

1.0 Description of my work

The learning algorithm can learn well. It can learn well on all the oracles with an error rate listed below. For these oracles, 500 is a good number of training examples to learn well. However, for oracle 5, the best training number might be 50 since after that the error goes up.

Oracle	#Training Sample	#Test Sample	Error Rate(%)
1	500	100	3
2	50	100	3
	100	100	3
	500	100	0
3	500	100	5
4	500	100	3
5	10	100	34
	50	100	29
	200	100	32
	500	100	35
	1000	100	39
	5000	100	43
6	500	100	3
7	50	100	0
	100	100	0
	500	100	0
8	500	100	0
	50	100	7
	100	100	2
9	500	100	0
	1000	100	0
	5000	100	0
	50000	100	0
10	500	100	0
11	500	100	0
12	500	100	0
	500	1000	0

2.0 pseudo-code

train function:

set initial_decision=-1, initial_decision_attribute=-1;

if (examples all 1)

 data splits well

 decision=1

 return to parent node

else if(examples all 0)

 data splits well

```

    decision=0
    return to parent node
else if (no available attributes)
    take majority vote
    decision= 0 or 1
else
    find decision attributes → go into function to find max information gain
    split attributes and class into two parts according to chosen attribute
    create left node, link the node with its parent node.
    create right node, link the node with its parent node.

    assign the part attribute=1 to new attribute in the left node
    assign corresponding class to new class for left node
    train left node with new attributes and class (recursive)
    (if goes to leaf node, return to parent node and continue with the following)

    assign the part attribute=0 to new attribute in the left node
    assign corresponding class to new class for left node
    train right node with new attributes and class (recursive)
end

```

3.0 Output of main.m:

```

命令窗口
>> main
Getting 50 training examples from Oracle 2
Training decision tree...
Getting 100 test examples from Oracle 2
Classifying with decision tree...
Checking accuracy with Oracle 2
The oracle says you have 97 out of 100 correct ( 97.0%)
----- Result -----
Oracle  #Training Sample    #Test Sample  Error Rate (%)
      2           50         100         3.00
>> main
Getting 100 training examples from Oracle 2
Training decision tree...
Getting 100 test examples from Oracle 2
Classifying with decision tree...
Checking accuracy with Oracle 2
The oracle says you have 97 out of 100 correct ( 97.0%)
----- Result -----
Oracle  #Training Sample    #Test Sample  Error Rate (%)
      2          100         100         3.00
>> main
Getting 500 training examples from Oracle 2
Training decision tree...
Getting 100 test examples from Oracle 2
Classifying with decision tree...
Checking accuracy with Oracle 2
The oracle says you have 100 out of 100 correct (100.0%)
----- Result -----
Oracle  #Training Sample    #Test Sample  Error Rate (%)
      2          500         100         0.00
fx >> |

```

```

>> main
Getting 50 training examples from Oracle 7
Training decision tree...
Getting 100 test examples from Oracle 7
Classifying with decision tree...
Checking accuracy with Oracle 7
The oracle says you have 100 out of 100 correct (100.0%)
----- Result -----
Oracle #Training Sample #Test Sample Error Rate (%)
7      50              100          0.00
>> main
Getting 100 training examples from Oracle 7
Training decision tree...
Getting 100 test examples from Oracle 7
Classifying with decision tree...
Checking accuracy with Oracle 7
The oracle says you have 100 out of 100 correct (100.0%)
----- Result -----
Oracle #Training Sample #Test Sample Error Rate (%)
7      100             100          0.00
>> main
Getting 500 training examples from Oracle 7
Training decision tree...
Getting 100 test examples from Oracle 7
Classifying with decision tree...
Checking accuracy with Oracle 7
The oracle says you have 100 out of 100 correct (100.0%)
----- Result -----
Oracle #Training Sample #Test Sample Error Rate (%)
7      500             100          0.00
fx >> |

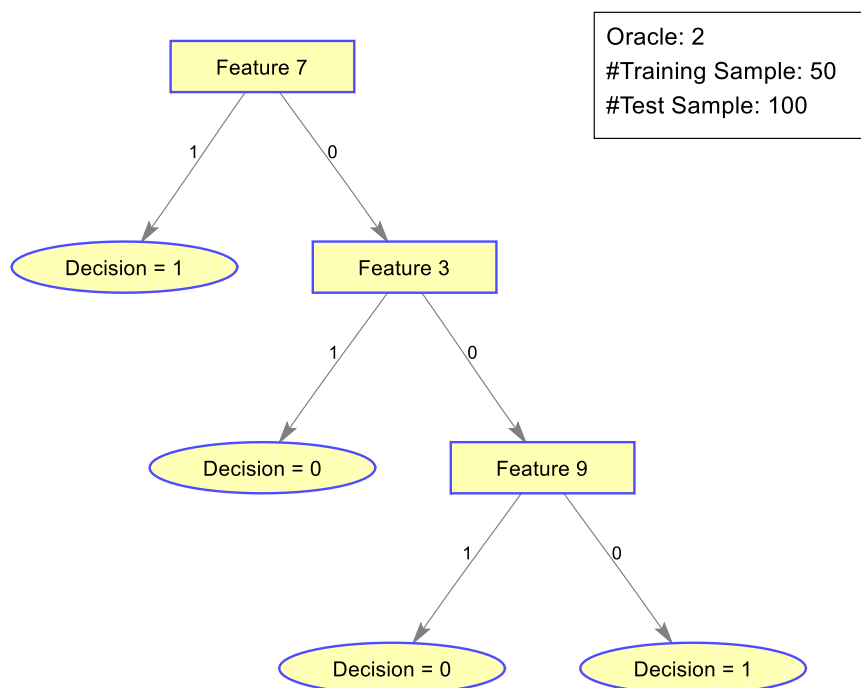
```

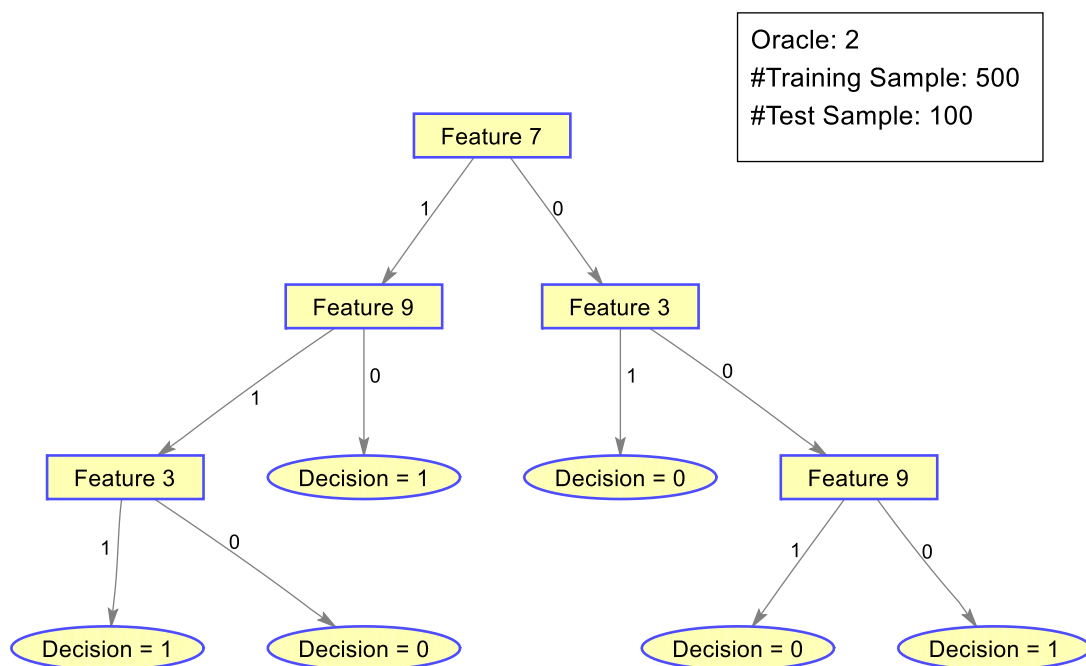
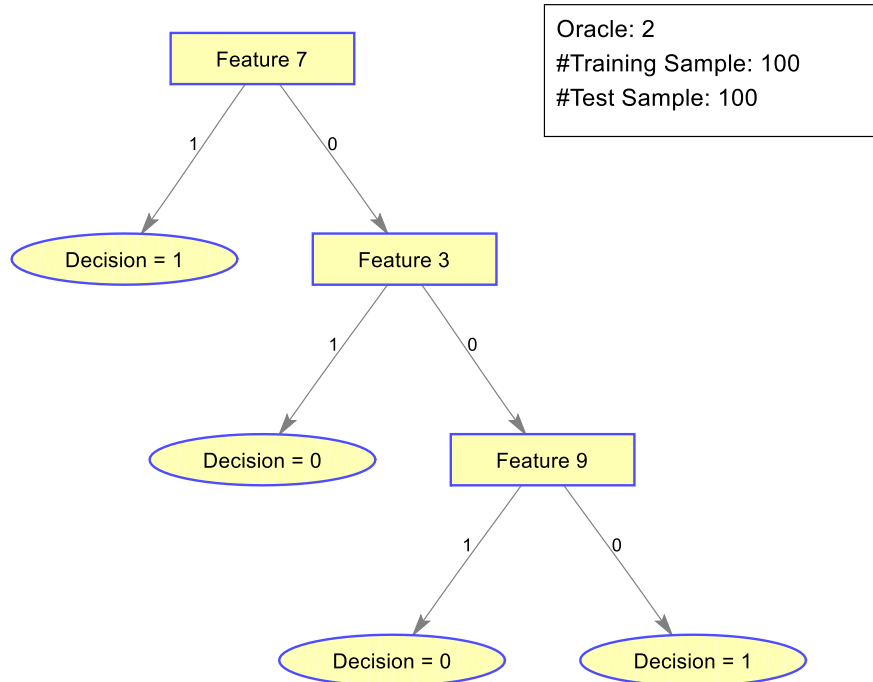
4.0 Tree structures under different circumstances:

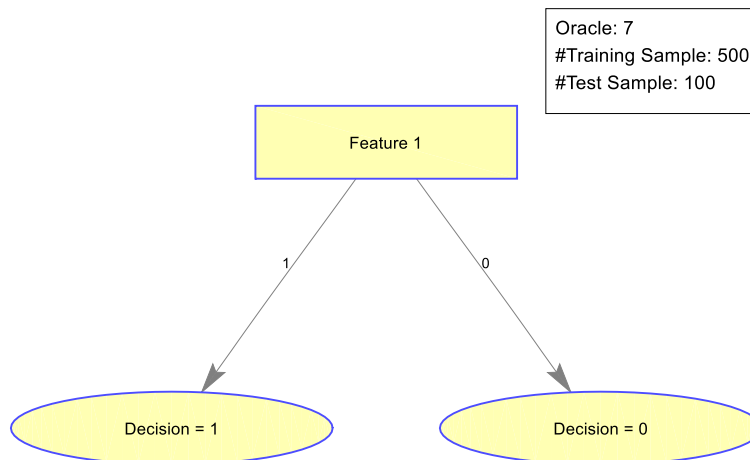
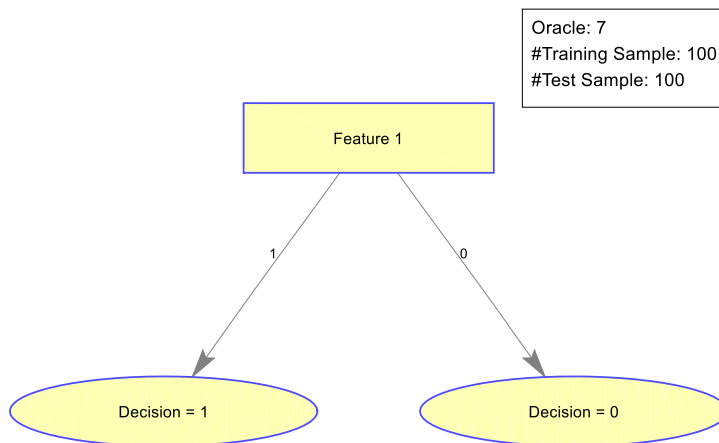
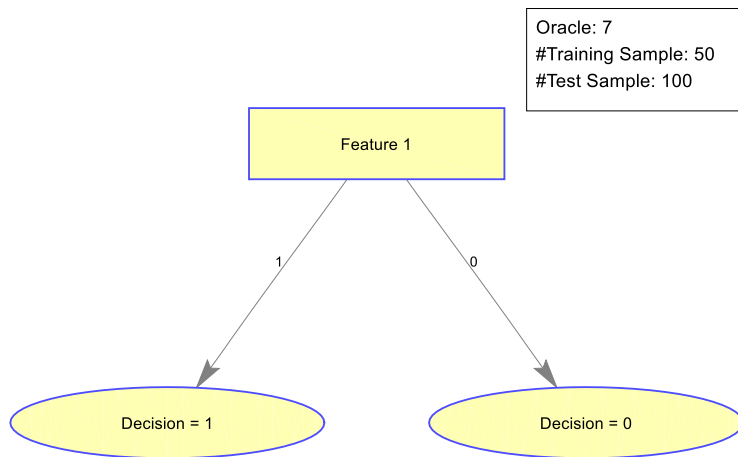
Oracle: 2, 7

#Training Sample: 50, 100, 500

#Test Sample: 100

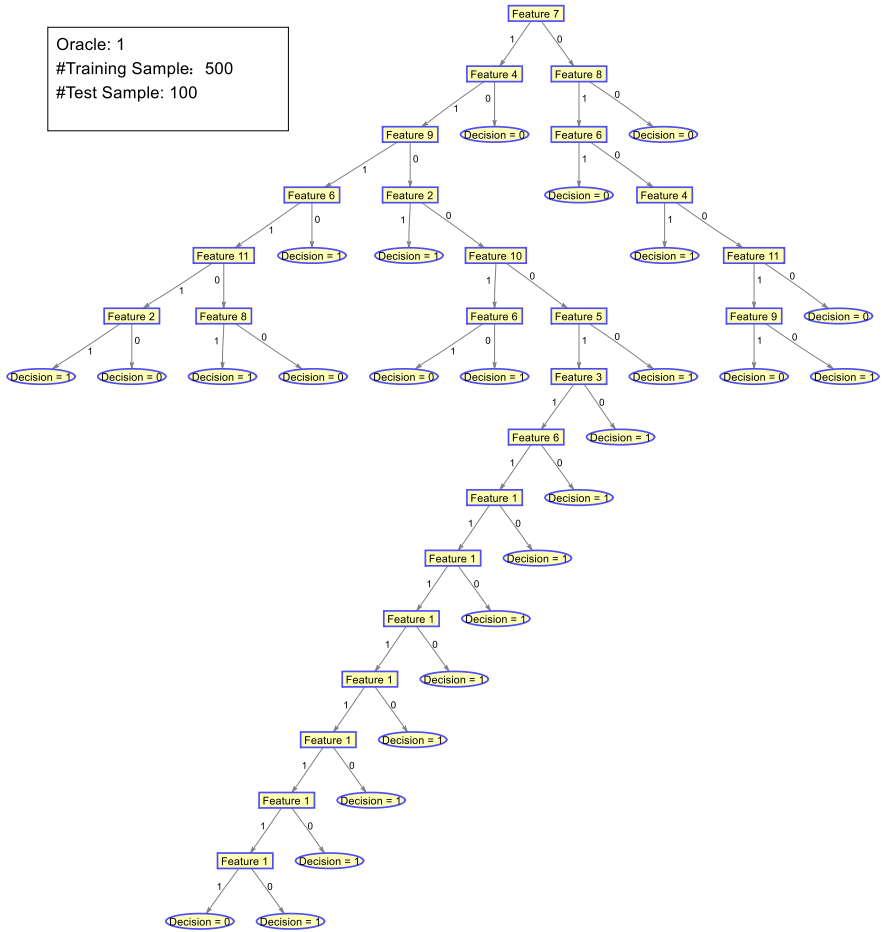




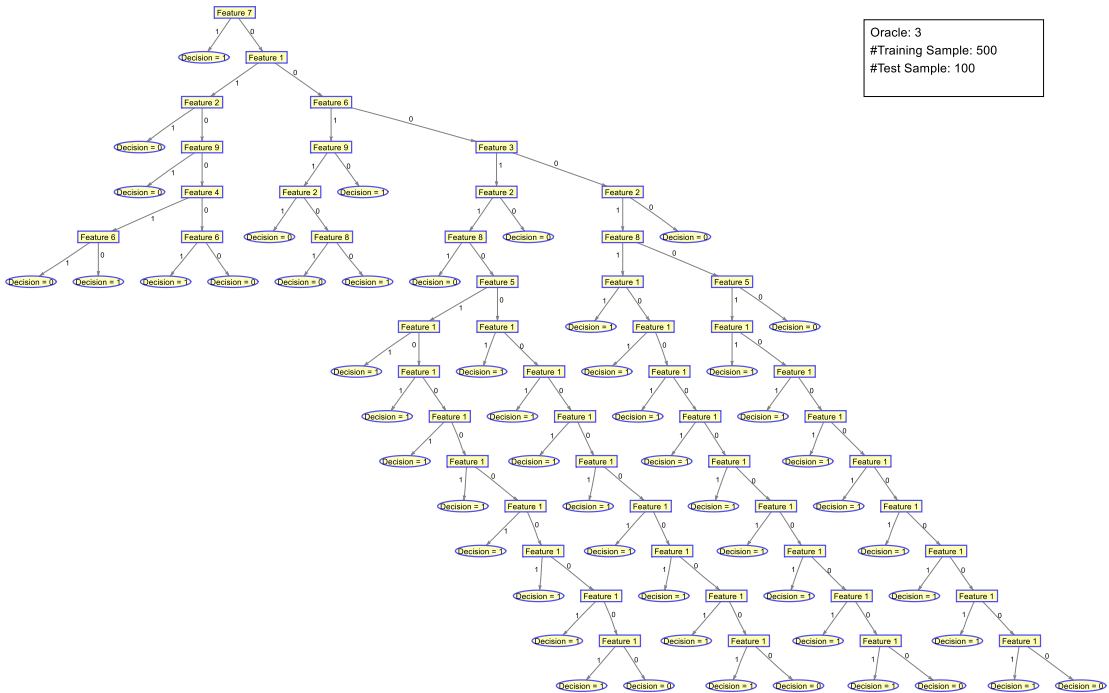


For other oracles:

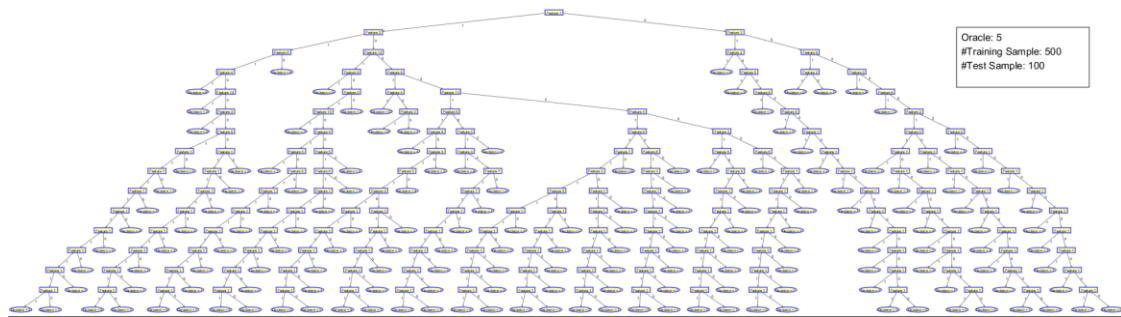
Oracle 1:



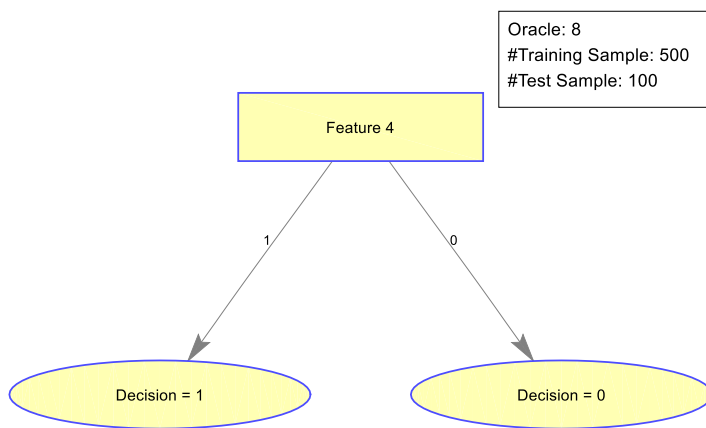
Oracle3:



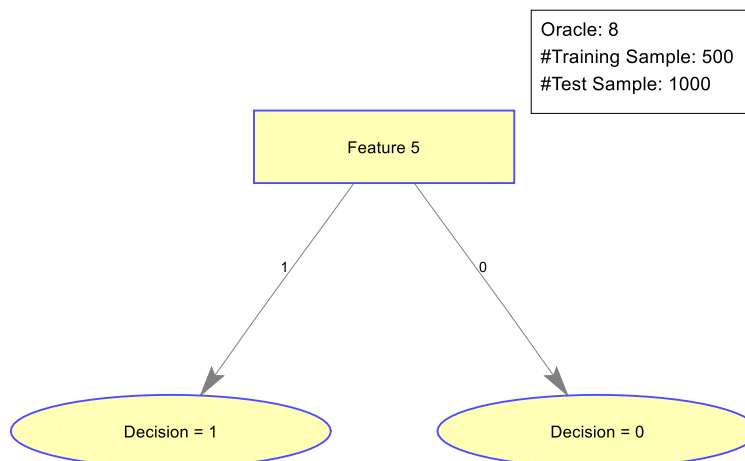
Oracle 5:



Oracle 8:



Oracle 12:



5.0 Discussion

There was no glaring implementation errors or bugs in my code when I run it. The outcomes look good for most of the samples. However, I'm not that satisfied with sample 5. And I found out when the training samples goes more than 50, the error increases. That might be caused by over-fitting. I could improve that by:

- (1) Generate several ID3 with different number of nodes

Since we have already built the decision algorithm and find decision attribute with max information gain every time. I think we can generate that in a simpler way. By adjusting the number of training samples between 10-500, we are able to find out the one with the least error, then choose that one as decision tree. As shown in the first page, 50 is the best value for training numbers among other choices.

- (2) Randomly choose 2/3 of the available data for training, use 1/3 of for testing (validation set).