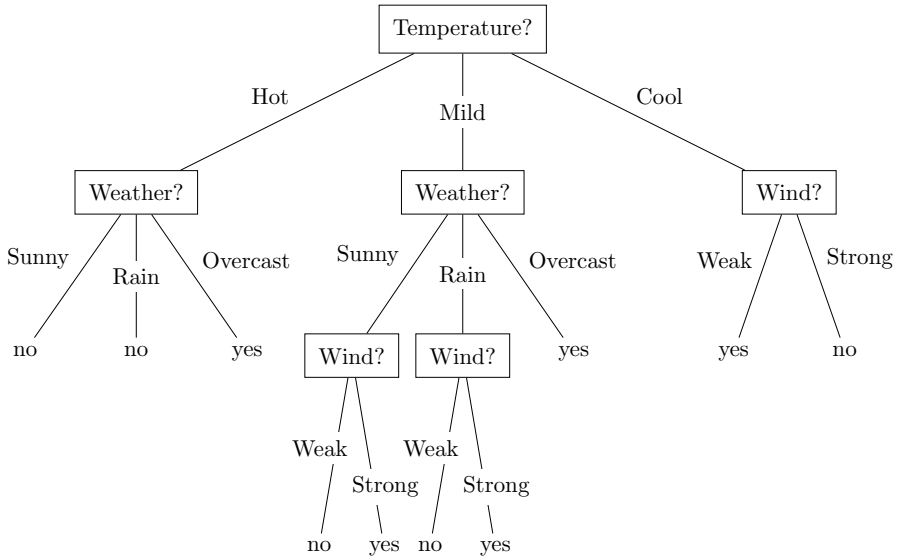
IA Solutions 5: Learning Decision Trees

Question 1

There are many examples that work.



Note here we could also have set the leaf reached after "Hot" and "Rain" to yes.

Question 2

We just need to insert the correct values into the formulas.

$$\mathsf{gain}(S, \text{Wind}) = \mathsf{entropy}(S) - \sum_{v \in \{\text{Weak,Strong}\}} \frac{\mid S_v \mid \cdot \mathsf{entropy}(S_v)}{\mid S \mid}$$

Question 3

By the table we see that asking for the state of the Weather will probably provide the most information. We now determine which question to ask next, given the three possible responses.

So first suppose that we learn that it is Sunny. We want to compare the following three values to determine which question to ask next:

$$\mathsf{gain}(S_{\text{Sunny}}, \text{Temperature}) \tag{1}$$
$$\mathsf{gain}(S_{\text{Sunny}}, \text{Humidity}) \tag{2}$$
$$\mathsf{gain}(S_{\text{Sunny}}, \text{Wind}) \tag{3}$$

Now, note that entropy is always positive. The entropy when we know that it is Sunny and the Humidity is High is zero. The entropy is also zero when it is Sunny and the Humidity is Normal. Thus

$$\sum_{v \in \{\text{High,Normal}\}} \frac{\mid (S_{\text{Sunny}})_v \mid \cdot \mathsf{entropy}((S_{\text{Sunny}})_v)}{\mid S_{\text{Sunny}} \mid} = 0$$

This means that the gain for Humidity in this case is maximal: in particular it is exactly $\mathsf{entropy}(S_{\text{Sunny}})$. Intuitively, this means that if we know it is Sunny, then after we know the state of the Humidity we know exactly whether we are in a "yes" or "no" case: the entropy is zero. Note that if we ask any other question here, we might not know for sure whether we are in a "yes" or "no" case: the expected entropy is greater than zero.

If we know that there is Rain, then asking the state of the Wind reduces the entropy to zero. Finally, in the Overcast case, we already have zero entropy, so we do not need to ask any more questions. Putting this altogether leads to the following decision tree: