

PL - Exam
John Efen V. Gannan

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

② What does it mean for a process to satisfy the Markov property? (3pts)

- 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 (3pts)

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 (3 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

$$\begin{bmatrix} 0.5 & 1.0 \\ 1.0 & 0.0 \end{bmatrix} = P$$

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

$$\text{sunny, } v_G = 0.5(2) + 0.5(0) = \underline{1}$$

(b) Avg. e-reward for rainy (2 pts)

$$\text{rainy, } v_G = 0.5(1) + 0.5(3) = \underline{2}$$

$$v_G = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

(c) using Bellman eq., solve for v_G (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_G (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)

$$1v_1 - 0.5v_2 = 1$$

$$1v_1 = 1 + 0.5v_2$$

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

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

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

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

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

$$0.75v_2 = 2.5$$

$$0.75$$

$$\boxed{v_2 = 3.33}$$

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

ABCDEFH = non-terminal S

D ~~E~~ F

E = wall

G H \textcircled{I}

I = terminal state

$V_k(s)$ $V_{k+1}(s)$

A	0	-1
B	0	-1
C	0	-1
D	0	-1
E	0	-1
F	0	-1
G	0	-1
H	0	-1

(a)

$\leftarrow A \rightarrow B \uparrow A \downarrow D$

$$V_{k+1}(A) = \frac{1}{4} [(-1+0) + (-1+0) + (-1+0) + (-1+0)]$$

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

$$V_{k+1}(B) = \frac{1}{4} [(-1+0) + (-1+0) + (-1+0) + (-1+0)]$$

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

$$V_{k+1}(C) = \frac{1}{4} [(-1+0) + (-1+0) + (-1+0) + (-1+0)]$$

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

$$V_{k+1}(D) = \frac{1}{4} [(-1+0) + (-1+0) + (-1+0) + (-1+0)]$$

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

$$V_{k+1}(F) = \frac{1}{4} [(-1+0) + (-1+0) + (-1+0) + (-1+0)]$$

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

$$V_{k+1}(G) = \frac{1}{4} [(-1+0) + (-1+0) + (-1+0) + (-1+0)]$$

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

$$V_{k+1}(H) = \frac{1}{4} [(-1+0) + (-1+0) + (-1+0) + (-1+0)]$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$q(I, \text{up}) = -1 + -2 = \boxed{-3}$$

$$\rightarrow \rightarrow b$$

nd a value
action
current state
value

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

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

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

$$q(F, \text{DOWN}) = -1 + 0 + 1 = -2$$

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

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

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

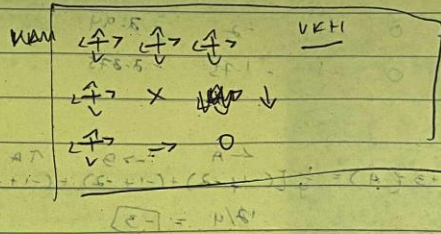
$$q(G, \text{DOWN}) = -1 + 1 = -2$$

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

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

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

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



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

$$V(B) = \frac{1}{4}$$

$$V_{k+2}(A) = \frac{1}{4} [(-1+1) + (-1+1) + (-1+1) + (-1+1)] = \frac{0}{4} = 0$$

$$V_{k+2}(B) = \frac{1}{4} [(-1+1) + (-1+1) + (-1+1) + (-1+1)] = \frac{0}{4} = 0$$

$$V_{k+2}(C) = \frac{1}{4} [(-1+1) + (-1+1) + (-1+1) + (-1+1)] = \frac{0}{4} = 0$$

$$V_{k+2}(D) = \frac{1}{4} [(-1+1) + (-1+1) + (-1+1) + (-1+1)] = \frac{0}{4} = 0$$

$$V_{k+2}(E) = \frac{1}{4} [(-1+1) + (-1+1) + (-1+1) + (-1+1)] = \frac{0}{4} = 0$$

$$V_{k+2}(F) = \frac{1}{4} [(-1+1) + (-1+1) + (-1+1) + (-1+1)] = \frac{0}{4} = 0$$

$$V_{k+2}(G) = \frac{1}{4} [(-1+1) + (-1+1) + (-1+1) + (-1+1)] = \frac{0}{4} = 0$$

$$V_{k+2}(H) = \frac{1}{4} [(-1+1) + (-1+1) + (-1+1) + (-1+1)] = \frac{0}{4} = 0$$

$$V_{k+2}(I) = \frac{1}{4} [(-1+1) + (-1+1) + (-1+1) + (-1+1)] = \frac{0}{4} = 0$$

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

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

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

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

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

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

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

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

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

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

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

$$q(C, \text{DOWN}) = -1 + -2 = -3$$

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

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

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

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

$$q(E, \text{LEFT}) = -1 + -2 = -3$$

$$q(E, \text{RIGHT}) = -1 + -2 = -3$$

$$q(E, \text{UP}) = -1 + -2 = -3$$

$$q(E, \text{DOWN}) = -1 + -2 = -3$$

$$q(F, \text{LEFT}) = -1 + -2 = -3$$

$$q(F, \text{RIGHT}) = -1 + -2 = -3$$

$$q(F, \text{UP}) = -1 + -2 = -3$$

$$q(F, \text{DOWN}) = -1 + -2 = -3$$

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

$$q(G, \text{RIGHT}) = -1 + -2 = -3$$

$$q(G, \text{UP}) = -1 + -2 = -3$$

$$q(G, \text{DOWN}) = -1 + -2 = -3$$

$$q(H, \text{LEFT}) = -1 + -2 = -3$$

$$q(H, \text{RIGHT}) = -1 + -2 = -3$$

$$q(H, \text{UP}) = -1 + -2 = -3$$

$$q(H, \text{DOWN}) = -1 + -2 = -3$$

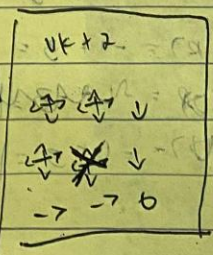
$$q(I, \text{LEFT}) = -1 + -2 = -3$$

$$q(I, \text{RIGHT}) = -1 + -2 = -3$$

$$q(I, \text{UP}) = -1 + -2 = -3$$

$$q(I, \text{DOWN}) = -1 + -2 = -3$$

	V_{k+0}	V_{k+1}	V_{k+2}
A	0	-1	-2
B	0	-1	-2
C	0	-1	-2
D	0	-1	-2
E	0	-1	-2
F	0	-1	-2
G	0	-1	-2
H	0	-1	-2
I	0	-1	-2



	DKS	VE+1	VE+2	VE+3	VE+4
A	0	-1	-2	-3	$=(1731, 0)$
B	0	-1	-2	-3	$=(1740, 0)$
C	0	-1	-2	-2.94	$=(90, 17)$
D	0	-1	-2	-3	$=(1740, 17)$
F	0	-1	-1.75	-2.375	
G	0	-1.1	-2	-2.94	
H	0	-1	-1.75	-2.375	

$$VE+3(A) = \frac{1}{4} [(-1+2) + (-1+2) + (-1+2) + (-1+2)] = 12/4 = \boxed{-3}$$

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

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

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

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

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

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

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

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

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

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

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

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

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

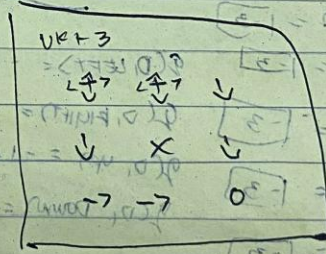
$$q(C, D) = -1 + 2.375 = \boxed{3.38}$$

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

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

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

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



(5 pts)

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

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

prediction 100 (b) optimal policy

→ ↘ → ↓
↓ x ↓
→ → 0