CS 540-1: Introduction to Artificial Intelligence

Final: 2:45-4:45pm, May 14, 1998
Room 180 Science Hall

CLOSED BOOK
(two sheets of notes and calculators allowed)

Write your answers on these pages and show your work. If you feel that a question is not fully specified, state any assumptions you need to make in order to solve the problem. You may use the backs of these sheets for scratch work.

Write your name on this and all other pages of this exam. Make sure your exam contains seven problems on ten pages.

Name ____

Student ID ____

Problem	**Score**	**Max Score**
1	_____	20
2	_____	25
3	_____	10
4	_____	10
5	_____	10
6	_____	15
7	_____	10
Total	_____	100

PROBLEM 1 - Bayesian Reasoning (20 points)

PART A. Define the following two variables about people:

F	=	had the *flu*
S	=	had a flu *shot*

Assume we know from medical studies that:

P(F)	=	0.75
P(S)	=	0.5
P(F | S)	=	0.1

Given you find out that someone had the flu, what's the probability they had a flu shot? Show and explain your calculations.

PART B. Robby the robot has just reported *danger*. We need to estimate the probability this is a ''false positive'' (i.e., a false alarm that can be ignored). Here is what we know:

i) If its batteries are low on power, the probability that Robby will (falsely) report danger on a safe day is 0.75 and the probability it will (correctly) report danger on a dangerous day is also 0.75.

ii) If its batteries are fine, Robby's reports are 100% accurate.

iii) We haven't charged Robby's batteries for awhile, so the probability they are low on power is 0.25.

iv) The probability that any arbitrary day is safe is 0.9.

Using the following symbols, write down the expression that we need to evaluate in order to decide if we should ignore Robby's report of danger.

safe	=	current situation is *safe* (ie, danger-free)
RD	=	Robby *reported* the situation is *dangerous*
BL	=	*batteries low* on power

Using the above numbers, determine whether we should ignore Robby's report. (We'll ignore it if the probability that the current situation is safe exceeds the probability it is dangerous.) Show and briefly explain all your steps.

PROBLEM 2 - Learning from Labelled Examples (25 points)

Consider the following training examples, where four Boolean-valued examples (X_1 through X_4) are used to predict a Boolean-valued output (Y_1):

$$
\begin{array}{cccccc}
0 & 1 & 1 & 0 & \Rightarrow & 1 \\
1 & 0 & 1 & 0 & \Rightarrow & 0 \\
1 & 0 & 0 & 1 & \Rightarrow & 1 \\
0 & 1 & 1 & 1 & \Rightarrow & 0
\end{array}
$$

A) Which feature would be chosen as the *root node* of a decision tree, if *information gain* was used as the scoring function? Show your calculations. (Note you need only select the root node, and need not create the complete tree.)

B) Assume you are given the following new (unlabelled) example:

$$1 \;\; 1 \;\; 1 \;\; 1$$

Explain how a case-based reasoning system might predict the output for this new example. Show the calculations it would go through.

C) Draw a perceptron for this task, initializing all the weights and biases to -0.1.

Show how the weights and biases in this perceptron would be changed after processing the first training example above. Use a learning rate of 0.5. Be sure to show the learning rule you used; don't just show the new weights.

D) Assume now that the weights in your perceptron are all 'frozen' at the value 1, and the only free (ie, adjustable) parameter is the bias (threshold) of the output unit. Draw and briefly explain the *weight space* for this case (ie, the x-axis is the setting of the bias).

PROBLEM 3 - Genetic Algorithms (10 points)

Assume you are trying to find the values for X_1 through X_4 that maximize the following function:

$$f = 5X_1 - 3X_2X_3 + X_3 - 2X_4$$

You decide to use a genetic algorithm and create the initial population:

0 1 1 0

1 1 0 0

1 0 1 1

0 0 0 1

Briefly show and explain how you might create the next generation of the population.

PROBLEM 4 - Search (10 points)

The search algorithm *A** does not terminate until a goal node is *removed* from the OPEN list. That goal state may have been in the OPEN list for thousands of node expansions, which seems inefficient.

Why not terminate *A** as soon as a goal node is added to OPEN? Justify your answer with a simple, concrete example that shows doing so would violate an important property of *A**. Be sure to explain your answer. (Don't use more than five nodes in your answer.)

PROBLEM 5 - Logic (10 points)

Assume the 'fuzzy' truth value of some statement *A* is *ftv(A)*. In fuzzy logic, what is the fuzzy truth value of:

$$A \lor \neg A$$

Assume the standard truth value of some statement *B* is *tv(B)*. In standard logic, what is the truth value of:

$$B \lor \neg B$$

Be sure to simplify your above answers as much as possible.

Express the following sentence in first-order predicate calculus:

For every book that Alice owns, Bob owns a book by the same author.

PROBLEM 6 - Miscellaneous Questions (15 points)

What do you feel is the most important difference between *simulated annealing* and *genetic algorithms*? Why?

What do you feel is the most similar aspect of *simulated annealing* and *genetic algorithms*?

Briefly explain how *interpretations* in logic can help us debug our English⇒FOPC conversions.

Describe the major weakness of *perceptrons*.

Draw a simple *partial-order plan* for going to the store to get groceries. (You needn't be too detailed - just show enough to demonstrate you understand what a partial-order plan is.)

PROBLEM 7 - Speech Recognition (10 points)

Assume that when read, the following two sentences are always pronounced identically by all speakers, producing the acoustic signal *AS*:

(i) *Recognize speech.*

(ii) *Wreck a nice beach.*

No other sentences produce acoustic signal *AS* when read.

Imagine we estimated the following probabilities about the word usage:

P('recognize' is current word)	=	10^{-4}
P('wreck' is current word)	=	$5 \, x \, 10^{-3}$
P('speech' is current word **|** 'recognize' was previous word)	=	10^{-4}
P('a' is current word **|** 'wreck' was previous word)	=	0.5
P('nice' is current word **|** 'a' was previous word)	=	$5 \, x \, 10^{-3}$
P('beach' is current word **|** 'nice' was previous word)	=	$2 \, x \, 10^{-3}$

Given that we heard acoustic signal *AS*, which of the above two sentences was most probably the one read? Show and *explain* your calculations (use scientific notation for your calculations).
[As shorthand, you may write *P('A' is current word* **|** *'B' was previous word)* as *P(A | B)*.]

(The end - have a good summer!)