# ISLAMIC UNIVERSITY OF TECHNOLOGY (IUT)

**ORGANISATION OF ISLAMIC COOPERATION (OIC)**

**Department of Computer Science and Engineering (CSE)**

**SEMESTER FINAL EXAMINATION** **WINTER SEMESTER, 2021-2022**

**DURATION: 3 HOURS** **FULL MARKS: 150**

## CSE 4711: Artificial Intelligence

Answer **all 6 (six)** questions. Marks of each question and corresponding CO and PO are written in the right margin with brackets. The symbols have their usual meanings.

---

1. Consider that a badly-designed website called 'BeFake' uses a simple encryption software to encrypt the passwords of its users. During registration, BeFake takes the password string as input, encrypts it using a hash function provided by the software, and stores it in a file. If any registered user tries to log in, their input password is again encrypted and checked against the stored password. Unfortunately, you use the same username and password on multiple sites, including BeFake.

   Mallory, a malicious hacker and your archenemy, hacked into the server of BeFake and got hold of the encryption software and the files containing the encrypted passwords. She knows your username but does not know the password. However, she knows that your password consists of a combination involving at most three different English letters, 'P', 'Q', and 'R'. She also knows that some letters are more likely to occur than others. So she encodes this by setting, $cost(\texttt{P}) = 1$, $cost(\texttt{Q}) = 2$, and $cost(\texttt{R}) = 3$.

   To find your password, Mallory will generate different strings consisting of those letters, provide it to the software to generate the encrypted string, and compare it with the stored password until the first match. She will use search algorithms, specifically, Breadth-First Search (BFS), Depth-First Search (DFS), and Uniform Cost Search (UCS), to generate the passwords in lexicographic order. Assume that all ties are broken in lexicographic order.

   a) Show how Mallory would formulate the search problem by identifying the states, start state, successor function, and goal test. 5 (CO2) (PO2)

   **Solution:**
   **State:** Each possible string containing the letters, $P$, $Q$, and $R$.
   **Start State:** An empty string.
   **Successor Function:** Appends one letter ($P$, $Q$, or $R$) to the string.
   **Goal Test:** Verify the candidate string using the encryption software.
   **Rubric:**

   - 2 points for state
   - 1 point each for start state, successor, and goal test.

   b) Assume that the search algorithms, used by Mallory, predict up to 6 letter combinations. Due to collision, the encryption software generates the same hash value for the strings: PPPRR, RQQPPP, RRR, and QQQQ. 20 (CO3) (PO2)

   Assume that your password is RQQPPP. With a brief explanation, assess how doomed you are.

   **Solution:**
   BFS will return the shortest possible password, which is RRR. There is no other string shorter than that.
   DFS will return the lexicographically smallest password, which is PPPRR.

UCS will return QQQQ since it has the lowest cost (8) among all the strings.

The passwords returned by the algorithms do not match the actual password. However, since the hash values match, Mallory will be able to log in to the badly-designed website using my credentials. The same will happen for all the websites that use the same hash function to encrypt strings. If not, then I am safe (for now).

**Rubric:**

- 2 points for the correct choice.
- 4 points for the explanation.
- 2 points for the final decision.

2. The St. Petersburg paradox, proposed by Nicolas Bernoulli in 1713, describes the following game: a fair coin is tossed repeatedly by the player until it comes up as tails. If the first tails appears on the $n^{\text{th}}$ toss, the player wins $2^n$ dollars.

   a) Given infinite resources, show that the expected monetary value of the lottery is infinite. 6 (CO1) (PO1)

**Solution:**

The probability that the first tails appears on the $n$th toss is $\frac{1}{2^n}$. So,

$$\begin{aligned} EMV(L) &= \sum_{n=1}^{\infty} \frac{1}{2^n} \times 2^n \\ &= \sum_{n=1}^{\infty} 1 \\ &= \infty \end{aligned}$$

**Rubric:**

- 2 points for probability.
- 2 points for EMV per game.
- 2 points for the conclusion.

   b) Assume that Elon Musk (whose net worth is 182.6 billion US Dollars) offers to cover the lottery payout. Show how that affects the expected monetary value of the lottery. 6 (CO3) (PO2)

**Solution:**

Since Elon Musk has finite resources, the game must end once those resources are exhausted. Then the maximum number of times a coin can be tossed before it no longer can fully be covered by Mr. Musk is $L = \lfloor \log_2(W) \rfloor$.

Then $EMV(L) = \sum_{n=1}^{L} = \frac{1}{2^n} \times 2^n = L$. So the expected value of the game is:

$$\left\lfloor \log_2 \left(1.826 \times 10^{11}\right) \right\rfloor = 37$$

. So the expected value of the game is \$37.

**Rubric:**

- 2 points for identifying the problem.

- 2 points for determining the equation to calculate the EMV of the lottery.
- 2 points for the actual value.

c) Nicolas's cousin, Daniel Bernoulli, resolved the paradox in 1738 by suggesting that the utility of money is measured on a logarithmic scale. For example, $U(S_m) = \log_2(m)$, where $S_m$ is the state of having $\$m$ and $m > 0$. 8 (CO3) (PO2)

Under these circumstances, determine the maximum amount of money a rational player would pay to play the game. Assume that the player has to give up all of his/her money. Use $\sum_{n=1}^{\infty} \frac{n}{2^n} = 2$, if necessary.

**Solution:**

Assume that the player has $\$k$ and has to give $\$c$ to play the game:

$$\begin{aligned} U(L) &= \sum_{n=1}^{\infty} \frac{1}{2^n} \times \log_2\left(k - c + 2^n\right) \\ &= \sum_{n=1}^{\infty} \frac{1}{2^n} \times \log_2\left(2^n\right) \text{ since } k - c = 0 \\ &= \sum_{n=1}^{\infty} \frac{n}{2^n} \\ &= 2 \end{aligned}$$

To pay the maximum amount to play the game, the player has to have the utility of the given money equal to the utility of the lottery:

$$\begin{aligned} \log_2(c) &= 2 \\ c &= 4 \end{aligned}$$

So, a rational player would pay $\$4$ to play the lottery.

**Rubric:**

- 5 points for the utility of the lottery.
- 2 points for identifying the indifference condition.
- 1 points for the correct amount.

3. The game tree for a game between Alice and Bob is shown in Figure 1. The players make choices to maximize their utility functions which are known to the players beforehand. Alice can move Left, Right, or Straight, whereas Bob can move Left or Right. Alice moves first.

   Let $(x, y)$ denote the pair of values in the terminal nodes.

   a) Let the utility functions for Alice and Bob be $U_A(x, y) = x$ and $U_B(x, y) = x - y$ respectively.

      i. Determine the optimal action for Alice in this scenario. 5 (CO3) (PO2)

**Solution:**

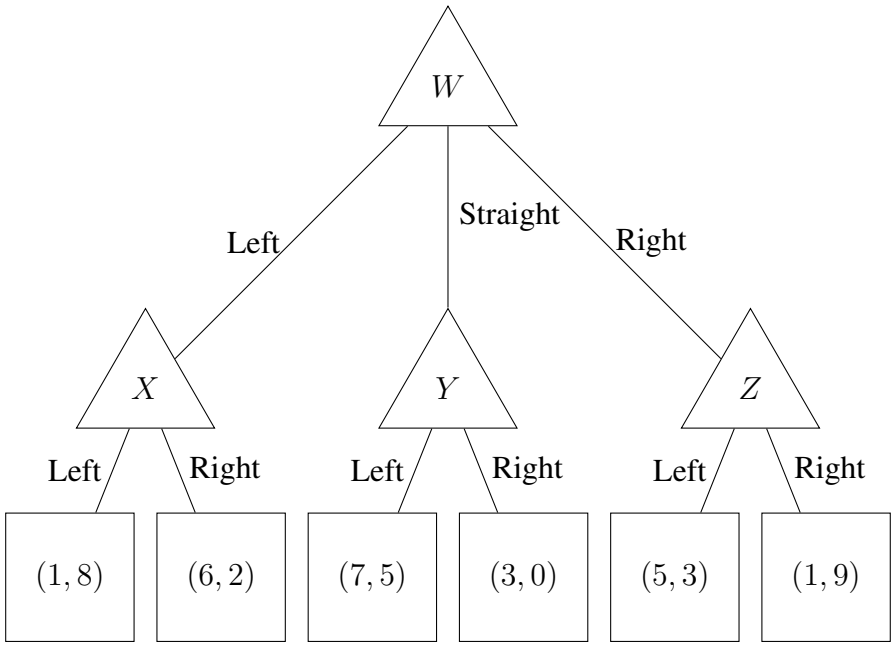


**Figure 1:** Game Tree for Question 3

$$X = (6, 2)$$
$$Y = (3, 0)$$
$$Z = (5, 3)$$
$$W = (6, 2)$$

Since $X$ gives Alice the maximum utility, she should move Left.

**Rubric:**

- 1 point for each value.
- 1 point for correct action.

ii. Identify the nodes that can be pruned considering the branches are explored from left to right. 1 (CO1) (PO1)

**Solution:**
None.

**Rubric:**

- 2 points for the correct answer.

b) Assume that Bob's utility function changes based on his mood. Table 1 denotes the marginal probability distribution associated with Bob's mood and Table 2 denotes the conditional probability distribution associated with Bob's utility function conditioned on his mood.

**Table 1:** Mood distribution

| $M$ | $P$ |
|---|---|
| Happy | 0.60 |
| Sad | 0.40 |

**Table 2:** Utility distribution

| $U_B$ | $P(U_B \mid M = \text{Happy})$ | $P(U_B \mid M = \text{Sad})$ |
|---|---|---|
| $-x$ | 0.45 | 0.30 |
| $x - y$ | 0.25 | 0.45 |
| $x^2 + y^2$ | 0.30 | 0.25 |

i. Calculate the likelihood of each utility functions for Bob. $3 \times 2$ (CO3) (PO2)

**Solution:**

$$\begin{aligned}
P(U_b(x,y) = -x) &= P(M = \text{Happy}) \times P(U_B = -x | M = \text{Happy}) \\
&\quad + P(M = \text{Sad}) \times P(U_B = -x | M = \text{Sad}) \\
&= 0.60 \times 0.45 + 0.40 \times 0.30 \\
&= 0.39 \\
P(U_b(x,y) = x - y) &= P(M = \text{Happy}) \times P(U_B = x - y | M = \text{Happy}) \\
&\quad + P(M = \text{Sad}) \times P(U_B = x - y | M = \text{Sad}) \\
&= 0.60 \times 0.25 + 0.40 \times 0.45 \\
&= 0.33 \\
P(U_b(x,y) = x^2 + y^2) &= P(M = \text{Happy}) \times P(U_B = x^2 + y^2 | M = \text{Happy}) \\
&\quad + P(M = \text{Sad}) \times P(U_B = x - y | M = \text{Sad}) \\
&= 0.60 \times 0.30 + 0.40 \times 0.25 \\
&= 0.28
\end{aligned}$$

**Rubric:**

- 1 point for each correct equation.
- 1 point for each correct value.

ii. Determine the optimal action for Alice in this scenario. 13 (CO3) (PO2)

**Solution:**
From $X$, Bob will choose:

- $(1, 8)$ when $U_B(x, y) = -x$, resulting in Alice getting 1.
- $(6, 2)$ when $U_B(x, y) = x - y$, resulting in Alice getting 6.
- $(1, 8)$ when $U_B(x, y) = x^2 + y^2$, resulting in Alice getting 1.

Expected utility of $X$ for Alice = $0.39 \times 1 + 0.33 \times 6 + 0.28 \times 1 = 2.65$.
From $Y$, Bob will choose:

- $(3, 0)$ when $U_B(x, y) = -x$, resulting in Alice getting 3.
- $(3, 0)$ when $U_B(x, y) = x - y$, resulting in Alice getting 3.
- $(7, 5)$ when $U_B(x, y) = x^2 + y^2$, resulting in Alice getting 7.

Expected utility of $Y$ for Alice = $0.39 \times 3 + 0.33 \times 3 + 0.28 \times 7 = 4.12$.
From $Z$, Bob will choose:

- $(1, 9)$ when $U_B(x, y) = -x$, resulting in Alice getting 1.
- $(5, 3)$ when $U_B(x, y) = x - y$, resulting in Alice getting 5.
- $(1, 9)$ when $U_B(x, y) = x^2 + y^2$, resulting in Alice getting 1.

Expected utility of $Z$ for Alice = $0.39 \times 1 + 0.33 \times 5 + 0.28 \times 1 = 2.32$.
Since $Y$ gives Alice the maximum expected utility, she should move Straight.
**Rubric:**

- 1 point for each correct choice for Bob and Alice.
- 1 point for each expected utility.
- 1 points for correct action.

4. Consider that a game of Pacman is being played on the grid shown in Figure 2.

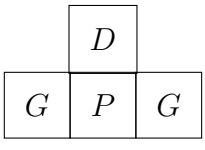


**Figure 2:** Pacman Grid for Question 4

Here, $P$ indicates the position of Pacman, $G$ indicates the position of the ghosts, and $D$ indicates the position of a food dot. For simplicity, the ghosts remain stationary throughout the game.

To determine the policy for playing the game, our feature-based Q-learning agent, Pacman, uses two features, $F_g$ and $F_d$ defined as follows:

$$F_g(s,a) = F_1(s) + F_2(s,a) + F_3(s,a)$$
$$F_d(s,a) = F_4(s) + 2 \times F_5(s,a)$$

where

$F_1(s)$ = number of ghosts within 1 step of state $s$
$F_2(s,a)$ = number of ghosts Pacman touches after taking action $a$ from state $s$
$F_3(s,a)$ = number of ghosts within 1 step of the state Pacman ends up in after taking action $a$
$F_4(s)$ = number of food dots within 1 step of state $s$
$F_5(s,a)$ = number of food dots eaten after taking action $a$ from state $s$

After a few episodes of Q-learning, the weights are $w_g$ = -10 and $w_d$ = 100. The discount factor, $\gamma = 0.5$ and learning rate, $\alpha = 0.5$. The action space of Pacman is $\{left, right, up, down, stay\}$. Pacman can take any actions from a state given it does not go beyond the grid.

a) Considering the actions, $a$ that are available from the current position, $s$ of Pacman: 12+6 (CO1) (PO1)
   i. Calculate $F_g(s,a)$ and $F_d(s,a)$.

**Solution:**

$$F_g(s,left) = 2+1+1 = 4$$
$$F_g(s,right) = 2+1+1 = 4$$
$$F_g(s,up) = 2+0+0 = 2$$
$$F_g(s,stay) = 2+0+2 = 4$$
$$F_d(s,left) = 1+2\times 0 = 1$$
$$F_d(s,right) = 1+2\times 0 = 1$$
$$F_d(s,up) = 1+2\times 1 = 3$$
$$F_d(s,stay) = 1+2\times 0 = 1$$

**Rubric:**

- 0.5 point for each correct equation.
- 1 point for each correct value.

ii. Calculate $\hat{Q}(s, a)$.

**Solution:**

$$\hat{Q}(s, up) = w_g F_g(s, up) + w_d F_d(s, up) = (-10) \times 2 + 100 \times 3 = 280$$
$$\hat{Q}(s, left) = w_g F_g(s, left) + w_d F_d(s, left) = (-10) \times 4 + 100 \times 1 = 60$$
$$\hat{Q}(s, right) = w_g F_g(s, right) + w_d F_d(s, right) = (-10) \times 4 + 100 \times 1 = 60$$
$$\hat{Q}(s, stay) = w_g F_g(s, stay) + w_d F_d(s, stay) = (-10) \times 4 + 100 \times 1 = 60$$

**Rubric:**

- 0.5 points for each correct equation.
- 1 point for each Q-Value.

b) Determine the optimal policy for Pacman from its current position following the $\hat{Q}(s, a)$ values calculated. Argue on how good the policy is considering the alternatives. $2 + 3$ (CO3) (PO2)

**Solution:**

The optimal policy for the current position of Pacman should be:

$$\pi(s) = \text{argmax}\left(\hat{Q}(s, up), \hat{Q}(s, left), \hat{Q}(s, right), \hat{Q}(s, stay)\right)$$
$$= up$$

Since going up results in Pacman eating the food dot, it is reasonable to decide to go up. Compared to other moves, where Pacman stays in the same position or gets eaten by the ghost, this is the best possible outcome for Pacman.

**Rubric:**

- 2 points for the correct decision.
- 3 points for the argument.

c) From its current position, $s$, Pacman moves up to go to the cell, $s'$ containing the food dot and eats it. We observe a reward, $R(s, a, s') = 250$. $4 + 3$ (CO1) (PO1)

Considering the actions, $a'$ that are available from $s'$:

i. Calculate $Q(s, up)$.

**Solution:**

$$Q(s, up) = R(s, a, s') + \gamma \times \max_{a'} \hat{Q}(s', a')$$
$$= 250 + 0.5 \times \max\left(\hat{Q}(s', down), \hat{Q}(s', stay)\right)$$
$$= 250 + 0.5 \times 0$$
$$= 250$$

where

$$\hat{Q}(s', down) = w_g F_g(s', down) + w_d F_d(s', down) = (\text{-}10) \times 2 + 100 \times 0 = -20$$
$$\hat{Q}(s', stay) = w_g F_g(s', stay) + w_d F_d(s', stay) = (\text{-}10) \times 0 + 100 \times 0 = 0$$

**Rubric:**

- 1 point for the correct equation.
- 1 point for each of the correct $Q(s', a')$.
- 1 point for the final result.

ii. Update $w_g$ and $w_d$.

**Solution:**

$$w_g = w_g + \alpha \times \Big(Q(s, up) - \hat{Q}(s, up)\Big) \times F_g(s, a) = \text{-}10 + 0.5 \times (250 - 280) \times 2 = \text{-}40$$
$$w_d = w_d + \alpha \times \Big(Q(s, up) - \hat{Q}(s, up)\Big) \times F_d(s, a) = 100 + 0.5 \times (250 - 280) \times 3 = 55$$

**Rubric:**

- 0.5 points for each equation.
- 1 point for each correct value.

5. a) Consider the following statements: 12 (CO3) (PO2)

- Surely computers cannot be intelligent - they can do only what their programmers tell them.
- Surely humans cannot be intelligent - they can do only what their genes tell them.

For each of the statements, is the latter clause true? And does it imply the former? Provide brief arguments.

**Solution:**

This depends on the definition of "intelligent" and "tell". In one sense computers/humans only do what their programmers/genes command them to do, but in another sense what the programmers/genes tells the computer/human to do often has very little to do with what the computer/human actually does.

For a computer, in one sense, a programmer can "tell" the computer "learn to play chess better than I do, and then play that way". But in another sense, the programmer told the computer "follow this learning algorithm" and it learned to play. So we are left in the situation where we may or may not consider learning to play chess to be a sign of intelligence, and we may think the intelligence resides int he programmer or in the computer.

The second statement is parallel with the first one. Whatever we decide about whether computers could be intelligent, we are committed to making the same conclusion about humans, unless our reasons for deciding whether something is intelligent take into account the mechanism.

**Rubric:** This question is intended to stimulate discussion. So there is no perfect answer. However, the argument of the student, based on what they have learnt throughout the semester, should discuss the following:

- The role of programmers/genes in AI

- The role of the behavior of computers/humans
- The parallel between these two
- Concluding remarks on their decision

3 points for each part.

b) Sometimes Markov Decision Processes (MDP) are formulated with a reward function $R(s, a)$ that depends on the action taken, or $R(s, a, s')$ that also depends on the outcome state. Modify the Bellman Equation to determine the value of a state based on these formulations. 5 (CO2) (PO2)

**Solution:**

For $R(s, a)$, we have:

$$V(s) = \max_a \left[ R(s, a) + \gamma \sum_{s'} T(s, a, s') V(s') \right]$$

And for $R(s, a, s')$, we have:

$$V(s) = \max_a \sum_{s'} T(s, a, s') [R(s, a, s') + \gamma V(s')]$$

**Rubric:**

- 2.5 points for each formulation.

c) Suppose that we define the value of a state to be the maximum (as opposed to summation) reward obtained from its future states. Does this result in stationary preferences? If yes, justify your position. If not, provide a counterexample. $2 + 6$ (CO1) (PO1)

**Solution:**

Stationary preferences require the agent to follow:

$$[a_1, a_2, \dots] \succ [b_1, b_2, \dots]$$
$$\Updownarrow$$
$$[r, a_1, a_2, \dots] \succ [r, b_1, b_2, \dots]$$

However, if the utility of a sequence is its maximum reward, we can easily violate the stationary preferences. For example,

$$[4, 3, 0, 0, \dots] \sim [4, 0, 0, 0, \dots]$$

but if we remove the first item from both sequences:

$$[3, 0, 0, \dots] \succ [0, 0, 0, \dots]$$

So it breaks stationary preferences.

**Rubric:**

- 2 points for decision.
- 3 points for the theorem.
- 3 points for the counterexample.

6. Consider that we have placed a robot in an unknown environment. It can perceive its current state,

$s_t$, and execute an action, $a_t$. Upon executing the action, the robot can move to another state, $s_{t+1}$ (possibly the same as before), and get an instant reward, $r_t$. The robot moves around for a while and comes up with the experiences shown in Table 3.

**Table 3:** Experience achieved in 5 iterations for Question 6

| $t$ | $s_t$ | $a_t$ | $s_{t+1}$ | $r_t$ |
|---|---|---|---|---|
| 0 | $A$ | $Up$ | $B$ | 2 |
| 1 | $B$ | $Up$ | $B$ | -4 |
| 2 | $B$ | $Down$ | $B$ | 0 |
| 3 | $B$ | $Down$ | $A$ | 3 |
| 4 | $A$ | $Down$ | $A$ | -1 |

Now, answer the following questions based on the scenario mentioned above:

a) Formulate the scenario as a sample-based Q-learning problem by determining Q-values for each state-action pair based on the experience of the robot. 22 (CO2) (PO2)

Assume that the discount factor, $\gamma = 0.5$ and the learning rate, $\alpha = 0.5$.

**Solution:**

At $t = 0$, we get:

$$\begin{aligned} Q(A, Up) &= (1-\alpha)Q(A, Up) + \alpha(r + \gamma \max_{a'} Q(B, a')) \\ &= (1 - 0.5) \times 0 + 0.5 \times (2 + 0.5 \times 0) \\ &= 1 \end{aligned}$$

At $t = 1$, we get:

$$\begin{aligned} Q(B, Up) &= (1-\alpha)Q(B, Up) + \alpha(r + \gamma \max_{a'} Q(B, a')) \\ &= (1 - 0.5) \times 0 + 0.5 \times (\text{-}4 + 0.5 \times 0) \\ &= \text{-}2 \end{aligned}$$

At $t = 2$, we get:

$$\begin{aligned} Q(B, Down) &= (1-\alpha)Q(B, Down) + \alpha(r + \gamma \max_{a'} Q(B, a')) \\ &= (1 - 0.5) \times 0 + 0.5 \times (0 + 0.5 \times 0) \\ &= 0 \end{aligned}$$

At $t = 3$, we get:

$$\begin{aligned} Q(B, Down) &= (1-\alpha)Q(B, Down) + \alpha(r + \gamma \max_{a'} Q(A, a')) \\ &= (1 - 0.5) \times 0 + 0.5 \times (3 + 0.5 \times 1) \\ &= 1.75 \end{aligned}$$

At $t = 4$, we get:

$$\begin{aligned} Q(A, Down) &= (1-\alpha)Q(A, Down) + \alpha(r + \gamma \max_{a'} Q(X, a')) \\ &= (1 - 0.5) \times 0 + 0.5 \times (\text{-}1 + 0.5 \times 1) \\ &= \text{-}0.25 \end{aligned}$$

So the final values are:

$Q(A, Down) = \text{-}0.25$

$Q(A, Up) = 1$

$Q(B, Down) = 1.75$

$Q(B, Up)$ = -2
**Rubric:**

- 4 points for each iteration.
- 0.5 points for each value.

b) Determine the optimal policy from the scenario. 3 (CO3) (PO2)

**Solution:**

$$\begin{aligned}\pi(A) &= \text{argmax}\left(Q(A, down), Q(A, up)\right)\\ &= Up\\ \pi(B) &= \text{argmax}\left(Q(B, down), Q(B, up)\right)\\ &= Down\end{aligned}$$

**Rubric:**

- 1.5 points for each correct policy.