Question 1

1. Find-S predicted negative; the true label was positive.

2. The most general hypothesis: \( \langle ?, ?, ?, ?, ?, ? \rangle \)

3. One mistake on the first instance. Then, in the worst case scenario each following instance on which Find-S makes a mistake will change one of the attribute constraints from a specific value to a \( ? \). This will add 6 more mistakes. So total 7 mistakes. An example sequence would be:

<table>
<thead>
<tr>
<th>Example</th>
<th>Updated Hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \langle Su, Wa, N, St, Wa, Sa \rangle, Yes )</td>
<td>( \langle Su, Wa, N, St, Wa, Sa \rangle )</td>
</tr>
<tr>
<td>( \langle Ra, Wa, N, St, Wa, Sa \rangle, Yes )</td>
<td>( \langle ?, Wa, N, St, Wa, Sa \rangle )</td>
</tr>
<tr>
<td>( \langle Ra, Co, N, St, Wa, Sa \rangle, Yes )</td>
<td>( \langle ?, ?, N, St, Wa, Sa \rangle )</td>
</tr>
<tr>
<td>( \langle Ra, Co, H, St, Wa, Sa \rangle, Yes )</td>
<td>( \langle ?, ?, ?, St, Wa, Sa \rangle )</td>
</tr>
<tr>
<td>( \langle Ra, Co, H, We, Wa, Sa \rangle, Yes )</td>
<td>( \langle ?, ?, ?, Wa, Sa \rangle )</td>
</tr>
<tr>
<td>( \langle Ra, Co, H, We, Co, Sa \rangle, Yes )</td>
<td>( \langle ?, ?, ?, ?, Sa \rangle )</td>
</tr>
<tr>
<td>( \langle Ra, Co, H, We, Co, Ch \rangle, Yes )</td>
<td>( \langle ?, ?, ?, ?, ? \rangle )</td>
</tr>
</tbody>
</table>

4. The first instance is an arbitrary instance. Each following instance disagrees with the current hypothesis in exactly one attribute. Total \( m + 1 \) mistakes.

Question 2

1. In all drawings, left branches correspond to True, and right branches correspond to False.

a)

```
     A
    / \   \
   B False
    / \   \
False True
```
b) 

```
A
 / \ 
 True   B
  / \ 
 C   False
 / \ 
 True False
```

c) 

```
A
 / \ 
 B   B
 / \ / \ 
False True True False
```

d) 

```
A
 / \ 
 B   C
 /   / 
True   D False
 / \ 
 True False
```

2. False. Counterexample:
Suppose each instance is described by two attributes, A and B. Let D1 and D2 be as follows (using same encoding scheme as above):

```
D1:   A
 / \ 
1 0
D2:   A
 / \ 
1 0
```

By the definition of more-general-than, D1 is more-general-than D2 iff whenever \( D2(X) = 1 \), it is also true that \( D1(X) = 1 \). The instance \( \langle A = False, B = True \rangle \) contradicts this. It is classified as 1 by D2 but as 0 by D1.

**Question 3** 5. The unpruned tree is a lot more complex and contains longer, bushier branches with more tests. It provides a better fit to the training data but its error on unseen instances might be higher. Pruning reduces overfitting and leads to simpler models. Main advantages of it are that the predictive accuracy of the learned hypothesis is increased, and the learned model is easier to interpret by a human.