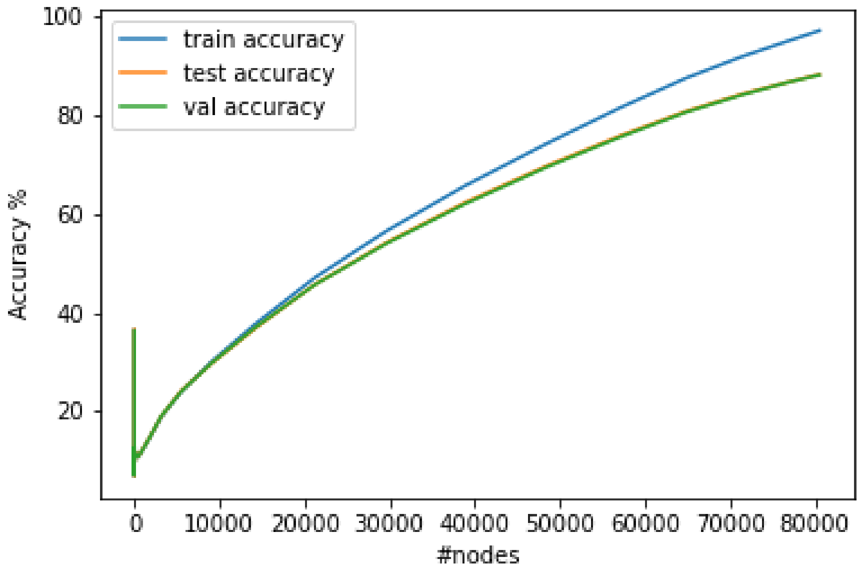
COL774 Assignment 3

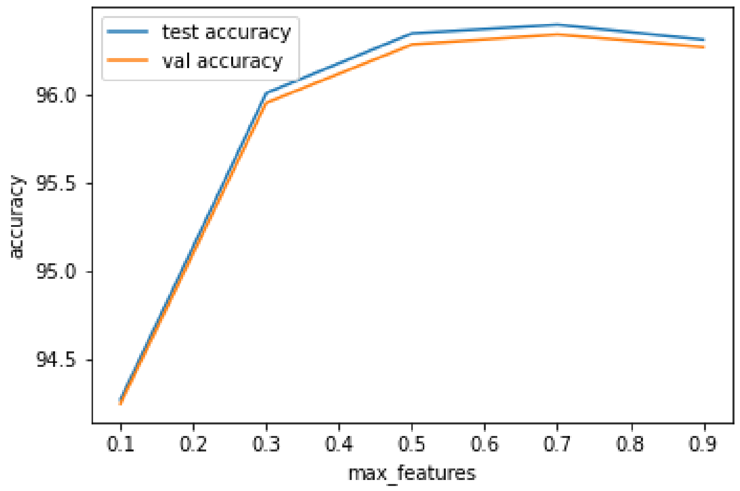
Soham Gaikwad, 2018CS10394

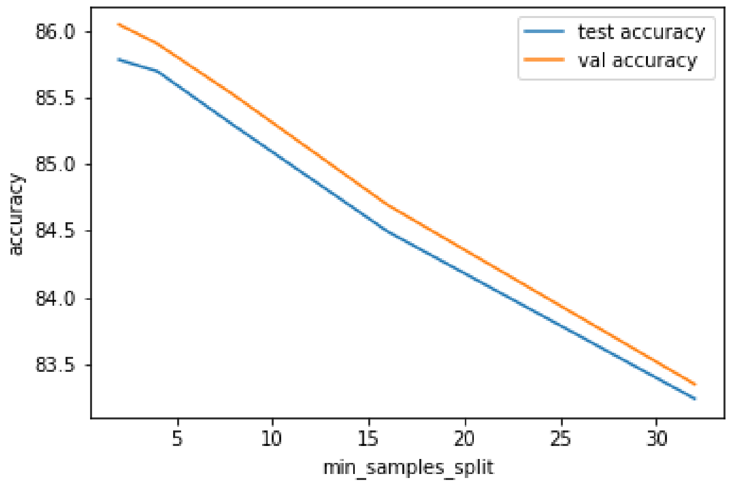
1. Decision Trees and Random Forests
   a. For choosing best attribute, maximum mutual information (calculated using entropy) was used. Node class is used to represent the nodes in the Decision tree. The tree is grown using the grow_tree function in the DecisionTree class.
      The tree was grown with different depths and the corresponding number of nodes and accuracies were noted as follows.

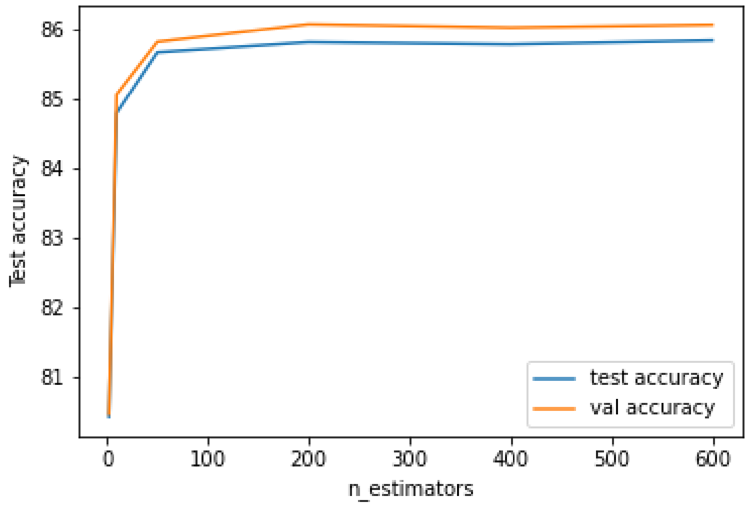
      

      Around 34% accuracy was observed when number of nodes was 1. In this case, the one node predicted according to the majority class of the total dataset, which is basically equivalent to majority prediction and after this accuracy kept increasing to converge near 99% as nodes were increased.
   b. Post pruning the Tree was done by looking through all current nodes and then pruning the node (subtree) which give max increase in val test accuracy until there was no node giving increase in val test accuracy. This was implemented but found to be too slow.
   c. Grid search was done over the parameter space and optimal parameters found were: n_estimators = 450, max_features = 0.7 and min_samples_split = 2.
      Training accuracy: 99.96%
      Test accuracy: 96.29%
      Validation accuracy: 96.24%
      Out-of-bag score: 96.24%
   d. After finding optimal parameters, they were varied in a range. Note that this was done on a subset of data as on the whole data, it was taking huge amounts of time even with enabling parallel tasks from sklearn.

It was observed that for n_estimators below 100, the accuracy was quite low. The accuracy kept increasing as n_estimators was increased but not by much.
For max_features, 0.7 was optimum and accuracy decreased by little on both sides.
As min_samples_split was increased from 2, accuracy kept decreasing. It seemed that model was quite sensitive to this parameter.

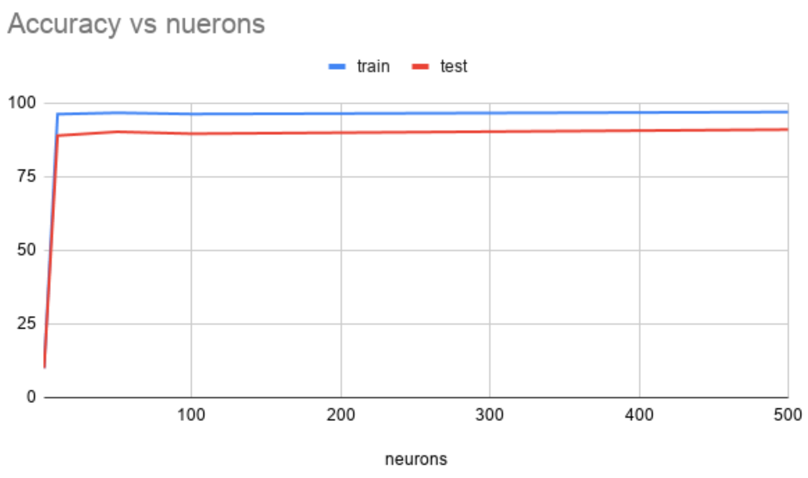


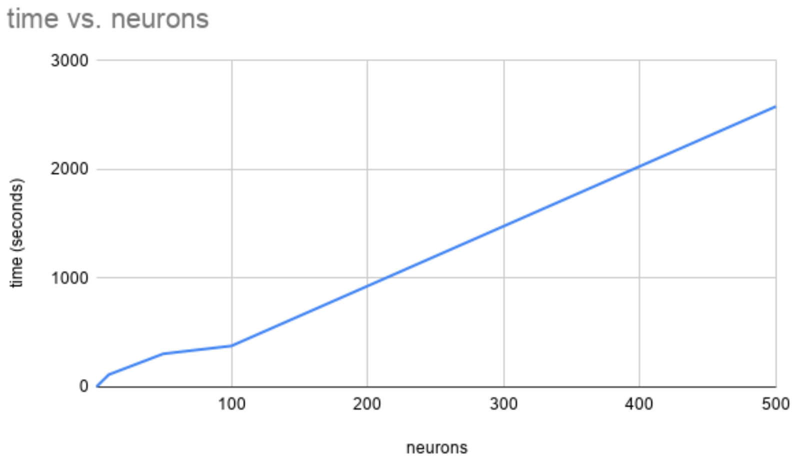




2. Neural Networks
   a. A generic nueral network implentation was done using the first principles. For initializing theta (parmeters of model), He initialization was found to give best results as compared with zero initialization and basic random initialization.
   b. Stopping criteria used was error threshold of 1e-4 for 2 consecutive times.

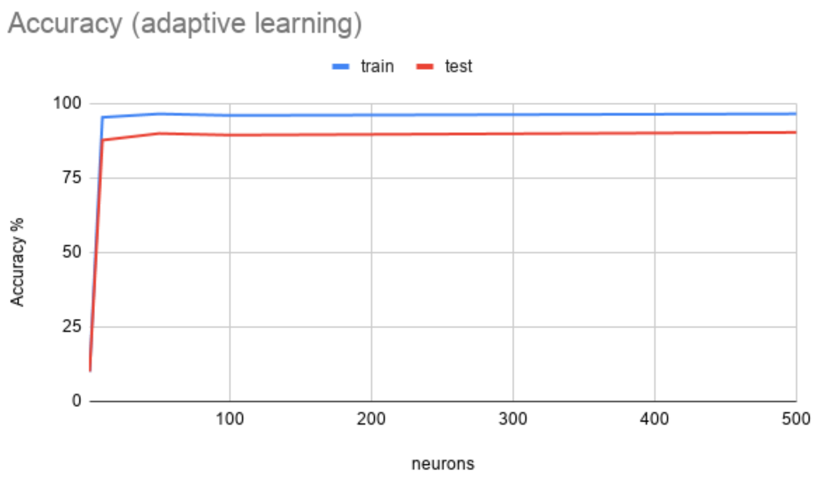
neurons	train	test	time
1	10	10	3.447994947
10	96.25	88.97	114.2791264
50	96.70667	90.26	305.7481751
100	96.29167	89.61	378.0230072

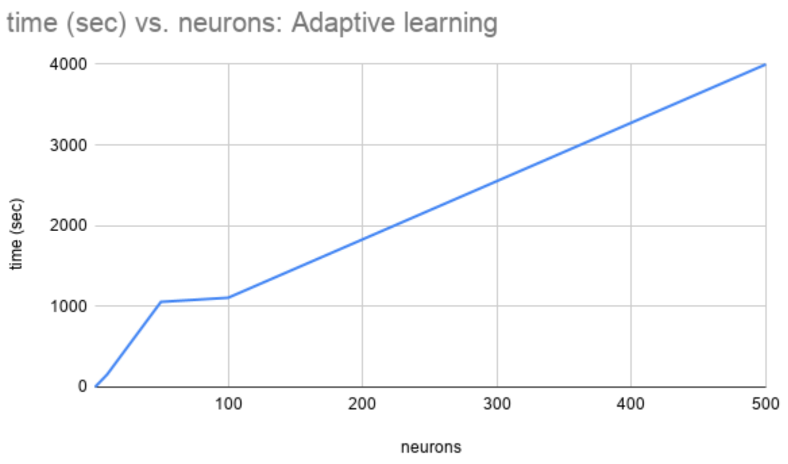




c. Needed to relax the stopping criteria to 5e-4 to converge in some cases. Also, it seemed that for this dataset, adaptive learning was slower and comparatively less accurate.

neurons	train	test	time
1	10	10	3.447994947
10	96.25	88.97	114.2791264
50	96.70667	90.26	305.7481751
100	96.29167	89.61	378.0230072
500	96.984	91.021	2580.35





d. With ReLU: (sigmoid on output layer)
Training accuracy: 99.32667, Test accuracy: 93.46, Time: 18 minutes

With Sigmoid:

Training accuracy: 99.0533, Test accuracy: 92.7, Time: 34 minutes

It seems that ReLU works well in increasing accuracy and in lesser time too (possible due to easier calculations involved in case with ReLU). Also these accuracies are much better than that of a single hidden layer.

e. It took around 3-5 mins to triain using sklearn.
Training accuracy: 99.38
Test accuracy: 92.96
In comparision to part d above, time taken is much less but accuracy is similar.