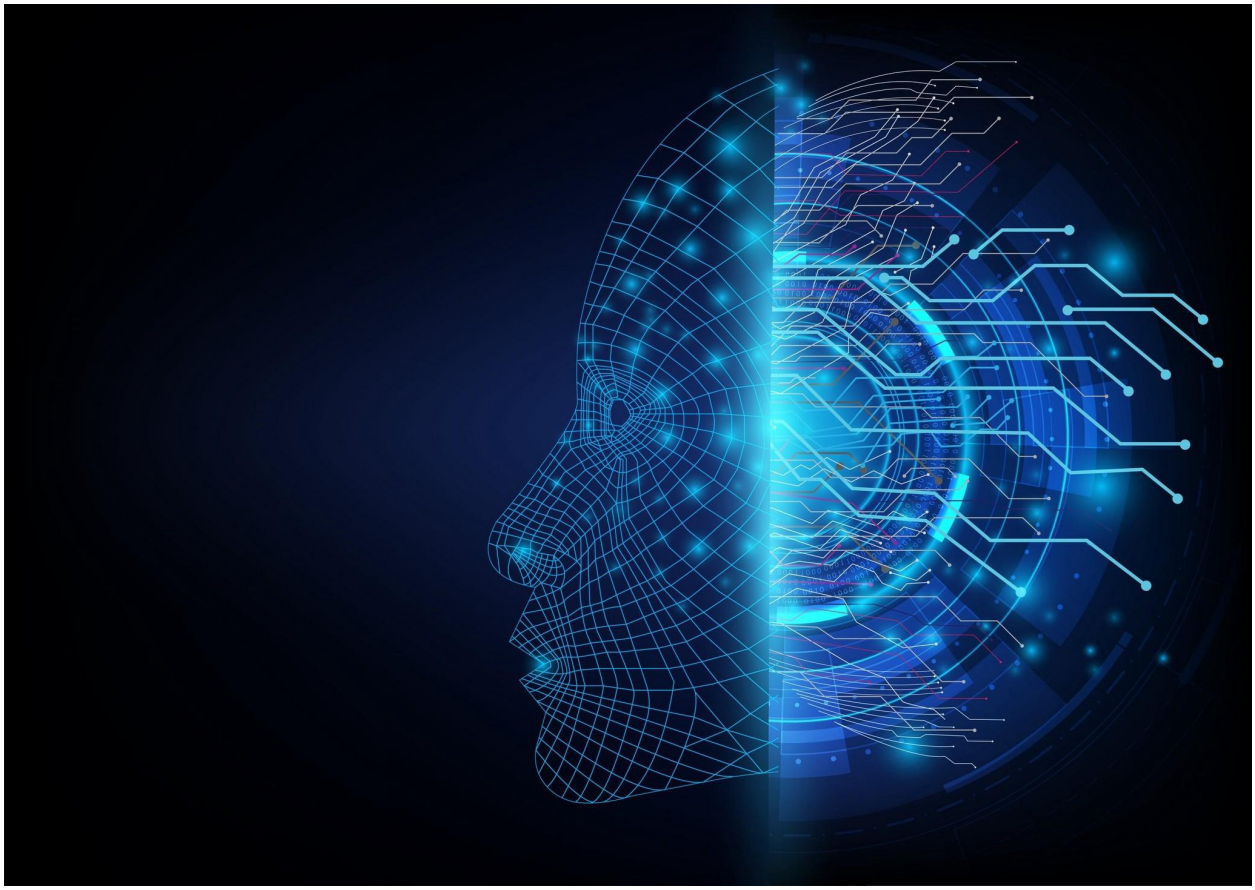
Question 1

Candidate Number: 59069

Artificial Intelligence - CS3AI18



1. Consider a scenario where a robotic agent, which is stuck in a building comprised of several rooms, is to be trained to escape the building to the "Free world". Given the building's Floor plan (shown in Figure Q1-1), a learning rate of 0.8, and a reward 100 points if the agent can escape from a room and zero otherwise, answer the following in the context of this scenario:

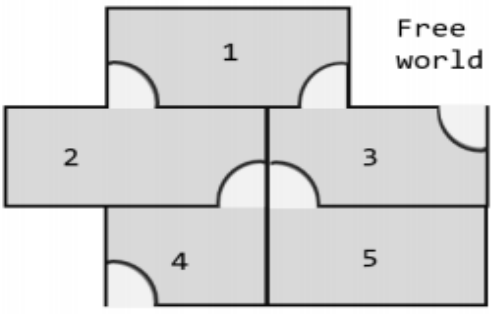


Figure Q1-1: Building's Floor Plan

(a) Explain the components of the reinforcement learning intelligent agent's interaction with the environment for the given scenario. (2 marks)

Reinforcement Learning is a method of machine learning implemented in such a way that focuses around the concept of rewarding actions which are positively regarded, while either punishing or simply not rewarding the actions which add no favour achieving the designated goal.

Each reinforcement learning system consists of 5 key aspects, a problem statement, an agent, an environment, possible actions and finally the rewards. Reinforcement learning uses the Markov decision process.

In the given situation, the reinforcement learning intelligent agent will interact with the environment by determining values for each of the rooms, updating them with each iteration, decreasing the probability of the agent returning into a less valuable room, and increasing the likelihood of the agent moving towards the target destination and escaping the building.

(b) Q-learning, a policy-based reinforcement learning, is to be used. Describe the Q-function's components for the above-mentioned agent.

(3 marks)

Q- Learning is a method of reinforced learning that operates on a policy-based methodology.

There are 5 main components to the process. First, a "Q-Value" table must be initialised, using the Q function to obtain the values where Q(s,a) represents applying the function to s, and a, which represent the state and action respectively.

The current state is then observed as $s = s_t$ where $t$represents the current iteration of the process.

An action is chosen for the new state, with respect to the action selection policy being implemented.

The reward from completing the action is then observed, combined with the new state returned s_{t+1}.

The respective value in for the current state is then updated with regards to the observed reward and the maximum reward possible for the next state.

Iterate through to the next state until the final state (termination goal) is reached.

In the above situation the reward is 100 points for a room which can lead to the free world, and 0 for a room that doesn't.

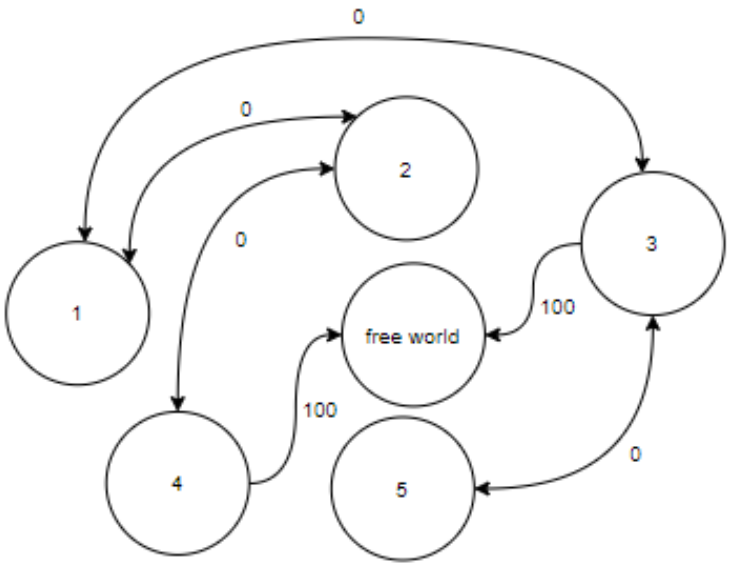
(c) Compute the following for this agent in the context reinforcement learning:

(i) State diagram and links labelled with Reward.
(ii) Initial Reward and Q matrices.
(iii) Q-value if the of the agent is in room 5.
(iv) Updated Q matrices.

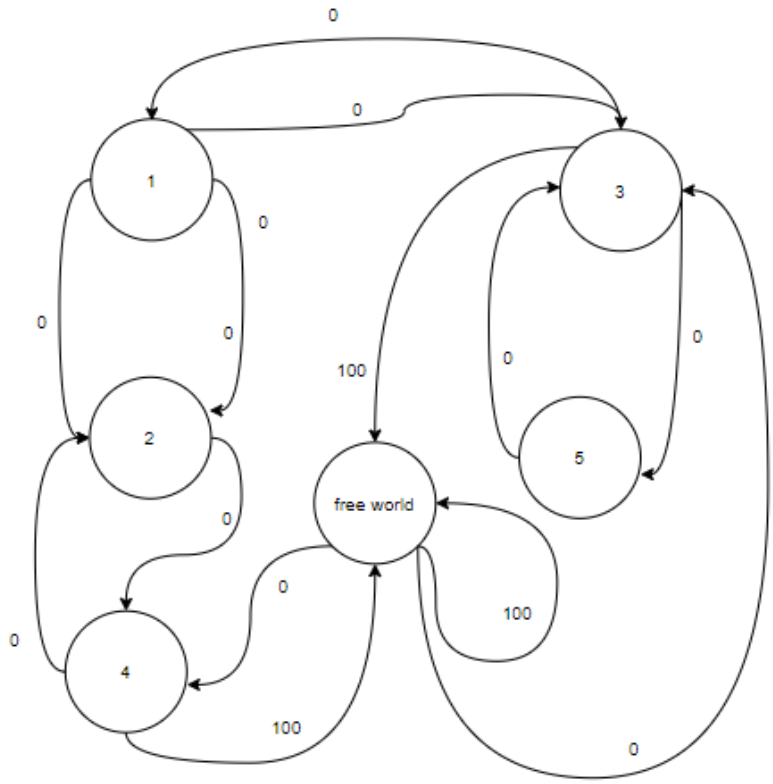
(12 marks)

i) State diagrams :

Simple:



Extended:



II)

Initial Reward Matrix:

R Values	1	2	3	4	5	Free world
1	-1	0	0	-1	-1	-1
2	0	-1	-1	0	-1	-1
3	0	-1	-1	-1	0	100
4	-1	0	-1	-1	-1	100
5	-1	-1	0	-1	-1	-1
Free	-1	-1	0	0	-1	100

Initial Q Matrix:

Initial Q Values	1	2	3	4	5	Free world
1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	0	0	0	0	0	0
4	0	0	0	0	0	0
5	0	0	0	0	0	0
Free	0	0	0	0	0	0

iii) Updated Q-Value for agent in Room 5

$$Q(5, \pi_{a_{5\to 3}}) = R(5, \pi_{a_{5\to 3}}) + \gamma\, max\{Q(3, \pi_{a_{3\to 1}}),\ Q(3, \pi_{a_{3\to 5}}), Q(3, \pi_{a_{3\to Free}})\}$$

Where Learning rate → γ = 0.8, πrepresents the policy, Q= respective Q value, R = respective R value.

$$Q(5, \pi_{a_{5\to 3}}) = 0 + 0.8(max\{0, 0, 100\}) = 0.8(100) = 80 \Rightarrow Q = 80$$

iv)

Updated Q Values	1	2	3	4	5	Free
1	0	64	80	0	0	0
2	51	0	0	80	0	0
3	80	0	0	0	64	100
4	0	80	0	0	0	100
5	0	0	80	0	0	0
Free	0	0	80	80	0	0

(d) Briefly explain how a supervised learning agent would help this robot escape the building.

(3 marks)

A supervised Learning agent could help this robot escape the building by using labels to determine the value of the room, and ranking them accordingly, it can provide feedback of whether entering or leaving a room is considered more valuable.

It is very unlikely that supervised learning would perform more effectively than reinforcement learning as this problem is extremely well-suited to reinforcement methods.