MIDTERM EXAM

Jessa Mae M. Labasan September 3, 2025

COM221ML

1. Define a Markov Decision Process (MDP). List its key components.

▷ Markov Decision Process (MDP) is an environment wherein all states are Markov. And it is fully observable which makes it best for reinforcement learning. It mainly consists of states, actions, transition probability matrix, reward function and discount factor.

2. What does it mean for a process to satisfy the Markov property?

▷ For a process to successfully satisfy the Markov property, the current state must be independent of the past states therefore, the history may be thrown away once the state is defined or known.

3. Explain the difference between a policy and a value function.

▷ A policy is the probability of what action to take in a state. This indicates or guides the agent's course of decision to achieve the highest reward possible. On the other hand, a value function measures the expected return. It is meant to assist agents in understanding the long-term value of state s

4. What is the role of the discount factor (γ) in an MDP?

▷ In an MDP, the role of the discount factor is to assist the agent to know the present value of future rewards. Simply put, we can understand and balance how much importance would be given to immediate reward and future reward.

- What happens when $\gamma = 0$ and when $\gamma \to 1$?

▷ The discount factor values are within 0 to 1. If the discount factor is equal to zero, it would encourage us to consider long-term rewards as it result to short-term rewards while if it is $\gamma \to 1$ which shows that it is closer to 1, it encourages the agent to focus more on long-term rewards.

5. Two-State weather MDP

	Go Out	Stay Inside
☀	+2	0
☁	+1	+3

s'

		☀	☁
S	☀	0.0	1.0
	☁	1.0	0.0

$\gamma = 0.5$

probability = 0.5

(a) Compute the average expected reward for Sunny

$r_\pi = 0.5 \times (2) + 0.5 \times (0) = 1 + 0 = 1$

(b) Compute the average expected reward for Rainy

$r_\pi = 0.5 \times (1) + 0.5 \times (3) = 0.5 + 1.5 = 2$

$\left.\right\} r_\pi = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

(c&d) Using the bellman expectation equation, solve for v_π (sunny) & (rainy)

Sunny:

$P_\pi(1) = 0.5 \times 0.0 = 0$

$\underline{P_\pi(1)} = 0.5 \times 1.0 = 0.5$

$v_1 = 1 + 0.5(0_{v_1} + 0.5_{v_2})$

$v_1 = 1 + 0_{v_1} + 0.25_{v_2}$

$v_1 - 0.25_{v_2} = 1$

Rainy:

$P_\pi(2) = 0.5 \times 1.0 = 0.5$

$\underline{P_\pi(2)} = 0.5 \times 0.0 = 0$

$v_2 = 2 + 0.5(0.5_{v_1} + 0_{v_2})$

$v_2 = 2 +$ ~~0.5~~ $0.25_{v_1} + 0v_2$

$-0.25_{v_1} + v_2 = 2$

$P_\pi = \begin{bmatrix} 0 & 0.5 \\ 0.5 & 0 \end{bmatrix}$

v_π(rainy):

$v_1 - 0.25_{v_2} = 1$

$\underline{v_1} = \underline{1 + 0.25v_2}$

$\underline{\underline{v_1 = 1 + 0.25_{v_2}}}$

$-0.25v_1 + v_2 = 2$

$-0.25(1 + 0.25_{v_2}) + v_2 = 2$

$-0.25 + (-0.0625_{v_2}) + v_2 = 2$

$-0.25 + 0.9375_{v_2} = 2$

$0.9375_{v_2} = 2 - 0.25$

// v_π(rainy) $= \dfrac{2 - 0.25}{0.9375} = \dfrac{1.75}{0.9375} = 1.87$

v_π(sunny):

$v_1 = 1 + 0.25_{v_2}$

$v_1 = 1 + 0.25(1.87) = 1 + 0.4675 = 1.4675$

// v_π(sunny) $= 1.47$

c. Consider the following gridworld MDP:

A	B	C
D	E	F
G	H	I

State E = wall

State I = terminal

The rest are non-terminal states

Entering I = 0 reward

Non-terminal = -1 reward

Actions = up/down/left/right

$\gamma = 1$

$p = 0.25$