CS 6375
Assignment 2-Part 1:
Decision Trees

Names of students in your group:
ADRITA DUTTA:axd172930
CHIRAG SHAHI:cxs180005

Number of free late days used: **0**
Note: You are allowed a total of 4 free late days for the entire semester. You can use at most 2 for each assignment. After that, there will be a penalty of 10% for each late day.

Please list clearly all the sources/references that you have used in this assignment.
References:
Chapter 3- Tom mitchell

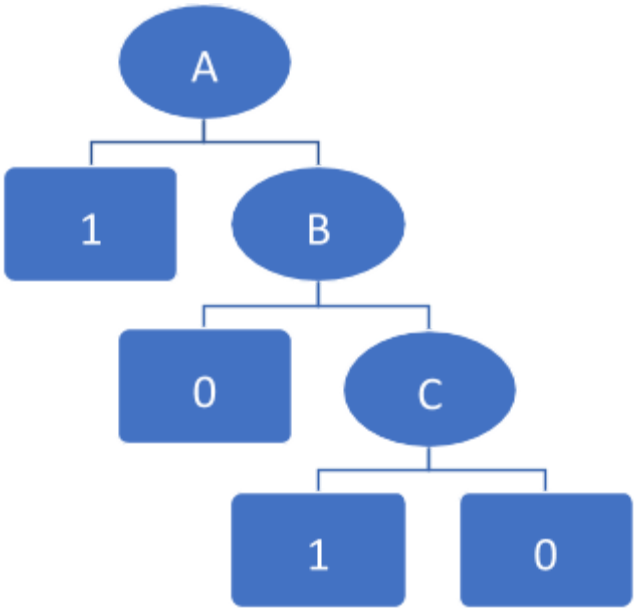
Part I: Written Problems (30 points)

1. Representing Boolean Functions (10 points)

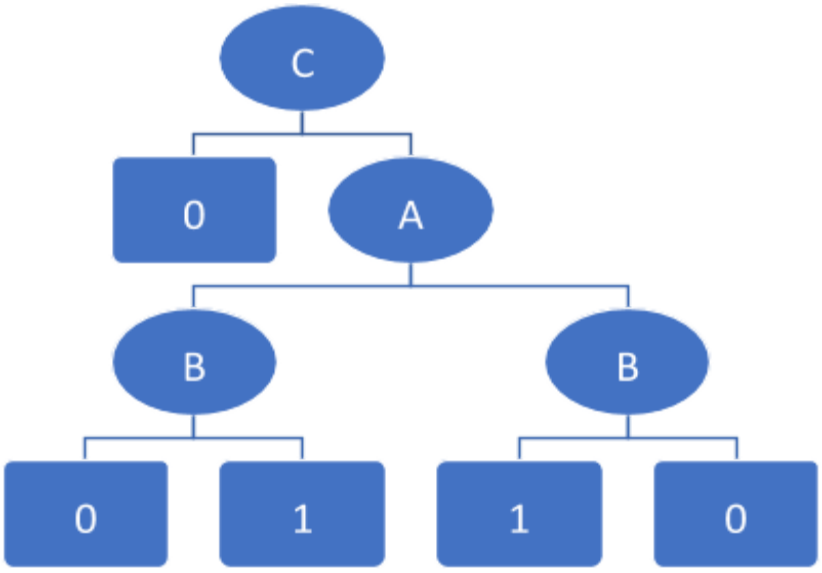
Give decision trees to represent the following concepts. Your decision tree must contain as few nodes as possible. You can assume A, B, and C are Boolean variables.

[FOR ALL GRAPHS: RIGHT BRANCH REPRESENTS 1, LEFT BRANCH REPRESENTS 0]

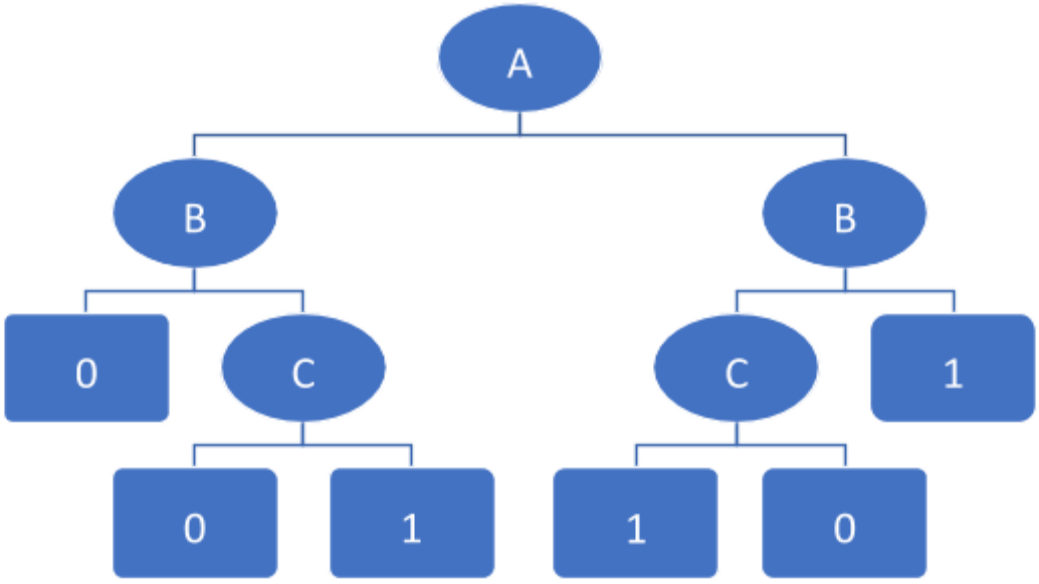
a) $Y = (\neg A \vee B) \wedge \neg(C \wedge A) = \neg A \vee (B \wedge \neg C)$



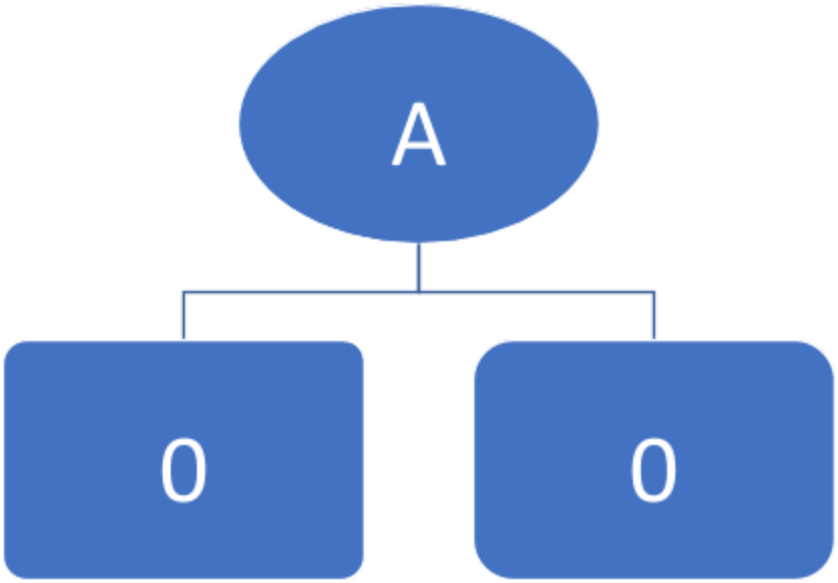
b) $Y = (A \oplus B) \wedge C = ((\neg A \wedge B) \vee (A \wedge \neg B)) \wedge C$



c) Y = (A ∨ B) ∧ (B ∨ C) ∧ (A ∨ C)



d) Y = (A ∨ B) ∧ ¬A ∧ ¬B = (A ∨ B) ∧ ¬ (A ∧ B)



Q2) Decision Trees (20 points)

In this question, you will use the ID3 algorithm to create a decision tree for the dataset given below. There are three Boolean attributes X1, X2, and X3 and a Boolean class attribute.
Be sure to show detailed calculations for each step including entropy and information gain values. Draw a plot of the final tree that you obtain and show the class labels for the leaf nodes.
Also indicate the set of instances that are associated with each leaf.

Instance	X1	X2	X3	Class
1	1	0	0	1
2	0	1	0	1
3	0	0	0	0
4	1	0	1	0
5	0	0	0	0
6	1	1	0	1
7	0	1	1	0
8	1	0	0	1
9	0	0	0	0
10	1	0	0	1

=>

STEP1:

Entropy (Root) $= -(5/10) * \log(5/10) - (5/10) * \log(5/10) = 1$

Gain (Root, x1) =

$1 - (5/10) * (-(4/5) * \log(4/5) - (1/5) * \log(1/5)) - (5/10) * (-(4/5) * \log(4/5) - (1/5) * \log(1/5)) = 0.278$

Gain (Root, x2) =

$1 - (3/10) * (-(2/3) * \log(2/3) - (1/3) * \log(1/3)) - (7/10) * (-(3/7) * \log(3/7) - (4/7) * \log(4/7)) = 0.035$

Gain (Root, x3) =

$1 - (2/10) * (-(0/2) * \log(0/2) - (2/2) * \log(2/2)) - (8/10) * (-(5/8) * \log(5/8) - (3/8) * \log(3/8)) = 0.236$

So x1 is the appropriate choice of attribute for the root node.

STEP2:

Entropy (x1 = 1) $= -(4/5) * \log(4/5) - (1/5) * \log(1/5) = 0.722$

Entropy (x1 = 0) $= -(4/5) * \log(4/5) - (1/5) * \log(1/5) = 0.722$

Gain (x1 = 1, x3) =

$0.722 - (1/5) * (-(0/1) * \log(0/1) - (1/1) * \log(1/1)) - (4/5) * (-(4/4) * \log(4/4) - (0/4) * \log(0/4))$
$=0.722$

Gain (x1 = 1, x2) =

$0.722 - (1/5) * (-(1/1) * \log(1/1) - (0/1) * \log(0/1)) - (4/5) * (-(3/4) * \log(3/4) - (1/4) * \log(1/4))$
$=0.073$

When x1 = 1, x3 is the appropriate choice of attribute to split on.

Both the children of x3 are pure nodes so we stop here.

Gain (x1 = 0, x3) =

$0.722 - (1/5) * (-(0/1) * \log(0/1) - (1/1) * \log(1/1)) - (4/5) * (-(1/4) * \log(1/4) - (3/4) * \log(3/4))$
$=0.073$

Gain (x1 = 0, x2) =

$0.722 - (2/5) * (-(1/2) * \log(1/2) - (1/2) * \log(1/2)) - (3/5) * (-(0/3) * \log(0/3) - (3/3) * \log(3/3))$
$=0.322$

When x1 = 0, x2 is the appropriate choice of attribute to split on.

When x1 = 0 and x2 = 0 the node is pure and so we stop there.

When x1 = 0 and x2 = 1, we split on the last remaining attribute i.e. on x3.

So, the resulting tree looks as follows:

[For Graph: RIGHT BRANCH REPRESENTS 1, LEFT BRANCH REPRESENTS 0]