

Homework:

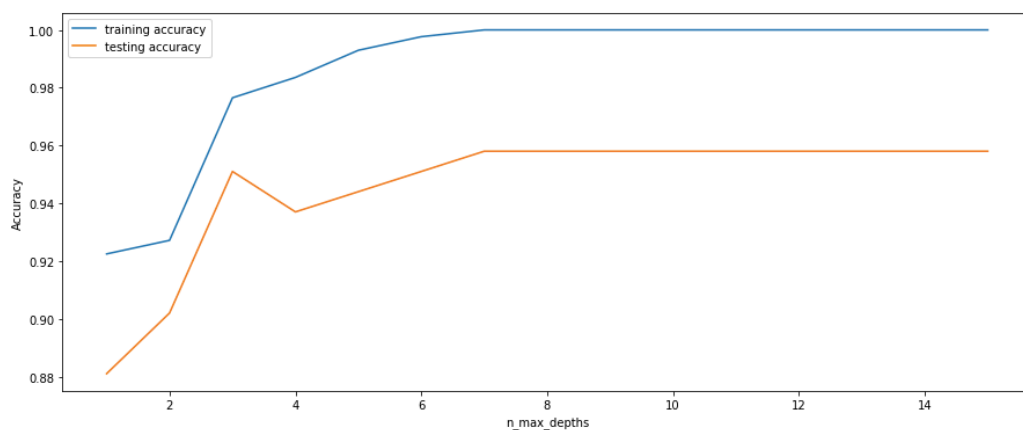
開啟iris.py，將第20行的`iris_dataset = datasets.load_iris()`換成`breast_cancer_dataset = datasets.load_breast_cancer()`

`breast_cancer_dataset`是一個關於乳癌的分類資料集，分類結果有良性跟惡性。

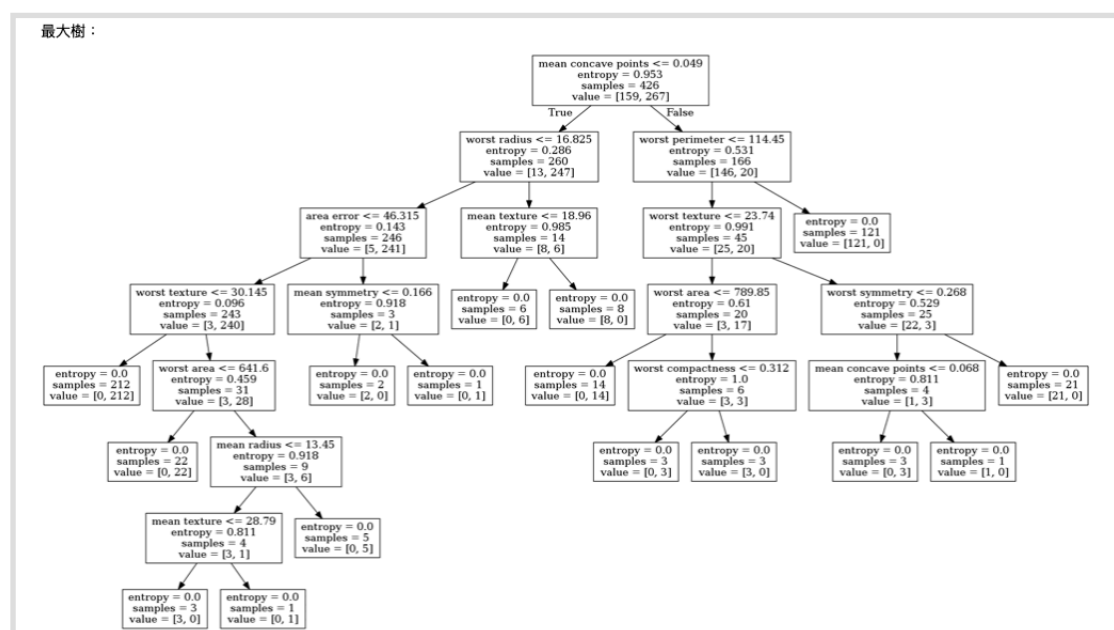
`breast_cancer_dataset`與`iris_dataset`同為類dict資料型態，`breast_cancer_dataset`中的key與對應的資料說明請對照iris.py的注釋說明(22行至28行)

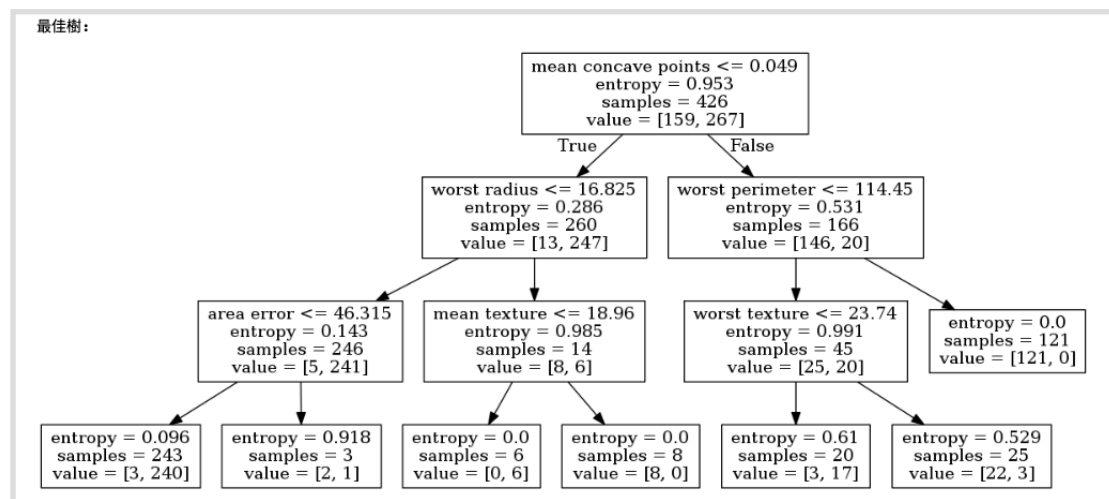
請依照iris.py的模型建立流程，以同樣的方式建立`breast_cancer_dataset`的分類樹，並回答以下問題：

1.顯示在不同深度下的分類樹準確率折線圖(最大深度為15)



2.顯示最大樹與最佳樹的樹狀圖





3.最大樹與最佳樹的子葉樹數目各為多少個?

最大樹: 15

最佳樹: 3

4.顯示最大樹與最佳樹的模型比較表

	Train	Test
最大樹	1.000000	0.958042
最佳樹	0.976526	0.951049

5.哪個模型較佳?如何判斷?

雖然從精確度來看, 最大樹較高, 但最佳樹訓練和測試的精確度較相近. 所以預測資料方面, 最佳樹會比最大樹的判斷還要正確些.

6.何謂Overfitting?

Overfitting代表訓練資料的結果和過去太過相似, 這樣的話反而對預測新的資料沒有參考性.

7.如何判斷Overfitting?

有很多方式可以判斷一個模型是否過度適配. 常見的特徵就是訓練和測試的精確度差異過大.