3.1 Give decision trees to represent the following boolean functions:
(a) A ∧ ¬B
(b) A ∨ [B ∧ C]
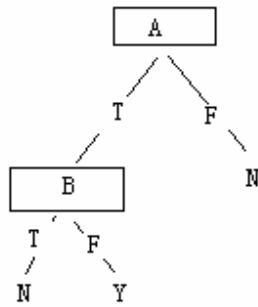(c) A XOR B
(d) [A∧ B] ∨ [C ∧ D]
Ans.
(a) A ∧ ¬B

(b) A ∨ [B ∧ C]
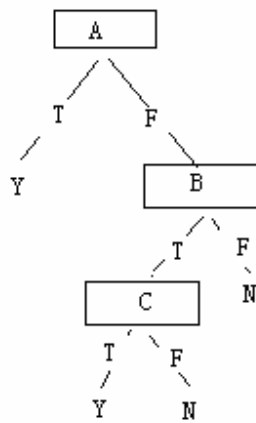
(c) A XOR B
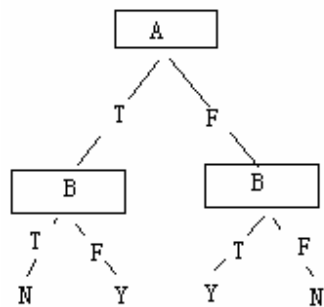
(d) [A∧ B] ∨ [C ∧ D]
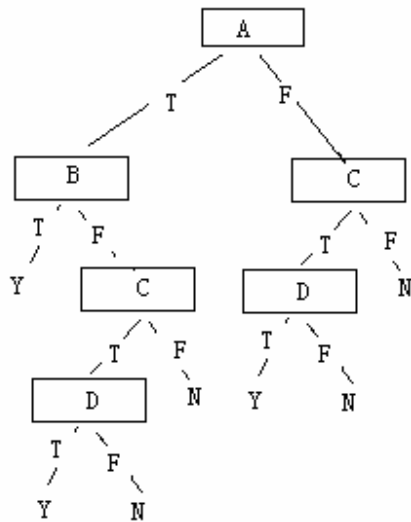
3.2 Consider the following set of training examples:

(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?

Instance	Classification	a_1	a_2
1	+	T	T
2	+	T	T
3	−	T	F
4	+	F	F
5	−	F	T
6	−	F	T

Ans.

(a) Entropy = 1

(b) Gain(a2) = 1-4/6*1-2/6*1 = 0

3.4. ID3 searches for just one consistent hypothesis, whereas the CANDIDATE-ELIMINATION algorithm finds all consistent hypotheses. Consider the correspondence between these two learning algorithms.

(a) Show the decision tree that would be learned by ID3 assuming it is given the four training examples for the *Enjoy Sport?* target concept shown in Table 2.1 of Chapter 2.

(b) What is the relationship between the learned decision tree and the version space (shown in Figure 2.3 of Chapter 2) that is learned from these same examples? Is the learned tree equivalent to one of the members of the version space?
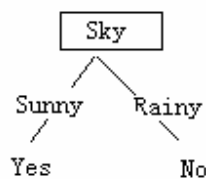
(c) Add the following training example, and compute the new decision tree. This time, show the value of the information gain for each candidate attribute at each step in growing the tree.

Sky Air-Temp Humidity Wind Water Forecast Enjoy-Sport?
Sunny Warm Normal Weak Warm Same No

Ans.

(a) Decision tree:

(b) Version space contains all hypotheses consistent with the training examples, whereas, the learned decision tree is one of the hypotheses (i.e., the first acceptable hypothesis with respect to the inductive bias) consistent with the training examples. Also, decision tree has a richer expression than hypothesis of version space which contains only conjunction forms of attribute constraints. If the target function is not contained in the hypothesis space (it may happen as {∧} is not a minimum complete basis), the version space will be empty. In this example, the learned decision tree "Sky = Sunny" is equivalent to <Sunny, ?,?,?,?,?> of G boundary set.

(c)
(1) First test:
Entropy(X) = -3/5*log2(3/5)-2/5*log2(2/5) = 0.971
Gain(X,Sky) = 0.971-4/5*(-3/4log2 (3/4)-(1/4)log2(1/4))-1/5*0 = 0.322
Gain(X,AirTemp) = 0.971-4/5*(-3/4log2 (3/4)-(1/4)log2(1/4))-1/5*0 = 0.322
Gain(X,Humidity) = 0.971-3/5*(-2/3log2 (2/3)-(1/3)log2(1/3))-2/5*1 = 0.02
Gain(X,Wind) = 0.971-4/5*(-3/4log2 (3/4)-(1/4)log2(1/4))-1/5*0 = 0.322
Gain(X,Water) = 0.971-4/5*(-2/4log2 (2/4)-(2/4)log2(2/4))-1/5*0 = 0.171
Gain(X,Forcast) = 0.971-3/5*(-2/3log2 (2/3)-(1/3)log2(1/3))-2/5*1 = 0.02
So, we choose "Sky" as the test attribute for the root. (note: You can also select AirTemp or Wind as the test attribute)
(2) Second test:
Entropy(X) = -3/4*log2(3/4)-1/4*log2(1/4) = 0.8113
Gain(X,AirTemp) = 0
Gain(X,Humidity) = 0.3113
Gain(X,Wind) = 0.8113
Gain(X,Water) = 0.1226
Gain(X,Forcast) = 0.1226
So, we choose "Wind" for test.
Decision tree: