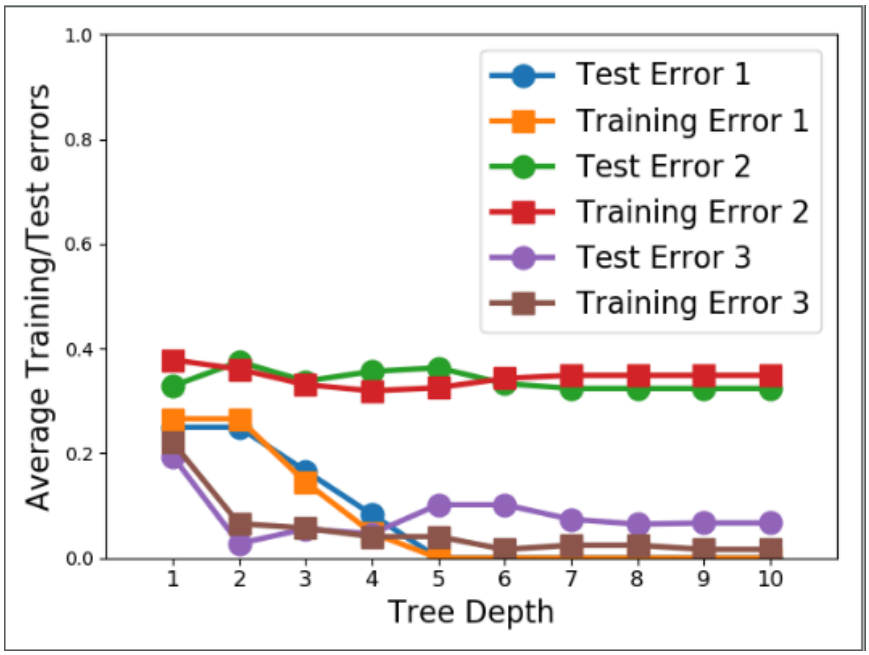
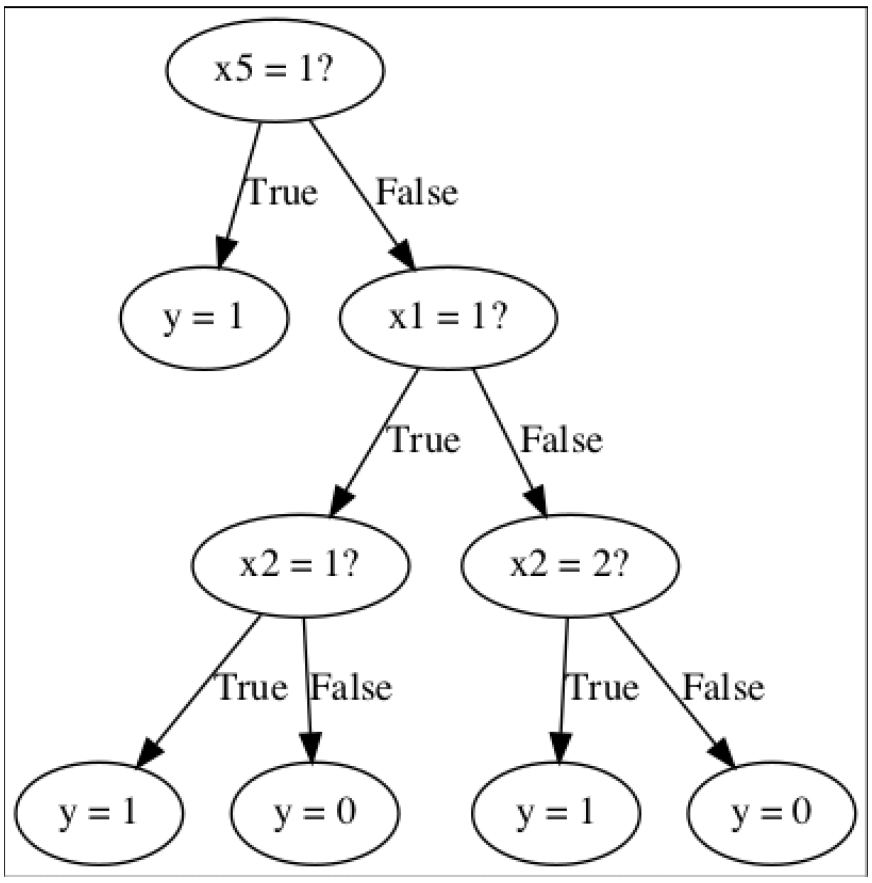
#Name: Xiaoran Guo
#Net ID: xxg180001
#UTD ID: 2021432123

Homework #3

b.



c.



#Name: Xiaoran Guo
#Net ID: xxg180001
#UTD ID: 2021432123

```
Confusion matrix on the test set for depth= 1
Test Error = 25.00%.
[[216   0]
 [108 108]]
Confusion matrix on the test set for depth= 3
Test Error = 16.67%.
[[180  36]
 [ 36 180]]
Confusion matrix on the test set for depth= 5
Test Error = 0.00%.
[[216   0]
 [  0 216]]
```

d.
scikit-learn

```
Confusion matrix on the test set for depth= 1
Test Error = 25.00%.
[[216   0]
 [108 108]]
Confusion matrix on the test set for depth= 3
Test Error = 16.67%.
[[144  72]
 [  0 216]]
Confusion matrix on the test set for depth= 5
Test Error = 16.67%.
[[168  48]
 [ 24 192]]
```

e.
Other data set from UCI repository about Breast Cancer Coimbra
The dataset has been pre-processed into binary features using the mean due to continuous features.

My own decisiontree's confusion matrix

```
Confusion matrix on the test set for depth= 1
Test Error = 53.66%.
[[12  6]
 [16  7]]
Confusion matrix on the test set for depth= 3
Test Error = 43.90%.
```

#Name: Xiaoran Guo
#Net ID: xxg180001
#UTD ID: 2021432123

```
[[15  3]
 [15  8]]
Confusion matrix on the test set for depth= 5
Test Error = 48.78%.
[[15  3]
 [17  6]]

DecisionTreeClassifier's confusion matrix
Confusion matrix on the test set for depth= 1
Test Error = 53.66%.
[[12  6]
 [16  7]]
Confusion matrix on the test set for depth= 3
Test Error = 36.59%.
[[15  3]
 [12 11]]
Confusion matrix on the test set for depth= 5
Test Error = 43.90%.
[[15  3]
 [15  8]]
```

The value of the test error and the tendency between my implementation and that of scikit-learn are basically the same. As the increase of the depth, the test error is decreasing. But when the depth is large enough, the tree may be overfit, then the error become bigger.
The features in scikit-learn are always randomly permuted at each split. Therefore, the best-found split may vary. So it may be the reason why the result are a little different.