RL - Exam

John Efren V. Gannaban

com-221

9/3/25

① Define a Markov Decision Process (MDP). List its key components (5 pts)

- A Markov Decision Process formally describes an environment where the current state completely characterizes the process.

Key components are:

a. state-value function
b. action-value function
c. optimal policy
d. optimal state-value function
e. optimal action-value function

② What does it mean for a process to satisfy the Markov property? (5pts)

- A "markov" process is defined as memoryless random process; sequence of random states. A process satisfies the Markov property when its state/s is able to capture all relevant information from its history up to current state; as the current state is enough of a statistic alone to determine the future.

③ Explain the difference between a policy and a value function (5pts)

Both are used to determine an agent's action, however a policy defines the behavior of an agent given the current state. The value function on the other hand, gives the long term value of a state (expected return starting from the state s)

④ What is the role of the discount factor (γ) in an MDP (5 pts)

- The discount factor appears in the MDP as a limiting factor to rewards given in order to avoid infinite returns and represent the uncertainty about the future.

when γ (discount) = 0, it leads to "myopic" evaluation.

when γ (discount) = 1, it leads to "far-sighted" evaluation.

(5) two-state weather MDP (15 pts) (sunny, Rainy) states

(Go out, stay inside) Actions

Reward function

$R(\text{Sunny}, G) = +2$

$R(\text{Sunny}, I) = 0$

$R(\text{Rainy}, G) = +1$

$R(\text{Rainy}, I) = +3$

Matrix

$$\begin{bmatrix} +2 & 0 \\ +1 & +3 \end{bmatrix}$$

Discount $\gamma = 0.5$

T-Matrix

$$\begin{bmatrix} 0.0 & 1.0 \\ 1.0 & 0.0 \end{bmatrix} = P\pi$$

(a) Avg e-reward for sunny (2 pts)

sunny, $r\pi = 0.5(2) + 0.5(0) = \underline{\underline{1}}$

$r\pi = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

(b) Avg e-reward for cloudy (2 pts)

cloudy, $r\pi = 0.5(1) + 0.5(3) = \underline{\underline{2}}$

(c) using bellman eq., solve for $v\pi$ (sunny)

$v_1 = 1 + 0.5(0.0v_1 + 1.0v_2)$

$v_1 = 1 + 0.0v_1 + 0.5v_2$

$v_1 - 0.0v_1 - 0.5v_2 = 1$

$\boxed{1v_1 - 0.5v_2 = 1}$

(d) using the bellman expectation eq. $v\pi$ (Rainy)

$v_2 = 2 + 0.5(1.0v_1 + 0.0v_2)$

$v_2 = 2 + 0.5v_1 + 0.0v_2$

$v_2 - 0.0v_2 - 0.5v_1 = 2$

$\boxed{-0.5v_1 + 1v_2 = 2}$

Bellman expectation eq. (Rainy)

$1v_1 - 0.5v_2 = 1$

$\frac{1v_1 = 1 + 0.5v_2}{1}$

$\boxed{v_1 = 1 + 0.5v_2}$

$-0.5(1 + 0.5v_2) + 1v_2 = 2$

$-0.5 - 0.25v_2 + 1v_2 = 2$

$-0.25v_2 + 1v_2 = 2 + 0.5$

$\frac{0.75v_2 = 2.5}{0.75}$

$\boxed{v_2 = 3.33}$

Bellman eq (Sunny)

$1v_1 - 0.5v_2 = 1$

$1v_1 - 0.5(3.33) = 1$

$1v_1 - 1.665 = 1$

$1v_1 = 1 + 1.665$

$1v_1 = 2.665$

$\boxed{v_1 = 2.665}$

(6.) consider the ft. gridworld MDP: (15 pts)

3x3 A - I

A B C

D ~~E~~ F

G H (1)

ABCDFGH = non-terminal S

→ E - wall

1 - terminal state

	vk(s)	vk+1(s)
A	0	-1
B	0	-1
C	0	-1
D	0	-1
F	0	-1
G	0	-1
H	0	-1

(a)

$$vk+1(A) = \frac{1}{4}[\overset{\leftarrow A}{(-1+0)} + \overset{\rightarrow B}{(-1+0)} + \overset{\uparrow A}{(-1+0)} + \overset{\downarrow D}{(-1+0)}]$$

$$-4/4 = \boxed{-1}$$

$$vk+1(B) = \frac{1}{4}[\overset{\leftarrow A}{(-1+0)} + \overset{\rightarrow C}{(-1+0)} + \overset{\uparrow B}{(-1+0)} + \overset{\downarrow B}{(-1+0)}$$

$$-4/4 = \boxed{-1}$$

$$vk+1(C) = \frac{1}{4}[\overset{\leftarrow B}{(-1+0)} + \overset{\rightarrow C}{(-1+0)} + \overset{\uparrow C}{(-1+0)} + \overset{\downarrow F}{(-1+0)}$$

$$-4/4 = \boxed{-1}$$

$$vk+1(D) = \frac{1}{4}[\overset{\leftarrow D}{(-1+0)} + \overset{D \rightarrow}{(-1+0)} + \overset{\uparrow A}{(-1+0)} + \overset{\downarrow G}{(-1+0)}$$

$$-4/4 = \boxed{-1}$$

$$vk+1(F) = \frac{1}{4}[\overset{\leftarrow F}{(-1+0)} + \overset{\rightarrow F}{(-1+0)} + \overset{\uparrow C}{(-1+0)} + \overset{\downarrow I}{(0+0)}$$

$$-4/4 = \boxed{-1}$$

$$vk+1(G) = \frac{1}{4}[\overset{\leftarrow G}{(-1+0)} + \overset{\rightarrow H}{(-1+0)} + \overset{\uparrow D}{(-1+0)} + \overset{\downarrow G}{(-1+)}$$

$$-4/4 = \boxed{-1}$$

$$vk+1(H) = \frac{1}{4}[\overset{\leftarrow G}{(-1+0)} + \overset{I \rightarrow}{(\cancel{1}+0)} + \overset{\uparrow H}{(-1+0)} + \overset{\downarrow H}{(-1+0)}]$$

$$-4/4 = \boxed{-1}$$

$q(A, LEFT) = -1 + (-1) = \boxed{-2}$

$q(A, RIGHT) = -1 + -1 = \boxed{-2}$

$q(A, UP) = -1 + -1 = \boxed{-2}$

$q(A, DOWN) = -1 + -1 = \boxed{-2}$

$q(B, LEFT) = -1 + -1 = \boxed{-2}$

$q(B, RIGHT) = -1 + -1 = \boxed{-2}$

$q(B, UP) = -1 + -1 = \boxed{-2}$

$q(B, DOWN) = -1 + -1 = \boxed{-2}$

$q(C, LEFT) = -1 + -1 = \boxed{-2}$

$q(C, RIGHT) = -1 + -1 = \boxed{-2}$

$q(C, UP) = -1 + -1 = \boxed{-2}$

$q(C, DOWN) = -1 + -1 = \boxed{-2}$

$q(D, LEFT) = -1 + -1 = \boxed{-2}$

$q(D, RIGHT) = -1 + -1 = \boxed{-2}$

$q(D, UP) = -1 + -1 = \boxed{-2}$

$q(D, DOWN) = -1 + -1 = \boxed{-2}$

$q(F, LEFT) = -1 + -1 = \boxed{-2}$

$q(F, RIGHT) = -1 + -1 = \boxed{-2}$

$q(F, UP) = -1 + -1 = \boxed{-2}$

$q(F, DOWN) = -1 + 0 = \boxed{-1}$

$q(G, LEFT) = -1 + -1 = \boxed{-2}$

$q(G, RIGHT) = -1 + -1 = \boxed{-2}$

$q(G, UP) = -1 + -1 = \boxed{-2}$

$q(G, DOWN) = -1 + -1 = \boxed{-2}$

$q(H, LEFT) = -1 + -1 = \boxed{-2}$

$q(H, RIGHT) = -1 + 0 = \boxed{-1}$

$q(H, UP) = -1 + -1 = \boxed{-2}$

$q(H, DOWN) = -1 + -1 = \boxed{-2}$

v_{k+1}

←↑→↓	←↑→↓	←↑→↓
←↑→↓	×	↓
←↑→↓	→	0

(A) = $\frac{1}{4}$

(B = $\frac{1}{4}$[

predic

$v_{k+2}(A) = \frac{1}{4}[(-1+-1) + (-1+-1) + (-1+-1) + (-1+-1)]$

$= \frac{-8}{4} = \boxed{-2}$

$v_{k+2}(B) = \frac{1}{4}[(-1+-1) + (-1+-1) + (-1+-1) + (-1+-1)]$

$= \frac{-8}{4} = \boxed{-2}$

$v_{k+2}(C) = \frac{1}{4}[(-1+-1) + (-1+-1) + (-1+-1) + (-1+-1)] = -8/4 = \boxed{-2}$

$v_{k+2}(D) = \frac{1}{4}[(-1+-1)] + (-1+-1) + (-1+-1) + (-1+-1)] = -8/4 = \boxed{-2}$

$v_{k+2}(F) = \frac{1}{4}[(-1+-1) + (-1+-1) + (-1+-1) + (-1+0)] = -7/4 = \boxed{-1.75}$

$v_{k+2}(G) = \frac{1}{4}[(-1+-1) + (-1+-1) + (-1+-1) + (-1+-1)] = -8/4 = \boxed{-2}$

[redacted] $(H) = \frac{1}{4}[(-1+-1) + (-1+0) + (-1+-1) + (-1+-1)] = -7/4 = \boxed{-1.75}$

	$v_k s_0$	$v_k s+1$	v_{k+2}
A	0	-1	-2
B	0	-1	-2
C	0	-1	-2
D	0	-1	-2
F	0	-1	-1.75
G	0	-1	-2
H	0	-1	-1.75

$q(A, LEFT) = -1 + -2 = \boxed{-3}$

$q(A, RIGHT) = -1 + -2 = \boxed{-3}$

$q(A, UP) = -1 + -2 = \boxed{-3}$

$q(A, DOWN) = -1 + -2 = \boxed{-3}$

$q(B, LEFT) = -1 + -2 = \boxed{-3}$

$q(B, RIGHT) = -1 + -2 = \boxed{-3}$

$q(B, UP) = -1 + -2 = \boxed{-3}$

$q(B, DOWN) = -1 + -2 = \boxed{-3}$

$q(C, LEFT) = -1 + -2 = \boxed{-3}$

$q(C, RIGHT) = -1 + -2 = \boxed{-3}$

$q(C, UP) = -1 + -2 = \boxed{-3}$

$q(C, DOWN) = -1 + -1.75$
$= \boxed{-2.75}$

$q(D, LEFT) = -1 + -2 = \boxed{-3}$

$q(D, RIGHT) = -1 + -2 = \boxed{-3}$

$q(D, UP) = -1 + -2 = \boxed{-3}$

$q(D, DOWN) = -1 + -2 = \boxed{-3}$

$q(G, LEFT) = -1 + -2 = \boxed{-3}$

$q(G, RIGHT) = -1 + -1.75 = \boxed{-2.75}$

v_{k+2}

←→	←↑→↓	↓
←↑→↓	~~×~~	↓
→	→	0

	VKS	VK+1	VK+2	VK+3	VK+4
A	0	-1	-2	-3	
B	0	-1	-2	-3	
C	0	-1	-2	-2.94	
D	0	-1	-2	-3	
F	0	-1	-1.75	-2.375	
G	0	-1	-2	-2.94	
H	0	-1	-1.75	-2.375	

A B C
D X F
G H I

VK+2

\+ \+ ↓
\+ ✗ ↓
→ → 0

$$\text{VK+3}(A) = \frac{1}{4}\big[\underbrace{(-1+-2)}_{\leftarrow A} + \underbrace{(-1+-2)}_{\rightarrow B} + \underbrace{(-1+-2)}_{\uparrow A} + \underbrace{(-1+-2)}_{\downarrow D}\big]$$

$$12/4 = \boxed{-3}$$

$$\text{VK+3}(B) = \frac{1}{4}\big[(-1+-2)+(-1+-2)+(-1+-2)+(-1+-2)\big] = -12/4 = \boxed{-3}$$

$$\text{VK+3}(C) = \frac{1}{4}\big[(-1+-2)+(-1+-2)+(-1+-2)+(-1+-1.75)\big] = \boxed{-2.94}$$

$$\text{VK+3}(D) = \frac{1}{4}\big[(-1+-2)+(-1+-2)+(-1+-2)+(-1+-2)\big] = -12/4 = \boxed{-3}$$

$$\text{VK+3}(F) = \frac{1}{4}\big[(-1+-1.75)+(-1+-1.75)+(-1+-2)+(-1+0)\big] = 9.5/4 = \boxed{-2.375}$$

$$\text{VK+3}(G) = \frac{1}{4}\big[(-1+-2)+(-1+-1.75)+(-1+-2)+(-1+-2)\big] = 11.75/4 = \boxed{-2.94}$$

$$\text{VK+3}(H) = \frac{1}{4}\big[(-1+-2)+(-1+0)+(-1+1.75)+(-1+-1.75)\big] = 9.5/4 = \boxed{-2.375}$$

$q(A, \text{LEFT}) = -1 + -3 = \boxed{-4}$ $\quad q(B, \text{LEFT}) = -1 + -3 = \boxed{-4}$

$q(A, \text{RIGHT}) = -1 + -3 = \boxed{-4}$ $\quad q(B, R) = -1 + -2.94 = \boxed{3.94}$

$q(A, \text{UP}) = -1 + -3 = \boxed{-4}$ $\quad q(B, U) = -1 + -3 = \boxed{-4}$

$q(A, \text{DOWN}) = -1 + -3 = \boxed{-4}$ $\quad q(B, D) = -1 + -3 = \boxed{-4}$

$q(C, L) = -1 + -3 = \boxed{-4}$

$q(C, R) = -1 + 2.94 = \boxed{3.94}$

$q(C, U) = -1 + 2.94 = \boxed{3.94}$

$q(C, D) = -1 + 2.375 = \boxed{3.38}$ ↓

VK+3

←↕→ ←↕→ ↓
↓ ✗ ↓
→ → 0

$q(D, L) = -1 + -3 = \boxed{-4}$

$q(D, R) = -1 + -3 = \boxed{-4}$

$q(D, U) = -1 + 2.94 = \boxed{3.94}$ ↓

$q(D, D) = -1 + 3 = \boxed{-4}$

(5 pts)

$v_{k+4}(A) = \frac{1}{4}[(-1+-3)+(-1+-3)+(-1+-3)+(-1+-3)] = 16/4 = \boxed{-4}$

$v_{k+4}(B = \frac{1}{4}[(-1+3)+(-1+2.94)+(-1+-3)+(-1+-3)]$ 15.94/4 $\boxed{-3.98}$

prediction k=0

b optimal policy

$\searrow$	$\downarrow\rightarrow$	$\rightarrow$	$\downarrow$
	$\downarrow$	x	$\downarrow$
	$\rightarrow$	$\rightarrow$	0