CMSC 421 Homework 5
OPTIONAL Homework

Due Date: Tuesday, December 13 at the start of class

1. **[20 points] Decision tree learning**

   The following table gives a data set for deciding whether or not to go sailing, depending on the weather conditions.

Outlook	Temp (F)	EngineOK?	Windy?	Class
sunny	75	yes	true	Sail
sunny	80	no	true	Don't Sail
sunny	85	yes	false	Don't Sail
sunny	72	yes	false	Don't Sail
sunny	69	yes	false	Sail
overcast	72	yes	true	Sail
overcast	83	yes	false	Sail
overcast	64	yes	true	Sail
overcast	81	yes	true	Sail
rain	71	no	true	Don't Sail
rain	65	no	true	Don't Sail
rain	75	yes	false	Don't Sail
rain	68	yes	false	Don't Sail
rain	70	yes	false	Don't Sail

   (a) At the root node for a decision tree in this domain, what are the information gains associated with the Outlook and EngineOK attributes?

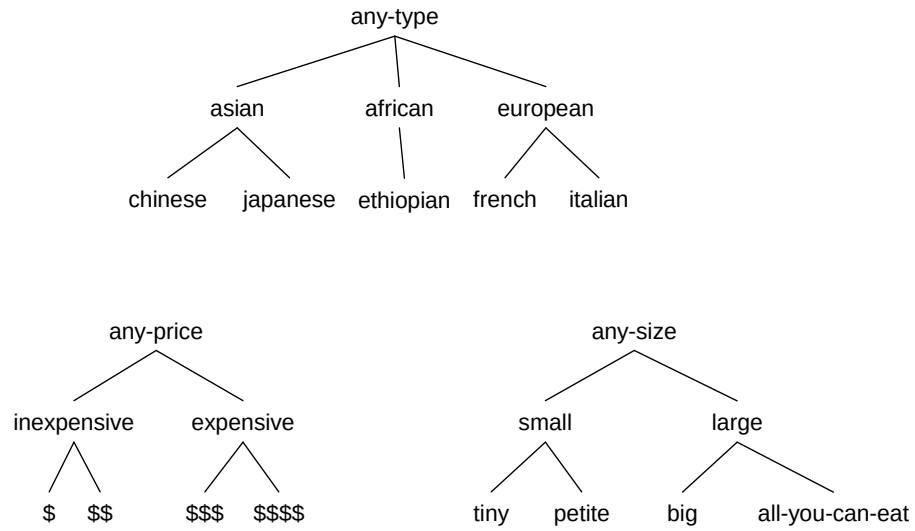
   (b) Again at the root node, what are the gain ratios associated with the Outlook and EngineOK attributes?

   (c) Suppose you build a decision tree that splits on the Outlook attribute at the root node. How many children nodes are there are at the first level of the decision tree? Which branches require a further split in order to create leaf nodes with instances belonging to a single class? For each of these branches, is there an attribute you can split on to complete the decision tree building process at the next level (i.e., so that at level 2, there are only leaf nodes)? Draw the resulting decision tree, showing the decisions (class predictions) at the leaves.

2. **[20 points] Version Spaces**

   Consider a domain describing restaurants and consider the version space defined by the concepts that are a conjunction of the following three attributes: cuisine type, price and portion size. Suppose we have the generalization hierachies for each of the attributes shown in the figure.

   Examples of legal concepts in this domain include [chinese $ all-you-can-eat], [european expensive small], and [any-type any-price any-size]. No disjunctive concepts are allowed, other than the implicit disjunctions represented by the internal nodes in the attribute hierarchies. (For example, the concept [[asian v african] $ large] isn't allowed.)



(a) Consider the initial version space for learning in this domain. What is the G set? How many elements are in the S set? Give one representative member of the initial S set.

(b) Suppose the first example I1 is a positive example, with attribute values [chinese $ all-you-can-eat]. After processing this instance, what are the G and S sets?

(c) Now suppose the learning algorithm receives example I2, a negative example, with attribute values [french $$$$ petite]. What are the G and S sets after processing this example?

(d) If learning ends at this point, how many possible concepts remain in the version space?

(e) For this question, you should start with the version space that remains after I1 and I2 are processed. For each of the following combinations of instance type and events, give a single instance of the specified type that would cause the indicated event, if such an instance exists. If no such instance exists, explain why.

i. A positive instance that causes the version space to collapse.

ii. A negative instance that causes the version space to collapse.

iii. A positive instance that reduces the size of the version space to a single concept.

iv. A negative instance that reduces the size of the version space to a single concept.

3. **[20 points] Statistical Learning and Utilities** R & N, 20.3

4. **[20 points] Neural Networks** R & N, 20.11