
4	0.877005	0.397094	0.546667	413.000
accuracy	0.445000	0.445000	0.445000	0.445
macro avg	0.431126	0.385997	0.350487	1000.000
weighted avg	0.740086	0.445000	0.528689	1000.000

(a) Metrics on test data

	precision	recall	f1-score	support
0	0.941654	0.512141	0.663450	3624.0000
1	0.132457	0.339378	0.190545	772.0000
2	0.100410	0.400000	0.160524	490.0000
3	0.156873	0.334395	0.213559	942.0000
4	0.892874	0.447507	0.596200	4172.0000
accuracy	0.449600	0.449600	0.449600	0.4496
macro avg	0.444854	0.406684	0.364856	10000.0000
weighted avg	0.743686	0.449600	0.531862	10000.0000

(b) Metrics on train data

Figure 55: Metrics for hidden layer : {512, 256}

	precision	recall	f1-score	support
0	0.934498	0.629412	0.752197	340.000
1	0.292929	0.491525	0.367089	118.000
2	0.140704	0.388889	0.206642	72.000
3	0.053476	0.238095	0.087336	42.000
4	0.898396	0.392523	0.546341	428.000
accuracy	0.478000	0.478000	0.478000	0.478
macro avg	0.464000	0.428089	0.391921	1000.000
weighted avg	0.749185	0.478000	0.551444	1000.000

(a) Metrics on test data

	precision	recall	f1-score	support
0	0.951801	0.571081	0.713851	3285.0000
1	0.263903	0.471119	0.338302	1108.0000
2	0.159836	0.376357	0.224380	829.0000
3	0.088645	0.335217	0.140213	531.0000
4	0.902439	0.444314	0.595456	4247.0000
accuracy	0.477500	0.477500	0.477500	0.4775
macro avg	0.473325	0.439617	0.402440	10000.0000
weighted avg	0.743130	0.477500	0.550920	10000.0000

(b) Metrics on train data

Figure 56: Metrics for hidden layer : {512, 256, 128}

	precision	recall	f1-score	support
0	0.973799	0.547912	0.701258	407.000
1	0.106061	0.466667	0.172840	45.000
2	0.045226	0.428571	0.081818	21.000
3	0.096257	0.163636	0.121212	110.000
4	0.930481	0.417266	0.576159	417.000
accuracy	0.445000	0.445000	0.445000	0.445
macro avg	0.430365	0.404810	0.330657	1000.000
weighted avg	0.800658	0.445000	0.548500	1000.000

(a) Metrics on test data

	precision	recall	f1-score	support
0	0.981228	0.500129	0.662556	3867.0000
1	0.076340	0.305051	0.122119	495.0000
2	0.056352	0.506912	0.101429	217.0000
3	0.121016	0.226257	0.157690	1074.0000
4	0.948350	0.456177	0.616030	4347.0000
accuracy	0.442100	0.442100	0.442100	0.4421
macro avg	0.436657	0.398905	0.331965	10000.0000
weighted avg	0.809687	0.442100	0.549180	10000.0000

(b) Metrics on train data

Figure 57: Metrics for hidden layer : {512, 256, 128, 64}

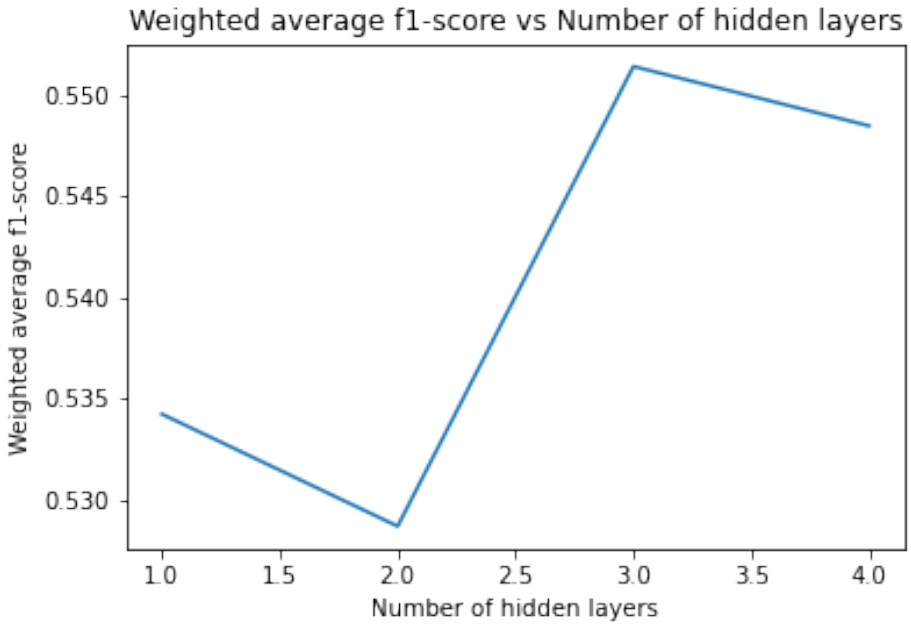


Figure 58: Plot of weighted average test F1 score vs network depth

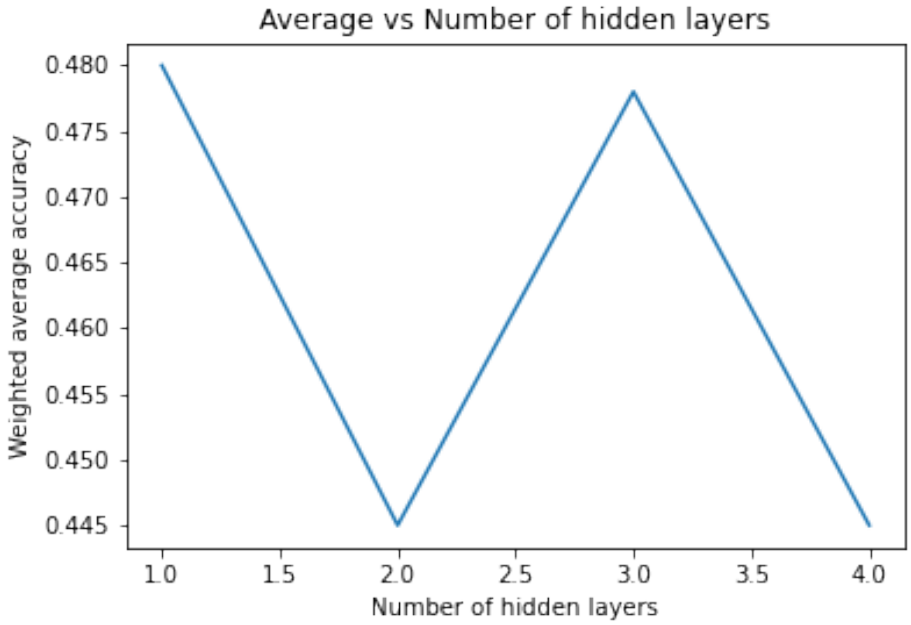


Figure 59: Plot of average test accuracy score vs network depth

We see that the accuracies and F1-score are fairly low as compared to the those in the previous part. This could be a result of difference in threshold. It seems that sklearn models have stopped early and as result don't show good results. Trainign with a smaller threshold will surely help.