Learning Decision Trees

Day	Attribute				Play
	Temperature	Humidity	Wind	Weather	
1	Hot	High	Weak	Sunny	no
2	Hot	High	Strong	Sunny	no
3	Hot	High	Weak	Overcast	yes
4	Mild	High	Weak	Rain	yes
5	Cool	Normal	Weak	Rain	yes
6	Cool	Normal	Strong	Rain	no
7	Cool	Normal	Strong	Overcast	yes
8	Mild	High	Weak	Sunny	no
9	Cool	Normal	Weak	Sunny	yes
10	Mild	Normal	Weak	Rain	yes
11	Mild	Normal	Strong	Sunny	yes
12	Mild	High	Strong	Overcast	yes
13	Hot	Normal	Weak	Overcast	yes
14	Mild	High	Strong	Rain	no

Question 1 : Find some decision tree that is coherent with the above *training set* and that always determines whether or not the user will play.

Question 2 : Let S be a set of instances for which the classification is binary (each instance is classified as positive or negative). In general, *entropy* is a measure of the disorder of a system. An information theoretic interpretation is that entropy is the minimum number of bits required, on average, to determine the classification of an arbitrary member of S. Assuming all instances have equal probability, the entropy of S is :

$$\text{entropy}(S) = -p_{\oplus} log_2(p_{\oplus}) - p_{\ominus} log_2(p_{\ominus})$$

where

- $p_{\oplus}$ is the proportion of positive examples in S
- $p_{\ominus}$ is the proportion of negative examples in S

In order to greedily define a *compact* decision tree, at each node we request the value of the attribute that we believe will give us the most information. Thus we need some measure of the amount of information gained when we learn the assignment of some attribute. We calculate this as the expected reduction in entropy. Write $\text{val}(A)$ for the set of values of an attribute A, for example $\text{val(Temperature)} = \{\text{Hot}, \text{Mild}, \text{Cool}\}$. For $v \in \text{val}(A)$, write S_v for the subset of instances that have value v, for example $S_{\text{Hot}} = \{1, 2, 3, 13\}$ (where each number indicates the day of the instance). For a set of instances S and an attribute A, the gain is defined by :

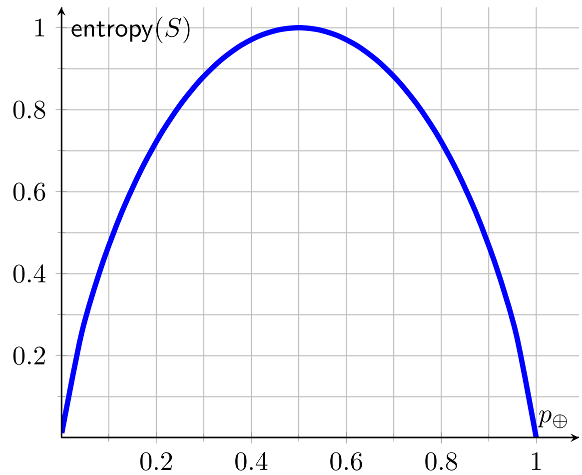
$$\text{gain}(S, A) = \text{entropy}(S) - \sum_{v \in \text{val}(A)} \frac{|S_v|}{|S|} \text{entropy}(S_v)$$

Apply the formula gain to the attributes `Wind` and `Temperature`.

Question 3 : For the full set of instances we have the following table :

	Weather	Humidity	Wind	Temperature
Gain	0.246	0.151	0.048	0.029

— Verify (at home) the values of this table.
— Consider now the case when the weather is sunny. Give expressions that calculate the gain in information for the remaining three attributes. Without fully calculating the answers, do you have a good idea of which attribute provides the best gain ?
— Can you determine the values for the attributes for the case when it rains ?
— Do the same for the case when the weather is overcast.
— Draw the decision tree produced by the greedy procedure.



entropy(S) for a set of instances S of which $p_\oplus$ is the proportion of positive instances and $1 - p_\oplus$ is the proportion of negative instances.