Machine Learning – HW2

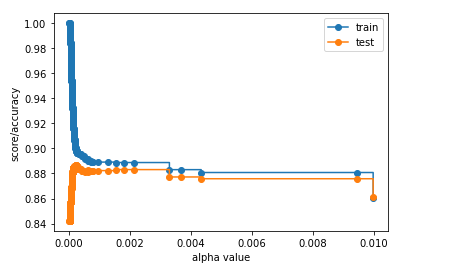
108062135 呂佳恩

How I built the tree

In the sklearn.tree package, I imported the DecisionTreeClassifier, it will automatically build the tree with the provided datas. I also imported the plot\_tree package to visualize the tree.

With the confusion matrix, I could visualize how my model was performing and could tune my model.

Improving the tree

I used the CPP (cost complexity pruning) method of pruning the tree, as we increase the alpha, the more of the tree will be pruned. As we can see from the graph below, as the alpha increases, the accuracy first increases then decreases. This could be explained by saying that when pruning, the model reduces the problem of overfitting. However, after pruning too many leaves, the model becomes too inaccurate since too many features are removed.

We know that the alpha varies when the training data is changed after doing some test, therefore, we us cross validation