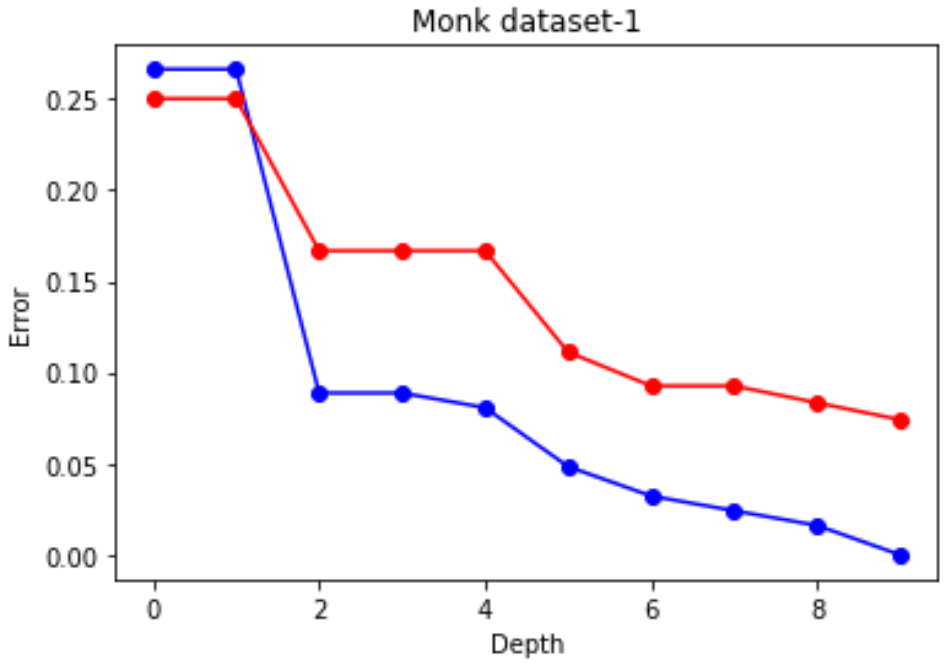
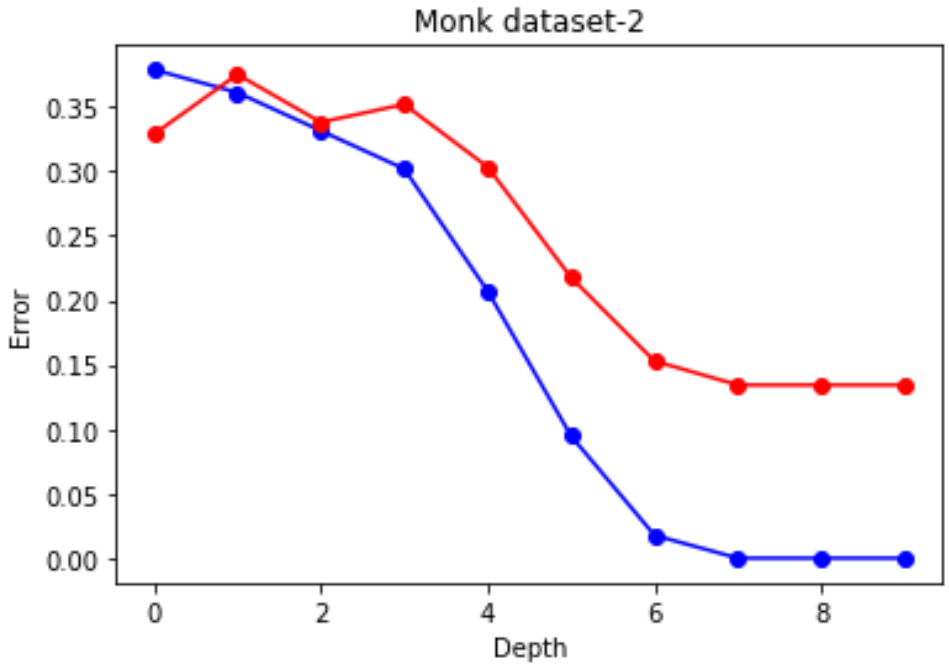
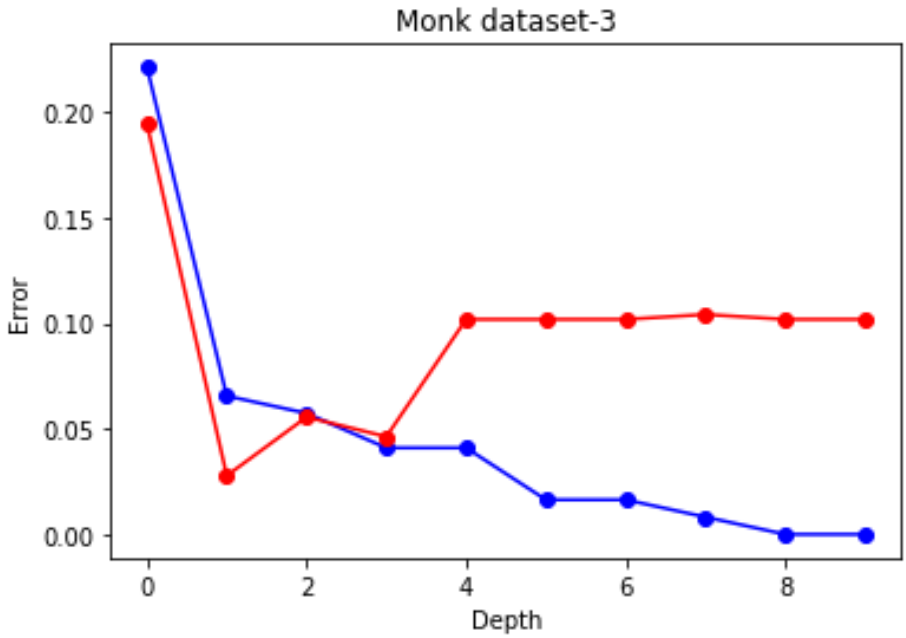
MACHIINE LEARNING
Assignment 2

Problem 2





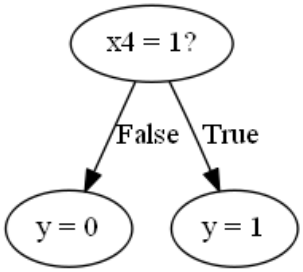


- The blue line is Train error.
- The red line is Test error.
- All 3 graphs shows the train and test error for different monk datasets for depth of decision tree varying from 1 to 10

Problem 3

Depth = 1

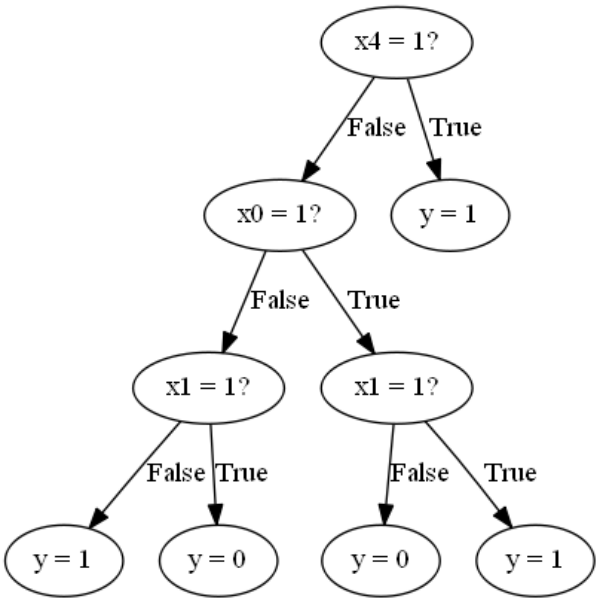
Output:
TREE
+-- [SPLIT: x4 = 1 False]
| +-- [LABEL = 0]
+-- [SPLIT: x4 = 1 True]
| +-- [LABEL = 1]
test error: 0.25
Confusion Matrix:
[[216 0]
 [108 108]]



Depth = 3

Output:

```
TREE
+-- [SPLIT: x4 = 1 False]
|       +-- [SPLIT: x0 = 1 False]
|       |       +-- [SPLIT: x1 = 1 False]
|       |       |       +-- [LABEL = 1]
|       |       +-- [SPLIT: x1 = 1 True]
|       |       |       +-- [LABEL = 0]
|       +-- [SPLIT: x0 = 1 True]
|       |       +-- [SPLIT: x1 = 1 False]
|       |       |       +-- [LABEL = 0]
|       |       +-- [SPLIT: x1 = 1 True]
|       |       |       +-- [LABEL = 1]
+-- [SPLIT: x4 = 1 True]
|       +-- [LABEL = 1]
test error: 0.16666666666666666
Confusion Matrix:
[[144  72]
 [  0 216]]
```

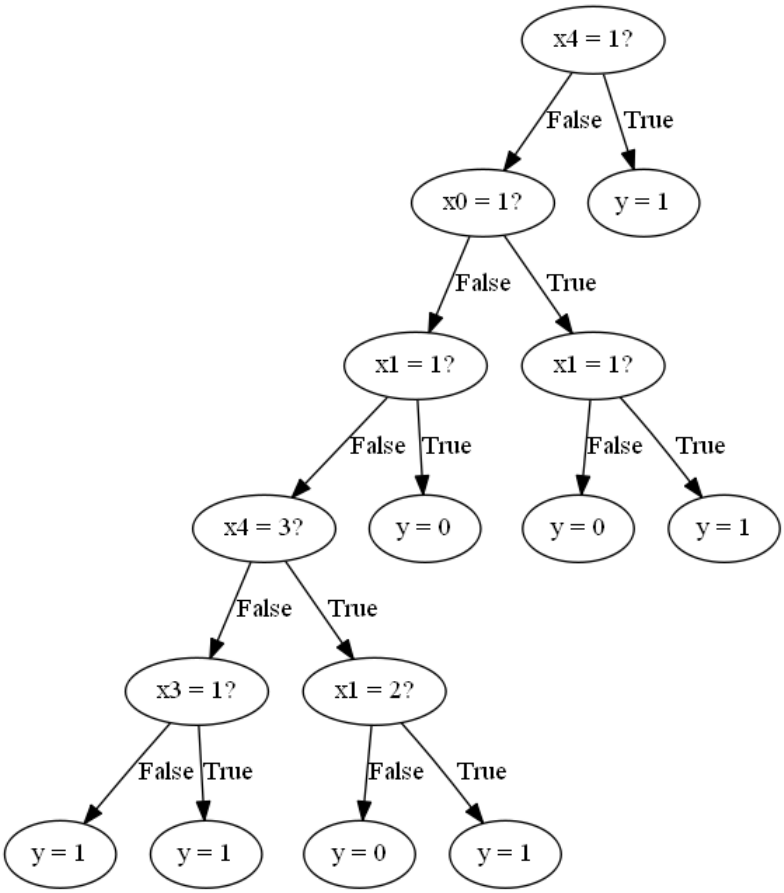


Depth = 5

Output:

```
TREE
+-- [SPLIT: x4 = 1 False]
|       +-- [SPLIT: x0 = 1 False]
|       |       +-- [SPLIT: x1 = 1 False]
|       |       |       +-- [SPLIT: x4 = 3 False]
|       |       |       |       +-- [SPLIT: x3 = 1 False]
```

```
|       |       |       |       |       +-- [LABEL = 1]
|       |       |       |       +-- [SPLIT: x3 = 1 True]
|       |       |       |       |       +-- [LABEL = 1]
|       |       |       |       +-- [SPLIT: x4 = 3 True]
|       |       |       |       |       +-- [SPLIT: x1 = 2 False]
|       |       |       |       |       |       +-- [LABEL = 0]
|       |       |       |       |       +-- [SPLIT: x1 = 2 True]
|       |       |       |       |       |       +-- [LABEL = 1]
|       |       |       +-- [SPLIT: x1 = 1 True]
|       |       |       |       +-- [LABEL = 0]
|       |       +-- [SPLIT: x0 = 1 True]
|       |       |       +-- [SPLIT: x1 = 1 False]
|       |       |       |       +-- [LABEL = 0]
|       |       |       +-- [SPLIT: x1 = 1 True]
|       |       |       |       +-- [LABEL = 1]
+-- [SPLIT: x4 = 1 True]
|       +-- [LABEL = 1]
test error: 0.16666666666666666
Confusion Matrix:
[[156  60]
 [ 12 204]]
```



test error decreases with increasing depth

Problem 4

Depth = 1

Output:
Sklearn test error0.25
Confusion Matrix for sklearn code:
[[216 0]
 [108 108]]

Depth =3

Output:
Sklearn test error0.16666666666666666
Confusion Matrix for sklearn code:
[[144 72]
 [0 216]]

Depth = 5

Output:
Sklearn test error0.16666666666666666
Confusion Matrix for sklearn code:
[[168 48]
 [24 192]]

Test error obtained from hand coded ID3 algorithm and decision tree algorithm from sklearn are same in this case. But algorithm in sklearn will be more optimized.

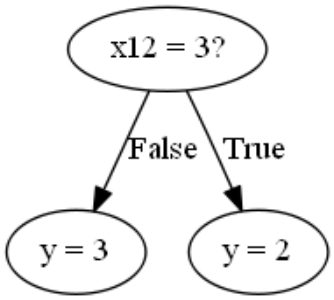
Problem 5

Using lymphography dataset from UCI repository

Id3

Depth = 1

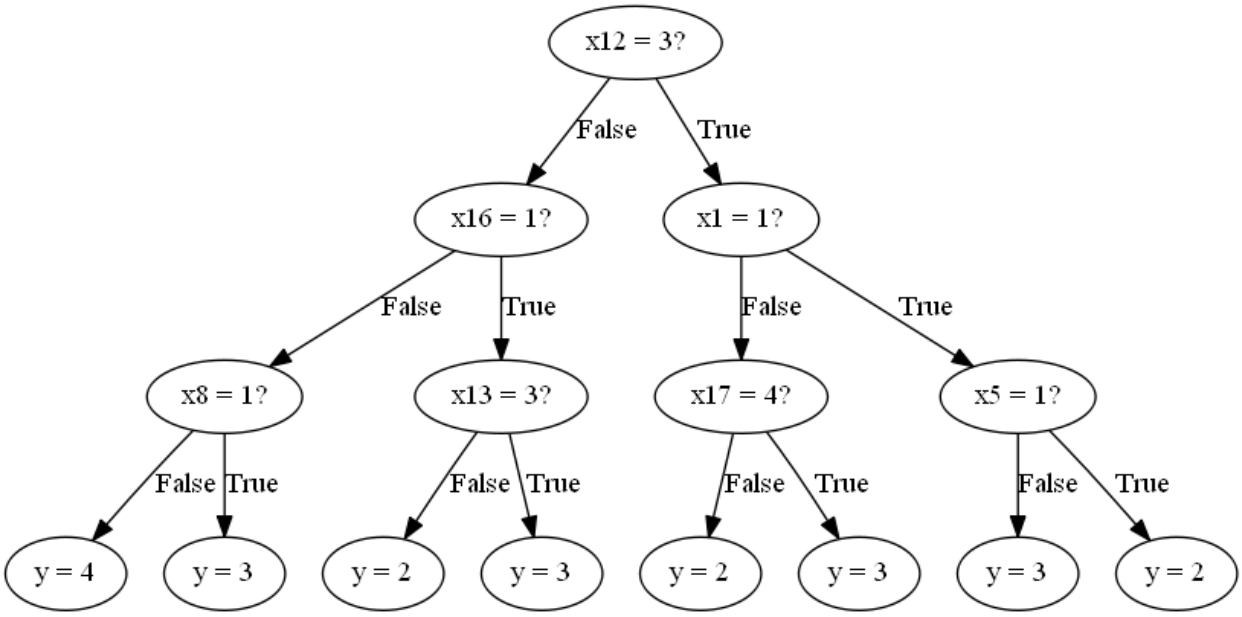
Output:
TREE
+-- [SPLIT: x12 = 3 False]
| +-- [LABEL = 3]
+-- [SPLIT: x12 = 3 True]
| +-- [LABEL = 2]
test error: 0.3111111111111111
Confusion Matrix:
[[0 0 1 0]
 [0 17 6 0]
 [0 5 14 0]
 [0 0 2 0]]



Depth =3

Output:

```
TREE
+-- [SPLIT: x12 = 3 False]
|       +-- [SPLIT: x16 = 1 False]
|       |       +-- [SPLIT: x8 = 1 False]
|       |       |       +-- [LABEL = 4]
|       |       +-- [SPLIT: x8 = 1 True]
|       |       |       +-- [LABEL = 3]
|       +-- [SPLIT: x16 = 1 True]
|       |       +-- [SPLIT: x13 = 3 False]
|       |       |       +-- [LABEL = 2]
|       |       +-- [SPLIT: x13 = 3 True]
|       |       |       +-- [LABEL = 3]
+-- [SPLIT: x12 = 3 True]
|       +-- [SPLIT: x1 = 1 False]
|       |       +-- [SPLIT: x17 = 4 False]
|       |       |       +-- [LABEL = 2]
|       |       +-- [SPLIT: x17 = 4 True]
|       |       |       +-- [LABEL = 3]
|       +-- [SPLIT: x1 = 1 True]
|       |       +-- [SPLIT: x5 = 1 False]
|       |       |       +-- [LABEL = 3]
|       |       +-- [SPLIT: x5 = 1 True]
|       |       |       +-- [LABEL = 2]
test error: 0.26666666666666666
Confusion Matrix:
[[ 0  1  0  0]
 [ 0 17  6  0]
 [ 0  4 15  0]
 [ 0  1  0  1]]
```

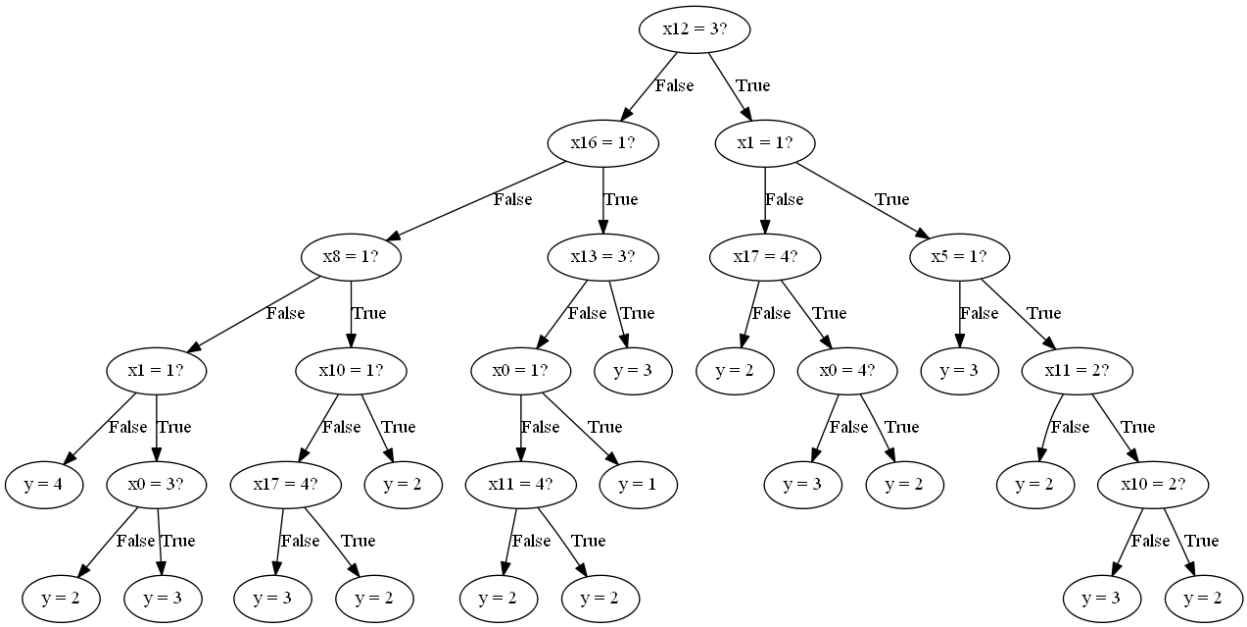


Depth = 5

Output:

```
TREE
+-- [SPLIT: x12 = 3 False]
|       +-- [SPLIT: x16 = 1 False]
|       |       +-- [SPLIT: x8 = 1 False]
|       |       |       +-- [SPLIT: x1 = 1 False]
|       |       |       |       +-- [LABEL = 4]
|       |       |       +-- [SPLIT: x1 = 1 True]
|       |       |       |       +-- [SPLIT: x0 = 3 False]
|       |       |       |       |       +-- [LABEL = 2]
|       |       |       |       +-- [SPLIT: x0 = 3 True]
|       |       |       |       |       +-- [LABEL = 3]
|       |       +-- [SPLIT: x8 = 1 True]
|       |       |       +-- [SPLIT: x10 = 1 False]
|       |       |       |       +-- [SPLIT: x17 = 4 False]
|       |       |       |       |       +-- [LABEL = 3]
|       |       |       |       +-- [SPLIT: x17 = 4 True]
|       |       |       |       |       +-- [LABEL = 2]
|       |       |       +-- [SPLIT: x10 = 1 True]
|       |       |       |       +-- [LABEL = 2]
|       +-- [SPLIT: x16 = 1 True]
|       |       +-- [SPLIT: x13 = 3 False]
|       |       |       +-- [SPLIT: x0 = 1 False]
|       |       |       |       +-- [SPLIT: x11 = 4 False]
|       |       |       |       |       +-- [LABEL = 2]
|       |       |       |       +-- [SPLIT: x11 = 4 True]
|       |       |       |       |       +-- [LABEL = 2]
|       |       |       +-- [SPLIT: x0 = 1 True]
```

```
|       |       |       |       +-- [LABEL = 1]
|       |       |       +-- [SPLIT: x13 = 3 True]
|       |       |       |       +-- [LABEL = 3]
+-- [SPLIT: x12 = 3 True]
|       +-- [SPLIT: x1 = 1 False]
|       |       +-- [SPLIT: x17 = 4 False]
|       |       |       +-- [LABEL = 2]
|       |       +-- [SPLIT: x17 = 4 True]
|       |       |       +-- [SPLIT: x0 = 4 False]
|       |       |       |       +-- [LABEL = 3]
|       |       |       +-- [SPLIT: x0 = 4 True]
|       |       |       |       +-- [LABEL = 2]
|       +-- [SPLIT: x1 = 1 True]
|       |       +-- [SPLIT: x5 = 1 False]
|       |       |       +-- [LABEL = 3]
|       |       +-- [SPLIT: x5 = 1 True]
|       |       |       +-- [SPLIT: x11 = 2 False]
|       |       |       |       +-- [LABEL = 2]
|       |       |       +-- [SPLIT: x11 = 2 True]
|       |       |       |       +-- [SPLIT: x10 = 2 False]
|       |       |       |       |       +-- [LABEL = 3]
|       |       |       |       +-- [SPLIT: x10 = 2 True]
|       |       |       |       |       +-- [LABEL = 2]
test error: 0.28888888888888886
Confusion Matrix:
[[ 1  0  0  0]
 [ 0 18  5  0]
 [ 0  6 13  0]
 [ 0  1  1  0]]
```



Test error decreases in depth = 3 compared to depth =1, but it increases at depth = 5. The tree with depth may be overfitting

Sklearn:

Depth = 1

Output:
Sklearn test error0.4222222222222222
Confusion Matrix for sklearn code:
[[0 1 0 0]
 [0 22 1 0]
 [0 15 4 0]
 [0 2 0 0]]

Depth =3

Output:
Sklearn test error0.28888888888888886
Confusion Matrix for sklearn code:
[[0 1 0 0]
 [0 18 5 0]
 [0 5 14 0]
 [0 1 1 0]]

Depth = 5

Output:
Sklearn test error0.24444444444444444
Confusion Matrix for sklearn code:
[[0 1 0 0]
 [1 21 1 0]
 [0 6 13 0]
 [0 1 1 0]]

In sklearn's algorithm test error decreases with increase in depth. ID3 performs better for depth = 1 and depth = 3 but Sklearn's decision tree algorithm works better with depth = 5.