University No.: ______

THE UNIVERSITY OF HONG KONG

FACULTY OF ENGINEERING
DEPARTMENT OF COMPUTER SCIENCE

COMP7404 Computational Intelligence and Machine Learning

Date: May 12, 2018 **Time: 9:30am - 11:30am**

Only approved calculators as announced by the Examinations Secretary can be used in this examination. It is the candidates' responsibility to ensure that their calculator operates satisfactorily, and candidates must record the name and type of the calculator used on the front page of the examination script.

Calculator Brand: ______________ *Calculator Model:* ______________

Write your University No. on every page.

Answer all questions in the space provided.

Special Note: Candidates are permitted to bring to the examination ONE sheet of A4-sized paper with printed or written notes on both sides.

Question No.	Available Marks	Obtained Marks
1	40	
2	12	
3	12	
4	12	
5	10	
6	10	
7	4	
Total	100	

1. True / False Questions

For each True / False question, write down True or False. A correct answer is worth +2, a missing answer is worth 0 and **a wrong answer is worth -2**. You may provide an explanation, but this is optional.

(a) [*True* or *False*] The greedy search algorithm is not complete.

(b) [*True* or *False*] The heuristic $h(n) = 0$ is consistent for every search problem.

(c) [*True* or *False*] The heuristic $h(n) = 1$ is admissible for every search problem.

(d) [*True* or *False*] The heuristic $h(n) = c(n)$, where $c(n)$ is the true cheapest cost to get from the node n to a goal state, is consistent for every search problem.

(e) [*True* or *False*] Alpha-beta pruning is always faster than minimax.

(f) [*True* or *False*] The tree-search version of A* will be optimal if a consistent heuristic function is used.

(g) [*True* or *False*] An admissible heuristic never underestimates the cost to the goal.

(h) [*True* or *False*] Local search is guaranteed to find a global optimum.

(i) [*True* or *False*] The most-constrained variable heuristic provides a way to select the next variable to assign in a backtracking search for solving a CSP.

(j) [*True* or *False*] Depth-first search with a depth limit can be made to return the same solution as breadth-first search by iteratively increasing the depth limit.

(k) [*True* or *False*] The starting position for the game tic-tac-toe has a minimax value of 0.

(l) [*True* or *False*] Assume *Agent* 1 has a utility function U_1 and *Agent* 2 has a utility function U_2. If $U_1 = k_1 U_2 + k_2$ with $k_1 > 0, k_2 > 0$ then *Agent* 1 and *Agent* 2 have the same preferences.

(m) [*True* or *False*] Assume $A \succ B$, $D \succ A$, $B \succ L$, where $L = [0.5, C; 0.5, D]$. With rational preferences it follows that $B \succ C$.

(n) [*True* or *False*] If the policy has converged in value iteration, the values must have converged as well.

(o) [*True* or *False*] If the values have converged in value iteration, the policy must have converged as well.

(p) [*True* or *False*] Value iteration is guaranteed to converge if the discount factor (γ) satisfies $0 < \gamma \leq 1$.

(q) [*True* or *False*] Q-learning can learn the optimal Q-function without ever executing the optimal policy.

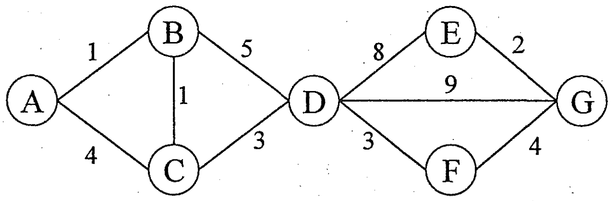
(r) [*True* or *False*] For an MDP, if we just change the reward function R the optimal policy is guaranteed to remain the same.

(s) [*True* or *False*] If the only difference between two MDPs is the value of the discount factor then they must have the same optimal policy.

(t) [*True* or *False*] For an MDP with finite number of states and actions with a discount factor γ, with $0 < \gamma < 1$, policy iteration is guaranteed to converge.

2. Search

Consider the following state space graph.



A is the start state and G is the goal state. The cost for each edge are shown on the graph. Each edge can be traversed in both directions.

Suppose you are completing the new heuristic function h shown below. All the values are fixed except $h(B)$.

State	A	B	C	D	E	F	G
h	10	?	9	7	1.5	4.5	0

For each of the following conditions, write the set of values that are possible for $h(B)$. For example, to denote all non-negative numbers, write $[0, \infty]$, to denote the empty set, write $\varnothing$, and so on.

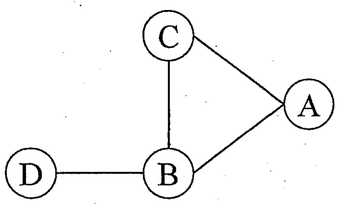
(a) What values of $h(b)$ make h admissible? *(2 marks)*

(b) What values of $h(b)$ make h consistent? *(4 marks)*

(c) What values of $h(B)$ will cause A^* Graph Search Algorithm (GSA) to expand node A, then node C, then node B and then node D in order? *(6 marks)*

3. Constraint Satisfaction Problems (CSPs)

Consider the following constraint graph for a Constraint Satisfaction Problem (CSP).



The domains of the four variables are indicated in the following table.

A	0	1	2	3
B	0	1	2	3
C	0	1	2	3
D	0	1	2	3

The binary constraints are as follows.

- $A > B$
- $A \neq C$
- $C > B$
- $D < B$

(a) Indicate what the domains of all variables are after arc consistency is enforced by crossing out eliminated values from the domains in the table below. *(6 marks)*

A	0	1	2	3
B	0	1	2	3
C	0	1	2	3
D	0	1	2	3

(b) Now suppose you are given a different CSP with variables still being A, B, C, D, but you are not given the constraints. The domains of variables remaining after enforcing arc consistency for this CSP are given to you below.

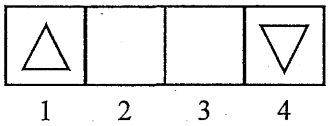
A			2	3
B			2	3
C	0	1	2	
D			2	3

Select all of the following options which can be inferred given only this information. *(6 marks)*

☐ The CSP may have no solution
☐ The CSP must have a solution
☐ The CSP must have exactly one solution
☐ The CSP may have more than one solution
☐ The CSP must have more than one solution
☐ None of the above

4. Adversarial Search

Two players, player $\triangle$ and player $\triangledown$, find themselves in the following start state of an adversarial game.



Player $\triangle$ moves first. The two players take turns moving, and each player must move to an open adjacent square in either direction. If the opponent occupies an adjacent space, then the player will jump over the opponent to the next open space, if any. For example, if player $\triangle$ is on square 3 and player $\triangledown$ on square 2, then player $\triangle$ may move back to square 1. The game ends when one player reaches the opposite end of the board. If player $\triangle$ reaches square 4 first, then the value of the game is +1; if player $\triangledown$ reaches square 1 first, then the value of the game is -1.

(a) Draw the minimax game tree and use the proper triangles to distinguish a *min* from a *max* node. Put terminal states in square boxes and loop states in double square boxes. Loop states are those that already appear on the path to the root. Write down the value of every state, you may write *"unknown"* for loop states. Can you indicate the optimal decision at the root of the tree? Explain. *(8 marks)*

(b) Consider a game with $n > 4$ squares instead of the above game with 4 squares. How would the value at the root of the tree change? *(4 marks)*

5. Markov Decision Processes

Consider the following Grid World introduced in our lecture where an agent may move *north*, *south*, *east* or *west* and exit states are $a4$ and $b4$.

-1.11	-1.10	-0.94	+1	a
-1.11		-1.10	−1	b
-1.11	-1.11	-1.11	-1.10	c
1	2	3	4	

Actions are unreliable: 80% of time each action achieves the desired effect and 20% of the time the action moves the agent at right angles (with equal probability) to the intended directions. If the agent runs into a wall, it stays in the same square. After running value iteration we have obtained the values shown in the figure. Note that values have been rounded to two decimal places. Let the discount factor $\gamma = 0.1$ and the reward R(s) = -1. Consider the state $b3$.

(a) Calculate the Q-values of the state $b3$ and perform policy extraction for the state. *(10 marks)*

6. **Reinforcement Learning**

Consider an MDP with three states, A, B and C; and two actions CW and CCW. We do not know the transition function or the reward function for the MDP, but instead, we are given samples of what an agent actually experiences when it interacts with the environment (although, we do know that we do not remain in the same state after taking an action). In this problem, instead of first estimating the transition and reward functions, we will directly estimate the Q function using Q-learning. Assume the discount factor $\gamma = 0.25$ and the step size for Q-learning $\alpha = 0.75$. Let the current Q function, $Q(s, a)$, be:

	A	B	C
CW	0.3	-0.5	-2.5
CCW	0.0	-1.3	-1.9

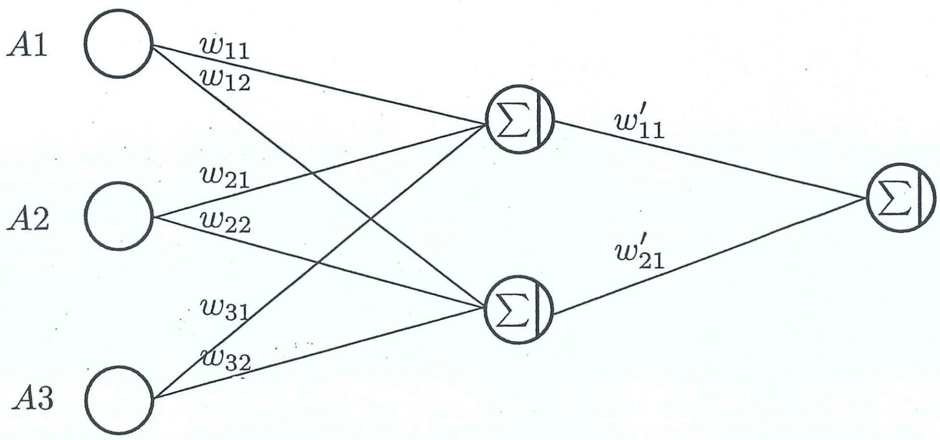
The agent encounters the following samples:

s	a	s'	r
A	CCW	C	9.0
B	CW	A	1.0

Process the samples given above and write down all Q-values of the three states after both samples have been accounted for. *(10 marks)*

7. Artificial Neural Networks

Consider the problem of predicting the performance of a student in a final exam based on that student's performance in three assignments: assignment 1 (A1), assignment 2 (A2) and assignment 3 (A3). We are going to use the following neural network.



Write down the number of

(a) input nodes

(b) output nodes

(c) neurons

(d) synapses

in the neural network above. *(4 marks)*

END OF PAPER

This page is intentionally left blank.
You may use it to draft your answers.