(PRINT) Name ____Student No__________

Signature ____Total Mark_____/100

University of Waterloo

Computer Science 486/686 – Introduction to Artificial Intelligence

Midterm Test
2023 June 23
Time: 4:30 pm – 5:50 pm

Time: 80 minutes
Total marks: 100

Answer all questions on this paper. No books or other materials may be used. Non-programmable calculators are permitted, but not personal computers.

This examination has 7 pages. Check that you have a complete paper.

1	**/ 20**
2	**/ 19**
3	**/ 20**
4	**/ 20**
5	**/ 21**
Total	**/ 100**

Question 1 [20 pts] Search techniques

Consider the following generic search procedure:

1. Let *PQ = {s}* i.e., priority queue consists of the start state
2. Loop until priority queue is empty
    a. Remove the first node *n* from *PQ*
    b. If *n* is a goal state then stop; report success
    c. Otherwise add each neighbour of *n* to *PQ*

Briefly explain how nodes should be added to the priority queue (step 2c) to emulate each of the following algorithms.

a) [5 pts] Depth-first search

insert new nodes at the beginning of the priority queue

b) [5 pts] Breadth-first search

insert new nodes at the end of the priority queue

c) [5 pts] Greedy best-first search

insert new nodes based on their h value, keeping the nodes sorted by ascending h value.

d) [5 pts] A* search

insert new nodes based on their f value, keeping the nodes sorted by ascending f value.

Question 2 [19 pts] Every term, the university must design a schedule for the final exams. Ideally the schedule should be conflict free, meaning that students should not have to write two exams simultaneously.

a) [9 pts] Consider 5 students (s_1, s_2, s_3, s_4 and s_5) and 5 courses (c_1, c_2, c_3, c_4 and c_5) such that s_1 and s_2 are taking c_1; s_1, s_3 and s_4 are taking c_2; s_2 and s_4 are taking c_3; s_3 is taking c_4; and s_4 and s_5 are taking c_5. Suppose that the 5 courses must be scheduled in 3 time slots t_1, t_2 and t_3. Describe how you would encode this scheduling problem as a constraint satisfaction problem. List the variables and their domain as well as the constraints.

variables: c_1, c_2, c_3, c_4, c_5

domain: $c_i \in \{t_1, t_2, t_3\}$

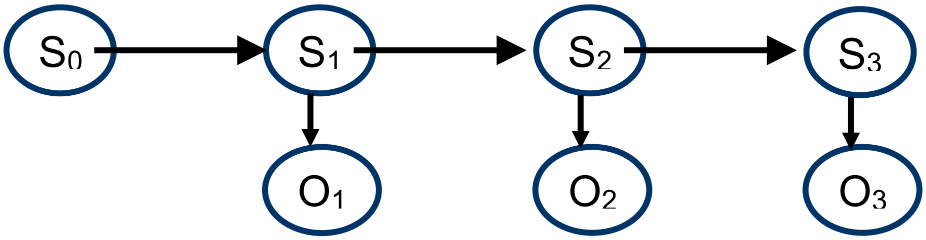
constraints : $c_1 \neq c_2$, $c_1 \neq c_3$, $c_2 \neq c_4$, $c_2 \neq c_3$, $c_2 \neq c_5$, $c_3 \neq c_5$

b) [10 pts] Suppose that you use backtracking search with the most constrained variable and least constraining value heuristics. Show the search tree expanded by backtracking search until a satisfying assignment is found for the CSP in a). Indicate in which order the nodes are expanded in the search tree.

c1 = t1 (all variables are equally constrained so start with any variable)
|
c2 = t2 (c2 and c3 can only take two values)
|
c3 = t3 (c3 can only take one value)
|
c5 = t1 (c5 can only take one value)
|
c4 = t1 (c4 is the only variable left)

no backtracking necessary

Question 3 [20 pts] Consider a first-order hidden Markov model.



a) **[6 pts]** A hidden Markov model can be viewed as a Bayesian network with the same structure that repeats at each time step. Normally, when specifying a Bayesian network, we need to provide a conditional probability table for each variable in the network. However, an HMM can be completely specified by providing $Pr(S_t|S_{t-1})$, $Pr(O_t|S_t)$ and $Pr(S_0)$. What assumption allows us to specify an HMM with those three distributions only? Explain briefly.

Stationarity, which means that $Pr(S_t|S_{t-1}) = Pr(S_{t+1}|S_t)$ and $Pr(O_t|S_t) = Pr(O_{t+1}|S_{t+1})$ for all t.

b) **[8 pts]** Conditional independence:

i) **[4 pts]** Is S_3 independent of S_1 given S_2? Explain briefly

Yes: the path between S_1 and S_3 is blocked by S_2

ii) **[4 pts]** Is O_3 independent of O_1 given O_2? Explain briefly.

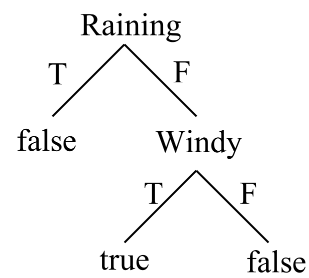
No: the path between O_1 and O_3 is open

c) **[6 pts]** Suppose you want to predict S_3 based on O_1 only (i.e., $Pr(S_3|O_1)$). What are the relevant and irrelevant variables for this query?

Relevant: S_0, S_1, S_2, S_3, O_1
Irrelevant: O_2, O_3

Question 4 [20 pts] Consider the problem of deciding whether or not to go on a picnic based on different attributes of the day. Here is a set of examples classified based on whether or not it was a good idea to go on a picnic. Besides the table is a possible decision tree for this problem.

Example	Rainy	Windy	Warm	Summer	Sunday	Picnic
X_1	T	F	F	F	F	False
X_2	F	T	F	F	T	True
X_3	F	T	T	T	T	False
X_4	F	T	T	F	T	False
X_5	T	F	F	F	T	False
X_6	F	T	F	F	T	False



a) **[6 pts]** Indicate the class of each example according to the decision tree. Indicate also whether each example is correctly classified.

Example	X_1	X_2	X_3	X_4	X_5	X_6
Picnic (true/false)	**F**	**T**	**T**	**T**	**F**	**T**
Correct class? (yes/no)	**Y**	**Y**	**N**	**N**	**Y**	**N**

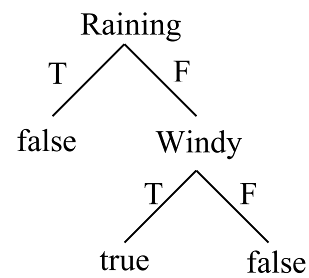
b) **[3 pts]** Suppose you could turn one leaf of the decision tree into a new attribute test, which leaf would that be? No justification required.

i) Rainy = T   ii) Windy = T   iii) Windy = F

Answer: ii)

Question 4 continued

Example	Rainy	Windy	Warm	Summer	Sunday	Picnic
X_1	T	F	F	F	F	False
X_2	F	T	F	F	T	True
X_3	F	T	T	T	T	False
X_4	F	T	T	F	T	False
X_5	T	F	F	F	T	False
X_6	F	T	F	F	T	False



c) **[8 pts]** For the leaf that you picked in b), compute the information gain of each possible attribute test. Which attribute test provides the highest information gain?

Hint: Entropy(p/(n+p), n/(n+p)) = - p/(n+p) $\log_2$ p/(n+p) – n/(p+n) $\log_2$ n/(p+n)
InformationGain(attribute) = leaf entropy - expected entropy of each attribute branch

Leaf entropy: -1/4 $\log_2$ ¼ - ¾ $\log_2$ ¾ = 0.811278

Expected remaining entropy for warm:
½ (- ½ $\log_2$ ½ - ½ $\log_2$ ½) + ½ (-0 $\log_2$ 0 – 1 $\log_2$ 1) = 0.5

Expected remaining entropy for summer:
¾ (- 1/3 $\log_2$ 1/3 – 2/3 $\log_2$ 2/3) + ¼ (-1 $\log_2$ 1 – 0 $\log_2$ 0) = 0.68872

Expected remaining entropy for Sunday:
1(- ¼ $\log_2$ ¼ - ¾ $\log_2$ ¾) = 0.811278

InfoGain(warm) = 0.811278 – 0.5 = 0.311278
InfoGain(summer) = 0.811278 – 0.68872 = 0.122556
InfoGain(Sunday) = 0.811278 – 0.811278 = 0

Best attribute: warm

d) **[3 pts]** There is *no* decision tree consistent with the data. True or false? No justification required.

True

Question 5 [21 pts] Are the following statements true or false? No justification required.

a) [3 pts] In Bayesian learning, the prior is subjective and therefore it is okay for different people to have different prior distributions.

True

b) [3 pts] Admissible heuristics are necessarily consistent.

False

c) [3 pts] All search algorithms take exponential time in the worst case.

True

d) [3 pts] Variable elimination takes exponential time in the worst case.

True

e) [3 pts] Bayesian learning, maximum a posteriori hypothesis and maximum likelihood hypothesis all make the same predictions when an infinitely large amount of training data is used.

True

f) [3 pts] Structural causal models encode only deterministic relations that are given by some equations.

True

g) [3 pts] Causal Bayesian networks can be converted into equivalent structural causal models, but not the other way around.

False