

Assignment 1: Machine Learning

- 1 For the following decision table:

	Attributes				Decision
	Size	Color	Feel	Material	Attitude
1	big	yellow	soft	fabric	positive
2	small	yellow	soft	fabric	positive
3	small	blue	hard	plastic	positive
4	small	yellow	soft	fabric	positive
5	small	red	soft	fabric	positive
6	small	red	hard	fabric	negative
7	big	yellow	hard	plastic	negative
8	big	blue	soft	plastic	negative

- (a) determine a decision tree using the ID3 (information gain) algorithm
 - (b) derive a set of rules from the decision tree
 - (c) simplify the rules using linear (not exponential) dropping conditions.
- 2 For the same table:
- (a) determine a decision tree using the C4.5 (information gain ratio) algorithm
 - (b) derive a set of rules from the decision tree
 - (c) simplify the rules using linear dropping conditions.
- 3 For the same table:
- (a) determine a decision tree using the ID3 (information gain) algorithm and windowing as follows:
 - (i) apply ID3 to a window consisting of only the last four examples in the table
 - (ii) check whether the resulting table correctly classifies all the other items in the original table
 - (iii) if not, add any incorrectly classified items to the window, and apply ID3 again to the extended window
 - (iv) iterate until all items in the original table are correctly classified
 - (b) derive a set of rules from the final decision tree
 - (c) simplify the rules using linear (not exponential) dropping conditions.
- 4 Prove that $H(X, Y) = H(X) + H(Y|X)$. (Do not assume X and Y are independent)

Notes: Show all workings and explain every step of the tree-building process.

If attributes are tied at any stage, break ties from left to right (i.e., prefer the left-most)

Question 4 requires a simple half-page proof from definitions – e-mail me if you need hints.

There will be a second part of this assignment that I am deferring for a week until after I finish discussing machine learning next class.