SOLUTIONS

Module:	Introduction to Artificial Intelligence		
Module Code	EBU4203	Paper	B
Time allowed	2hrs	Filename	**Solutions_2324_EBU4203_B**
Rubric	ANSWER ALL FOUR QUESTIONS		
Examiners	Dr Yuanwei Liu	Dr Muhammad Salman Haleem	

Question 1

a) State THREE ways that an experiment can fail when uncertainties are not considered appropriately. **(3 marks)**

b) State and discuss THREE ethical challenges associated with the development and use of Artificial Intelligence (AI). **(6 marks)**

c) In a high school classroom, 30% of students study History, 45% study Maths, and 15% study both History and Maths. **[16 marks]**

- **i)** If two students are randomly selected independently, what is the probability that both of them study Maths? **(3 marks)**
- **ii)** If a student is randomly selected, what is the probability that he/she studies History but not Maths? **(3 marks)**
- **iii)** If a student is randomly selected, what is the probability that he/she studies either History or maths? **(6 marks)**
- iv) If a student is randomly selected, what is the probability that he/she does not study History or Maths? **(4 marks)**

Answers:

a)

1. LAZINES
2. THEORETICAL ignorance:
3. PRACTICAL ignorance:

1 mark for each answer

b)

Ethical considerations: Any three of the following are acceptable.

- Privacy [1 mark] : AI applications often rely on vast amounts of sensitive data, particularly in fields like banking, healthcare, and law. [1 mark]
- Bias [1 mark]: AI systems can perpetuate biases present in the training data, which can lead to unfair or discriminatory outcomes. [1 mark]
- job displacement [1 mark]: The automation enabled by AI can lead to job losses and significant disruptions in the workforce. [1 mark]
- Misuse [1 mark]: AI technology can be misused for malicious purposes, such as creating deepfakes or engaging in phishing attacks. [1 mark]
- Interpretability [1 mark]: AI algorithms, particularly in deep learning and generative adversarial network (GAN) applications, can be difficult to interpret. [1 mark]

c)

i) Probability that a student studies Maths (P(M)) = 45% = 0.45

To find the probability that both students study Maths:

P(Both students study Maths) = P(M) * P(M) [2 mark]

P(Both students study Maths) = 0.45 * 0.45 = 0.2025 [1 mark]

So, the probability that both of the two randomly selected students study Maths is 0.2025 or 20.25%.

ii)
Probability that a student studies History (P(H)) = 30% = 0.30
Probability that a student studies both History and Maths ($P(H \cap M)$) = 15% = 0.15
Now, calculate the probability of studying History but not Maths:

P(H and not M) = P(H) - P(H and M) [2 marks for formula]

P(H and not M) = 0.30 - 0.15 = 0.15 [1 mark]

So, the probability that a randomly selected student studies History but not Maths is 0.15 or 15%.

iii)

Given that 45% study History, i.e., P(H) = 45/100 = 9/20 [1 mark]
30% study Maths, i.e., P(M) = 30/100 = 6/20 [1 mark]
15% study both Hindi and Maths, i.e., P(H and M) = 15/100 = 3/20 [1 mark]
So, P(H or M) = P(H) + P(M) - P(H and M) [2 marks for formula]
= 9/20 + 6/20 - 3/20
= 12/20
= 3/5
(=60%)
[1 marks for final answer]

iv)
P(not H or M) = 1 - P(H or M) [2 marks for formula]

= 1 - 0.60
= 0.40 [2 mark]

Question 2

a) Answer the following questions on Reinforcement Learning (RL): **[11 marks]**
   i) Discuss why a discount factor is necessary in a Markov decision process. **(3 marks)**
   ii) Explain what is a model-free RL. **(3 marks)**
   iii) The actor-critic architecture is a popular RL approach, please give the functionalities of actor and critic, and explain the main idea of this architecture. **(5 marks)**

b) **Figure 1** illustrates a deterministic Markov Decision Process (MDP). The discount factor $\boldsymbol{\gamma}$ is 0.5. States are represented as A, B, C, and D. Arrows indicate state transition with corresponding actions. The action probability and immediate rewards are **P** and **R**, labelled next to the arrows, respectively. The MDP starts with an initial value function of $V_0(A)=V_0(B)=V_0(C)=V_0(D)=2$. **[14 marks]**

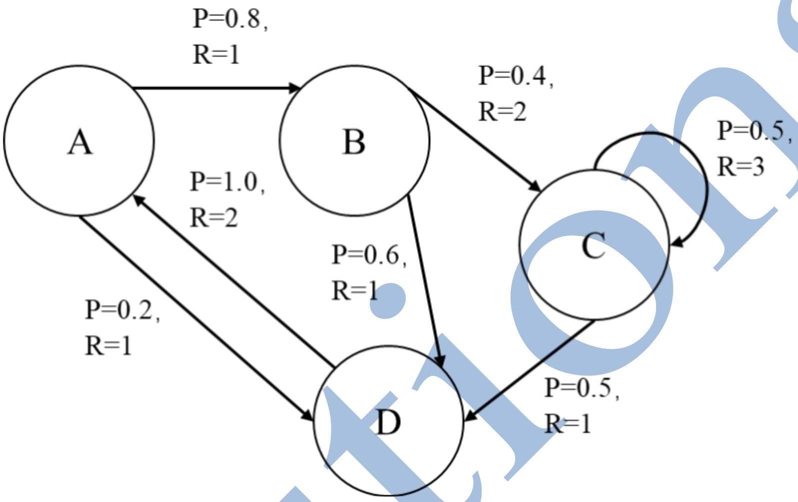


Figure 1

   i) Complete the Markov Chain transition matrix for the given problem. **(4 marks)**
   ii) For **one iteration**, calculate the value function $V_1(D)$. **(4 marks)**
   iii) For **one iteration**, calculate the value function V_1 of states A, B, and C. **(6 marks)**

Answer:

a)
i)
Avoiding the problem of unlimited cumulative rewards [1 mark]
Control the balance between long-term and short-term rewards [1 mark]
Promoting finite horizon decisions [1 mark]

ii)
model-free RL does not require knowledge or a model of the environment 1 mark. Instead, it learns by interacting with the environment 1 mark and receiving feedback in the form of rewards or penalties. 1 mark

iii)
Actor: deciding which action to take in a given state (1 mark)

Critic: evaluating the value of the action at the state (1 mark)

The Actor-Critic model consists of both an Actor network and a Critic network.
The Actor network learns the policy, deciding which action to take in a given state.
The Critic network estimates the value function, evaluating the value of the action at the state.
Motivation: This combination allows the agent to learn from both policy and value function aspects, making full use of the advantages of both methods. (3 mark for any reasonable explanation)

b)

i)

$$\mathbb{p} = \begin{bmatrix} 0 & 0.8 & 0 & 0.2 \\ 0 & 0 & 0.4 & 0.6 \\ 0 & 0 & 0.5 & 0.5 \\ 1.0 & 0 & 0 & 0 \end{bmatrix}$$

[1 mark for each row]

ii) Bellman equation: $V(s) = E\,[r_{t+1} + \gamma\, V(s_{t+1}) \mid S_t = s]$

$V_1(D) = 1.0 * [2 + 0.5V_0(A)] = 3$

[1 mark for equation, 1 mark for substituting the values, 2 mark for correct answer]

iii)

$V_1(C) = 0.5 * [1 + 0.5V_0(D)] + 0.5 * [3 + 0.5V_0(C)] = 3$ [1 mark for equation, 1 mark for correct answer]

$V_1(B) = 0.4 * [2 + 0.5V_0(C)] + 0.6 * [1 + 0.5V_0(D)] = 2.4$

[1 mark for equation, 1 mark for correct answer]

$V_1(A) = 0.8 * [1 + 0.5V_0(B)] + 0.2 * [1 + 0.5V_0(D)] = 2$

[1 mark for equation,1 mark for correct answer]

Question 3

a) State and explain the FIVE fundamental steps involved in processing an image in Computer Vision. **(10 marks)**

b) Consider an input image with dimensions of 3x3 pixels. The values of the input image are as follows:

[[1, 2, 1],
[0, 1, 2],
[1, 0, 2]]

i) Is this a coloured image? Explain your answer. **(5 marks)**

ii) Using valid padding and a stride of 1, calculate the result of applying the 2x2 convolutional kernel with the following weights: **(5 marks)**

[[1, -1],
[0, 1]]

iii) Based on your answer in part ii), perform max pooling with a 2x2 pooling window and a stride of 2. Calculate the resulting pooled feature map and show its values. **(3 marks)**

iv) What is the purpose of adding zero-padding to an input image before convolution, and how does it affect the size of the output feature map? **(2 marks)**

Answer:
a)
The fundamental steps in image processing in Computer Vision include
1. data collection (capturing an image)
2. preprocessing (noise reduction, resizing, etc.)
3. feature extraction (identifying relevant characteristics)
4. feature selection (choosing important features)
5. classification or recognition (making sense of the visual information).

. 2 marks for each answer (1 for stating the step, 1 for explanation)

b)
i)
No 2 marks
If it were a colored image, it would typically have multiple channels (e.g., RGB channels), where each channel represents a color component (Red, Green, and Blue), and each pixel would have multiple values associated with it to specify its color. 1 mark

In the provided input image, there is only one value (intensity level) per pixel 2 mark , which makes it a grayscale image.

ii)
The answer is
[[2, 3],
[-1, 1]]

1 mark for each value
1 mark for showing the calculations below

Steps:

1. Place the 2x2 kernel at the top-left corner of the input image.
   Input Portion:
   [[1, 2],
    [0, 1]]
   Kernel:
   [[1, -1],
    [0, 1]]
   Element-wise multiplication:
   [[1*1, 2*(-1)],
    [0*0, 1*1]]
   Sum the results:(1*1 + 2*(-1) + 0*0 + 1*1) = 2

2. Place the 2x2 kernel at the top-right corner of the input image.
   Input Portion:
   [[2, 1],
    [1, 2]]
   Kernel:
   [[1, -1],
    [0, 1]]
   Element-wise multiplication:
   [[2*1, 1*(-1)],
    [1*0, 2*1]]
   Sum the results:(2*1 + 1*(-1) + 1*0 + 2*1) = 3
3. Place the 2x2 kernel at the down-left corner of the input image.
   Input Portion:
   [[0, 1],
    [1, 0]]
   Kernel:
   [[1, -1],
    [0, 1]]
   Element-wise multiplication:
   [[0*1, 1*(-1)],
    [1*0, 0*1]]
   Sum the results:(0*1 + 1*(-1) + 1*0 + 0*1) = -1
4. Place the 2x2 kernel at the down-right corner of the input image.
   Input Portion:
   [[1, 2],
    [0, 2]]
   Kernel:
   [[1, -1],
    [0, 1]]
   Element-wise multiplication:
   [[1*1, 2*(-1)],
    [0*0, 2*1]]
   Sum the results:(1*1 + 2*(-1) + 0*0 + 2*1) = 1

So, the final result of applying the 2x2 convolutional kernel with the given weights to the input image using "valid" padding and a stride of 1 is as follows:

[[2, 3],
 [-1, 1]]

iii)

Start with the 2x2 window:

 [[2, 3],
 [-1, 1]]

The maximum value in this window is 3. 1 mark

The answer is:
[[3]]

2 marks for the correct answer

iv)

Padding is used to control the size of the output feature map. 1 marks

Adding zero-padding to the input image before convolution preserves its spatial dimensions. Without padding, the output feature map is smaller than the input. 1 marks

Question 4

This question is about Natural Language Processing (NLP).

a) The block diagram of 'Classical' NLP Pipeline is illustrated in **Figure 1**. State the names of the missing components, which are labelled as '?' in the diagram. Then, briefly describe their functionalities. **[6 marks]**

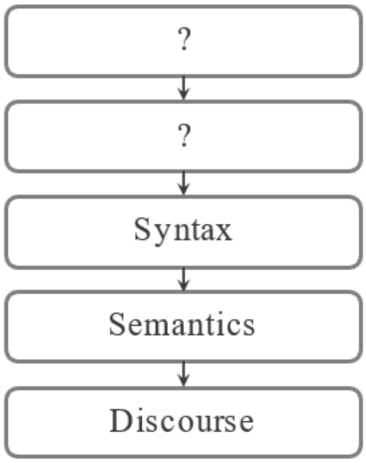


Figure 1

b) Consider a simplified Hidden Markov Model (HMM)-based part-of-speech tagging system with three part-of-speech tags: Noun (N), Verb (V), and Model (M). There are three sentences in the training set

1. Jane will eat apple
2. Will can meet Jane
3. Will Jane meet Will?

Table 1 shows the co-occurrence table, which can be used to analyze how different parts-of-speech interact within a text corpus. Fill in the missing values labelled by "?" in **Table 1.**

[7 marks]

	Noun	**Model**	**Verb**	**<End>**
<Start>	2	?	?	?
Noun	?	?	?	?
Model	?	?	?	0
Verb	?	?	?	?

Table 1

c) Consider the following sentence **[8 marks]**

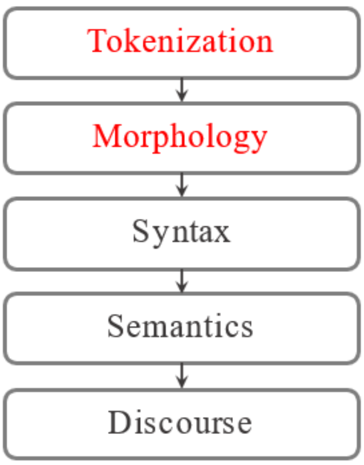
"The burglar jumps off the compound wall"

You are required to develop the training set for generating word embedding. When the index of the centre word is $i = 2$ and the window size is $W = 2$, the source text and training samples will look like this

Source Text								Training Samples
The	burglar	**jumps**	off	the	compound	wall	⇨	(? , ?)

i) Define the training set when $i = 2, W = 1$ **(2 marks)**
ii) Define the training set when $i = 1, W = 2$ **(3 marks)**
iii) Define the training set when $i = 0, W = 6$ **(3 marks)**
iv) State and explain FOUR main advantages of Long Short Term Memory over traditional Recurrent Neural Networks architecture **[4 marks]**

Answer:
a)



1 mark each for naming the blocks and 2 marks for putting in right order

Tokenization: break text into sentences and words and lemmatize [1 mark]
Morphology: Part of speech (POS) tagging, stemming, NER [1 mark]

b)

	Noun	**Model**	**Verb**	**<End>**
<Start>	2	1	0	0
Noun	0	2	1	3
Model	1	0	2	0
Verb	3	0	0	0

Each cell of the table worth 0.5 mark

c)
i)
(jumps, burglar) 1 mark
(jumps, off) 1 mark

ii)
(burglar, The) 1 mark
(burglar, jumps) 1 mark
(burglar, off) 1 mark

iii)
(The, burglar) 0.5 mark
(The, jumps) 0.5 mark
(The, off) 0.5 mark
(The, the) 0.5 mark
(The, compound) 0.5 mark
(The, wall) 0.5 mark

d)
Vanishing Gradient Problem: LSTM were introduced to mitigate vanishing gradient problem as they incorporate specialized memory cells amd gating mechanisms which allow them to learn and store information over long sequences [1 mark]

Handling Long Term Dependencies: LSTM are better at capturing long term dependencies because of their ability to maintain and update cell states over time [1 mark]

Gating Mechanism: LSTM have gating mechanism (input gate, forget gate and output gate) that regulate the flow of information, allowing them to selectively update and use information from previous time steps[1 mark]

Memory and Information Retention: LSTM have an explicit memory cell that allows them to store and retrieve information over long sequences, making them better suited for tasks that require maintaining context. [1 mark]