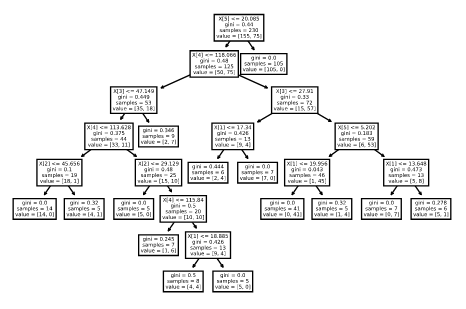
1. (a) The decision trees were constructed by DecisionTreeClassifier from Scikit.

The five trees are shown below, ordered from left to right, and top to bottom.

A close up of a piece of paper

Description automatically generated

A screenshot of a cell phone

Description automatically generatedA black sign with white text

Description automatically generated

A black sign with white text

Description automatically generated

The accuracy scores are shown below.

A screenshot of a cell phone

Description automatically generated

The value of min\_samples\_leaf affects the depth of the tree. Since “a split point at any depth will only be considered if it leaves at least min\_samples\_leaf training samples in each of the left and right branches”. The min\_samples\_leaf = 15 is the best choice among all the options. The min = 5 might has over-fitting. The min = 25, 40 and 50 are under-fitting since not all features are used in splitting.

(b) The precision values are plotted below.

A screenshot of a cell phone

Description automatically generated

The abnormal precision is less affected by the under fitting than the normal precision. The reason could be the features selected for splitting split the abnormal class better than the normal case.

The recall values are plotted below.

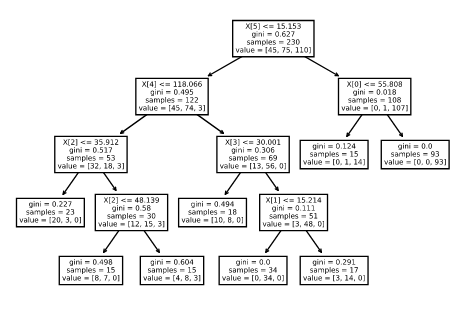
A screenshot of a cell phone

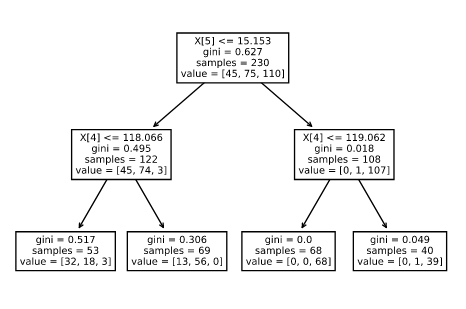
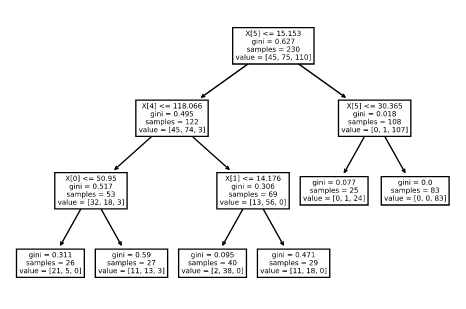
Description automatically generated

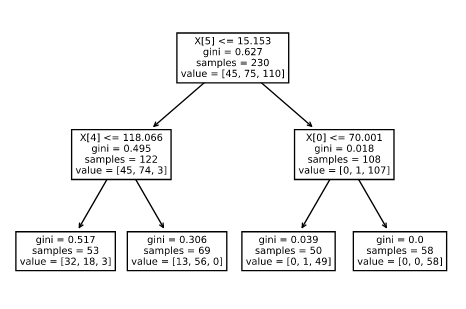
The precision/recall for normal and abnormal have opposite trend with respect to min\_sample\_per\_leaf.

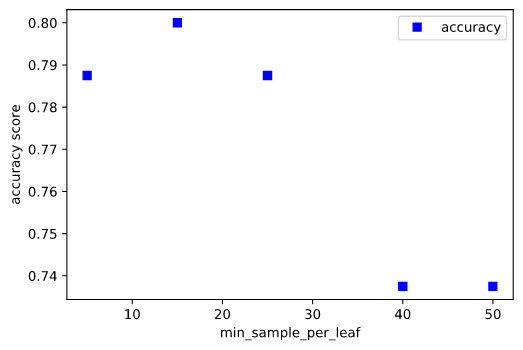
1. (a)

A screenshot of a cell phone

Description automatically generated

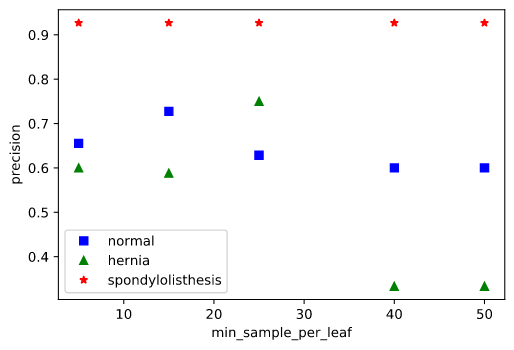


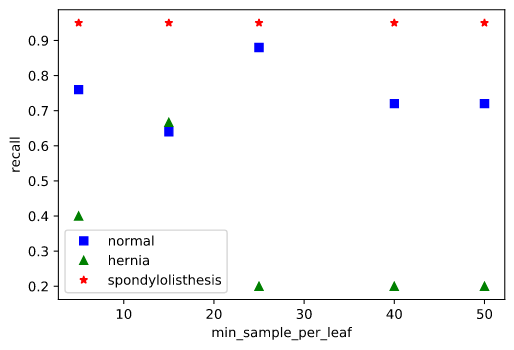




The best choice is 15. The depth of trees are similar to the 2-class trees. The over fitting is more significant than 2-class tree. However, the under fitting occurs later than that of the 2-class tree due to the increased classes. (The 3-class tree classifies the data better than 2-class tree for the 25 case.)

(b)





The Spondylolisthesis is easy to classify. The Hernia only correlates to certain features because it is sensitive to under fitting. The precision of normal is similar to the 2-class tree. The recall of normal is slightly higher than the 2-class tree. (False negative is reduced.)

1. (a)The correlation matrix is shown below. The “degree\_spond” has the highest correlation values. So it is dropped out of the data set.



A screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

Description automatically generatedA screenshot of a cell phone

Description automatically generated

A screenshot of a cell phone

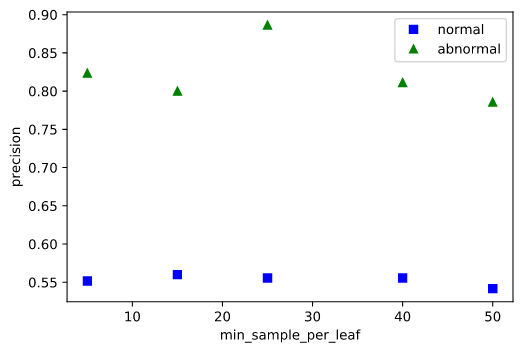
Description automatically generated

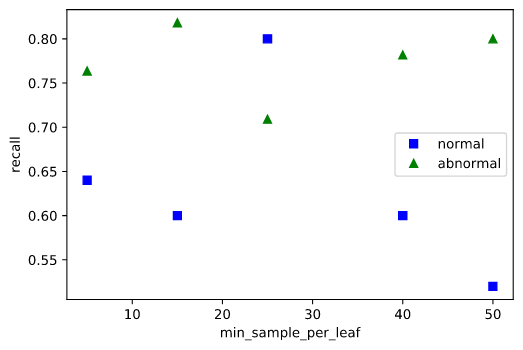
A screenshot of a cell phone

Description automatically generated

The best choice is still min = 15. Over fitting for min = 5 and under fitting for min = 25 and above.

(b)





The abnormal has a higher classification accuracy, and it is less affected by the over/under fitting. For the recall, normal and abnormal has opposite trend.

(c) The drop out simplified the tree at the cost of accuracy.

The drop out reduces the depth of the tree.

The precision accuracy is lower in 3 compared with 1. The best case in 3 is about the same as the under fitting case in 1.

The overall precision and recall are lower than that of the case in (1b). However, the trend with respect to min\_per\_leaf is similar.