CS 6375

ASSIGNMENT _2

DECISION TREE

Names of students in your group:

Mahasweta Sarma (mxs175831)

Maleeha Shabeer Koul (msk180001)

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.

- eLearning

ANSWER FOR QUESTION 1:

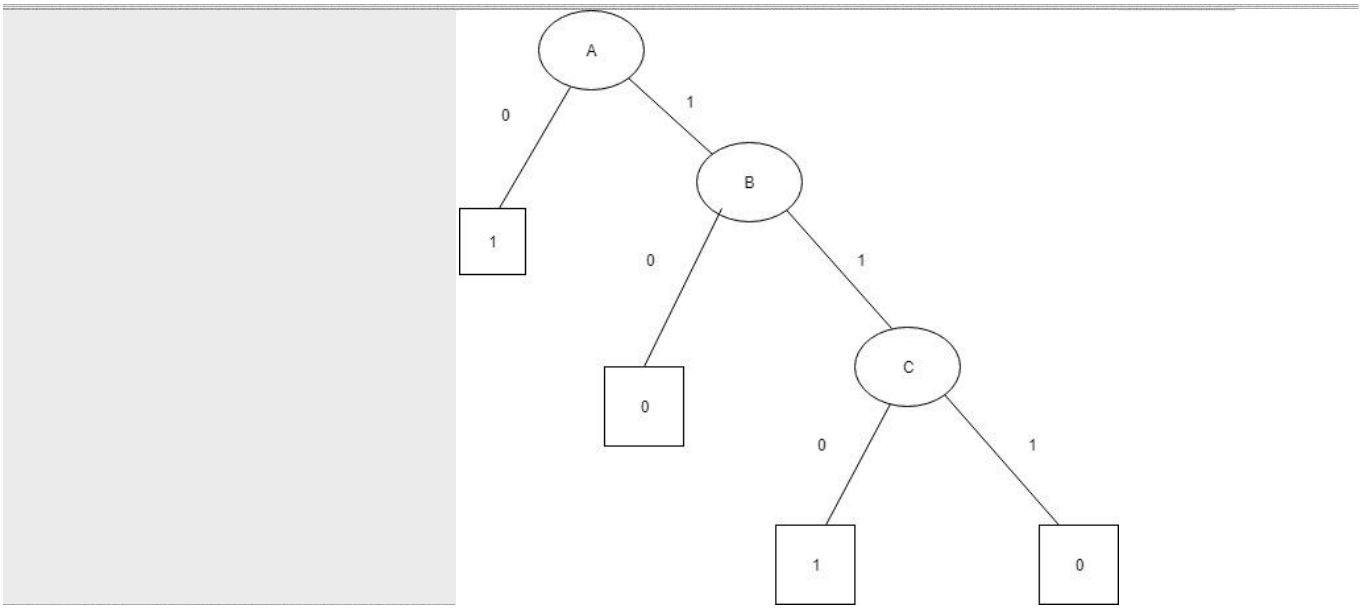
1. Representing Boolean Functions

a) Y = (~A V B) ^ ~(C^ A)

Truth table for the above expression :

A	B	C	~A	~A v B	C^ A	~(C^ A)	Y
0	0	0	1	1	0	1	1
0	0	1	1	1	0	1	1
0	1	0	1	1	0	1	1
0	1	1	1	1	0	1	1
1	0	0	0	0	0	1	0
1	0	1	0	0	1	0	0
1	1	0	0	1	0	1	1
1	1	1	0	1	1	0	0

Required Decision tree with the least number of nodes:

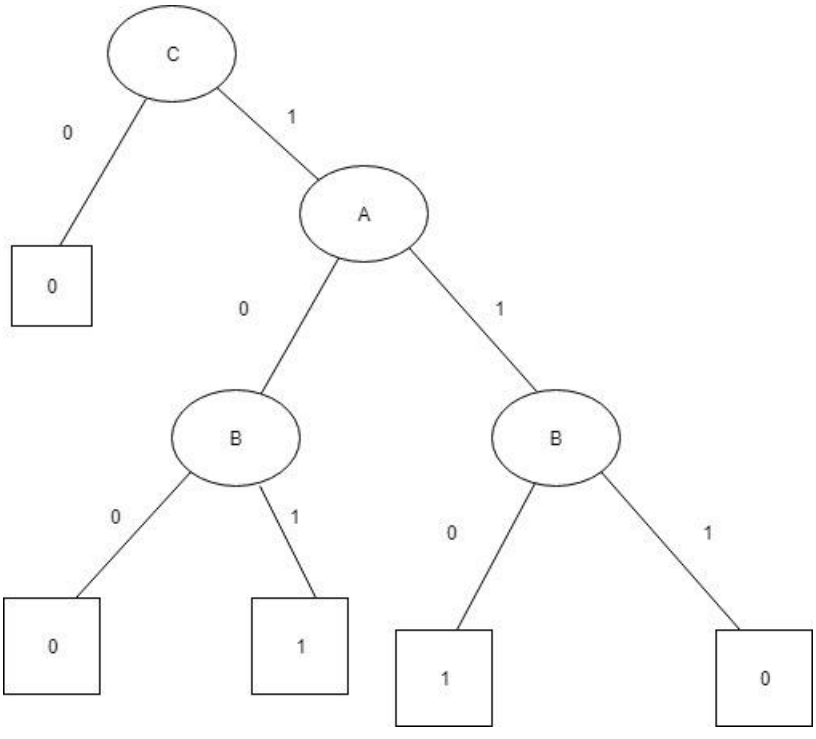


b) Y = (A xor B) ^ C

Truth table for the above expression:

A	B	C	A xor b	Y
0	0	0	0	0
0	0	1	0	0
0	1	0	1	0
0	1	1	1	1
1	0	0	1	0
1	0	1	1	1
1	1	0	0	0
1	1	1	0	0

Required Decision Tree with the least number of nodes :

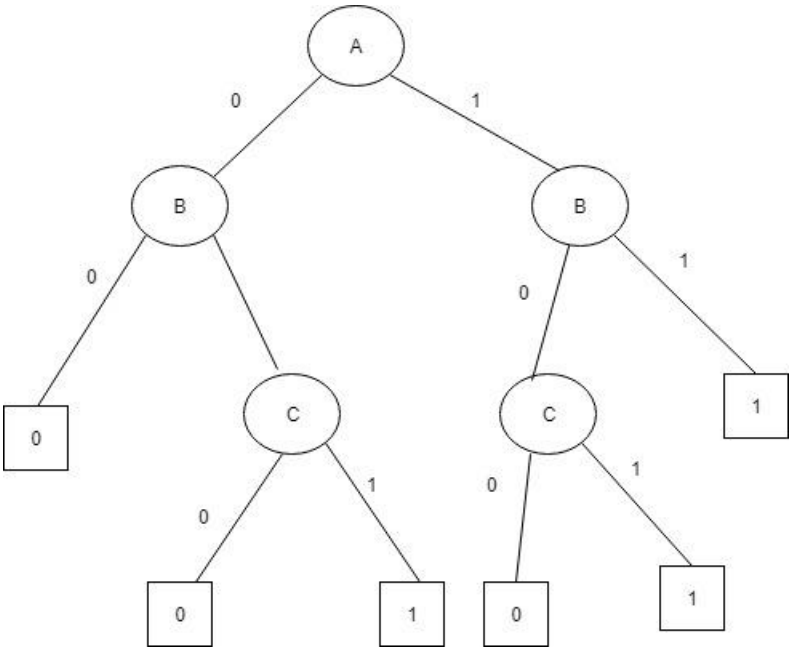


c) Y = (A v B) ^ (B v C) ^ (A v C)

Truth table for the above expression :

A	B	C	(A v B)	(B v C)	(A v C)	Y
0	0	0	0	0	0	0
0	0	1	0	1	1	0
0	1	0	1	1	0	0
0	1	1	1	1	1	1
1	0	0	1	0	1	0
1	0	1	1	1	1	1
1	1	0	1	1	1	1
1	1	1	1	1	1	1

The decision tree with least number of nodes :

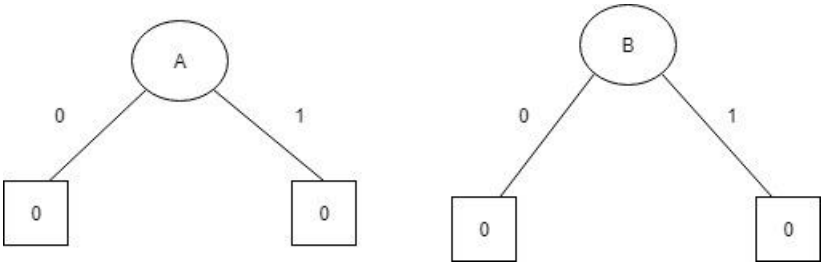


d) Y= (A V B) ^ ~A ^ ~B

Truth table for the above expression :

A	A	~A	~B	A V B	Y
0	0	1	1	0	0
0	1	1	0	1	0
1	0	0	1	1	0
1	1	0	0	1	0

Required decision tree with least number of nodes :



ANSWER FOR QUESTION 2:

Class: (5+, 5-)

H: Entropy = $\sum_i -(p_i \log_2 p_i)$

I.G. : Information Gain = H(Y) – H(Y | X)

H(class) = - 5/10 log 5/10 - 5/10 log 5/10 = 1

X1|1: (4+, 1-)

H(X1 | 1) = - 4/5 log 4/5 - 1/5 log 1/5 = 0.7219

X1|0: (1+, 4-)

H(X1 | 0) = - 1/5 log 1/5 - 4/5 log 4/5 = 0.7219

H (Class | X1) = (5/10 * 0.7219) + (5/10 * 0.7219) = **0.7219**

I.G.$_{x1}$ = 1 – 0.7219 = **0.2781**

X2|1: (2+, 1-)

H(X2 | 1) = - 2/3 log 2/3 - 1/3 log 1/3 = 0.9182

X2|0: (3+, 4-)

H(X1 | 0) = - 3/7 log 3/7 - 4/7 log 4/7 = 0.9852

H (Class | X2) = (3/10 * 0.9182) + (7/10 * 0.9852) = **0.9651**

I.G.$_{x2}$ = 1 – 0.9651 = **0.0349**

X3|1: (0+, 2-)

H(X3 | 1) = - 0/2 log 0/2 - 2/2 log 2/2 = 0

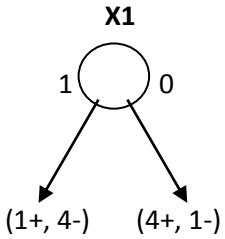
X3|0: (5+, 3-)

H(X3 | 0) = - 5/8 log 5/8 - 3/8 log 3/8 = 0.9544

H (Class | X3) = (2/10 * 0) + (8/10 * 0.9544) = **0.76352**

I.G.$_{x2}$ = 1 – 0.76352 = **0.23648**

Since Information Gain (I.G) is maximum for X1, we will split the gain based on X1.



For left child of X1 node:

H (class) = 0.7219

H (class |X2 = 0) = (0+, 3-) = 0

H (class |X2 = 1) = (1+, 1-) = 1

H (class | X2) = (3/5 * 0) + (2/5 * 1) = **0.4**

I.G.$_{x2}$ = 0.7219 – 0.4 = **0.3219**

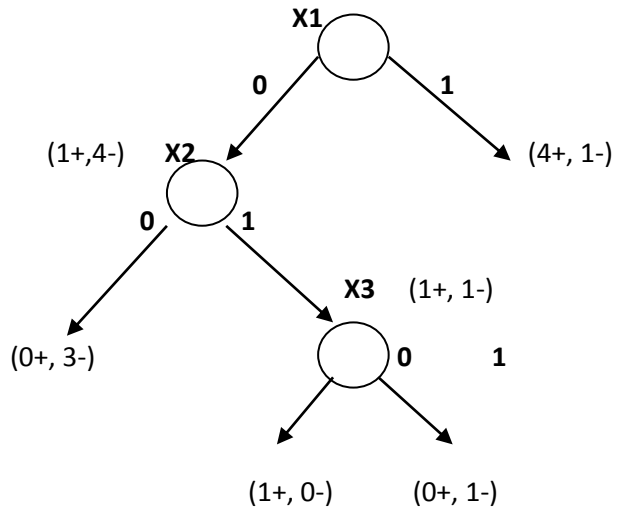
H (class |X3 = 0) = (1+, 3-) = - 1/4 log 1/4 - 3/4 log 3/4 = 0.8112

H (class |X3 = 1) = (0+, 1-) = 0

H (class | X3) = (4/5 * 0.8112) + (1/5 * 0) = **0.64896**

I.G.$_{x3}$ = 0.7219 – 0.64896 = **0.07294**

Since Information Gain is greater for X2, we split the data based on X2 attribute for left child of X1.



For right child of X1 node:

H (class) = 0.7219

H (class |X2 = 0) = (3+, 1-) = -3/4 log 3/4 – 1/4 log 1/4 = 0.8112

H (class |X2 = 1) = (1+, 0-) = 0

H (class | X2) = (4/5 * 0.8112) + (1/5 * 0) = **0.64896**

I.G.$_{x2}$ = 0.7219 – 0.64896 = **0.07294**

H (class |X3 = 1) = (0+, 1-) = 0

H (class |X3 = 0) = (4+, 0-) = 0

H (class | X3) = 0

I.G.$_{x3}$ = 0.7219 – 0= **0.7219**

Since Information gain is greater for X3, we will split the data based on X3 attribute for right child of X1.