Take Home Quiz 2 (Dr. Winslow's Lectures): Classification Trees
~~Due THURSDAY, October 17th, 11:59 pm.~~
Extended to FRIDAY, October 18th, 11:59 pm.

Please submit a typeset PDF (compiled LaTeX, exported Word Doc, etc.).
BUT if you choose to draw a tree, it is okay to import a image drawn by hand.

To reduce my email load, I want to implement a machine learning algorithm to decide whether or not I should read an email, or simply file it away instead. To train my model, I obtain the following data set of binary-valued features about each email, including:

- whether or not I know the author
- whether the email is long or short
- whether it has any of several key words

and my final decision about whether to read it (y = +1 for "read", y = -1 for "discard").

x_1 know author?	x_2 is long?	x_3 has 'research'	x_4 has 'grade'	x_5 has 'lottery'	y $\Rightarrow$ read?
0	0	1	1	0	-1
1	1	0	1	0	-1
0	1	1	1	1	-1
1	1	1	1	0	-1
0	1	0	0	0	-1
1	0	1	1	1	1
0	0	1	0	0	1
1	0	0	0	0	1
1	0	1	1	0	1
1	1	1	1	1	-1

In the case of ties where both classes have equal probability, predict class +1.

1. Calculate the entropy H(y) of the binary class variable y

2. Calculate the information gain for each feature xi . Which feature should I split on for the root node of the decision tree?

3. Determine the complete decision tree that will be learned from these data. (The tree should perfectly classify all training data). Specify the tree by drawing it, or with a set of nested if-then-else statements.