FIT3080 Exercise 1 (b) discussion

Name: Takayuki Kimura

ID: 28575393

- Which value of depth gave the best test set accuracy?

Figure 1 represents plot results of test set accuracy from depth 0 to 30 where train file is train.txt and test file is test.txt. The highest accuracy is indicated in depth 11 which is 0.987…. and it continues until depth 30. It means that all dataset, or examples are classified after depth 11 and no new nodes are added into decision tree of train.txt. Even though the train dataset is different from test data, the accuracy is quite high especially the increase from depth 2 and depth 3 is remarkable, which is around 15% increase.

- Does increasing the depth lead to over fitting on this dataset?

Figure 2 shows results of training set accuracy in contrast to Figure 1. The highest accuracy is 1.0 after depth 12 (including 12) until depth 30. It is theoretically reasonable as the decision tree is trained and tested with same dataset, which should have perfect prediction.

Does increasing the depth lead to over-fitting on this dataset?

Over-fitting or over-training commonly happens when the machine is excessively trained with a particular test set and fits the data set too deeply. As a result of that, it can't work with other data set. As far as I can observe, increasing the depth itself does not lead to over-fitting as the test set accuracy results marked high prediction rate.
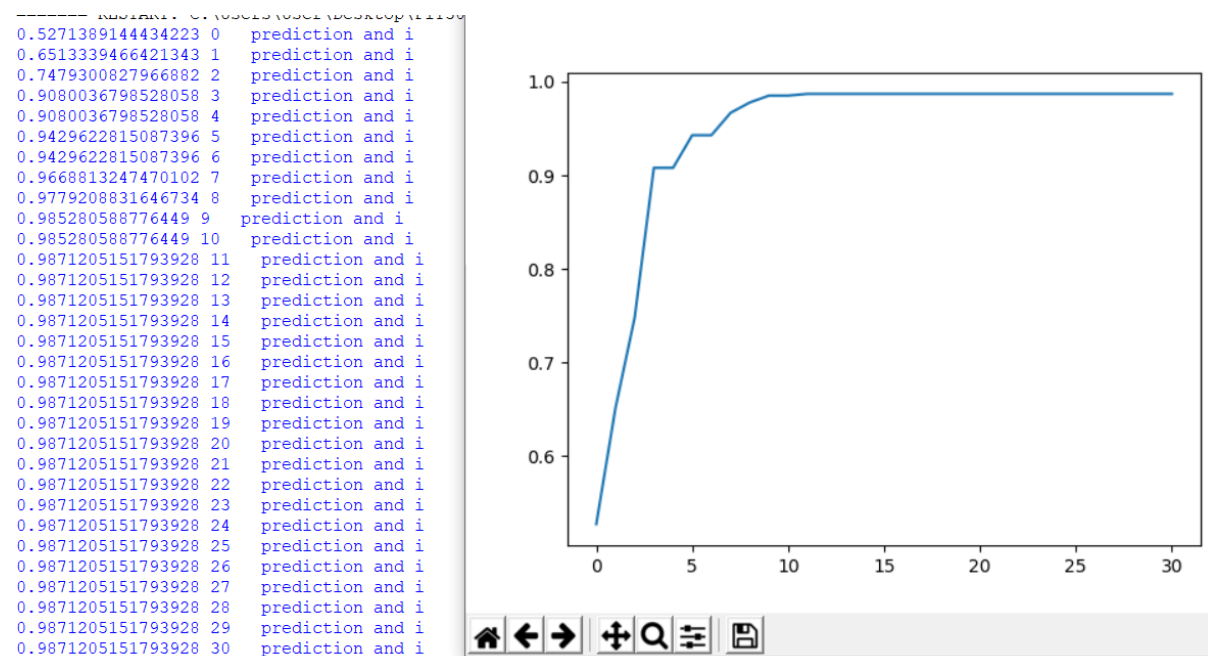

Fig 1: test set accuracy

```
Type "help", "copyright", "credits" or "license()" for more information.
>>>
======= RESTART: C:\Users\User\Desktop\FIT3(
0.5196775723091512 0    prediction and i
0.665244191559981 1    prediction and i
0.7581792318634424 2    prediction and i
0.9023233760075865 3    prediction and i
0.9023233760075865 4    prediction and i
0.9421526789947843 5    prediction and i
0.9449976292081556 6    prediction and i
0.968705547652916 7    prediction and i
0.9886201991465149 8    prediction and i
0.9947842579421526 9    prediction and i
0.9966808914177335 10    prediction and i
0.9985775248933144 11    prediction and i
1.0 12    prediction and i
1.0 13    prediction and i
1.0 14    prediction and i
1.0 15    prediction and i
1.0 16    prediction and i
1.0 17    prediction and i
1.0 18    prediction and i
1.0 19    prediction and i
1.0 20    prediction and i
1.0 21    prediction and i
1.0 22    prediction and i
1.0 23    prediction and i
1.0 24    prediction and i
1.0 25    prediction and i
1.0 26    prediction and i
1.0 27    prediction and i
1.0 28    prediction and i
1.0 29    prediction and i
1.0 30    prediction and i
```
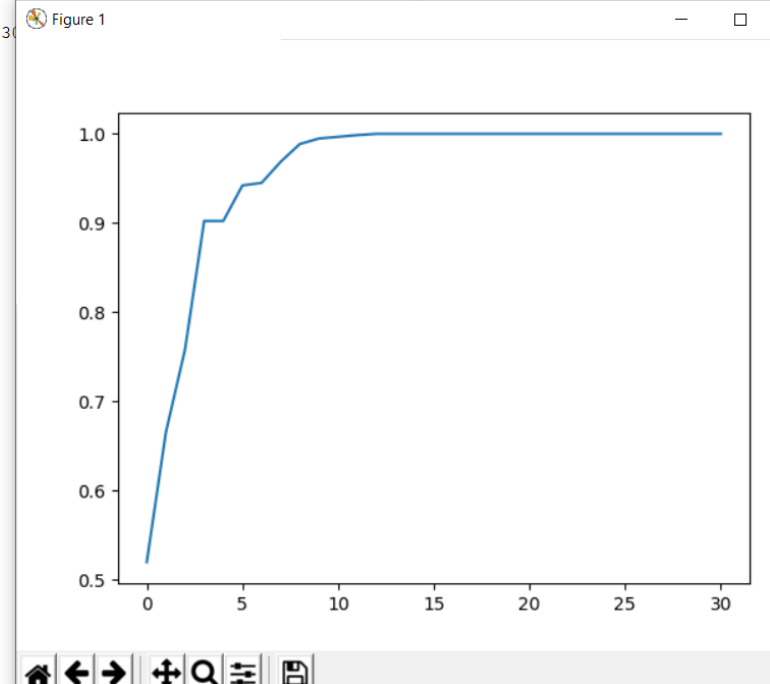

Fig 2: training set accuracy