COEN 240 Machine Learning

Homework #5

Problem 1

You are a robot in a lumber yard, and must learn to discriminate Oak wood from Pine wood. You choose to learn a Decision Tree classifier. You are given the following examples:

Example	Density	Grain	Hardness	Class
Example #1	Heavy	Small	Hard	Oak
Example #2	Heavy	Large	Hard	Oak
Example #3	Heavy	Small	Hard	Oak
Example #4	Light	Large	Soft	Oak
Example #5	Light	Large	Hard	Pine
Example #6	Heavy	Small	Soft	Pine
Example #7	Heavy	Large	Soft	Pine
Example #8	Heavy	Small	Soft	Pine

1.1 Which attribute will be chosen as the root of the tree (show derivations)?

1.2 Derive the complete decision tree by recursively applying the smallest entropy criterion to select root nodes of sub-trees (show derivations). Then draw the complete decision tree.

Problem 2

NASA wants to discriminate Martians (M) from Humans (H) based on these features (attributes): Green $\in$ {N, Y}, Legs$\in$ {2,3}, Height$\in$ {S, T}, Smelly $\in$ {N, Y}. Your available training data is as follows (N=No, Y=Yes, S=Short, T=Tall):

Example Number	Height	Green	Legs	Smelly	Target: Species
1	S	Y	3	Y	M
2	T	Y	3	N	M
3	S	Y	3	N	M
4	T	Y	3	N	M
5	T	N	2	Y	M
6	T	Y	2	Y	H
7	S	N	2	N	H
8	T	N	3	N	H
9	S	N	3	N	H
10	T	N	3	N	H

Please note:
A human might be green or have three legs for many possible reasons, e.g., if they were an actor playing a Martian as a role in a film or play. Anyway, it's a made-up problem

(a) What is the entropy of the target species before testing any attribute?

(b) For each possible choice of root attribute below, show the resulting species distribution. Give your answer as M over H. The first one is done for you, as an example.

(example) Height
S — M M / H H
T — M M M / H H H

Green
Y —
N —

Legs
2 —
3 —

Smelly
Y —
N —
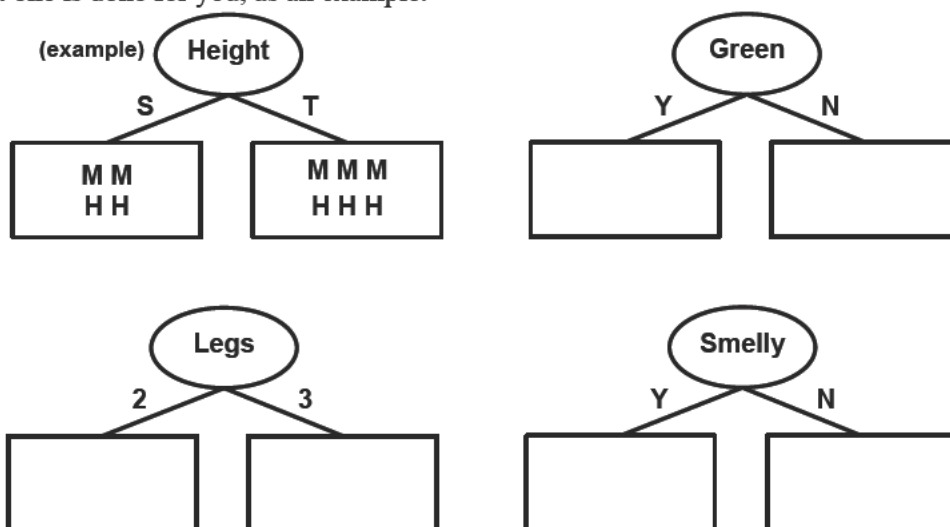
(c) What is the conditional entropy of species under attribute Height?

(d) What is the conditional entropy of species under attribute Green?

(e) What is the conditional entropy of species under attribute Legs?

(f) What is the conditional entropy of species under attribute Smelly?

(g) Which attribute would you select as the root attribute (i.e., the attribute to test first)? Why?

(h) Derive the complete decision tree by recursively applying the smallest entropy criterion to select root nodes of sub-trees (show derivations). Then draw the complete decision tree.