198:440
Introduction to Artificial Intelligence
Professor McCarty

Sample Final Exam

This is a closed book examination, but you may have with you a "cheat sheet" consisting of one (8.5 in. × 11 in.) piece of paper with writing on both sides.

Name:____

ID: ____

The following space is for the use of the grader:

1. ________________________
2. ________________________
3. ________________________
4. ________________________
5. ________________________

Question 1 (20 points):

In your first programming project, you solved the 8-Puzzle using two heuristic functions: *misplaced-tiles* and *manhattan-distance*. Here are two additional heuristic functions for the 8-Puzzle:

1. Consider a function called *xy-misplaced-tiles*, which counts separately the number of tiles that are in the wrong row and the number of tiles that are in the wrong column, and adds these two numbers together. Is this heuristic function admissible? Explain why or why not.

   Does *xy-misplaced-tiles* dominate either of the heuristic functions used in your programming project? Explain why or why not.

   Is *xy-misplaced-tiles* dominated by either of the heuristic functions used in your programming project? Explain why or why not.

2. Consider a function called *out-of-sequence*. This function counts 1 if a state has a central tile, and counts 2 for each tile on the perimeter that is not followed in clockwise order by its proper successor. For example, here is the start state and the goal state used as an illustration in our text:

5	**4**	
6	**1**	**8**
7	**3**	**2**

start

1	**2**	**3**
8		**4**
7	**6**	**5**

goal

The start state in this example has a central tile (= +1), and the only tile on the perimeter that is followed in clockwise order by its proper successor is 2. Thus 6 tiles are out of sequence, and the *out-of-sequence* score is 13. (Note: Tile 4 is followed in clockwise order by tile 8, not by the blank.)

Is the heuristic function *out-of-sequence* admissible? Explain why or why not.

There is some empirical evidence in the literature that a search routine for the 8-Puzzle achieves very good performance using as a heuristic function:

$$h(n) = \textit{manhattan-distance}(n) + 3 * \textit{out-of-sequence}(n)$$

Can you explain why this choice of h might work better than *manhattan-distance* alone?

Question 2 (20 points):

Here is a classic example of a problem that could not be solved using Aristotle's syllogisms. The premise is: "All horses are animals." The conclusion is: "The head of a horse is the head of an animal." Can you derive this conclusion from this premise using first-order logic?

You should use the predicates *Horse(x)* and *Animal(x)*, with their obvious meanings, and the predicate *HeadOf(h,x)* with the meaning "h is the head of x." Follow these steps:

1. Write the premise in first-order logic.

2. Convert the premise to conjunctive normal form.

3. Write the conclusion in first-order logic. (Hint: To say that "h is the head of a horse" you can say that "there exists a horse and h is its head," and similarly for "h is the head of an animal.")

4. Negate the conclusion and convert it to conjunctive normal form.

5. Construct a resolution proof from these clauses that terminates with the empty clause.

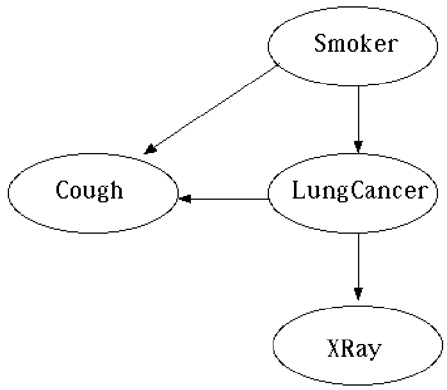

Figure 1: Bayesian Network

Question 3 (20 points):

Figure 1 shows a Bayesion network (or belief network) with Boolean variables: *Smoker*, *Cough*, *LungCancer* and *XRay*. Intuitively, smokers are more likely than non-smokers to have a cough and to have lung cancer, but lung cancer is also likely to cause a cough. The variable *XRay* is true if a spot appears on the X-ray of a patient's lung, and this tends to happen if the patient has lung cancer.

Answer the following questions:

1. How many independent parameters are needed to specify the conditional probability tables associated with this network?

2. Write out a reasonable conditional probability table for the node labelled *Cough*. Explain briefly your reasons for choosing parameters with these relative magnitudes.

3. Write down a symbolic expression for the joint probability $P(S \wedge C \wedge \neg L \wedge \neg X)$, using probabilities that would be available in the tables attached to the network. (We are using here the obvious single-letter abbreviations for the Boolean variables.)

4. Write down a symbolic expression for $P(\neg X|S)$, using probabilities that would be available in the tables attached to the network. (Hint: Use conditioning.)

5. Write down a symbolic expression for $P(L|X)$, using probabilities that would be available in the tables attached to the network. (Hint: Use Bayes' Rule as well as conditioning.)

Question 4 (20 points):

Consider the language *Buffalo*^{n}, which is very much like English except that the only word in its lexicon is *buffalo*.

Here are two sentences from the language:

- Buffalo buffalo buffalo Buffalo buffalo.
- Buffalo Buffalo buffalo buffalo buffalo Buffalo buffalo.

In case you don't believe these are sentences, here are two English sentences with corresponding syntactic structure:

- Dallas cattle bewilder Denver cattle.
- Chefs London critics admire cook French food.

Here is another hint: (i) there is a city in upstate New York named "Buffalo"; (ii) there are wild animals in the American West called "buffalo"; (iii) the expression "to buffalo" in colloquial American English means "to fool" or "to confuse".

1. Write a grammar for *Buffalo*^{n}. The lexical categories are *adjective*, *noun*, and (transitive) *verb*. There should be one grammar rule for *sentences*, one for *verb phrases*, and three rules for *noun phrases*: a bare noun, an adjective modifier, and a reduced relative clause (i.e., a relative clause without the word "that").

2. How many distinct parse trees are there for *Buffalo*^{5}, the subset of *Buffalo*^{n} consisting of sentences with exactly 5 words? Draw them.

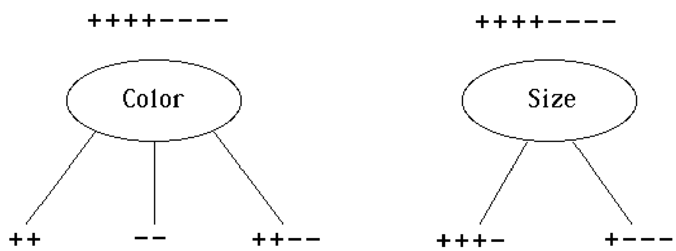

Figure 2: Decision Tree Splits

Question 5 (20 points):

Figure 2 shows two possible splits in the initial construction of a decision tree. The first split uses the attribute *Color* for the test at the root of the tree, while the second split uses the attribute *Size*. The $+$'s and $-$'s indicate the positive and negative examples at each node in the tree.

Recall that the *information content* of the answer to a decision problem is defined as:

$$I(p^+, p^-) = -p^+ \log_2(p^+) - p^- \log_2(p^-)$$

where p^+ is the percentage of positive examples and p^- is the percentage of negative examples in the population. Here are some typical values of this function: $I(0.0, 1.0) = 0.0$ bits, $I(0.25, 0.75) = 0.8$ bits, $I(0.5, 0.5) = 1.0$ bits.

Answer the following questions:

1. What is the information content of a correct answer to the decision problem shown in Figure 2?

2. What is the information gain that would be achieved by splitting first on the attribute *Color*? Show your calculations.

3. What is the information gain that would be achieved by splitting first on the attribute *Size*? Show your calculations.

4. Which attribute would be the better choice for the initial split? Explain why.