Homework II

1. Consider the training dataset given below. A, B, and C are the attributes and Y is the class variable.

[10 Points]

| A	B	C	Y
0	1	0	Yes
1	0	1	Yes
0	0	0	No
1	0	1	No
0	1	1	No
1	1	0	Yes

a. Can you draw a decision tree having 100% accuracy on this training set? If you answer is yes, draw the decision tree in the space provided below. If your answer is no, explain why?

b. Which attribute among A, B and C has the highest information gain? Explain your answer.

2. Interpreting a decision tree: Consider the decision boundary in Fig. and draw the equivalent decision tree. Red circle are Class +1 and blue squares are class -1.

[5 Points]
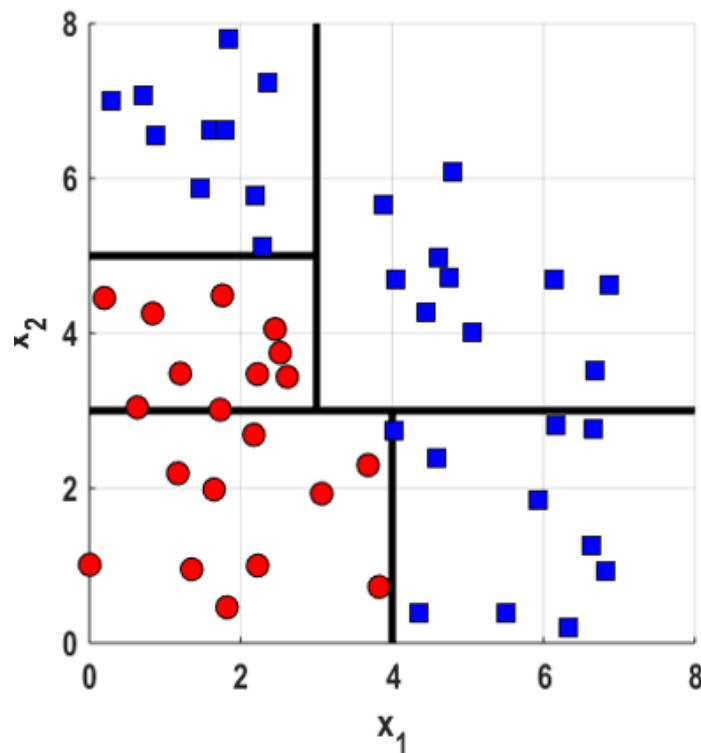
3. Visualizing a decision tree: Consider the decision in Fig and draw the equivalent decision boundary. Make sure to label each decision region with the corresponding leaf node from the decision tree.

   [5 Points]
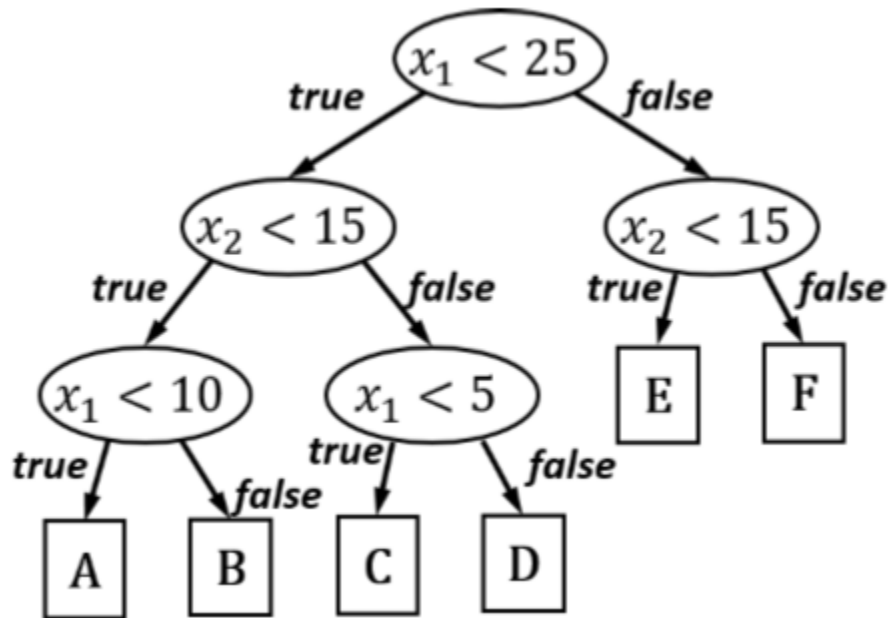
4. Learning a decision tree: Consider the synthetic data set shown in the table below, where the task is to predict if a car is going to be bought based on three features: its color, type and origin. Draw the decision tree that will be learned if you use accuracy as the splitting criterion.

Color	Type	Origin	Buy?
Red	Sports	Domestic	Yes
Red	Sports	Domestic	No
Red	Sports	Domestic	Yes
Blue	Sports	Domestic	No
Blue	Sports	Imported	Yes
Blue	SUV	Imported	No
Blue	SUV	Imported	Yes
Blue	SUV	Domestic	No
Red	SUV	Imported	No
Red	Sports	Imported	Yes

Analyze the overfitting nature of the decision tree.   **[10 Points]**

5. Bayes rule for medical diagnosis.
   After your yearly checkup, the doctor has some bad news and some good news. The bad news is that you tested positive for a serious disease, and the test is 99% accurate(i.e. that probability of testing positive given that you have the disease is .99, as is the probability of testing negative given that you don't have the disease). The good news is that this is a rare disease, striking only one in 10,000 people. What are the changes that you actually have the disease? (Show the calculation as well as giving the final result)

   [10 Points]

6. Express the mutual information in terms of the entropies. Show that
   I[X,Y] = H[X] − H[X|Y] = H[Y] − H[Y|X]   **[10 Points]**