1. Give decision trees to represent the following Boolean expressions
    a. $A \wedge \neg B$
    b. $A \vee [\neg B \wedge C]$
    c. $A \otimes B$
    d. $[A \vee B] \wedge [\neg C \wedge D]$
2. Consider the following set of training examples

Instance	Classification	A1	A2
1	+	T	T
2	+	T	T
3	-	T	F
4	+	F	F
5	-	F	T
6	-	F	T

    a) What is the entropy of this collection of training examples with respect to the target function classification?
    b) What is the information gain of A2 relative to these training examples?
3. Suppose we generate a training set from a decision tree and then apply decision-tree learning to that training set. Is it the case that the learning algorithm will eventually return the correct tree as the training-set size goes to infinity? Why or why not?
4. Suppose that an attribute splits the set of examples E into subsets E_k and that each subset has p_k positive examples and n_k negative examples. Show that the attribute has strictly positive information gain unless the ratio $p_k/(p_k + n_k)$ is the same for all k.
5. A *decision graph* is a generalization of a *decision tree* that allows nodes (i.e., attributes used for splits) to have multiple parents, rather than just a single parent. The resulting graph must still be acyclic. Now, consider the XOR function of three binary input attributes, which produces the value 1 if and only if an odd number of the three input attributes has value 1.
    a. Draw a minimal-sized *decision tree* for the three-input XOR function.
    b. Draw a minimal-sized *decision graph* for the three-input XOR function.