ARTIFICIAL INTELLIGENCE

ASSIGNMENT-2

Name: A. Afreen
Roll.NO: 2248IA0508
Branch: CSE - A

(1) What is MDP? Explain the components that define MDP.

(a) MDP stands for Markov Decision Process. It is a mathematical framework used for modelling decision-making situations where outcomes are partly under the control of a decision-maker and partly determined by chance. MDPs are widely used in fields like reinforcement learning, operations research, and economics.

Components of MDP:-

An MDP is defined by a tuple (S, A, P, R, γ) where:

1) States (S):
- A set of states ~~respon~~ representing all possible situations the system can be in.
- Example: In a grid-world game, each position on the grid is a state.

2) Actions (A):-
- A set of actions available to the agent in each state.
- Example:

3) Transition probability:
- $P(s'|s,a)$: The probability of transitioning to state s' from state s after taking action a.
- This captures the stochastic nature of the environment

4) Reward function (R):
- $R(s,a,s')$: The immediate reward received after transitioning from state s to state s' by taking action a.
- Rewards can be deterministic or probabilistic

5) Discount factor (γ):
- A scalar value $\gamma \in [0,1]$ that represents the importance of future rewards.

(1) b) What is optimal policy? Describe the optimal policy for the stochastic environment with $R(s) = -0.04$

The optimal policy is a policy that yields highest expected utility.

$\pi(s)$ is the action recommended by policy π for state s.

The optimal policy is represented by π^*

given by $\pi^* : S \to A$

optimal policy with $R(s) = -0.04$ for non terminal states.

3	→	→	→	+1
2	↑		↑	−1
1	↑	←	←	←
	1	2	3	4

→	→	→	+1
↑		→	−1
→	→	→	↑

$R(s) < -1.6284$

→	→	→	+1
↑		↑	−1
↑	→	↑	←

$-0.4278 < R(s) < -0.0850$

→	→	→	+1
↑		←	−1
Start	←	←	↓

$-0.0221 < R(s) < 0$

☆	☆	☆	+1
☆		←	−1
☆	☆	☆	↓

$R(s) > 0$

When $R(s) \leq -1.6284$, life is so painful that the agent heads straight for the nearest exit even if the exit is worth −1.

When $-0.4278 < R(s) < -0.0850$, life is quite unpleasant that the agent takes the shortest route to +1 State and is willing to risk into -1 state by accident

The agent takes the shortcut from state (3,1). When $-0.0221 < R(S) < 0$. The life is slightly deary where the policy takes no risks at all. In (4,1) & (3,2) the agent heads directly away from -1 state so, that is cannot fall in by accident.

finally, if $R(S) > 0$, then life is positively enjoyable where the agent avoids both exits. and the agent obtains infinite total reward as it never enters the terminal state.

2) 1(a) Explain the following:

(i) Definition of POMDP:

The Description of MDP assumes that the environment is fully observable. Which means that the agent always knows in which state it is in. The optimal policy depends on only the current state

However when the environment is partially observable the situation is less clear. The agent does not necessarily know which state it is in, so it cannot execute the action $\pi(s)$. recommended for that state in the optimal policy π.

POMDP's are difficult than ordinary MDP's but the real world is partially observable.

The POMDP has same elements as MDP:

(i) Set of states (S)

(ii) Actions A(S)

(iii) Rewards R(S)

(iv) Transition Model $P(s'|s,a)$

(i) Decision cycle of a POMDP agent:

let $b(s)$ is the previous belief state and the agent does action A and perceives evidence e then the new belief state is given by

$$b'(s') = \alpha P(e|s') \sum_s P(s'|s,a) b(s).$$

The simplified form of this equation would be

$$b' = \text{FORWARD}(b,a,e)$$

→ The optimal action in POMDP depends only on agents current belief state which is, given by $\pi^*(b)$ which maps belief states to actions. It does not depend on the actual state the agent is in because the agent does not known its actual state, all its knows is only the belief state

→ The POMDP decision cycle can be broken into 3 steps.

(1) Given the current belief state b, execute the action $\boxed{a = \pi^* b}$

(2) Receives evidence 'e'.

(3) set of current belief state to FORWARD (b,a,e) and REPEATE

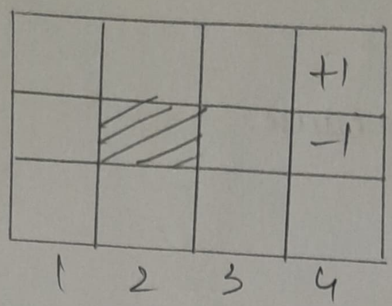
(ii) Calculating the probability that an agent in belief state b reaches final state after executing action a

In POMDP's the belief state b is a probability distribution over all possible states that the agent Might be in.

→ The initial Belief state for the 4x3 POMDP Could be the uniform distribution over the 9 non terminal states as follows

$$b = \langle \tfrac{1}{9}, \tfrac{1}{9}, \tfrac{1}{9}, \tfrac{1}{9}, \tfrac{1}{9}, \tfrac{1}{9}, \tfrac{1}{9}, \tfrac{1}{9}, \tfrac{1}{9}, 0, 0 \rangle$$

$$b(1,1) = 1/9$$



b(s) is the probability assign through the actual state s by the belief state b.

The agent can calculate its current belief state as the conditional probability distribution over the actual states given the sequence of action so far. This is called "filtering Task".

→ The following is the filtering equation which shows how to calculate the new belief state from the previous belief state and new evidence

$$P(X_{t+1} | e_{1:t+1}) = \alpha P(e_{t+1} | X_{t+1}) \sum_{x_t} (X_{t+1} | x_t, e_{1:t}) P(X_t | e_{1:t})$$

where α is constant that makes the Belief state sum to one

(2) b) Write short notes on the following with examples:

(i) full joint probability distribution.

The full joint probability distribution specifies the probability of every possible combination of values for a set of random variables. It is a comprehensive representation of the probabilistic relationships among the variables in a domain.

Example:

Density: Cavity, ¬Cavity } cause

toothache, catch.

	Toothache		¬Toothache	
	Catch	¬Catch	Catch	¬Catch
Cavity	0.018	0.012	0.072	0.008
¬Cavity	0.016	0.064	0.144	0.576

full joint probability.

Each row represents a possible outcome, and the probabilities sum to 1

(ii) Bayesian Networks:-

for Bayesian Network, we will represent

$$P(X_1, X_2, \ldots X_n) = \prod_{i=1}^{n} P(x_i \mid parent(x_i)).$$

Where $X_1, X_2 \ldots X_n$ random variables that are represented by nodes in the network

examples:

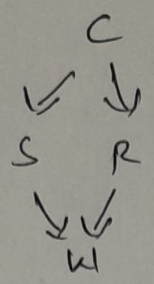
Consider the following Bayesian Networks which consists of following Variables

Cloudy (C) - Whether it is cloudy or not

sprinkless (S) - Whether sprinkless are ON/OFF

Rain (R) - whether its raining/not

Wet Grass (W) - Whether the grass is wet or Not



$P(C,S,R,W) = P(C) * P(S|C) * P(R|C) * P(W|S,R)$

P(C)

C	P(C)
T	0.5
F	0.5

P(S|C).

S	C	P(S|C)
T	T	0.1
F	T	0.9
T	F	0.5
F	F	0.5

P(R|C).

R	C	P(R|C)
T	T	0.8
F	T	0.2
T	F	0.2
F	F	0.8

P(W|S,R)

W	S	R	P(W|S,R)
T	T	T	0.99
F	T	T	0.01
T	T	F	0.9
F	T	F	0.1
T	F	T	0.8
F	F	T	0.2
T	F	F	0.0
F	F	F	1.0

<u>case(i):</u>

$C=T, S=T, R=T, W=T$

$P(C=T, S=T, R=T, W=T)$

$\Rightarrow P(C=T) * P(S=T|C=T) * P(R=T|C=T) * P(W=T|S=T, R=T)$

$\Rightarrow 0.5 * 0.1 * 0.8 * 0.99$

$\Rightarrow 0.0396$

Case (ii) :-

$(C, S, R, W) = (F, T, F, T)$

$P(C=F, S=T, R=F, W=T)$

$\Rightarrow P(C=F) * P(S=T|C=F) * P(R=F|C=F) * P(W=T|S=T, R=F)$

$\Rightarrow 0.5 * 0.5 * 0.8 * 0.9$

$\Rightarrow 0.18$

Case (iii) :-

$(C, S, R, W) = P(C=F, S=F, R=T, W=T)$

$\Rightarrow P(C=F) * P(S=F|C=F) * P(R=T|C=F) * P(W=T|S=F, R=T)$

$\Rightarrow 0.5 * 0.5 * 0.2 * 0.8$

$\Rightarrow 0.04.$

Case (iv) :-

$(C, S, R, W) = P(C=F, S=F, R=F, W=F)$

$P(C=F) * P(S=F|C=F) * P(R=F|C=F) * P(W=F|S=F, R=F)$

$\Rightarrow 0.5 * 0.5 * 0.8 * 1.0$

$\Rightarrow 0.2$