INTRODUCTION TO ARTIFICIAL INTELLIGENCE
Lab 3 – Decision Trees

1) **ENTROPY:** Given is a data set consisting of 8 objects. These objects are characterized by one conditional attribute X and one decision attribute Y. For each attribute, calculate its entropy (i.e., H(X) and H(Y)). Finally, compute a conditional entropy H(Y|X).

ID	X	Y
1	1	1
2	1	0
3	0	1
4	0	0
5	0	0
6	1	1
7	0	0
8	1	0

	$Y=0$	$Y=1$	$\sum$
$X=0$			
$X=1$			
$\sum$			

$log_2(3) = 1.58$

$log_2(5) = 2.32$

Entropy

$H(X) =$

$H(Y) =$

Conditional entropy

$H(Y|X=0) =$

$H(Y|X=0) =$

$H(Y|X) =$

2) **ID3 ALGORITHM:** Given is a data set consisting of 8 objects characterized by three conditional attributes X and one decision attribute Y. For these alternatives calculate first three iterations of ID3 algorithm, choose attribute to split data based on information gain IG. Compute majority class and classification error rate. Assume that the default class is zero.

ID	X_1	X_2	X_3	Y
1	1	1	0	1
2	0	1	1	0
3	0	0	0	1
4	1	0	0	0
5	0	0	0	0
6	0	1	0	1
7	0	0	1	0
8	1	1	1	0

0th ITERATION

Majority class:

Classification error rate:

1th ITERATION

$H(Y) =$

$H(Y|X_1) = 0.95$

$H(Y|X_2) =$

$H(Y|X_3) = 0.60$

$IG(Y|X_1) =$

$IG(Y|X_2) =$

$IH(Y|X_3) =$

Split on: $\boldsymbol{X}_{\_}$

Left node: $\boldsymbol{X}_{\_} =$

ID	X_1	X_2	X_3	Y
1	1	1	0	1
3	0	0	0	1
4	1	0	0	0
5	0	0	0	0
6	0	1	0	1

Right node: $\boldsymbol{X}_{\_} =$

ID	X_1	X_2	X_3	Y
2	0	1	1	0
7	0	0	1	0
8	1	1	1	0

Majority class for left node:

Majority class for right node:

Classification error rate:

2nd ITERATION

$H(Y) =$

$H(Y|X_1) = 0.95$

$H(Y|X_2) = 0.55$

$H(Y|X_3) =$

$IG(Y|X_1) =$

$IG(Y|X_2) =$

$IH(Y|X_3) =$

Split on: $\boldsymbol{X}_{\_}$

Left node: $\boldsymbol{X}_{\_} =$

ID	X_1	X_2	X_3	Y
3	0	0	0	1
4	1	0	0	0
5	0	0	0	0

Right node: $\boldsymbol{X}_{\_} =$

ID	X_1	X_2	X_3	Y
1	1	1	0	1
6	0	1	0	1

Majority class for left node:

Majority class for right node:

Classification error rate:

3rd ITERATION

$H(Y) = 0.91$
$H(Y|X_1) = 0.67$
$H(Y|X_2) = 0.92$
$H(Y|X_3) = 0.92$

$IG(Y|X_1) = 0.25$
$IG(Y|X_2) = 0.00$
$IH(Y|X_3) = 0.00$
Split on: X_1

Left node: $X\_ =$

ID	X_1	X_2	X_3	Y
3	0	0	0	1
5	0	0	0	0

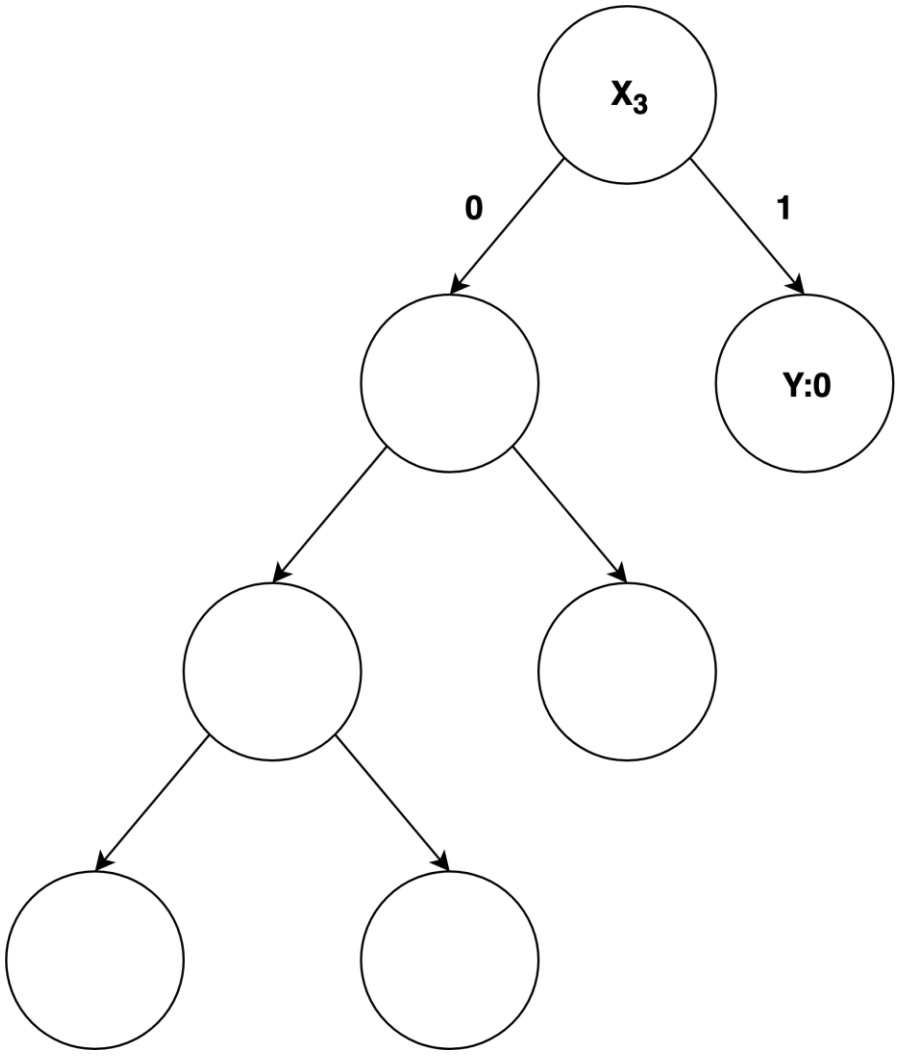
Right node: $X\_ =$

ID	X_1	X_2	X_3	Y
4	1	0	0	0

Majority class for left node:
Majority class for right node:

Classification error rate:

3) TREE BULDING: Finish the decision tree below based on the splits done in the previous exercise. Put attribute names in nodes, decisions in leaves, and attribute values on graph edges.



Classification of New objects

ID	X_1	X_2	X_3	Y
9	1	0	1	
10	1	0	0	