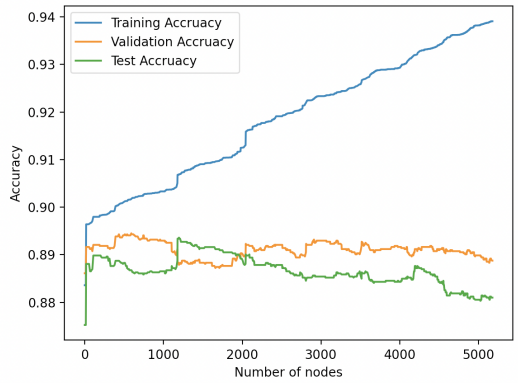
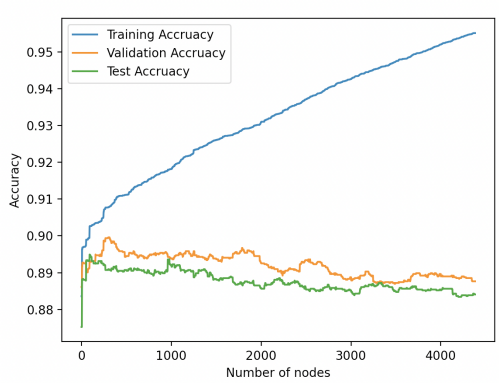
COL774 Assignment 3

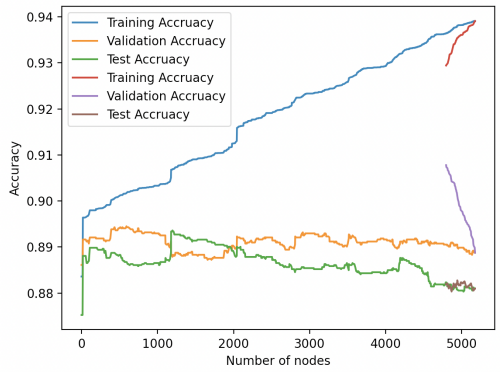
Harshil Vagadia - 2019CS10356

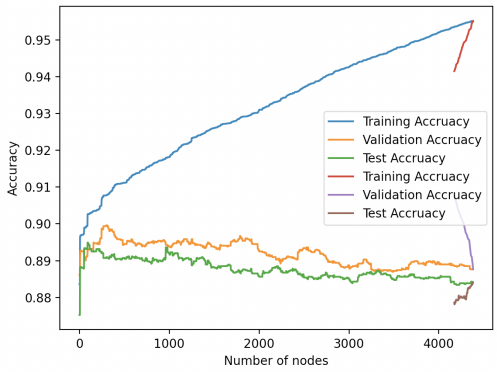
Decision Trees

A. As seen, the tree expands to just more than **5000** nodes. The training accuracy increases, but the validation and test accuracy do not increase. The left graphs are for normal encoding while the right graphs are for one hot encoding.



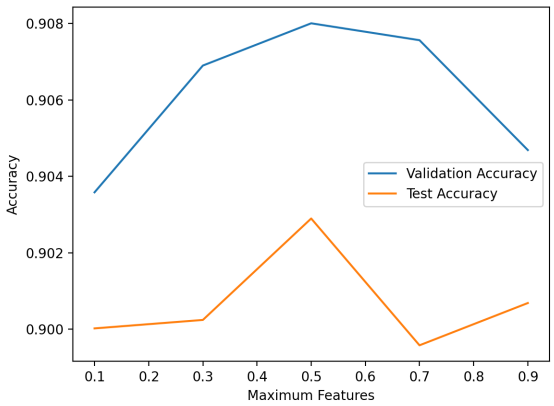


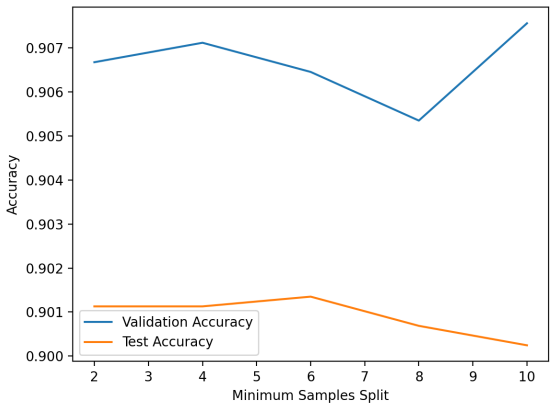


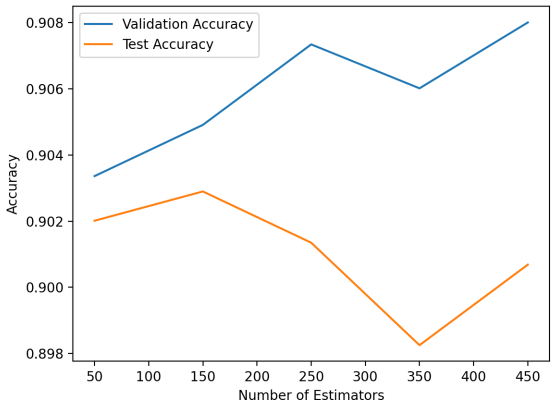


B. As visible in the above graph, pruning some nodes will reduce overfitting. The validation accuracy increases, but the training accuracy (and in case of one hot encoding, the test accuracy) decreases. The increase in validation accuracy is drastic as we aggressively prune best nodes from the tree.

C. Grid Search was done over the range of parameters as asked in the question. For each set of parameters, the Random Forest Classifier was trained on training data. The classifier with maximum Out Of Bag Error was selected. This classifier had parameters as: **n_estimators = 450, max_features = 0.7** and **min_samples_split = 10**. The training accuracy was **0.98036**, validation accuracy was **0.90667**, test accuracy was **0.90068** and the out of bag error was **0.90881**. This model performs only marginally better to part A and B.







D. As visible from the graph, the maximum variation in accuracy is due to the **Number of Estimators** parameter. Hence it is the most sensitive parameter. Maximum features also has significant deviation in validation accuracy but relatively less deviation in test accuracy. The accuracy almost does not change with minimum samples split.

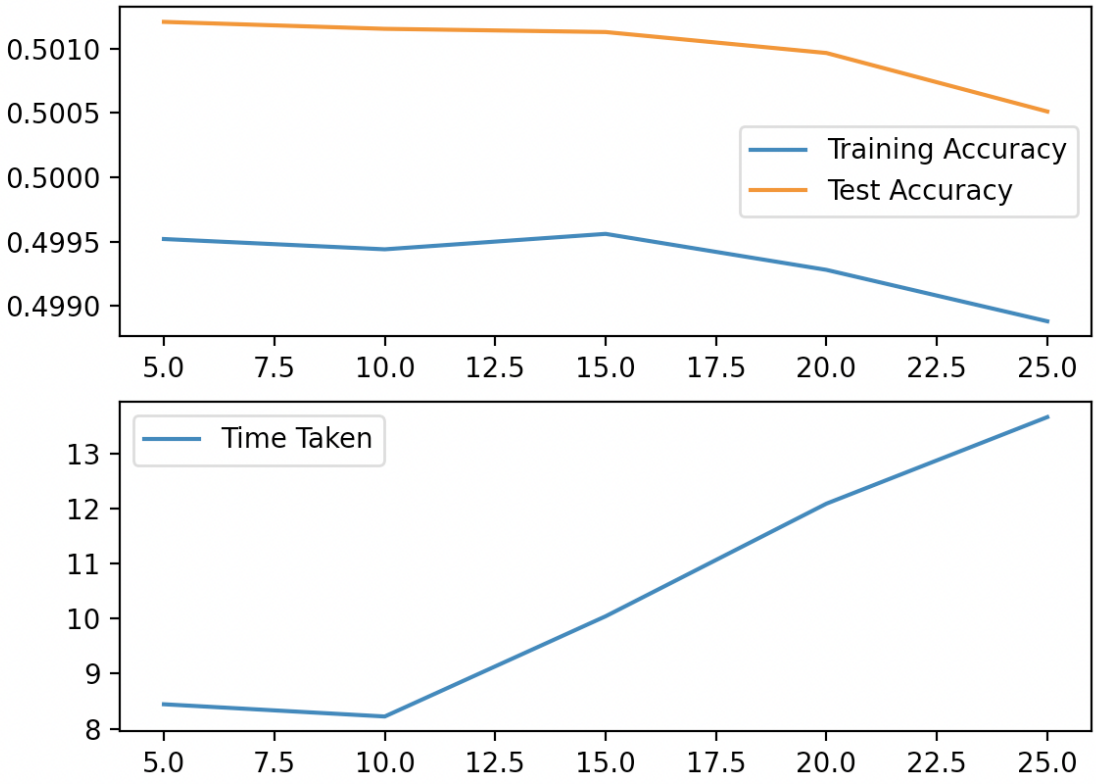
Neural Networks

A. Both input and output features were encoded as mentioned in the question statement. Input data contained 85 features (5 * 13 + 5 * 4) and the output data contained 10 features.

B. A generic implementation was created which could be configured on parameters like number of hidden layers, size of hidden layers, activation of each hidden layer, batch size, maximum epochs, learning rate, adaptive learning rate etc.

C. Given below is the plot of training and test accuracies with hidden layers.The accuracy does not vary much. Further, as it can be seen from the confusion matrix, the model almost always predicts 0. Thus it seems that this model is not expressive enough to capture the dataset. The time taken to train roughly increases with hidden layers. The convergence criteria was chosen as the difference between previous and current losses

to be less than 0.0001.



```
Hidden Layer:  5
Epochs:  189
Training Accuracy:  0.4995201919232307
Test Accuracy:  0.501209
[[12493.     0.     0.     0.     0.     0.     0.     0.     0.     0.]
 [10599.     0.     0.     0.     0.     0.     0.     0.     0.     0.]
 [ 1206.     0.     0.     0.     0.     0.     0.     0.     0.     0.]
 [  513.     0.     0.     0.     0.     0.     0.     0.     0.     0.]
 [   93.     0.     0.     0.     0.     0.     0.     0.     0.     0.]
 [   54.     0.     0.     0.     0.     0.     0.     0.     0.     0.]
 [   36.     0.     0.     0.     0.     0.     0.     0.     0.     0.]
 [    6.     0.     0.     0.     0.     0.     0.     0.     0.     0.]
 [    5.     0.     0.     0.     0.     0.     0.     0.     0.     0.]
 [    5.     0.     0.     0.     0.     0.     0.     0.     0.     0.]]
```

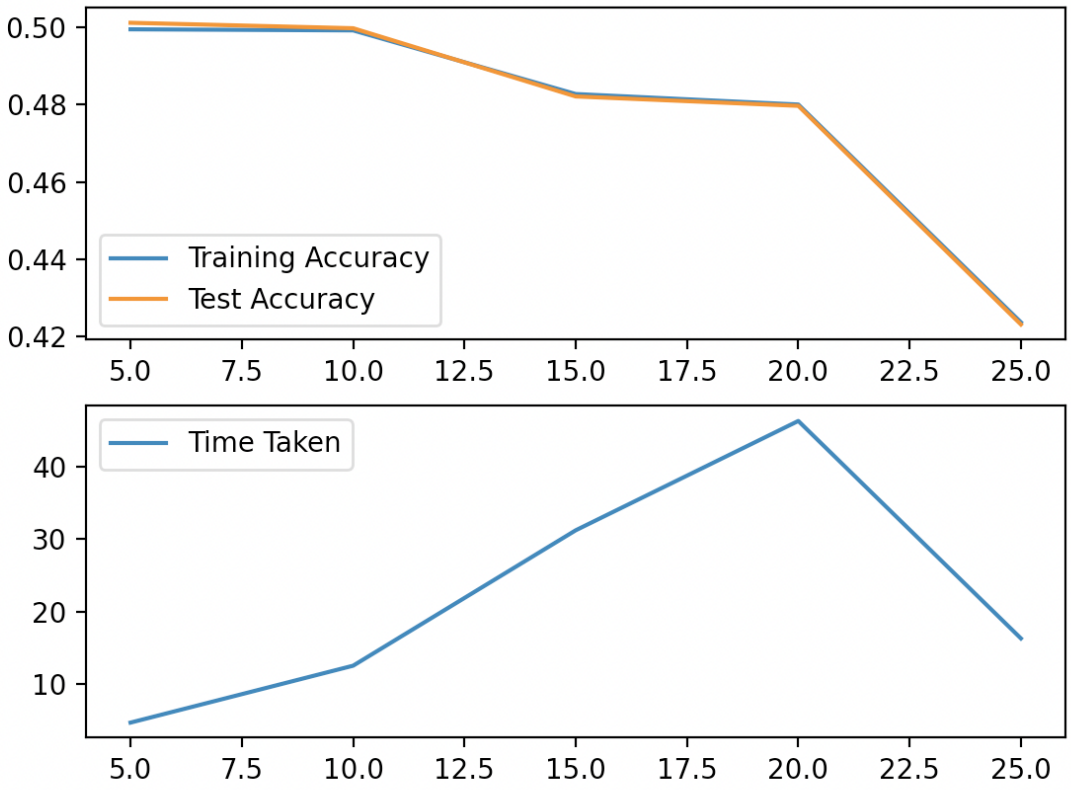
```
Hidden Layer:  10
Epochs:  176
Training Accuracy:  0.4994402239104358
Test Accuracy:  0.501155
[[12490.     0.     3.     0.     0.     0.     0.     0.     0.     0.]
 [10597.     0.     2.     0.     0.     0.     0.     0.     0.     0.]
 [ 1205.     0.     1.     0.     0.     0.     0.     0.     0.     0.]
 [  513.     0.     0.     0.     0.     0.     0.     0.     0.     0.]
 [   93.     0.     0.     0.     0.     0.     0.     0.     0.     0.]
 [   54.     0.     0.     0.     0.     0.     0.     0.     0.     0.]
 [   36.     0.     0.     0.     0.     0.     0.     0.     0.     0.]
 [    6.     0.     0.     0.     0.     0.     0.     0.     0.     0.]
 [    5.     0.     0.     0.     0.     0.     0.     0.     0.     0.]
 [    5.     0.     0.     0.     0.     0.     0.     0.     0.     0.]]
```

```
Hidden Layer:  15
Epochs:  210
Training Accuracy:  0.49956017592962815
Test Accuracy:  0.50113
[[12493.     0.     0.     0.     0.     0.     0.     0.     0.     0.]
 [10597.     0.     1.     0.     0.     0.     1.     0.     0.     0.]
 [ 1205.     0.     1.     0.     0.     0.     0.     0.     0.     0.]
 [  513.     0.     0.     0.     0.     0.     0.     0.     0.     0.]
 [   93.     0.     0.     0.     0.     0.     0.     0.     0.     0.]
 [   54.     0.     0.     0.     0.     0.     0.     0.     0.     0.]
 [   36.     0.     0.     0.     0.     0.     0.     0.     0.     0.]
 [    6.     0.     0.     0.     0.     0.     0.     0.     0.     0.]
 [    5.     0.     0.     0.     0.     0.     0.     0.     0.     0.]
 [    5.     0.     0.     0.     0.     0.     0.     0.     0.     0.]]
```

```
Hidden Layer:  20
Epochs:  245
Training Accuracy:  0.4992802878848461
Test Accuracy:  0.500967
[[12487.     0.     1.     1.     0.     0.     1.     1.     2.     0.]
 [10596.     0.     0.     0.     1.     0.     0.     0.     2.     0.]
 [ 1206.     0.     0.     0.     0.     0.     0.     0.     0.     0.]
 [  513.     0.     0.     0.     0.     0.     0.     0.     0.     0.]
 [   93.     0.     0.     0.     0.     0.     0.     0.     0.     0.]
 [   54.     0.     0.     0.     0.     0.     0.     0.     0.     0.]
 [   36.     0.     0.     0.     0.     0.     0.     0.     0.     0.]
 [    6.     0.     0.     0.     0.     0.     0.     0.     0.     0.]
 [    5.     0.     0.     0.     0.     0.     0.     0.     0.     0.]
 [    5.     0.     0.     0.     0.     0.     0.     0.     0.     0.]]
```

```
Hidden Layer:  25
Epochs:  263
Training Accuracy:  0.49888044782087165
Test Accuracy:  0.500512
[[12460.    27.     1.     0.     1.     0.     0.     0.     0.     4.]
 [10572.    17.     1.     3.     0.     1.     0.     1.     4.     0.]
 [ 1202.     3.     0.     0.     0.     0.     1.     0.     0.     0.]
 [  510.     3.     0.     0.     0.     0.     0.     0.     0.     0.]
 [   93.     0.     0.     0.     0.     0.     0.     0.     0.     0.]
 [   54.     0.     0.     0.     0.     0.     0.     0.     0.     0.]
 [   36.     0.     0.     0.     0.     0.     0.     0.     0.     0.]
 [    6.     0.     0.     0.     0.     0.     0.     0.     0.     0.]
 [    5.     0.     0.     0.     0.     0.     0.     0.     0.     0.]
 [    5.     0.     0.     0.     0.     0.     0.     0.     0.     0.]]
```

D. Using adaptive learning rate does not seem to benefit this model. The accuracy decreases with increase in hidden layers. Also the time for convergence increases. This is probably because the learning rate decreases as we train, thus making the progress slower. One trend to notice is that the model predicts more 1s as the hidden layers increases.


0.50
0.48
0.46
0.44
0.42
Training Accuracy
Test Accuracy
5.0
7.5
10.0
12.5
15.0
17.5
20.0
22.5
25.0
Time Taken
40
30
20
10
5.0
7.5
10.0
12.5
15.0
17.5
20.0
22.5
25.0

```
Hidden Layer:  5
Epochs:  87
Training Accuracy:  0.4995201919232307
Test Accuracy:  0.501209
[[12493.     0.     0.     0.     0.     0.     0.     0.     0.     0.]
 [10599.     0.     0.     0.     0.     0.     0.     0.     0.     0.]
 [ 1206.     0.     0.     0.     0.     0.     0.     0.     0.     0.]
 [  513.     0.     0.     0.     0.     0.     0.     0.     0.     0.]
 [   93.     0.     0.     0.     0.     0.     0.     0.     0.     0.]
 [   54.     0.     0.     0.     0.     0.     0.     0.     0.     0.]
 [   36.     0.     0.     0.     0.     0.     0.     0.     0.     0.]
 [    6.     0.     0.     0.     0.     0.     0.     0.     0.     0.]
 [    5.     0.     0.     0.     0.     0.     0.     0.     0.     0.]
 [    5.     0.     0.     0.     0.     0.     0.     0.     0.     0.]]
```

```
Hidden Layer:  10
Epochs:  280
Training Accuracy:  0.49924030387844864
Test Accuracy:  0.499801
[[12359.   120.    10.     0.     1.     0.     0.     1.     0.     2.]
 [10461.   124.    10.     0.     2.     0.     0.     1.     0.     1.]
 [ 1185.    18.     3.     0.     0.     0.     0.     0.     0.     0.]
 [  506.     7.     0.     0.     0.     0.     0.     0.     0.     0.]
 [   93.     0.     0.     0.     0.     0.     0.     0.     0.     0.]
 [   54.     0.     0.     0.     0.     0.     0.     0.     0.     0.]
 [   36.     0.     0.     0.     0.     0.     0.     0.     0.     0.]
 [    5.     1.     0.     0.     0.     0.     0.     0.     0.     0.]
 [    5.     0.     0.     0.     0.     0.     0.     0.     0.     0.]
 [    5.     0.     0.     0.     0.     0.     0.     0.     0.     0.]]
```

```
Hidden Layer:  15
Epochs:  686
Training Accuracy:  0.48276689324270294
Test Accuracy:  0.482144
[[9587. 2885.    0.    4.    4.    7.    0.    3.    3.    0.]
 [8099. 2487.    0.    6.    2.    5.    0.    0.    0.    0.]
 [ 939.  264.    0.    1.    1.    0.    0.    1.    0.    0.]
 [ 408.  105.    0.    0.    0.    0.    0.    0.    0.    0.]
 [  69.   23.    0.    1.    0.    0.    0.    0.    0.    0.]
 [  40.   14.    0.    0.    0.    0.    0.    0.    0.    0.]
 [  31.    5.    0.    0.    0.    0.    0.    0.    0.    0.]
 [   3.    3.    0.    0.    0.    0.    0.    0.    0.    0.]
 [   4.    1.    0.    0.    0.    0.    0.    0.    0.    0.]
 [   3.    1.    0.    1.    0.    0.    0.    0.    0.    0.]]
```

```
Hidden Layer:  20
Epochs:  1000
Training Accuracy:  0.4800479808076769
Test Accuracy:  0.479711
[[8747. 3726.    6.    0.    2.    6.    3.    0.    2.    1.]
 [7330. 3259.    5.    0.    2.    1.    1.    0.    0.    1.]
 [ 826.  378.    0.    0.    0.    0.    0.    0.    0.    2.]
 [ 357.  155.    0.    0.    0.    0.    1.    0.    0.    0.]
 [  70.   23.    0.    0.    0.    0.    0.    0.    0.    0.]
 [  39.   15.    0.    0.    0.    0.    0.    0.    0.    0.]
 [  27.    9.    0.    0.    0.    0.    0.    0.    0.    0.]
 [   4.    2.    0.    0.    0.    0.    0.    0.    0.    0.]
 [   3.    2.    0.    0.    0.    0.    0.    0.    0.    0.]
 [   3.    2.    0.    0.    0.    0.    0.    0.    0.    0.]]
```

```
Hidden Layer:  25
Epochs:  307
Training Accuracy:  0.4236705317872851
Test Accuracy:  0.423121
[[  107. 12382.     0.     1.     0.     1.     0.     1.     1.     0.]
 [  107. 10489.     1.     0.     1.     0.     0.     0.     0.     1.]
 [   18.  1188.     0.     0.     0.     0.     0.     0.     0.     0.]
 [    8.   505.     0.     0.     0.     0.     0.     0.     0.     0.]
 [    1.    92.     0.     0.     0.     0.     0.     0.     0.     0.]
 [    0.    54.     0.     0.     0.     0.     0.     0.     0.     0.]
 [    0.    36.     0.     0.     0.     0.     0.     0.     0.     0.]
 [    0.     6.     0.     0.     0.     0.     0.     0.     0.     0.]
 [    0.     5.     0.     0.     0.     0.     0.     0.     0.     0.]
 [    0.     5.     0.     0.     0.     0.     0.     0.     0.     0.]]
```

E. As can be seen from the results below, relu performs slightly better than sigmoid. The accuracies are still almost the same as the other parts. However, as seen from the confusion matrix, the model now seems to predict values other than 0s also. This seems to imply that the model is learning better representation of the data, even though it is not able to learn it to completion.

```
Epochs:  655
Training Accuracy for sigmoid: 0.47616953218712516
Test Accuracy for sigmoid: 0.471203
[[8868. 3487.   16.   26.   46.    9.    9.   11.   10.   11.]
 [7445. 3036.   15.   15.   40.   11.    7.    8.   12.   10.]
 [ 876.  325.    1.    1.    1.    1.    1.    0.    0.    0.]
 [ 350.  154.    1.    3.    1.    0.    1.    0.    2.    1.]
 [  58.   35.    0.    0.    0.    0.    0.    0.    0.    0.]
 [  25.   28.    0.    0.    0.    1.    0.    0.    0.    0.]
 [  24.   11.    0.    0.    0.    0.    0.    0.    0.    1.]
 [   2.    4.    0.    0.    0.    0.    0.    0.    0.    0.]
 [   4.    1.    0.    0.    0.    0.    0.    0.    0.    0.]
 [   2.    3.    0.    0.    0.    0.    0.    0.    0.    0.]]
```

```
Epochs:  102
Training Accuracy for relu: 0.4478208716513395
Test Accuracy for relu: 0.44792
[[10889.   272.   185.   167.    39.   111.    98.   571.    34.   127.]
 [ 9181.   279.   157.   129.    43.   106.    88.   470.    24.   122.]
 [ 1045.    28.    22.    14.     4.    18.    10.    43.     0.    22.]
 [  450.    18.     3.     9.     0.     4.     5.    13.     2.     9.]
 [   81.     1.     2.     1.     0.     2.     1.     3.     0.     2.]
 [   50.     0.     1.     0.     0.     0.     0.     2.     0.     1.]
 [   31.     3.     0.     0.     0.     0.     0.     2.     0.     0.]
 [    5.     0.     0.     0.     0.     0.     0.     1.     0.     0.]
 [    5.     0.     0.     0.     0.     0.     0.     0.     0.     0.]
 [    5.     0.     0.     0.     0.     0.     0.     0.     0.     0.]]
```

F. The MLP classifier was set up as close to the configuration in part E. However, the MLP classifier allows only a single activation function across all the hidden layers as opposed to the part E implementation. This activation function was set to relu. The training accuracy was 0.99964 and the test accuracy was 0.97490. This model performs much better than the previous model. This maybe a combination of different activation function, stricter convergence criteria or different weight initialisation.