Machine Learning exercise 4

Eden Dupont 204808596

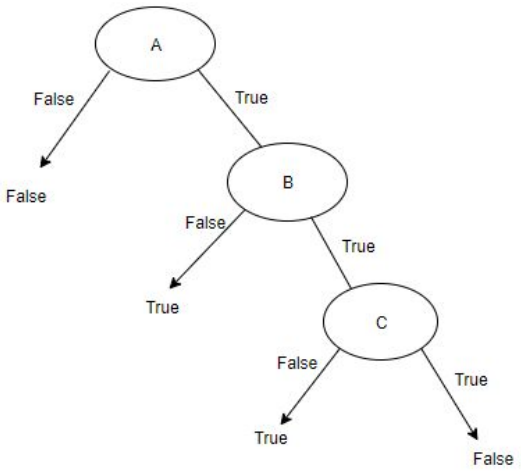
Theoretical Part

dataset				
sample number	A	B	C	Y
1	F	F	F	F
2	T	F	T	T
3	T	T	F	T
4	T	T	T	F

1.

Question: Using the dataset above, we want to build a decision tree which classifies Y as True (T) or False (F) given the binary variables A, B, C.
Draw the tree that would be learned by the greedy algorithm with zero training error. You do need to show computations.

Final solution (full solution below)



Full solution:

Using the greedy algorithm, I will choose the feature which has the least errors
In this case:
For A -
A(F)=F values expected, 0 errors
A(T)=T values expected, 1 errors (1 out of 3 False values)
Total error = 1

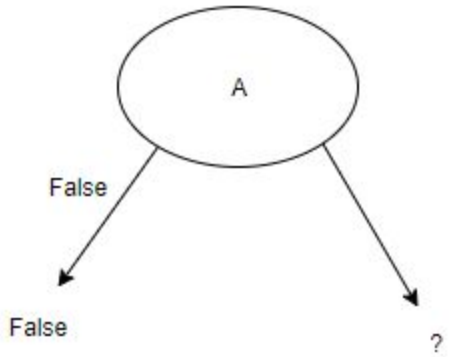
For B -
B(F)=F values expected, 1 errors (1 out of 2 True values)
B(T)=T values expected, 1 errors (1 out of 2 False values)
Total error = 2

For C -
C(F)=F values expected, 1 errors (1 out of 2 True values)
C(T)=T values expected, 1 errors (1 out of 2 False values)
Total error = 2

We will take the feature with the lowest error as the parent node for the next steps
Now, if:
A=False

dataset			
sample number	B	C	Y
1	F	F	F

From here we cannot split the Y values anymore and therefore we create a leaf where A=False



Next, for A=True

dataset			
sample number	B	C	Y
2	F	T	T
3	T	F	T
4	T	T	F

B(F)=T values, 0 error
B(T)=F values, 1 error
1 error total

C(F)=T values, 0 error
C(T)=F values, 1 error (1 of 2 is false)
1 error total

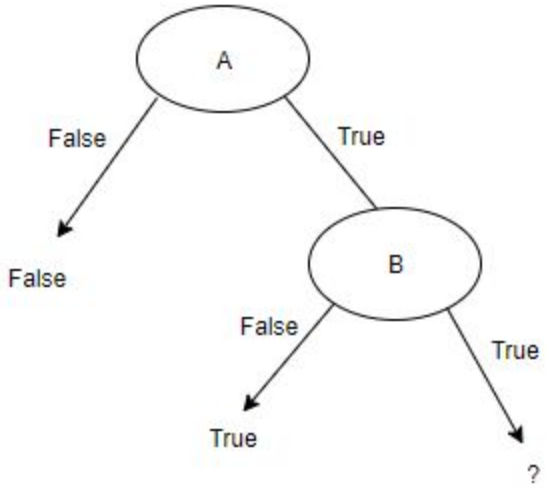
In this case, we will choose feature B arbitrarily

New tables

B=False

dataset		
sample number	C	Y
2	T	T

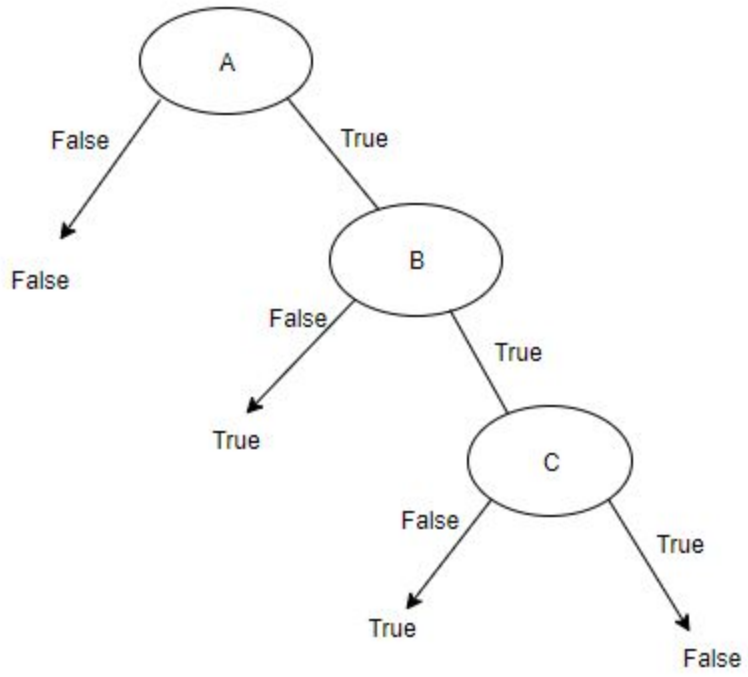
We cannot split the Y values and therefore we put a leaf of Y=T where B is False:



B=True

dataset		
sample number	C	Y
3	F	T
4	T	F

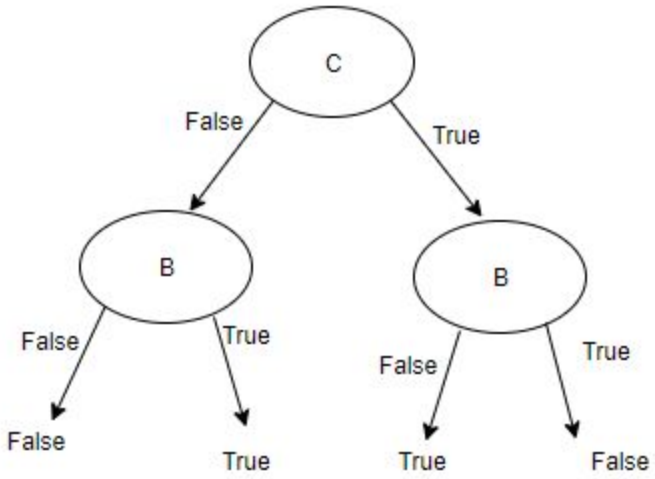
Here we can easily split C to two with zero error and the following tree is generated:



2. Question: Is this tree optimal (i.e.,does it get zero training error with minimal depth)? Explain in less than two sentences. If it is not optimal, draw the optimal tree as well.

Solution:

There is zero training error but the tree is not minimal, feature A can be discarded to create a tree with only B and C to correctly decide Y



In this case, the tree is of smaller height

3. Question: We will use the dataset below to learn a decision tree which predicts if people pass machine learning (Yes or No), based on their previous GPA (High, Medium, or Low) and whether or not they studied. On the following calculations use log on a base of 2.

dataset			
sample number	GPA	Studied	Passed
1	L	F	No
2	L	T	Yes
3	M	F	No
4	M	T	Yes
5	H	F	Yes
6	H	T	Yes

4. What is the entropy of H(Passed)?

P(Passed=True) = 4/6
P(Passed=False) = 2/6

H(Passed) = -[P(passed)*log(P(passed)) + P(not passed)*log(P(not passed))] = 0.27643

5. What is the entropy of H(Passed | GPA)?

H(Passed | GPA=H) = $-\frac{2}{2} * log(\frac{2}{2}) - \frac{0}{2} * log(\frac{0}{2}) = 0$

H(Passed | GPA=L) = $-\frac{1}{2} * log(\frac{1}{2}) - \frac{1}{2} * log(\frac{1}{2}) = 1$

H(Passed | GPA=M) = $-\frac{1}{2} * log(\frac{1}{2}) - \frac{1}{2} * log(\frac{1}{2}) = 1$

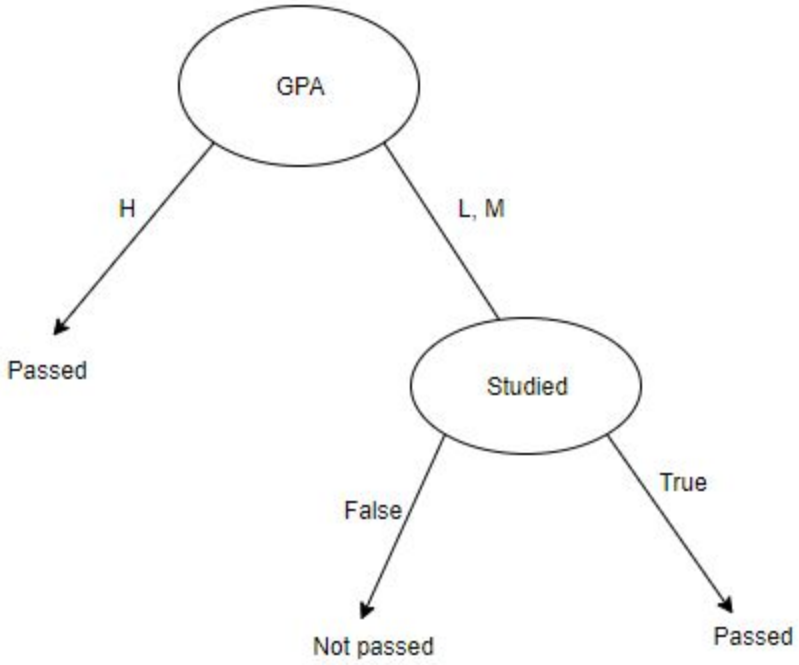
6. What is the entropy of H(Passed | Studied)?

H(Passed | Studied=True) = $-\frac{3}{3} * log(\frac{3}{3}) - \frac{0}{3} * log(\frac{0}{3}) = 0$

H(Passed | Studied=False) =

$$-\frac{1}{3} * log(\frac{1}{3}) - \frac{2}{3} * log(\frac{2}{3}) = 0.27643$$

7. Draw the full decision tree, that would be learned for this dataset. You do not need to show any calculations.



8. Suggest three different improvements to the algorithm that could improve

the result (add your answer to the theoretical part).

Answer:

1. Use weighted features, to force order change in the tree (bypassing the feature choice with the information gains)
2. Limit the model with max_depth, to force the tree to be shorter and search for a more optimal model