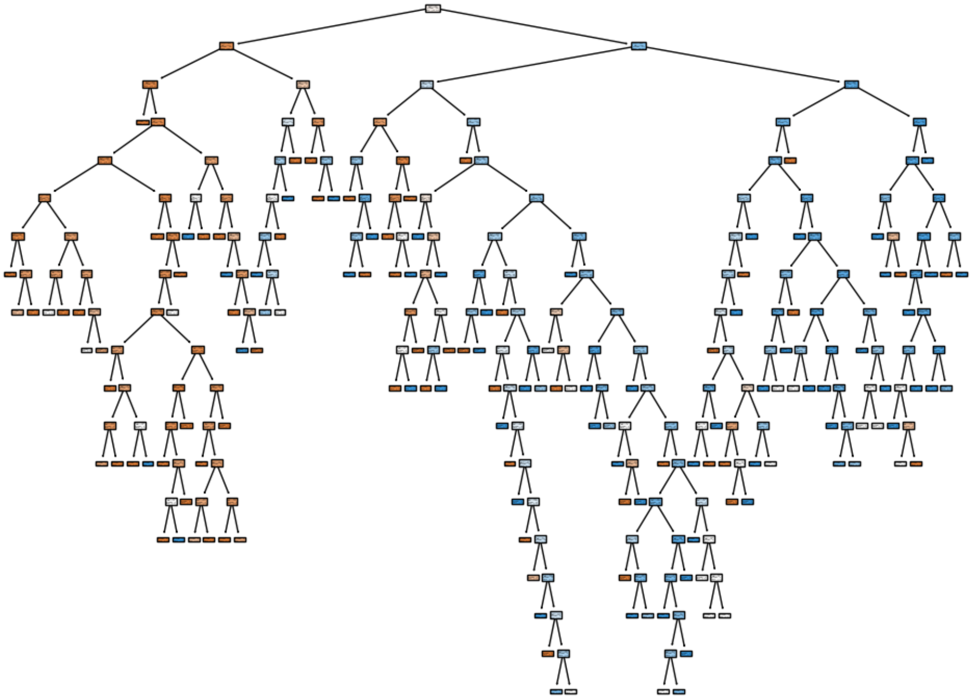
Assignment 3

1 Decision Tree Dataset:1

a)

Accuracy of Training data: **92.53%**
Accuracy of Test data: **69.17%**
Accuracy of Validation data: **76.03%**

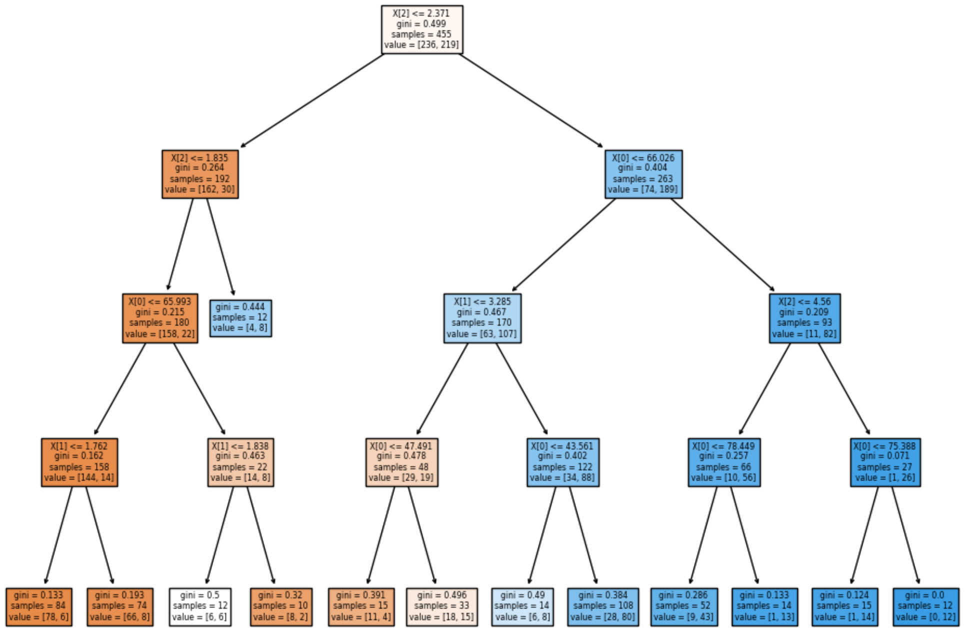


b)

Best Parameters

```
{'criterion': 'gini',
'max_depth': 5,
'min_samples_leaf': 2,
'min_samples_split': 7,
'splitter': 'random'}
```

Training data accuracy: **81.32%**
Test data accuracy: **76.23%**
Validation data accuracy: **86.77%**



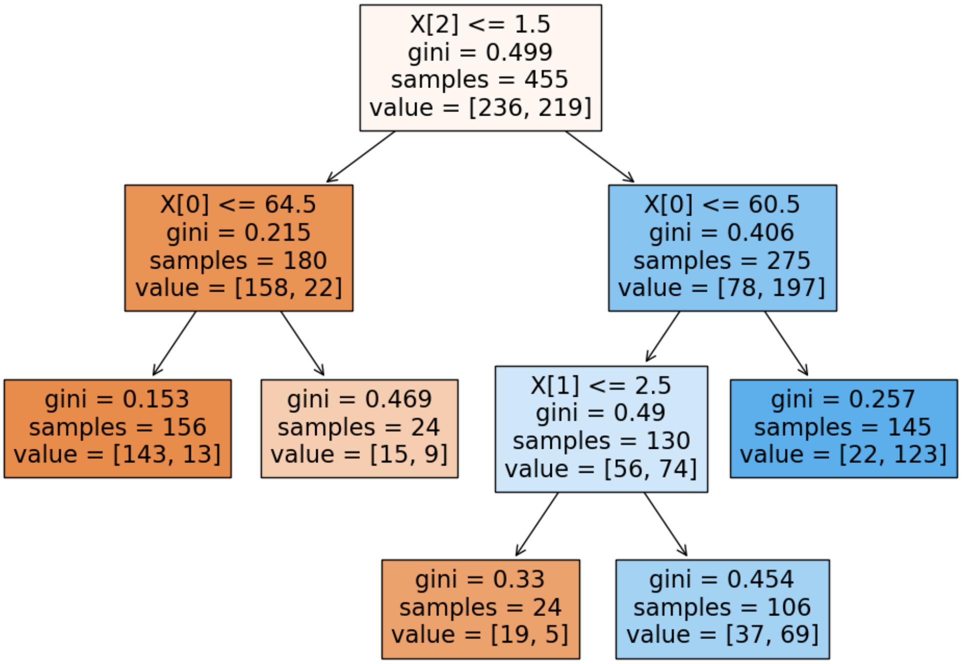
c)

Best performing pruned tree

Training data accuracy: **81.32%**

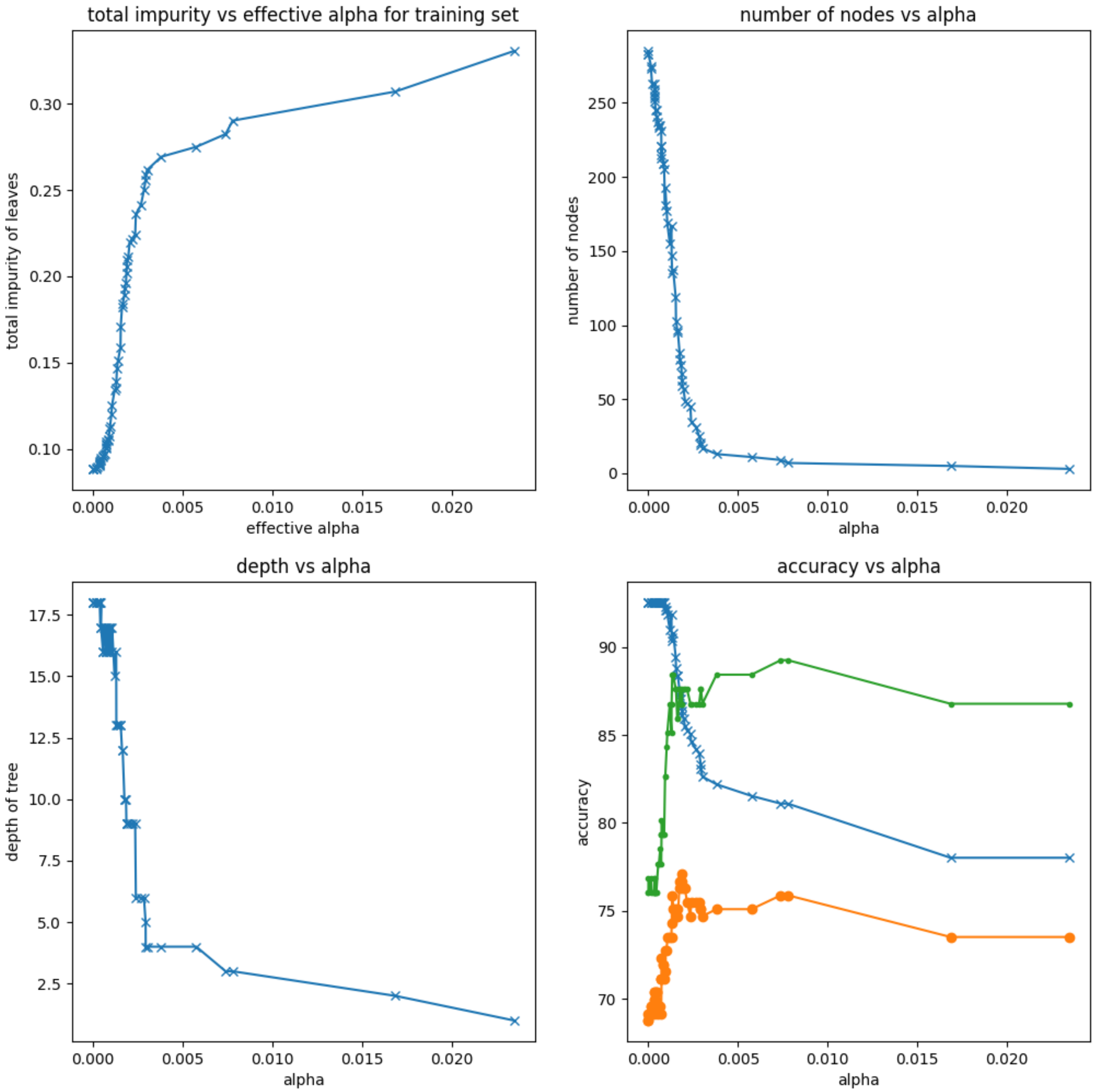
Test data accuracy: **76.88%**

Validation data accuracy: **89.25%**



Best alpha parameter

ccp_alpha for best fitting model based on validation split**: 0.0073871**



d) Random Forest

Best parameters

{'criterion': 'gini',

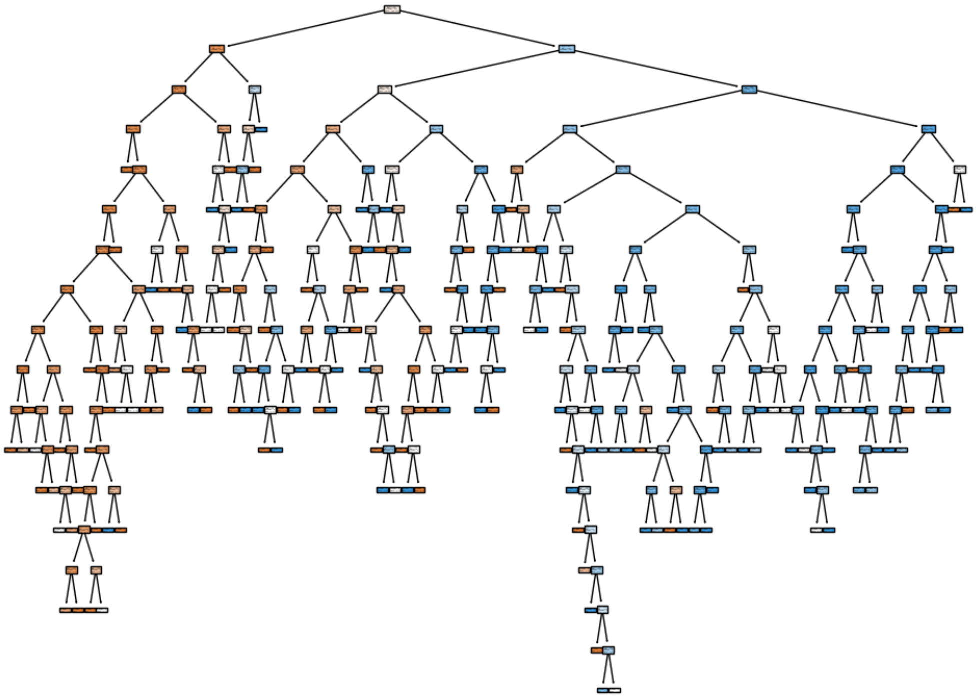
```
'max_features': 'sqrt',
'min_samples_leaf': 5,
'min_samples_split': 2,
'n_estimators': 30}
```

OOB Score: **78.46%**
Training data accuracy: **83.08%**
Test data accuracy: **77.07%**
Validation data accuracy: **88.43%**

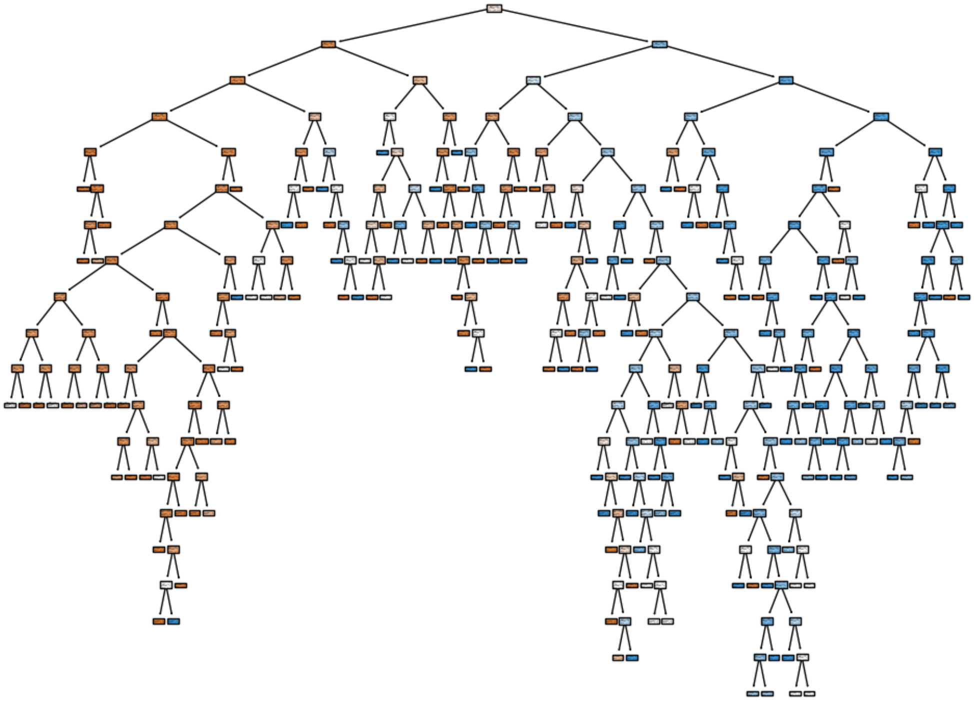
e) Replacing missing values with mode and median

a)

Median
Accuracy of Training data: **91.81%**
Accuracy of Test data: **72.57%**
Accuracy of Validation data: **73.34%**



Mode
Accuracy of Training data: **90.67%**
Accuracy of Test data: **71.19%**
Accuracy of Validation data: **77.04%**



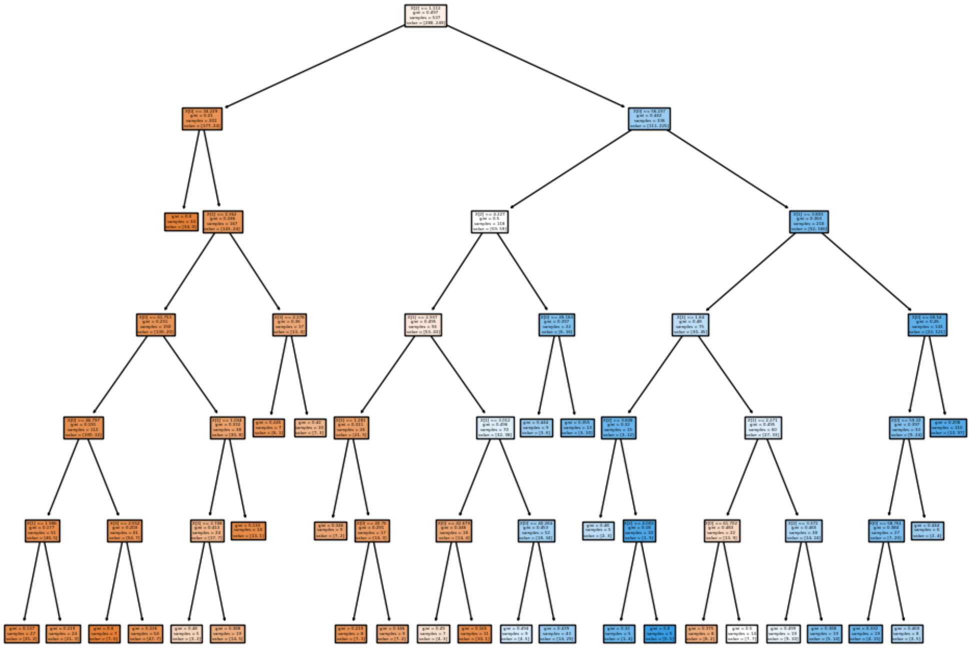
b)

Median

Training data accuracy: **80.45%**

Test data accuracy: **80.55%**

Validation data accuracy: **85.93%**

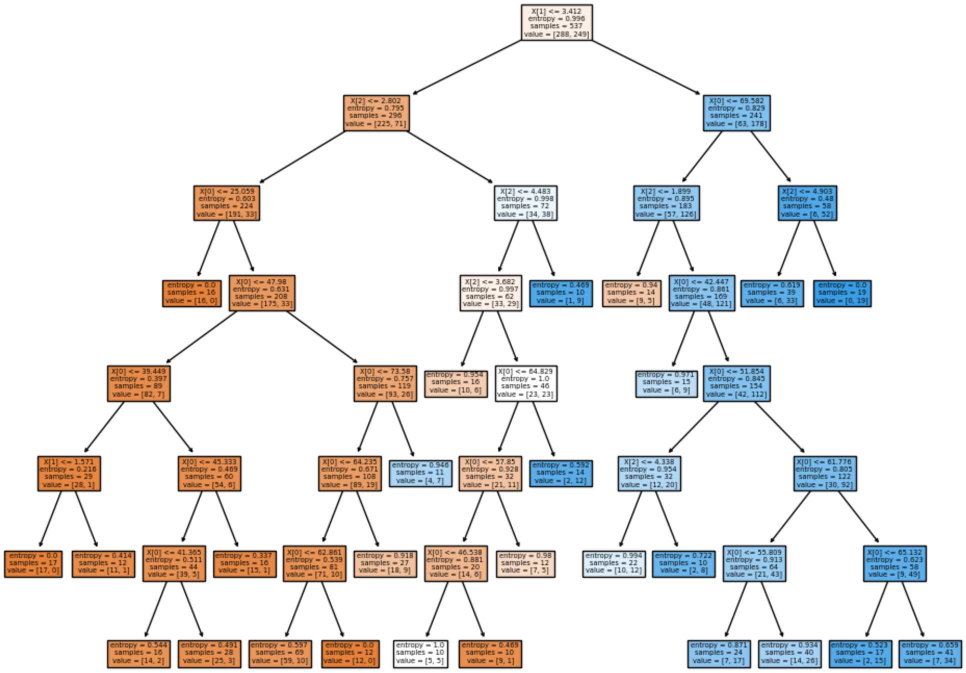


Mode

Training data accuracy: **78.58%**

Test data accuracy: **78.12%**

Validation data accuracy: **85.18%**


X[1] <= 3.412 entropy = 0.996 samples = 537 value = [288, 249]
X[2] <= 2.802 entropy = 0.795 samples = 296 value = [225, 71]
X[0] <= 69.582 entropy = 0.829 samples = 241 value = [63, 178]
X[0] <= 25.059 entropy = 0.603 samples = 224 value = [191, 33]
X[2] <= 4.483 entropy = 0.998 samples = 72 value = [34, 38]
X[2] <= 1.899 entropy = 0.895 samples = 183 value = [57, 126]
X[2] <= 4.903 entropy = 0.48 samples = 58 value = [6, 52]
entropy = 0.0 samples = 16 value = [16, 0]
X[0] <= 47.98 entropy = 0.631 samples = 208 value = [175, 33]
X[2] <= 3.682 entropy = 0.997 samples = 62 value = [33, 29]
entropy = 0.469 samples = 10 value = [1, 9]
entropy = 0.94 samples = 14 value = [9, 5]
X[0] <= 42.447 entropy = 0.861 samples = 169 value = [48, 121]
entropy = 0.619 samples = 39 value = [6, 33]
entropy = 0.0 samples = 19 value = [0, 19]
X[0] <= 39.449 entropy = 0.397 samples = 89 value = [82, 7]
X[0] <= 73.58 entropy = 0.757 samples = 119 value = [93, 26]
entropy = 0.954 samples = 16 value = [10, 6]
X[0] <= 64.829 entropy = 1.0 samples = 46 value = [23, 23]
entropy = 0.971 samples = 15 value = [6, 9]
X[0] <= 51.854 entropy = 0.845 samples = 154 value = [42, 112]
X[1] <= 1.571 entropy = 0.216 samples = 29 value = [28, 1]
X[0] <= 45.333 entropy = 0.469 samples = 60 value = [54, 6]
X[0] <= 64.235 entropy = 0.671 samples = 108 value = [89, 19]
entropy = 0.946 samples = 11 value = [4, 7]
X[0] <= 57.85 entropy = 0.928 samples = 32 value = [21, 11]
entropy = 0.592 samples = 14 value = [2, 12]
X[2] <= 4.338 entropy = 0.954 samples = 32 value = [12, 20]
X[0] <= 61.776 entropy = 0.805 samples = 122 value = [30, 92]
entropy = 0.0 samples = 17 value = [17, 0]
entropy = 0.414 samples = 12 value = [11, 1]
X[0] <= 41.365 entropy = 0.511 samples = 44 value = [39, 5]
entropy = 0.337 samples = 16 value = [15, 1]
X[0] <= 62.861 entropy = 0.539 samples = 81 value = [71, 10]
entropy = 0.918 samples = 27 value = [18, 9]
X[0] <= 46.538 entropy = 0.881 samples = 20 value = [14, 6]
entropy = 0.98 samples = 12 value = [7, 5]
entropy = 0.994 samples = 22 value = [10, 12]
entropy = 0.722 samples = 10 value = [2, 8]
X[0] <= 55.809 entropy = 0.913 samples = 64 value = [21, 43]
X[0] <= 65.132 entropy = 0.623 samples = 58 value = [9, 49]
entropy = 0.544 samples = 16 value = [14, 2]
entropy = 0.491 samples = 28 value = [25, 3]
entropy = 0.597 samples = 69 value = [59, 10]
entropy = 0.0 samples = 12 value = [12, 0]
entropy = 1.0 samples = 10 value = [5, 5]
entropy = 0.469 samples = 10 value = [9, 1]
entropy = 0.871 samples = 24 value = [7, 17]
entropy = 0.934 samples = 40 value = [14, 26]
entropy = 0.523 samples = 17 value = [2, 15]
entropy = 0.659 samples = 41 value = [7, 34]

c)

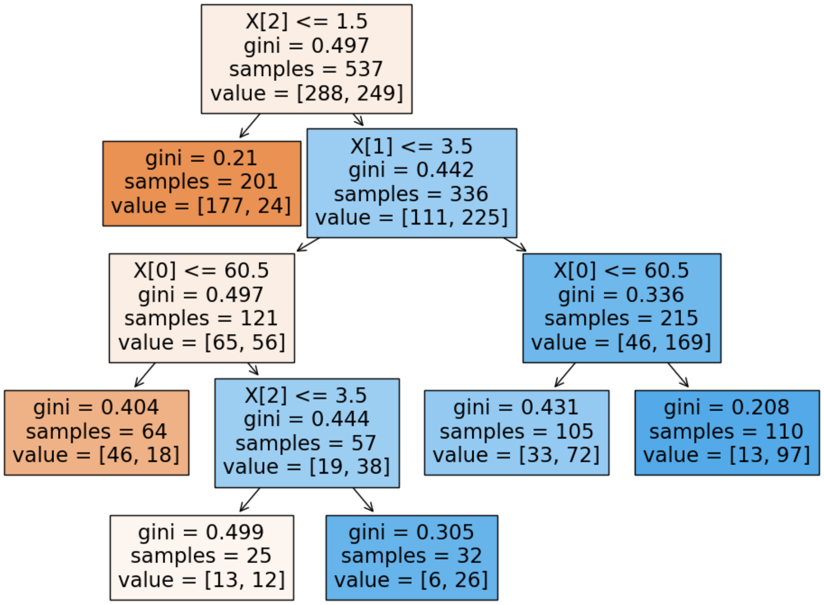
Median

ccp_alpha for best fitting model based on validation split: **0.005519**

Train accuracy of best tree: **80.26%**

Test accuracy of best tree: **79.16%**

Validation accuracy of best tree: **87.41%**



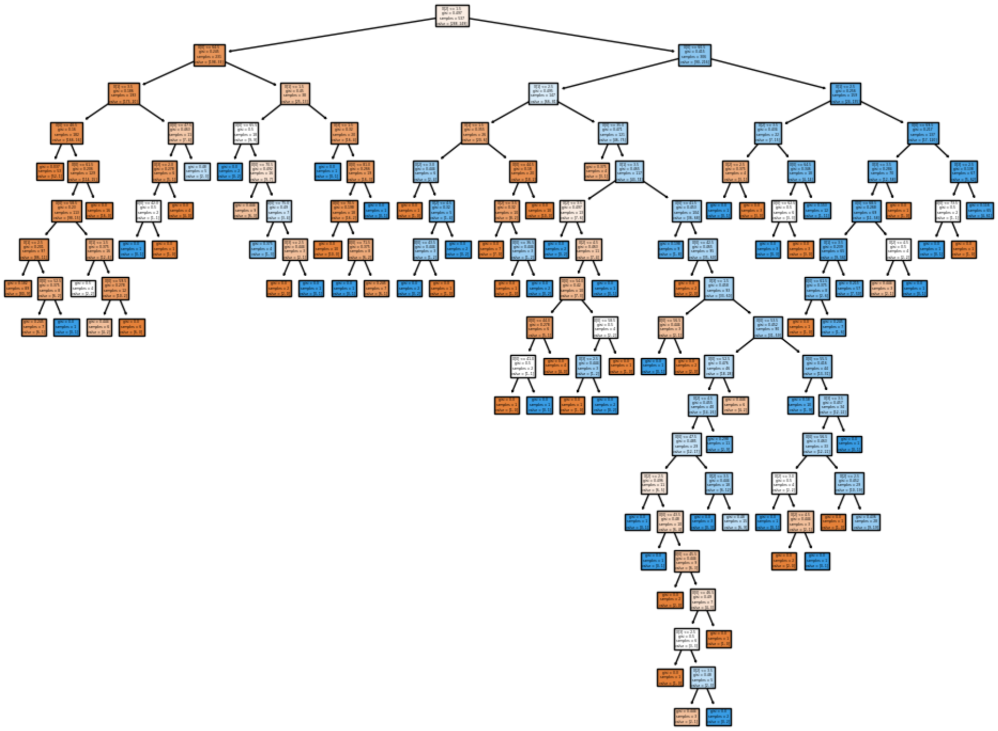
Mode

ccp_alpha for best fitting model based on validation split: **0.001474**

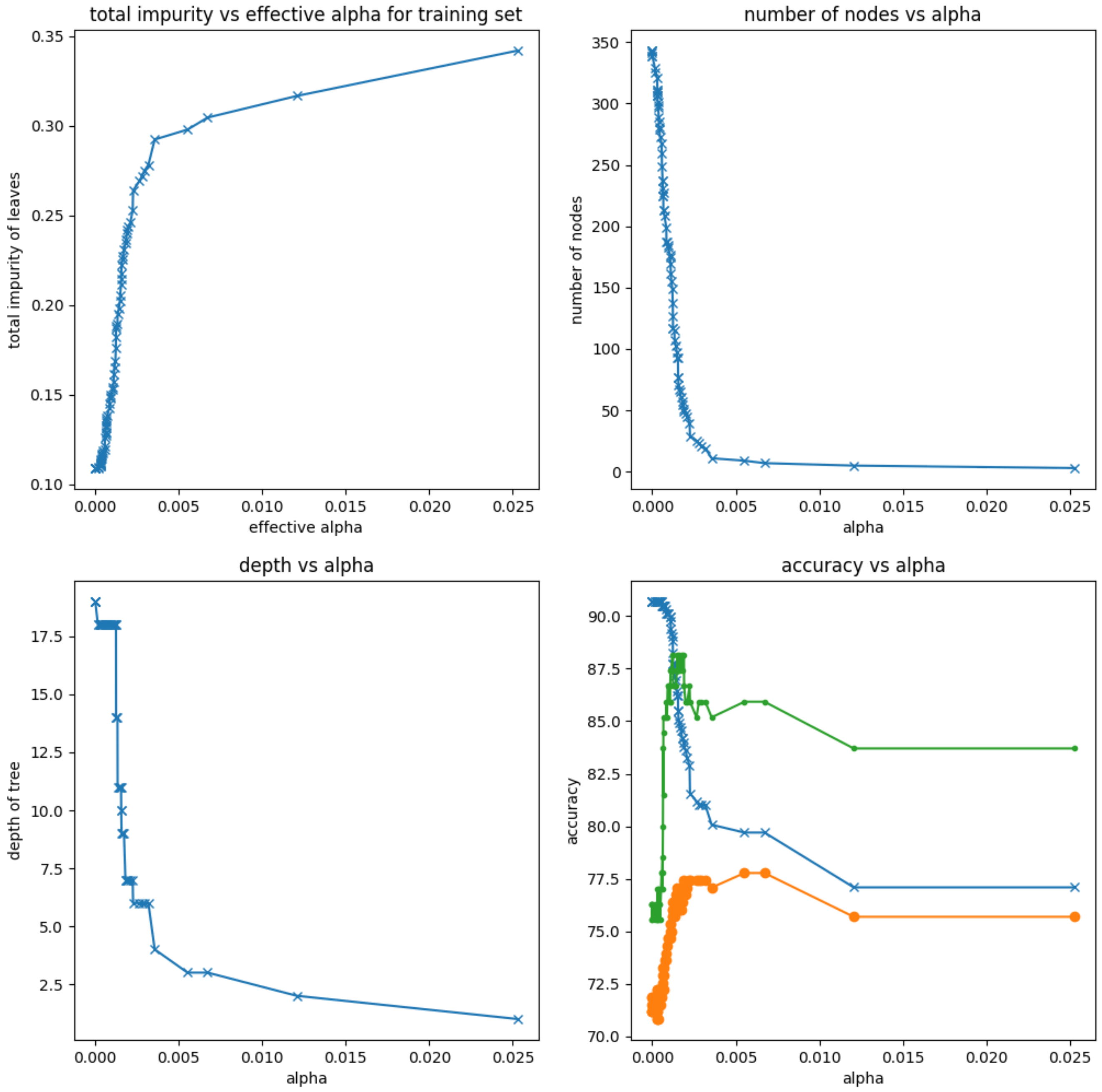
Train accuracy of best tree: **86.41%**

Test accuracy of best tree: **80.08%**

Validation accuracy of best tree: **88.14%**



Graphs for Best value for ccp_alpha parameter for mode model which is comparatively a better model



d)

Median

best parameters

```
{'criterion': 'entropy',
'max_features': 'sqrt',
'min_samples_leaf': 5,
'min_samples_split': 5,
'n_estimators': 50}
```

OOB Score: **77.47%**
Training data accuracy: **82.12%**
Test data accuracy: **79.16%**
Validation data accuracy: **85.93%**

Mode

best parameters

```
{'criterion': 'gini',
'max_features': 2,
'min_samples_leaf': 4,
'min_samples_split': 5, '
n_estimators': 70}
```

OOB Score: **77.28%**
Training data accuracy: **82.31%**
Test data accuracy: **80.21%**
Validation data accuracy: **85.18%**

f) XGBoost

Best Parameters:

```
{'max_depth': 10,
 'n_estimators': 10,
 'subsample': 0.5}
```

Training data accuracy: **83.61%**
Test data accuracy: **77.08%**
Validation data accuracy: **84.44%**

2 Decision Tree Dataset:2

a)

Training accuracy: **99.88%**
Testing accuracy: **57.24%**
Validation accuracy: **57.50%**

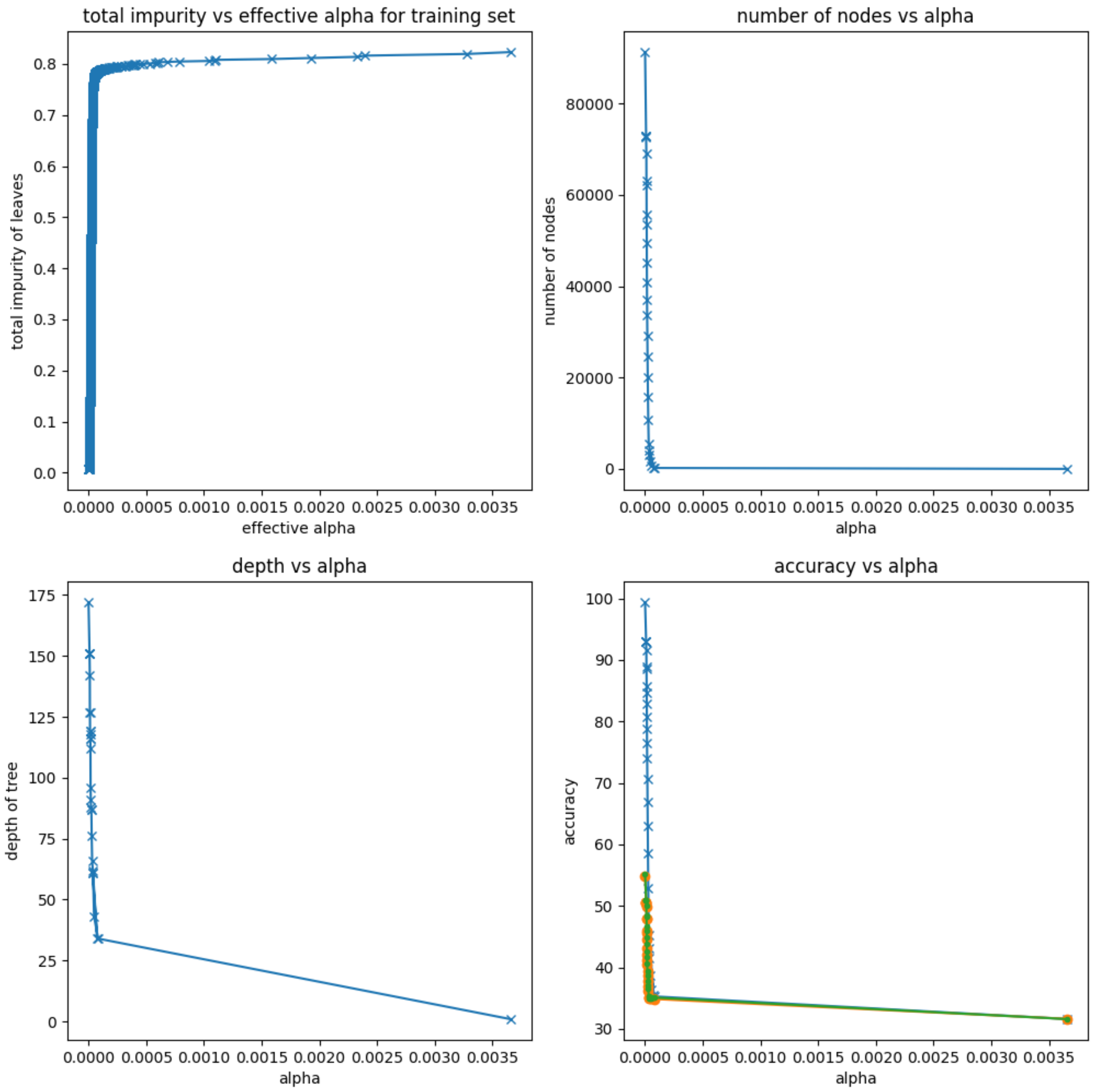
b)

Best Parameters:
{'max_depth': 100, 'min_samples_leaf': 1, 'min_samples_split': 2}

Training data accuracy: **99.04%**
Test data accuracy: **55.85%**
Validation data accuracy: **56.07%**

c)

ccp_alpha for best fitting model based on validation split: 0.0
Train accuracy of best tree: **99.34%**
Test accuracy of best tree: **54.92%**
Validation accuracy of best tree: **55.21%**



d)

Best Parameters:

{'max_features': 0.4, 'min_samples_split': 6, 'n_estimators': 50}

OOB Score: **62.67%**

Training data accuracy: **98.37%**

Test data accuracy: **62.32%**

Validation data accuracy: **63.10%**

e)

f)

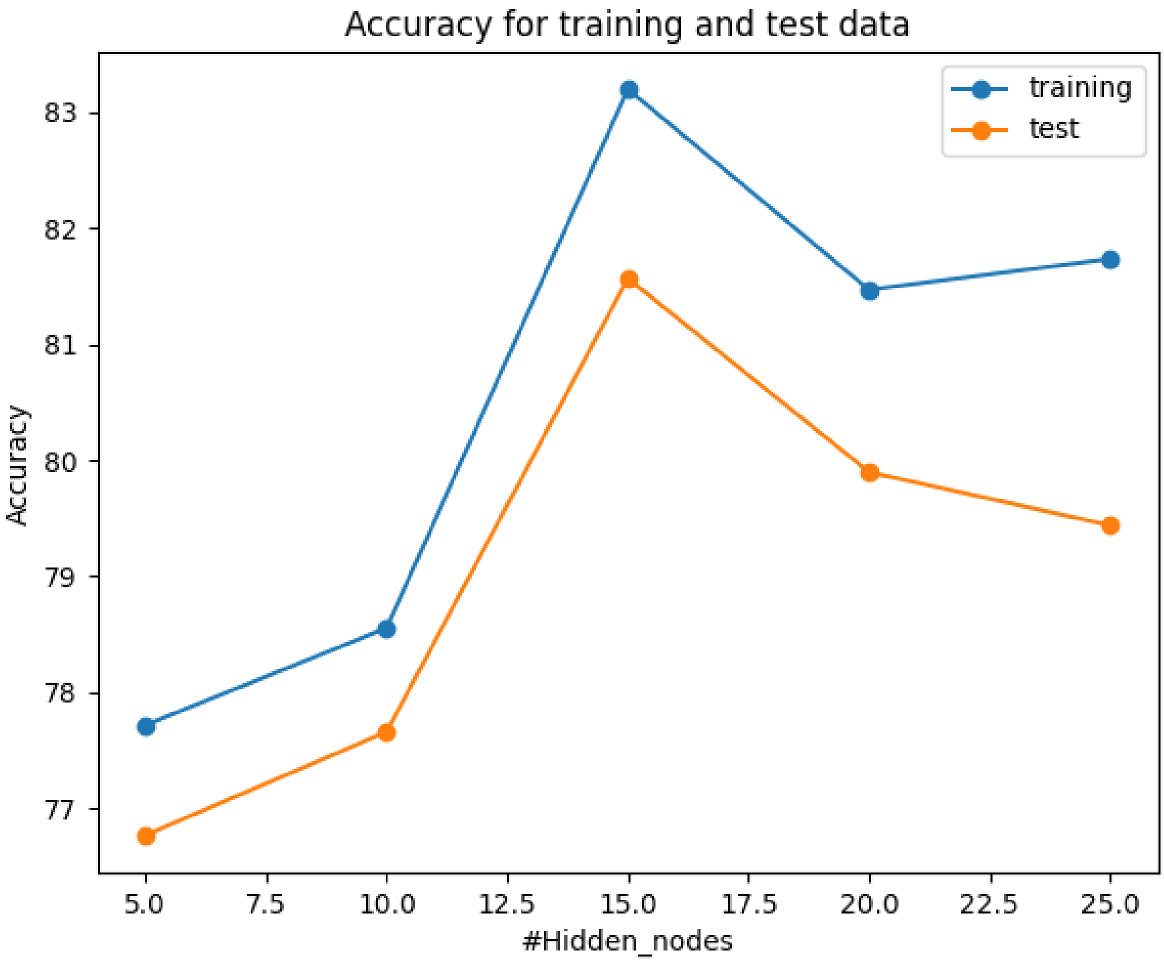
g)

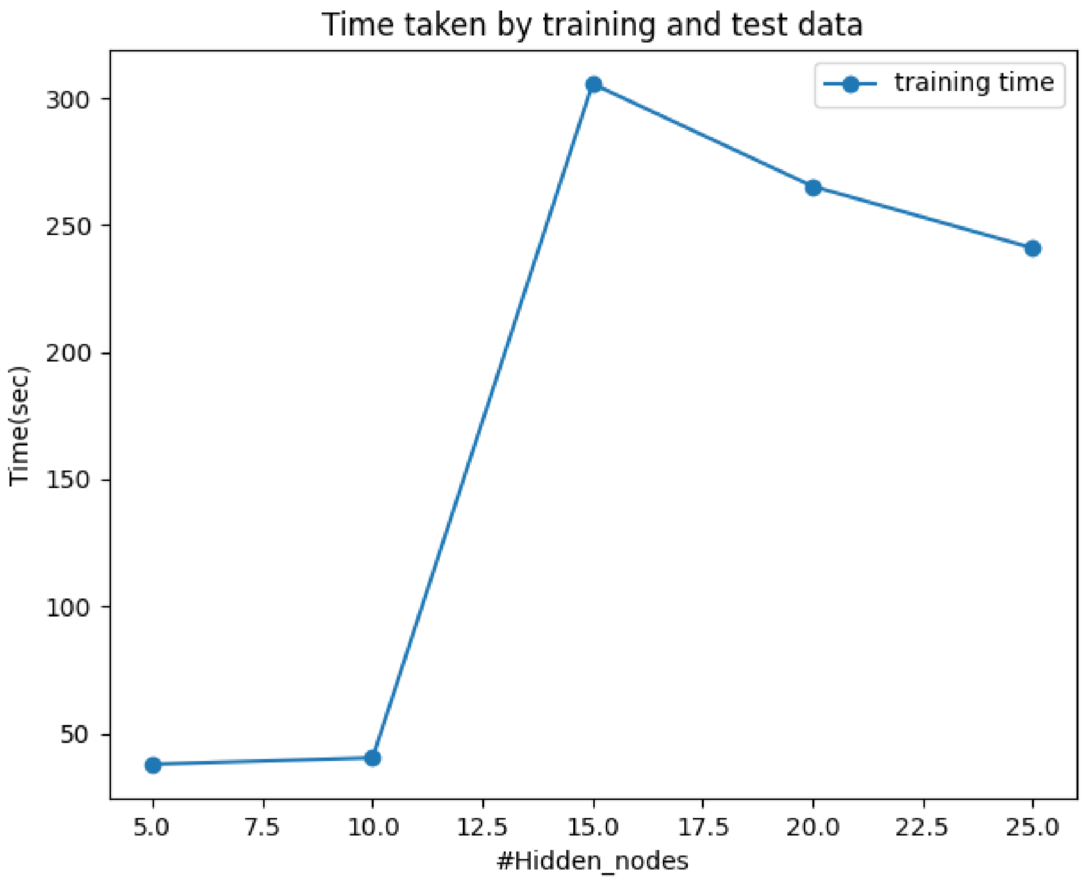
3 Neural Network:

a) Implemented the neural network interface as mentioned in the question
nn(
feature_layer=<integer>,
hidden_layer=<list>,
costFunction=['bce',’sgd’],
output_layer=<integer>,
batchSize=<integer>,
activationFunctions=dict{layer_number:[‘sigmoid’,’relu’]})

Convergence criteria: we set aside a fixed validation set of 5000 instance and after each epoch we calculate the network’s performance, when there is no significant change in the accuracy of the subsequent epoch we stop the learning

b) Hidden layer nodes: [5,10,15,20,25]





model 5 hidden layers

```
------------------------------------------------
Time taken to train: 37.8497838973999
Training Accuracy: 77.71629527158785%
Test Accuracy: 76.76767676767676%
----------------confusion matrix----------------
[[791  14  26 105  12   3 201   0   2   0]
 [  2 904   0   8   0   0   0   0   1   0]
 [ 10  12 503   5  70   0 106   0   9   0]
 [ 37  50  12 741  22   0  42   0   2   0]
 [ 33  15 276 104 724   0 119   0   8   2]
 [  2   0   1   1   0 887   0 140   6  43]
 [ 92   4 168  28 160   0 488   0  11   0]
 [  0   0   0   0   0  57   0 835   6  96]
 [ 32   1  13   8  11  13  44   1 946   1]
 [  1   0   1   0   1  40   0  24   9 857]]
---------------------------------------------
```

model 10 hidden layers

```
---------------------------------------------
Time taken to train: 40.40649962425232
Training Accuracy: 78.55297588293138%
Test Accuracy: 77.65776577657766%
----------------confusion matrix----------------
[[838  12  34  65   7   0 259   0  12   0]
 [  5 941   1  12   5   1   2   0   0   0]
 [ 17  10 581   7  66   0 121   0  19   1]
 [ 82  25  19 837  47   0  60   0   3   0]
 [ 11   8 309  38 824   0 395   0   7   0]
 [  2   0   1   1   1 839   1  25  12  18]
 [ 28   3  44  33  44   0 130   0   2   0]
 [  0   0   0   0   0  92   0 942   7  79]
 [ 17   1  11   7   6  14  32   0 935   3]
 [  0   0   0   0   0  54   0  33   3 898]]
---------------------------------------------
```

model 15 hidden layers

```
---------------------------------------------
Time taken to train: 305.5929822921753
Training Accuracy: 83.19971999533325%
Test Accuracy: 81.56815681568156%
----------------confusion matrix----------------
[[719   3   6  34   1   0 127   0   2   0]
 [  6 927   0   9   1   1   1   0   0   0]
 [  9  15 627  11  72   0  76   0  11   0]
 [ 84  41  15 850  24   2  62   0   9   1]
 [  9   8 197  52 771   0 129   0   8   0]
 [  2   0   0   0   0 906   1  48  10  13]
 [159   5 148  39 123   0 582   0  26   1]
 [  0   0   0   0   0  52   1 915   4  53]
 [ 12   1   7   3   8   4  21   0 929   1]
 [  0   0   0   2   0  35   0  37   1 930]]
---------------------------------------------
```

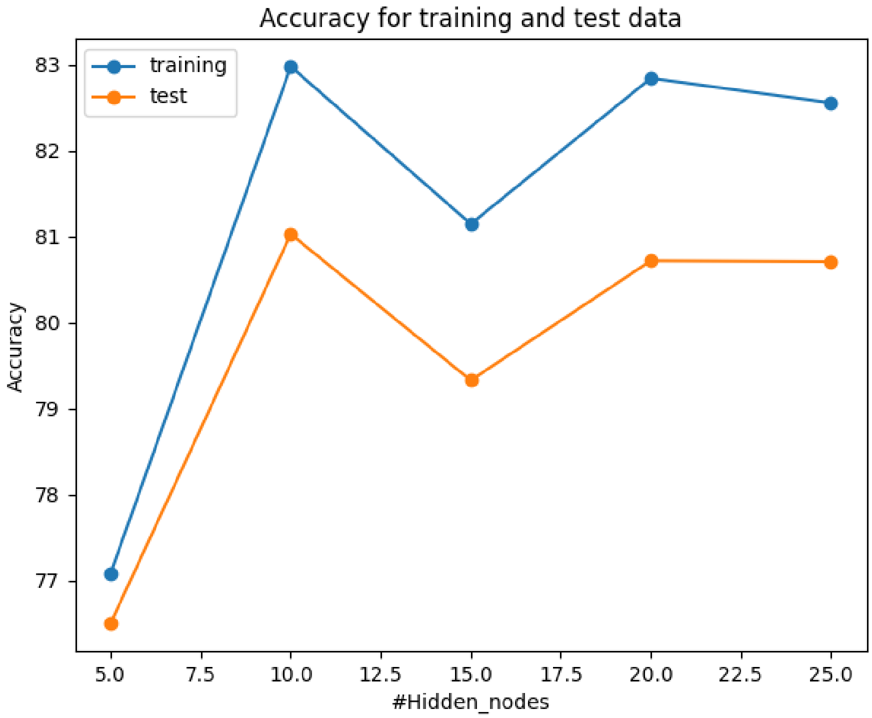
model 20 hidden layers

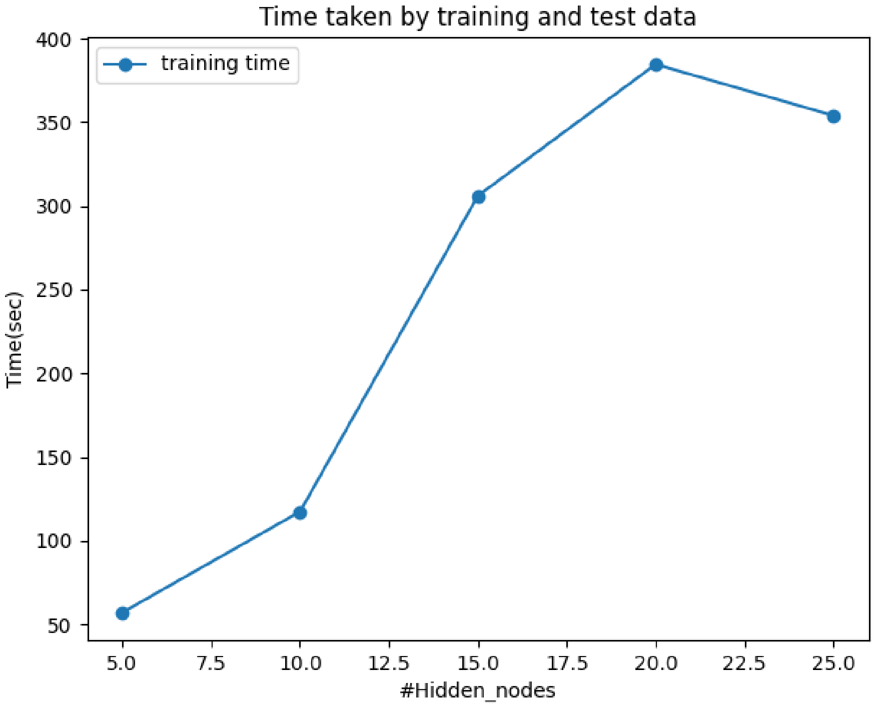
```
---------------------------------------------
Time taken to train: 265.26688027381897
Training Accuracy: 81.46969116151936%
Test Accuracy: 79.89798979897989%
----------------confusion matrix----------------
[[810   5  45  51   9   0 233   0  12   0]
 [  1 936   5  17   0   0   5   0   1   0]
 [ 21  16 784  16 229   0 197   0  13   0]
 [ 57  28  11 796  37   1  45   0   3   0]
 [ 19   9 118  49 653   0 107   0   5   0]
 [  2   0   0   2   0 894   1  69   6  30]
 [ 73   6  25  64  64   0 384   0   4   0]
 [  0   0   1   0   0  44   0 864   4  43]
 [ 17   0   8   3   8  17  25   1 943   1]
 [  0   0   3   2   0  44   3  66   9 925]]
---------------------------------------------
```

model 25 hidden layers

```
---------------------------------------------
Time taken to train: 241.0044286251068
Training Accuracy: 81.73636227270454%
Test Accuracy: 79.43794379437944%
----------------confusion matrix----------------
[[756   4  12  45  10   0 200   0   4   0]
 [  3 933   2  19   5   1   5   0   1   0]
 [ 78  15 842  12 277   0 246   0  56   1]
 [ 57  32  12 818  39   0  39   0   5   0]
 [ 19  13  82  43 585   0 105   0   6   0]
 [  2   0   2   4   0 899   1  58   9  24]
 [ 73   3  42  56  75   0 393   0  15   0]
 [  0   0   0   0   0  44   0 886   5  35]
 [ 11   0   5   2   9   9  11   0 892   0]
 [  1   0   1   1   0  47   0  56   7 939]]
---------------------------------------------
```

c) Adaptive learning





model 5 hidden layers

```
----------------------------------------------
Time taken to train: 57.03181791305542
Training Accuracy: 77.08961816030268%
Test Accuracy: 76.5076507650765%
----------------confusion matrix----------------
[[857   5  48  91  13   0 335   0   8   0]
 [  4 926   7   8   3   1   2   0   2   0]
 [ 18  16 691  25 136   0 227   0   7   0]
 [ 45  40  10 823  45   0  29   0   7   1]
 [ 14   6 214  10 692   0 271   0  14   0]
 [  2   0   0   1   0 863   1  78  12  33]
 [ 39   5  13  38 104   0  94   0   5   0]
 [  0   0   0   1   0  79   0 895   7  91]
 [ 21   2  17   3   7  16  41   0 937   2]
 [  0   0   0   0   0  41   0  27   1 872]]
----------------------------------------------
```

model 10 hidden layers

```
----------------------------------------------
Time taken to train: 117.32068991661072
Training Accuracy: 82.97638293971566%
Test Accuracy: 81.02810281028103%
----------------confusion matrix----------------
[[803   4  45  47   4   0 237   0  14   0]
 [  2 929   7  16   4   0   2   0   0   0]
 [ 31  18 719  16 122   0 140   0  12   0]
 [ 78  34  14 854  28   1  61   1   9   0]
 [  8   4 138  37 752   0 128   0   4   0]
 [  3   0   2   1   1 911   0  67  10  26]
 [ 51   7  58  24  83   1 392   0  12   0]
 [  0   0   0   0   0  39   0 871   7  28]
 [ 24   2  17   5   6  12  40   2 932   6]
 [  0   2   0   0   0  36   0  59   0 939]]
----------------------------------------------
```

model 15 hidden layers

```
---------------------------------------------
Time taken to train: 305.9366545677185
Training Accuracy: 81.14635243920732%
Test Accuracy: 79.32793279327933%
----------------confusion matrix----------------
[[764   2  19  30   1   1 172   0   3   0]
 [  2 934   4  18   1   0   0   0   3   1]
 [ 28  10 682  17 156   0 122   0  15   0]
 [ 81  38  16 862  69   1  65   0   7   0]
 [ 15  11 103  33 659   1 147   0  25   0]
 [  2   1   0   5   1 872   1  49  10  37]
 [ 94   0 172  27  95   2 476   0  20   2]
 [  1   1   0   3   3  67   0 853   3  40]
 [ 12   3   4   5  15   8  16   3 912   1]
 [  1   0   0   0   0  48   1  95   2 918]]
---------------------------------------------
```

model 20 hidden layers

```
---------------------------------------------
Time taken to train: 384.5056827068329
Training Accuracy: 82.83971399523325%
Test Accuracy: 80.71807180718072%
----------------confusion matrix----------------
[[765  10  12  60  21   0 193   0   2   0]
 [  5 932   3  14   1   0   0   0   1   0]
 [ 15  10 675  10 116   0  99   0   8   1]
 [ 38  29   9 796  28   1  30   0   3   0]
 [ 15  11 144  47 725   0 141   0   7   0]
 [  3   0   1   4   0 890   2  28  16  29]
 [141   6 145  65 100   0 511   0  26   0]
 [  0   0   1   0   0  62   0 923   3  47]
 [ 18   2  10   2   9   7  24   0 933   1]
 [  0   0   0   2   0  40   0  49   1 921]]
---------------------------------------------
```

model 25 hidden layers

```
---------------------------------------------
Time taken to train: 353.88060450553894
Training Accuracy: 82.55470924515409%
Test Accuracy: 80.7080708070807%
```

```
----------------confusion matrix----------------
[[822   3  43  72  25   0 241   0  10   0]
 [  1 939   4  12   6   0   4   0   1   0]
 [ 25  20 791  14 191   0 173   0  20   0]
 [ 54  30  10 816  33   3  35   0   4   1]
 [  8   7  97  40 659   0 120   0   8   0]
 [  2   1   0   2   0 893   2  58   8  24]
 [ 72   0  47  32  74   2 394   0   5   0]
 [  0   0   0   1   1  45   0 889   6  41]
 [ 16   0   7  11  10  15  31   2 937   3]
 [  0   0   1   0   1  42   0  51   1 930]]
------------------------------------------------
```

d) Several activation units: Sigmoid and ReLU

model sigmoid [100, 100] hidden layers

```
------------------------------------------------
Time taken to train: 2944.3284707069397
Training Accuracy: 77.20628677144619%
Test Accuracy: 76.58765876587658%
----------------confusion matrix----------------
[[811   4  36  39   1   0 263   0   5   0]
 [  7 931   5  12   8   1   3   0   1   0]
 [ 12  14 678   7 197   0 223   0   9   0]
 [111  42  12 870  60   1  73   0  17   0]
 [ 12   7 209  30 682   0 240   0   2   0]
 [ 10   0   6   4   5 836  17  64  16  28]
 [ 14   0  34  34  37   0 138   0   1   0]
 [  0   0   0   0   0 103   1 852   8  48]
 [ 23   2  20   4  10   8  42   1 938   1]
 [  0   0   0   0   0  51   0  83   3 922]]
------------------------------------------------
```

ReLU model [100, 100] hidden layers

```
--------------------------------------------------------
Time taken to train: 573.2644910812378
Training Accuracy: 86.76311271854532%
Test Accuracy: 84.85848584858486%
```

```
----------------confusion matrix----------------
[[815   4  15  29   0   0 139   0   0   0]
 [  4 955   3  19   1   1   5   0   1   0]
 [ 10   5 762  13 103   0 103   0   7   0]
 [ 48  27  10 861  34   1  42   0   5   0]
 [  4   6 125  37 763   0  79   0   6   0]
 [  6   0   1   1   0 912   0  36   7  21]
 [101   2  78  35  92   0 610   0  17   0]
 [  0   0   0   0   0  55   1 922   4  44]
 [ 12   1   6   4   7   5  21   1 952   1]
 [  0   0   0   1   0  26   0  41   1 933]]
------------------------------------------------
```

e) Varying the number of hidden layers

ReLU

model [50, 50] hidden layers

```
------------------------------------------------
Time taken to train: 687.1628496646881
Training Accuracy: 83.64806080101334%
Test Accuracy: 82.34823482348234%
------------------------------------------------
```

model [50, 50, 50] hidden layers

```
------------------------------------------------
Time taken to train: 949.8115842342377
Training Accuracy: 84.51474191236521%
Test Accuracy: 83.1983198319832%
------------------------------------------------
```

model [50, 50, 50, 50] hidden layers

```
------------------------------------------------
Time taken to train: 807.3541448116302
Training Accuracy: 84.23973732895548%
Test Accuracy: 82.67826782678267%
------------------------------------------------
```

model [50, 50, 50, 50, 50] hidden layers

```
---------------------------------------------
Time taken to train: 971.501805305481
Training Accuracy: 84.59640994016567%
Test Accuracy: 83.1083108310831%
---------------------------------------------
```

Sigmoid

model [50, 50] hidden layers

```
---------------------------------------------
Time taken to train: 1145.8908560276031
Training Accuracy: 74.76291271521193%
Test Accuracy: 74.22742274227423%
---------------------------------------------
```

model [50, 50, 50] hidden layers

```
---------------------------------------------
Time taken to train: 2317.100964307785
Training Accuracy: 47.79412990216504%
Test Accuracy: 47.814781478147815%
---------------------------------------------
```

model [50, 50, 50, 50] hidden layers

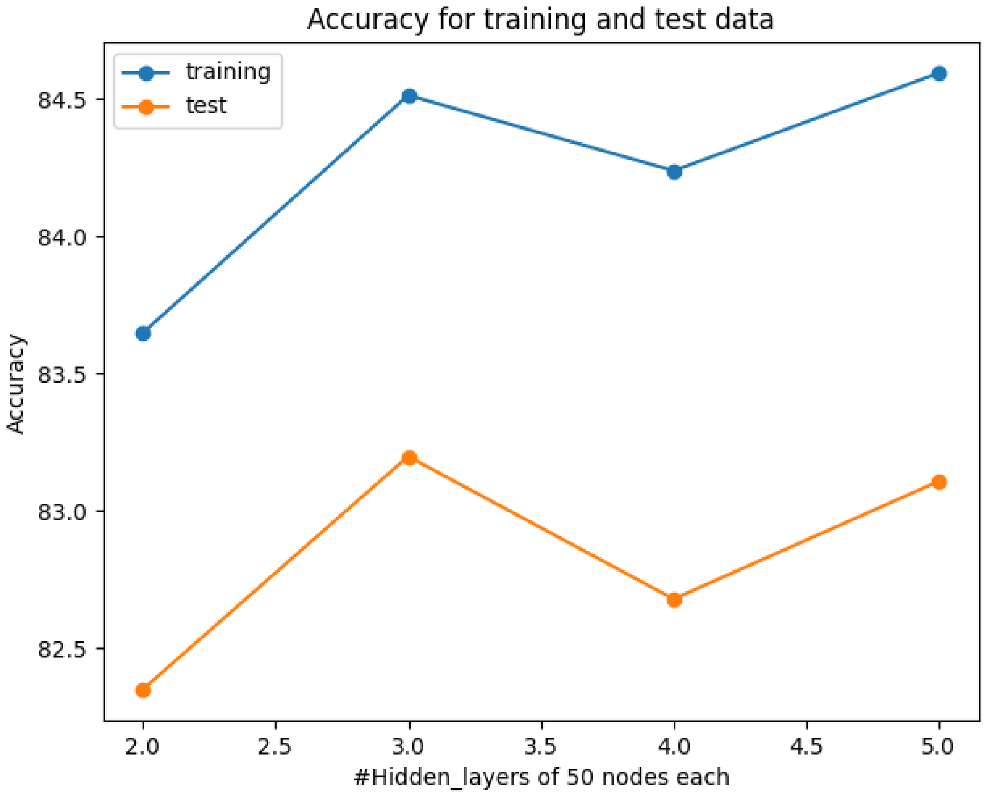
```
---------------------------------------------
Time taken to train: 2650.133682012558
Training Accuracy: 27.13378556309272%
Test Accuracy: 27.322732273227324%
---------------------------------------------
```

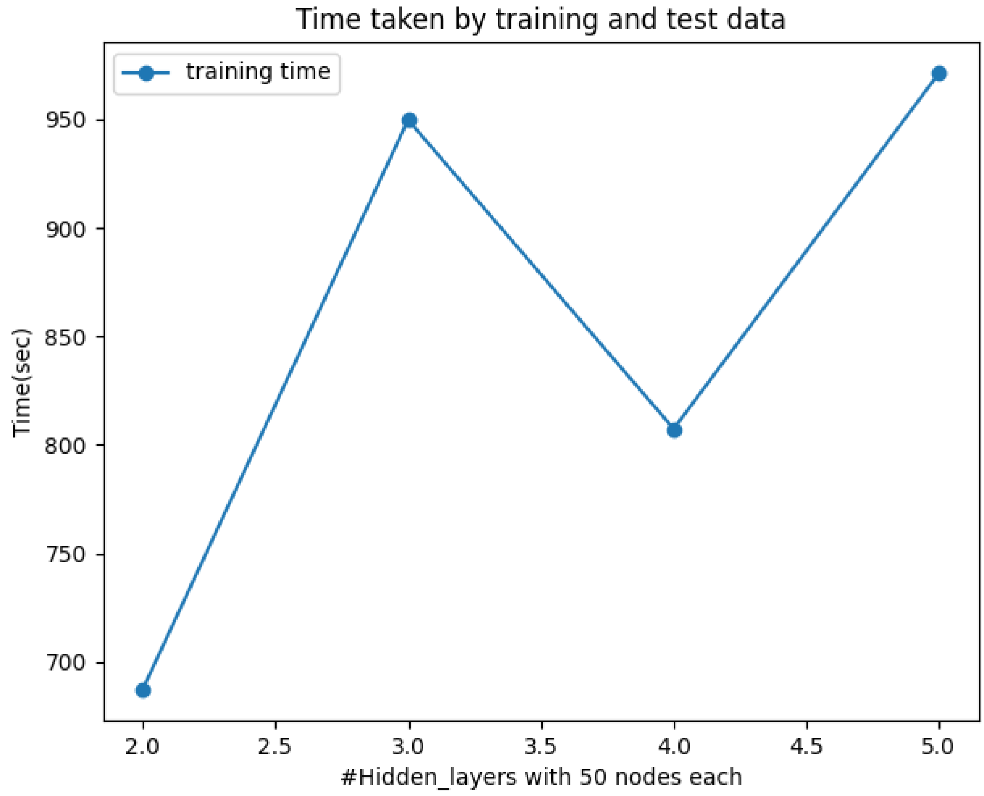
model [50, 50, 50, 50, 50] hidden layers

```
---------------------------------------------
Time taken to train: 2852.7096073627472
Training Accuracy: 10.051834197236621%
Test Accuracy: 10.021002100210021%
---------------------------------------------
```

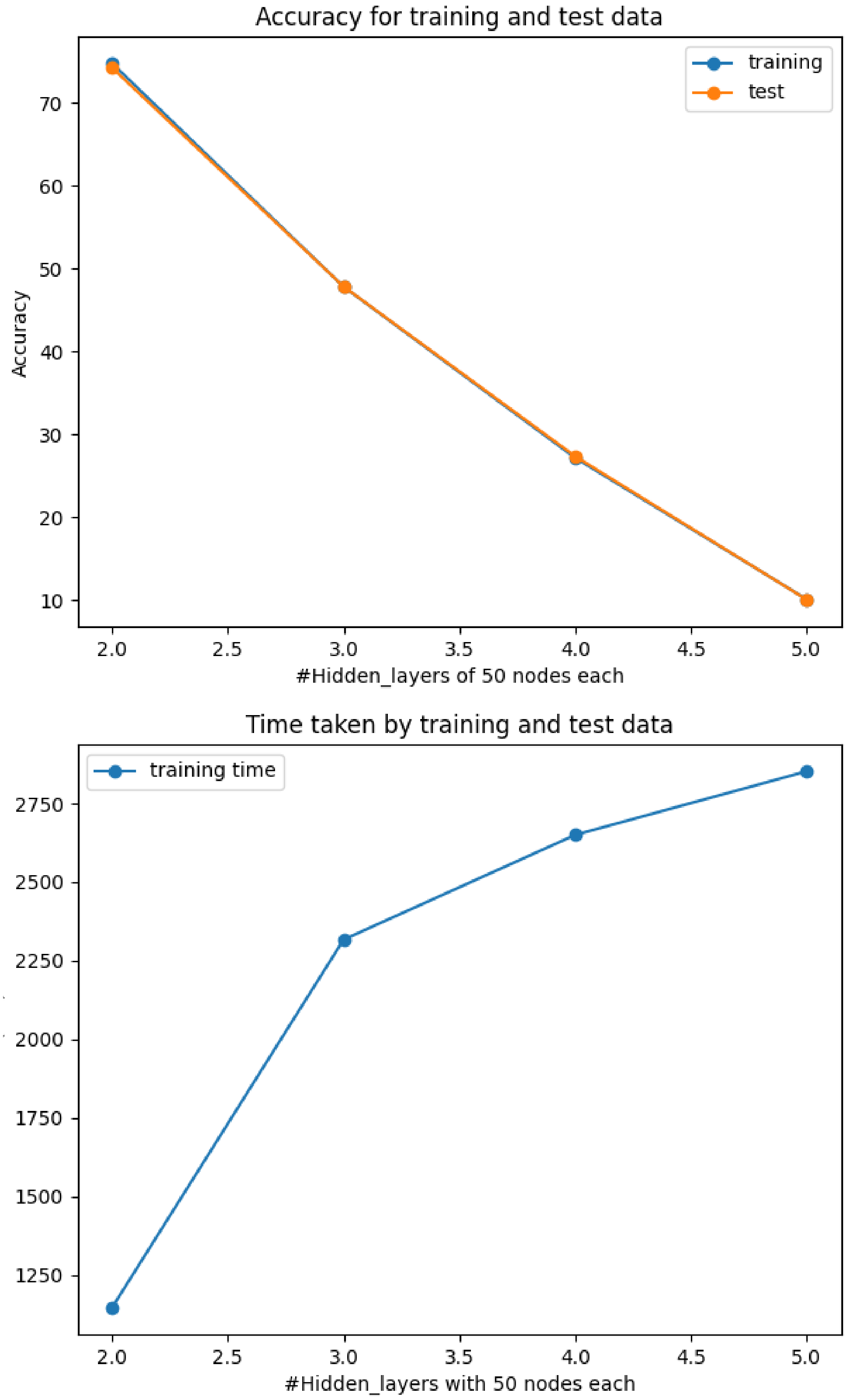
Best performing model is **ReLU** with 3 **[50, 50, 50]** hidden layers. As we start increasing the number of layers in sigmoid model it’s performance started deteriorating.

Graphs for **ReLU** activation function in Hidden layers





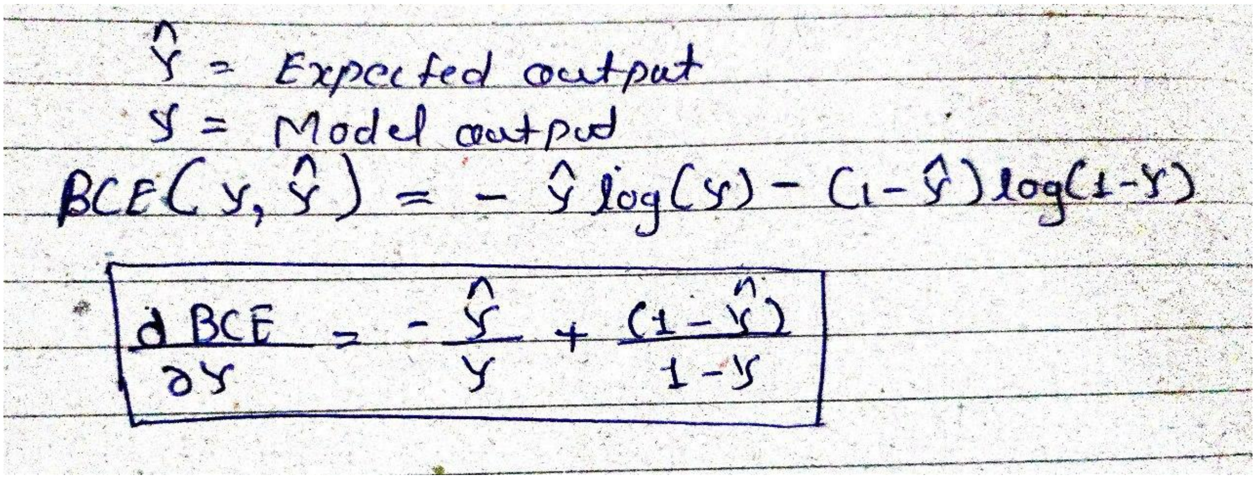
Graphs for **Sigmoid** activation function in Hidden layers



Accuracy for training and test data

Time taken by training and test data

f) Binary-Cross Entropy (BCE) loss



$\hat{Y}$ = Expected output

Y = Model output

$$BCE(Y, \hat{Y}) = -\hat{Y}\log(Y) - (1-\hat{Y})\log(1-Y)$$

$$* \boxed{\frac{d\,BCE}{\partial Y} = -\frac{\hat{Y}}{Y} + \frac{(1-\hat{Y})}{1-Y}}$$

Best model of 2e) three [50, 50, 50] hidden layers

```
-----------------------------------------------
Time taken to train: 1273.2066054344177
Training Accuracy: 86.60811013516891%
Test Accuracy: 84.74847484748474%
-----------------------------------------------
```

Training set

```
----------------confusion matrix----------------
[[4940   10   68  169 8       1  793 0   15  0]
 [  16 5783   11   63   10     0   18  0       3        2]
 [  95   39 4715   43  517    1  656 0   23  0]
 [ 330  135   68 5378  212  5  225 0   42  0]
 [  16   14  680  197 4837   0  544 0   20  0]
 [   8    0        4        0       1 5566 0  139   19   96]
 [ 541   16  427  133  390   3 3687 0  103          1]
 [   0    0        1        0       0  275 3 5603   24  189]
 [  53   3   25   15   24   29   73   17 5747 3]
 [   1    0        1        2       1  120 1  241 4 5708]]
-----------------------------------------------
```

Test set

```
----------------confusion matrix----------------
[[802   2  16  24   0   0 145   0   0   0]
 [  2 953   3  14   2   0   1   0   0   0]
 [ 14   9 760   5 109   0 121   0   5   0]
 [ 59  29  17 874  40   1  49   0   6   0]
 [  6   4 115  36 779   0  94   0  10   0]
 [  1   0   1   1   0 911   2  28  10  15]
 [104   3  85  41  64   0 573   0  16   1]
 [  0   0   0   0   0  52   0 933   4  43]
 [ 11   0   3   4   6   5  15   1 949   0]
 [  1   0   0   1   0  31   0  38   0 940]]
-----------------------------------------------
```

g) MLPClassifier from scikit-learn library

```
Iteration 200, loss = 0.00206089
Time taken to train: 906.4067075252533
model [50, 50, 50] hidden layers
Training Accuracy: 99.99%
Test Accuracy: 86.75%
```