Practice Problem on Decision Trees: Happy Halloween!

In the U.S. and many countries around the world, children in costumes go from place to place in the evening of October 31st, asking for treats with the phrase "Trick or treat". The "treat" is usually candy, while the "trick" refers to a "trickery" performed to the homeowner or their property if no treat is given. We would like to predict whether a trick or a treat will be performed for different houses given the availability of candy (Candy) and the degree of "spookiness" of one's costume (Spooky). The data that we have collected from previous years are the following:

	Sample	**Features**		**Outcome**
		Candy	**Spooky**	**Treat**
Train	**S1**	No	Little	No
	S2	No	Very	No
	S3	Moderate	Little	Yes
	S4	Great	Little	Yes
	S5	Great	Little	Yes
	S6	Great	Very	No
	S7	Moderate	Very	Yes

Based on the above data, we will build a decision tree using the information entropy as a splitting criterion. The input features are **Candy** and **Spooky**, while the outcome variable is **Treat**.

(a) Compute the entropy splitting criterion of the outcome **Treat** conditioned on the **Candy** and **Spooky** features. Which feature will be used as the splitting attribute in root of the tree? Show all your calculations.
Note: You **do not** need to perform arithmetic calculations for logarithms, e.g. if one of your equations contains $\log(\frac{1}{3})$, you can leave it like that and still solve the problem.

We will compute the partial information entropy for each feature value. Then we will obtain the information entropy of each feature using a weighted mean, where the weights correspond to the probability of each value. In the following calculations, by convention, we ignore that $log0 \to \infty$, and assume that the corresponding terms contribute with zero.

$$H(Treat|Candy = No) = -\left[\frac{2}{0+2}log\frac{2}{0+2} + \frac{0}{0+2}log\frac{0}{0+2}\right] = 0$$

$$H(Treat|Candy = Moderate) = -\left[\frac{0}{2+0}log\frac{0}{2+0} + \frac{2}{2+0}log\frac{2}{2+0}\right] = 0$$

$$H(Treat|Candy = Great) = -\left[\frac{1}{1+2}log\frac{1}{1+2} + \frac{2}{1+2}log\frac{2}{1+2}\right] = -\frac{1}{3}log\frac{1}{3} - \frac{2}{3}log\frac{2}{3}$$

$$
\begin{aligned}
H(Treat|Candy) &= \frac{2}{7}H(Treat|Candy = No) + \frac{2}{7}H(Treat|Candy = Moderate) \\
&+ \frac{3}{7}H(Treat|Candy = Great) \\
&= -\frac{3}{7}\left(\frac{1}{3}log\frac{1}{3} + \frac{2}{3}log\frac{2}{3}\right)
\end{aligned}
$$

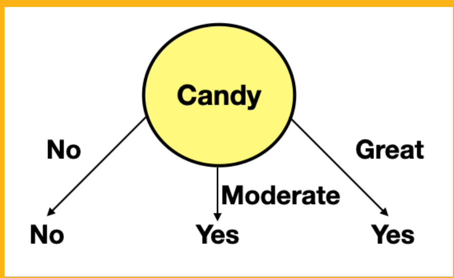
$$H(Treat|Spooky = Little) = -\left[\tfrac{1}{1+3}log\tfrac{1}{1+3} + \tfrac{3}{1+3}log\tfrac{3}{1+3}\right] = -\tfrac{1}{4}log\tfrac{1}{4} - \tfrac{3}{4}log\tfrac{3}{4}$$

$$H(Treat|Spooky = Very) = -\left[\tfrac{2}{2+1}log\tfrac{2}{2+1} + \tfrac{1}{2+1}log\tfrac{1}{2+1}\right] = -\tfrac{2}{3}log\tfrac{2}{3} - \tfrac{1}{3}log\tfrac{1}{3}$$

$$
\begin{aligned}
H(Treat|Spooky) &= \frac{4}{7}H(Treat|Spooky = Little) + \frac{3}{7}H(Treat|Spooky = Very) \\
&= -\frac{4}{7}\left(\frac{1}{4}log\frac{1}{4} + \frac{3}{4}log\frac{3}{4}\right) - \frac{3}{7}\left(\frac{2}{3}log\frac{2}{3} + \frac{1}{3}log\frac{1}{3}\right)
\end{aligned}
$$

We observe that $H(Treat|Candy) < H(Treat|Spooky)$, therefore **Candy** will be the splitting criterion for the root node.

(b) Create the decision tree using only one node, i.e., the tree will only have the root. Please show the **splitting criterion of the node**, as well as **the decisions from each possible outcome of the corresponding criterion**. Please **describe how the decisions were made**.



All samples with $Candy = No$ correspond to $Treat = No$.

All samples with $Candy = Moderate$ correspond to $Treat = Yes$.

The majority of samples (i.e., 2 out of 3) with $Candy = Great$ correspond to $Treat = Yes$, therefore we will be taking this decision if we make this as a terminal node. If we select not to make this as a terminal node, then we will keep expanding the tree.

(c) Which of the training samples will be classified correctly only using the above tree and which not?

Classified correctly: $S1$, $S2$, $S3$, $S4$, $S5$, $S7$

Classified incorrectly: $S6$

(d) Expand the tree until all samples are correctly classified.

We will compute the information entropy for both features only for the samples that have reached the node $Candy = Great$ (i.e., $S4$, $S5$, $S6$).

$H(Treat|Candy) = H(Treat|Candy = Great) = -\frac{1}{3}log\frac{1}{3} - \frac{2}{3}log\frac{2}{3}$

$H(Treat|Spooky = Little) = 0$, since the outcome is Yes for both $S4$ and $S5$, for which $Spooky = Little$

$H(Treat|Spooky = Very) = 0$, since the outcome is No for $S6$, for which $Spooky = Very$

Therefore $H(Treat|Spooky) < H(Treat|Candy)$, so we will pick feature **Spooky** in this node, and the tree will looks as follows: