

Computer Science Department  
Stanford University  
Comprehensive Examination in Artificial Intelligence  
Autumn 2000

October 31, 2000

PLEASE READ THIS FIRST

- a. You should write your answers for this part of the Comprehensive Examination in a BLUE BOOK. Be sure to write your MAGIC NUMBER on the cover of every blue book that you use.
  - b. Be sure you have all the pages of this exam. There are 4 pages in addition to this cover sheet.
  - c. This exam is OPEN BOOK. You may use notes, articles, or books – but no help from people or computers.
  - d. Show your work, since PARTIAL CREDIT will be given for incomplete answers. For example, you can get credit for making a reasonable start on a problem even if the idea or arithmetic does not work out. You can also get credit for realizing that certain approaches are incorrect.
- 
- e. Points in this exam add up to 60. Points are allocated according to the number of minutes we believe a student familiar with the material should take to answer the questions. If you are somewhat less familiar with the material, a question may easily take you longer than the number of points it's worth. Therefore be careful: **IF YOU ARE TAKING TOO LONG ON A QUESTION, WRITE DOWN WHATEVER YOU HAVE AND MOVE ON.**

## 1 Search (8 points)

### a. Uninformed Search

- (i) (1 point) Describe or give an example of a search space where depth-first search will perform much better than iterative deepening search
- (ii) (1 point) Describe or give an example of a search space where breadth-first search will perform much better than depth-first search
- (iii) (1 point) Describe or give an example of a search space where depth-first search will perform much better than breadth-first search

### b. Heuristic Search

(5 points) A\* search involves evaluating search paths via  $\hat{f} = g + \hat{h}$ , where  $g$  is the lowest cost path to the current search state, and  $\hat{h}$  is the heuristic function which provides an optimistic estimate of the cost to a goal state. Now assume that the heuristic function  $\hat{h}$  is induced from a cost function between nodes  $h'(x, y)$  which obeys the triangle inequality. (The triangle inequality says that the sum of the costs from  $x$  to  $z$  and  $z$  to  $y$  must not be less than the cost from  $x$  to  $y$  directly.) Prove that the  $\hat{f}$ -cost along any path in the search tree is nondecreasing.

## 2 Logic: resolution (12 points)

(This question comes from (the late) Jon Barwise and (our new provost) John Etchemendy. It also appears in the exercises of Russell and Norvig.)

Consider the following statements:

If a unicorn is mythical, then it is immortal, but if it is not mythical, then it is a mortal mammal. If the unicorn is either immortal or a mammal, then it is horned. The unicorn is magical, if it is horned.

From the above, can you prove that the unicorn is mythical? How about magical? Horned? Use resolution for your proofs (using propositional logic is possible and acceptable). Show:

- a. (2 points) the basic logical translation of this text
- b. (2 points) the translations of these into a form suitable for resolution theorem proving
- c. (8 points) For each of the three unicorn properties (mythical, magical, horned), if it can be proved, show your proof. If it can't be proved, explain the justification for being able to conclude that the fact that unicorns have this property doesn't follow from the information given.

### 3 Probabilistic models (12 points)

- a. (1 point) A scientist tells us that  $1/3$  of all kangaroos have blue eyes, and  $1/3$  of all kangaroos are left-handed. On the basis of this information alone, what bounds can we place on the proportion of kangaroos that are both blue-eyed and left-handed?
- b. (2 points) A loglinear model for the joint distribution of some number of categorical variables  $X_1, \dots, X_n$  takes the following general form:

$$\log P(X_1 = x_1, \dots, X_n = x_n) = \sum_{C \in \mathcal{C}} \lambda_C(x_C)$$

where each  $C \subseteq \{1, \dots, n\}$  (so  $\mathcal{C} \subseteq 2^{\{1, \dots, n\}}$ ). That is,  $x_C$  is some subset of the  $x_i$  variables, and  $\lambda_C(\cdot)$  is a function of this subset.

Show that all Bayesian networks ("graphical models") are loglinear models.

- c. (9 points) Consider this situation: You saw John talking to his boss. Later you saw John looking upset. John may have been upset because his boss gave him a warning. Or he may be upset because his boyfriend left him.
- (i) (1 point) Draw a suitable Bayesian network to represent the causal structure among the 4 random variables B for boss talking to John, W for boss warning John, U for John being upset, and L for his boyfriend leaving him.
- (ii) (1 point) Make up conditional probability tables for each node. [HINT: keep the numbers simple. Use quarters, thirds, and halves!!]
- (iii) (4 points) Given the information above (B and U), calculate the probability that John's boss warned him.
- (iv) (3 points) Calculate the probability that John's boss warned him given the above information and that you know that his boyfriend left him.

Your answers may be approximate, but you should show your work

### 4 Learning (14 points)

- a. (2 points) A circuit has two input values A and B whose values are either +1 for 'true' or -1 for 'false'. Design a perceptron (Linear Threshold Unit) network which computes the function not (A or B). Draw the network and indicate clearly all weights and threshold values (assume the network outputs 1 if the dotproduct of the weights and the inputs is greater than some threshold  $t$  specified by each unit).
- b. (3 points) We are trying to predict whether Comps questions are easy or difficult ( $D = +$  if they are difficult) based on two features:

$L$  Whether they are long (1 = long)

$M$  Whether they have a lot of math (1 = yes)

For training data, we have examined 12 Comps questions, and have collected the following statistics, which we show twice: on the left are counts for the different data patterns, and on the right the data is shown in a contingency table showing  $- : +$  counts for each combination of the classificatory variables.

$L$	$M$	$D$	count
0	0	-	4
0	0	+	1
0	1	-	0
0	1	+	3
1	0	-	1
1	0	+	2
1	1	-	1
1	1	+	0

		$M$	
		0	1
$L$	0	4:1	0:3
	1	1:2	1:0

Draw a decision tree for this data (using information gain for node splitting, and no stopping criterion or pruning, so that the tree is grown so long as there is some classificatory feature that appears to have information about the target feature).

- c. (4 points) A Naive Bayes classifier for this problem predicts the target feature from the prior and independently from the classificatory features as follows:

$$\text{Choose } \hat{d} = \arg \max_{d \in \{-, +\}} P(d)P(L = l|d)P(M = m|d)$$

(That is, it calculates the expression shown for both  $d = +$  and  $d = -$  and chooses the value of  $d$  that gives the expression higher probability.) What classificatory decisions would a Naive Bayes classifier make for each combination of classificatory variables?

- d. (3 points) Suppose we have 3 test data instances, whose correct classification we know, as follows:

$L$	$M$	$D$
0	0	-
1	1	+
0	1	+

What is the decision of each classifier on each datum? Which does better overall?

- e. (2 points) Is the better performance of one learner reasonable or surprising here?

## 5 Vision and Natural Language Processing (14 points)

- a. (2 points) One can easily use a segmentation algorithm to find edges in images. However, it is generally hard to use an edge detector to segment an image into regions. Why doesn't the output of an edge detector segment an image into regions?



- b. (2 points) You want to design a vision system that uses shading information to recover the three-dimensional geometry of the scene from pictures taken in the real world. You don't know what will be in front of the camera. What key difficulties are you going to face?
- c. (10 points) Write:
- (i) (1 point) a set of phrase structure grammar rules
  - (ii) (1 point) a lexicon
  - (iii) (3 points) a set of meanings for words in the lexicon, and
  - (iv) (3 points) semantic combination rules for the grammar rules, in the form of a definite clause grammar

for the sentence:

*Sydney is a beautiful city.*

You can assume that the adjective *beautiful* has intersective semantics, and that *Sydney* is a constant – that is, a suitable semantic form for the sentence is: *beautiful(sydney) & city(sydney)*. You may find it helpful to look at, but will need to extend, the grammar of Russell & Norvig pp. 665ff. In particular, the semantic forms in your lexicon will need to include lambda-expressions, as shown there.

- (v) (2 points) Discuss the most salient difficulties involving the semantics of this sentence and your grammar for it which would make extending your grammar difficult.