3.1 Give decision trees to represent the following boolean functions:
(a) \( A \land \neg B \)
(b) \( A \lor (B \land C) \)
(c) \( A \text{ XOR } B \)
(d) \( (A \land B) \lor (C \land D) \)
Ans.
(a) \( A \land \neg B \)

(b) \( A \lor (B \land C) \)

(c) \( A \text{ XOR } B \)

(d) \( (A \land B) \lor (C \land D) \)
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 \( a_2 \) relative to these training examples?

\[
\begin{align*}
\text{Ans.} \\
(a) \text{Entropy} &= 1 \\
(b) \text{Gain}(a_2) &= 1 - \frac{4}{6} \cdot 1 - \frac{2}{6} \cdot 1 = 0
\end{align*}
\]

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.

<table>
<thead>
<tr>
<th>Sky</th>
<th>Air-Temp</th>
<th>Humidity</th>
<th>Wind</th>
<th>Water</th>
<th>Forecast</th>
<th>Enjoy-Sport?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[
\begin{align*}
\text{Ans.} \\
(a) \text{Decision tree:}
\end{align*}
\]
(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 $\land$ is not a minimum complete basis), the version space will be empty. In this example, the learned decision tree “Sky = Sunny” is equivalent to $<\text{Sunny}, ?, ?, ?, ?, ?>$ of G boundary set.

(c)

(1) First test:

Entropy($X$) = $-3/5 \cdot \log_2(3/5) - 2/5 \cdot \log_2(2/5) = 0.971$

Gain($X,\text{Sky}$) = $0.971 - 4/5 \cdot (-3/4 \cdot \log_2(3/4) - (1/4) \log_2(1/4)) - 1/5 \cdot 0 = 0.322$

Gain($X,\text{AirTemp}$) = $0.971 - 4/5 \cdot (-3/4 \cdot \log_2(3/4) - (1/4) \log_2(1/4)) - 1/5 \cdot 0 = 0.322$

Gain($X,\text{Humidity}$) = $0.971 - 3/5 \cdot (-2/3 \cdot \log_2(2/3) - (1/3) \log_2(1/3)) - 2/5 \cdot 1 = 0.02$

Gain($X,\text{Wind}$) = $0.971 - 3/5 \cdot (-2/3 \cdot \log_2(2/3) - (1/3) \log_2(1/3)) - 1/5 \cdot 0 = 0.322$

Gain($X,\text{Water}$) = $0.971 - 4/5 \cdot (-2/4 \cdot \log_2(2/4) - (2/4) \log_2(2/4)) - 1/5 \cdot 0 = 0.171$

Gain($X,\text{Forcast}$) = $0.971 - 3/5 \cdot (-2/3 \cdot \log_2(2/3) - (1/3) \log_2(1/3)) - 2/5 \cdot 1 = 0.02$

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($X$) = $-3/4 \cdot \log_2(3/4) - 1/4 \cdot \log_2(1/4) = 0.8113$

Gain($X,\text{AirTemp}$) = 0

Gain($X,\text{Humidity}$) = 0.3113

Gain($X,\text{Wind}$) = 0.8113

Gain($X,\text{Water}$) = 0.1226

Gain($X,\text{Forcast}$) = 0.1226

So, we choose “Wind” for test.

Decision tree: