

ISLAMIC UNIVERSITY OF TECHNOLOGY (IUT)
ORGANISATION OF ISLAMIC COOPERATION (OIC)
Department of Computer Science and Engineering (CSE)

SEMESTER FINAL EXAMINATION
DURATION: 1 Hour 30 Minutes

WINTER SEMESTER, 2020-2021
FULL MARKS: 75

CSE 4711: Artificial Intelligence

Programmable calculators are not allowed.

There are **3 (three)** questions. Answer all **3 (three)** of them.

Figures in the right margin indicate marks of each question.

The square brackets on the start of each question denotes the corresponding CO and PO.

1. Consider that an agent is moving in the Gridworld shown in Figure 1. The only action available in cells *A*, *F*, *G*, and *H* is `Exit` with rewards $2X$, X , 0 , and 0 , respectively. Here, X = The last digit of your student ID + 1.

When the agent exits from a cell, it gets the reward for that cell as specified. From the other cells, the agent can only take the action `Left` or `Right`, which results in the agent moving to the immediate left or right cell, respectively.

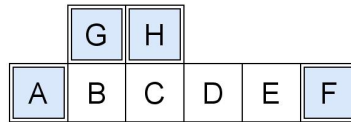


Figure 1: Gridworld for Question 1

If the agent is in cell *B* or *C*, and it takes a `Left` or `Right` action, it might fail sometimes. The agent will move in the desired direction with probability p , and it will fail and move up with probability $1 - p$. If the agent is in any other cell (*A*, *D*, *E*, *F*, *G*, or *H*), the action will always be successful.

Assume that the agent plays optimally, there is no living reward/penalty, and the discount factor is $\gamma \in [0, 1]$.

- a) [CO1, PO1] For each of the following policies, determine the value of each non-terminal state. 6 × 2
- i. $\pi_{\text{Right}}(S) = \begin{cases} \text{Exit}, & \text{if } S \text{ is a Terminal State} \\ \text{Right}, & \text{otherwise} \end{cases}$
- ii. $\pi_{\text{Left}}(S) = \begin{cases} \text{Exit}, & \text{if } S \text{ is a Terminal State} \\ \text{Left}, & \text{otherwise} \end{cases}$
- b) [CO3, PO2, PO3] For what range of value for p will the agent choose π_{Left} over π_{Right} ? 7
- c) [CO3, PO2, PO3] For what range of value for p will the agent choose π_{Right} over π_{Left} ? 6
2. a) [CO1, PO1] Two students, Rabbi and Sajal, are on their way to their class. The class is starting right now, so both of them are running late. It takes 40 minutes to get to the classroom by walking, and only 10 minutes to get there by bus. Sajal prefers to go on a bus given the choice, so the bus incurs an additional utility bonus of 5. However, to get on the bus, they need to wait at the bus stand. Due to traffic congestion, the bus might come 20, 40, or 60 minutes, each with equal probability. For each of the following utility functions, determine whether that person should walk or take the bus: 8 × 2
- i. Rabbi does not like being late. For him, the utility function for being late by time t is:

$$U_R(t) = \begin{cases} 0 & t \leq 0 \\ -2^{t/5} & t > 0 \end{cases}$$

- ii. Once Sajal is late, it does not matter to him how late he is. For him, the utility function for being late by time t is:

$$U_S(t) = \begin{cases} 0 & t \leq 0 \\ -50 & t > 0 \end{cases}$$

- b) [CO3, PO2, PO3] Assume that you have designed a deep learning model for determining whether a tomato leaf is infected by disease or not. Your system takes tomato leaf images as input and outputs “Positive” denoting that the leaf is infected, and “Negative” otherwise. After testing your model with a dataset, you have generated the confusion matrix to see how your model performs. You see that in 97% of the cases your model identifies a leaf as positive when the leaf is actually infected. And in 95% of the cases, your model identifies a leaf as negative, when the leaf is not infected. Suppose that, we know that 0.5% of the tomato leaf images in your dataset were infected by diseases. Explain why you should or should not trust the prediction of your system.
3. [CO2, PO4] Consider that Pacman is moving in an unknown grid with only two cells P and Q . From each state, the Pacman can either move Left or Right. According to some policy π , Pacman got the following sequence of actions and rewards:

t	s_t	a_t	s_{t+1}	r_t
0	X	<i>Right</i>	Y	2
1	Y	<i>Right</i>	Y	-4
2	Y	<i>Left</i>	Y	0
3	Y	<i>Left</i>	X	3
4	X	<i>Left</i>	X	-1

Table 1: Samples table for Question 3

Model the scenario as Reinforcement Learning problem. Mention the states, actions, and Q-values for each state-action pair after performing sample-based Q-Learning. Recall the update function for sample-based Q-Learning is:

$$Q(s_t, a_t) \leftarrow (1 - \alpha)Q(s_t, a_t) + \alpha(r_t + \gamma \max_{a'} Q(s_{t+1}, a'))$$

Initially all Q-values are set to 0. The discount factor $\gamma = 0.5$ and the learning rate $\alpha = 0.5$.