*University No.:*______________________________

THE UNIVERSITY OF HONG KONG

FACULTY OF ENGINEERING
DEPARTMENT OF COMPUTER SCIENCE

Quiz 1

COMP3270 Artificial Intelligence

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:*_____________________

Answer questions in the space provided.
Write your University No. on every page.

Chapter	Max. Mark	Your Mark (examiner use only)
1	10	
2	5	
3	5	
Total	20	

1: Search

1.1 (1 mark for correct answer, -1 mark for incorrect answer): (True/False) If one search heuristic $h_1(s)$ is admissible and another one $h_2(s)$ is admissible, then $h_3(s) = (h_1(s)+h_2(s))/2$ will be admissible.

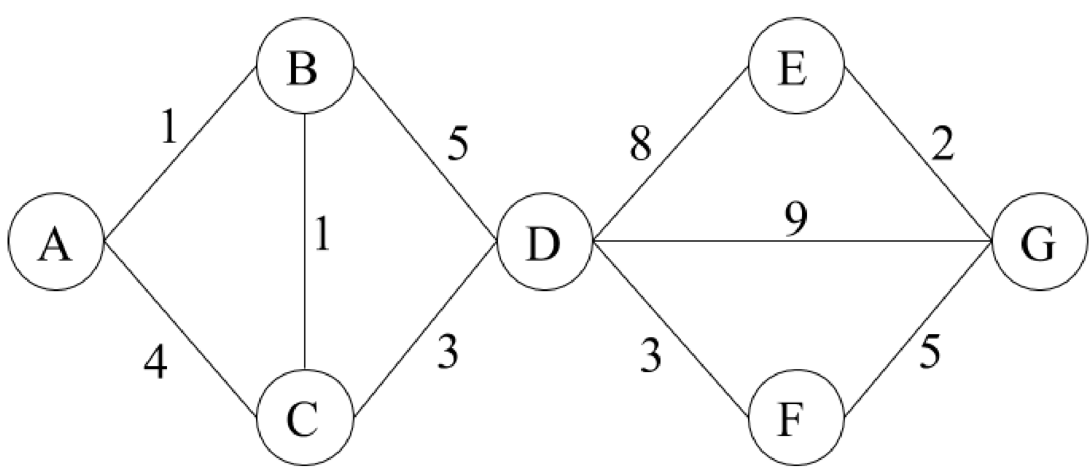
1.2 (1 mark for correct answer, -1 mark for incorrect answer): (True/False) If a CSP is arc consistent, it can be solved without backtracking.

1.3 (1 mark for correct answer, -1 mark for incorrect answer): (True/False) The heuristic $h(s) = 0$ is consistent for every search problem.

1.4 (1 mark for correct answer, -1 mark for incorrect answer): (True/False) If the policy has converged in value iteration, the values must have converged as well.

1.5: Consider the state space graph below. A is the start state and G is the goal state. The costs for each edge are shown on the graph. Each edge can be traversed in both directions. Two heuristics are shown in the table.

Node	h_1	h_2
A	9.5	10
B	9	12
C	8	10
D	7	8
E	1.5	1
F	4	4.5
G	0	0



1.5.1 (2 mark): For each of the following graph search strategies, write down the solution path that it returns. Assume that nodes are added to the frontier in alphabetical order.

Search Algorithm	Solution Path
BFS	
UCS	
A* with h_1	
A* with h_2	

1.5.2: Suppose you are completing the new heuristic function h_3 shown below. All values are fixed except $h_3(B)$.

Node	A	B	C	D	E	F	G
h_3	10	?	9	7	1.5	4.5	0

For each of the following conditions, write the set of values that are possible for $h_3(B)$.

1.5.2.1 (1 mark): What values of $h_3(B)$ make h_3 admissible?

1.5.2.2 (1 mark): What values of $h_3(B)$ make h_3 consistent?

1.5.2.3 (2 mark): What values of $h_3(B)$ will cause A* graph search to expand node A, then node C, then node B, then node D in order?

3: MDP

Consider an MDP with states {4, 3, 2, 1, 0}, where 4 is the starting state.
In states k >= 1, you can walk (W) and T(k, W, k-1) = 1. In states k >= 2, you can also jump (J) and T(k, J, k-2) = T(k, J, k) = 1/2. State 0 is a terminal state. The reward R(s, a, s') = $(s-s')^2$ for all (s, a, s'). Use a discount of gamma = 1/2.

3.1 (3 marks): Compute $V^*(2)$.

3.2 (2 marks): Compute $Q^*(4, W)$.

4: MDP+RL

Consider a gridworld MDP that operates like the one we saw in class. The states are grid squares, identified by their row and column number (row first). The agent always starts in state (1, 1), marked with the letter S. There are two states where only the exit action is available, (2, 3) with reward +5 and (1, 3) with reward -5. The agent will get the +5 and -5 when taking the exit action and leaving the state. In other states the transition function is such that the intended agent movement (North, East, South, West) happens with probability .8. With probability .1 each, the agent ends up in one of the states perpendicular to the intended direction. If a collision with a wall happens, the agent stays in the same state. Rewards are 0 in non-terminal states. Here is a figure of the grid world MDP.

		+5
S		-5

4.1 (2 marks): Suppose the agent knows the transition probabilities. Give the first two rounds of value iteration updates for each state, with a discount of gamma = 0.9, i.e., assume V_0 is 0 everywhere and compute V_i for all states for i = 1, 2.

4.2 (1 mark): Now, suppose the agent does not know the transition probabilities. The agent starts with the policy that always chooses to go right, and executes the following three trials:

1. (1, 1) - (1, 2) - (1, 3) - terminal state
2. (1, 1) - (1, 2) - (2, 3) - terminal state
3. (1, 1) - (2, 1) - (2, 2) - (2, 3) - terminal state

Perform direct evaluation and compute the values for states (1, 1) and (2, 2). Assume no discounting, i.e., gamma = 1.0.

4.3 (2 marks): Using a learning rate of .5 and assuming initial values of 0, what updates does a TD-learning agent make after trials 1 and 2, above? Assume no discounting, i.e., gamma = 1.0.

END OF PAPER

You may use this page to draft your answer