NO.:
DATE: 09/03/2025

Baldeo, John Vernon B.
COM221 : Midterm Exam

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

- is a Markov reward process with decisions. It is an environment in which all states are Markov.
- A Markov Decision Process is a tuple ⟨S, A, P, R, γ⟩
  - S is a finite set of states
  - A is a finite set of actions
  - P is a state transition probabilty matrix.
  - R is a reward function
  - γ is a discount factor

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

- A process must have succeded in retaining all relevant information from the history.

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

- A policy is a distribution over actions given states. It defines an agent's behavior
- A value function estimates the expected future reward of taking an action in a state.
- In summary, a policy defines the agent's behavior while a value function$ examines a state and provides an optimal policy to take.

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

- It controls the importance of future rewards compared to immediate ones.

4.1) What happens when γ = 0 and y = 1?

- γ = 0 priortizes only immediate rewards
- γ = 1 gives weight on cumulative rewards.

5) Two - State weather MDP (15 pt's)

	G	I
Sunny	+2	0
Rainy	+1	+3

S \ S'	Sunny	Rainy
Sunny	0.0	1.0
Rainy	1.0	0.0

- $\gamma = 0.5$
- uniform random policy, in each state it chooses Go out or stay Inside with probability 0.5

① Compute state-wise average reward under the policy π

▷ Sunny

$\Rightarrow 0.5 \times (2) + 0.5\,(0)$

$\boxed{\pi \Rightarrow 1 \quad (a)}$ //

▷ Rainy

$\Rightarrow 0.5\,(1) + 0.5(3)$

$\boxed{\pi = 2 \quad (b)}$ //

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

② Compute the policy transition matrix

▷ Row 1 (Sunny)

$P_\pi(1,1) = 0.5\,(0) + 0.5\,(0) = 0$

$P_\pi(1,2) = 0.5\,(1) + 0.5(1) = 1$

▷ Row 2 (Rainy)

$P_\pi(2,1) = 0.5\,(1) + 0.5(1) = 1$

$P_\pi(2,2) = 0.5\,(0) + 0.5(0) = 0$

$$P_\pi = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

③ Write the Bellman expectation equations

▷ V_π (Sunny)

$= r_\pi(s) + \gamma \sum P_\pi(s,s')\, v_\pi(s')$

$v_1 = 1 + 0.5\,(0v_1 + 1v_2)$

$v_1 = 1 + 0.5v_2$

▷ V_π (Rainy)

$\Rightarrow 2 + 0.5\,(1v_1 + 0v_2)$

$v_2 = 2 + 0.5\,v_1$

④ Solve for V_π (Rainy) [V_2]

$v_2 = 2 + 0.5\,v_1$

$v_2 = 2 + 0.5\,(1 + 0.5v_2)$

$v_2 = 2 + 0.5 + 0.25v_2$

$v_2 - 0.25v_2 = 2.5$

$\boxed{v_2 = 3.\overline{3333} \quad (d)}$ //

④ Solve for V_π (Sunny) [V_1]

$v_1 = 1 + 0.5v_2$

$= 1 + 0.5(3.\overline{3333})$

$\boxed{v_1 = 2.\overline{6666} \quad (c)}$ //

① compute value functions at ~~k~~ k+1

$$V_{k+1}(A) = \frac{1}{4}\left[(-1+V(A)) + (-1+V(B)) + (-1+V(D)) + (-1+V(A))\right]$$

$$= 1/4\left[(-1+0)+(-1+0)+(-1+0)+(-1+0)\right]$$

$$V_{k+1}(A) = -1$$

$$V_{k+1}(B) = 1/4\left[(-1+V(A)) + (-1+V(B)) + (-1+V(C)) + (-1+V(B))\right] = -1$$

$$V_{k+1}(C) = 1/4\left[(-1+V(C)) + (-1+V(C)) + (-1+V(B)) + (-1+V(F))\right] = -1$$

$$V_{k+1}(D) = 1/4\left[(-1+V(D)) + (-1+V(D)) + (-1+V(G)) + (-1+V(A))\right] = -1$$

$$V_{k+1}(F) = 1/4\left[(-1+V(F)) + (-1+V(F)) + (-1+V(I)) + (-1+V(C))\right] = -1$$

$$V_{k+1}(G) = 1/4\left[(-1+V(G)) + (-1+V(H)) + (-1+V(G)) + (-1+V(D))\right] = -1$$

$$(H) = 1/4\left[(-1+V(G)) + (-1+V(I)) + (-1+V(H)) + (-1+V(H))\right] = -1$$

② Compute the action-value function and update the policy of states at k+1

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

▷ A

$q_{k+1}(A, \text{Left}) = -1 + (-1) = -2$

$(A, \text{Right}) = -1 + (-1) = -2$

$(A, \text{Up}) = (-1) + (-1) = -2$

$(A, \text{Down}) = -1 + (-1) = -2$

$\pi_{k+1}(A) = \{\text{Left, Right, Up, Down}\}$

▷ B

$q_{k+1}(B, L) = -1 + (-1) = -2$

$(B, R) = -1 + (-1) = -2$

$(B, U) = -1 + (-1) = -2$

$(B, D) = -1 + (-1) = -2$

$\pi_{k+1}(B) = \{\text{Left, Right, Up, Down}\}$

▷ C

$q_{k+1}(C, L) = -1 + (-1) = -2$

$(C, R) = -1 + (-1) = -2$

$(C, U) = -1 + (-1) = -2$

$(C, D) = -1 + (-1) = -2$

$\pi_{k+1}(C) = \{\text{Left, Right, Up, Down}\}$

NO.:
DATE:

▷ D

$q_{k+1}(D,L) = -1 + (-1) = -2$
$(D,R) = -1 + (-1) = -2$
$(D,U) = -1 + (-1) = -2$
$(D,D) = -1 + (-1) = -2$

$\pi_{k+1}(D) = \{\text{Left, Right, Up, Down}\}$

▷ G

$q_{k+1}(G,L) = -1 + (-1) = -2$
$(G,R) = -1 + (-1) = -2$
$(G,U) = -1 + (-1) = -2$
$(G,D) = -1 + (-1) = -2$

$\pi_{k+1}(G) = \{\text{Left, Right, Up, Down}\}$

▷ F

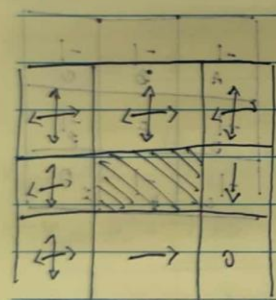
$q_{k+1}(F,L) = -1 + (-1) = -2$
$(F,R) = -1 + (-1) = -2$
$(F,U) = -1 + (-1) = -2$
$(F,D) = -1 + (0) = -1$

$\boxed{\pi_{k+1}(F) = \{\text{Down}\}}$

▷ H

$q_{k+1}(H,L) = -1 + (-1) = -2$
$(H,R) = -1 + 0 = -1$
$(H,U) = -1 + (-1) = -2$
$(H,D) = -1 + (-1) = -2$

$\boxed{\pi_{k+1}(H) = \{\text{Down Right}\}}$



	k	k+1	k+2
A	0	-1	
B	0	-1	
C	0	-1	
D	0	-1	
F	0	-1	
G	0	-1	
H	0	-1	

(1.2) Value functions at k+2

$v_{k+2}(A) = \frac{1}{4}\left[(-1 + v(A)) + (-1 + v(B)) + (-1 + v(A)) + (-1 + v(D))\right]$
$= \frac{1}{4}\left[-1 + (-1) + ((-1) + (-1)) + (-1 + (-1)) + (-1 + (-1))\right]$

$v_{k+2}(A) = -2$

$v_{k+2}(B) = \frac{1}{4}\left[(-1 + v(A)) + (-1 + v(C)) + (-1 + v(C)) + (-1 + v(D))\right]$
$= \frac{1}{4}\left[(-1 + (-1)) + (-1 + (-1)) + (-1 + (-1)) + (-1 + (-1))\right]$

$v_{k+2}(B) = -2$

VICTORY

NO.:
DATE:

$$V_{k+2}(C) = \frac{1}{4}\left[(-1+V(B)) + (-1+V(C)) + (-1+V(C)) + (-1+V(F))\right]$$
$$= \frac{1}{4}\left[(-1+(-1)) + (-1+(-1)) + (-1+(-1)) + (-1+(-1))\right]$$
$$V_{k+2}(C) = -2$$

$$V_{k+2}(D) = \frac{1}{4}\left[(-1+V(D)) + (-1+V(D)) + (-1+V(A)) + (-1+V(G))\right]$$
$$= \frac{1}{4}\left[(-1+(-1)) + (-1+(-1)) + (-1+(-1)) + (-1+(-1))\right]$$
$$V_{k+2}(D) = -2$$

$$V_{k+2}(F) = \frac{1}{4}\left[(-1+V(F)) + (-1+V(F)) + (-1+V(C)) + (-1+V(I))\right]$$
$$= \frac{1}{4}\left[(-1+(-1)) + (-1+(-1)) + (-1+(-1)) + (-1+0)\right]$$
$$V_{k+2}(F) = -1.75$$

$$V_{k+2}(G) = \frac{1}{4}\left[(-1+V(G)) + (-1+V(H)) + (-1+V(D)) + (-1+V(G))\right]$$
$$= \frac{1}{4}\left[(-1+(-1)) + (-1+(-1)) + (-1+(-1)) + (-1+(-1))\right]$$
$$V_{k+2}(G) = -2$$

$$V_{k+2}(H) = \frac{1}{4}\left[(-1+V(G)) + (-1+V(I)) + (-1+V(H)) + (-1+V(H))\right]$$
$$V_{k+2}(H) = -1.75$$

	k	k+1	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

-2	-2	-2
-2		-1.75
-2	-1.75	0

2.2 Compute the action-value function at k+2

▷ A

q_{k+2} (A, L) = -1 + (-2) = -3
(A, R) = -1 + (-2) = -3
(A, U) = -1 + (-2) = -3
(A, D) = -1 + (-2) = -3

π_{k+2} (A) = {Left, R, U, D}

▷ B

q_{k+2} (B, L) = -1 + (-2) = -3
(B, R) = -1 + (-2) = -3
(B, U) = -1 + (-2) = -3
(B, D) = -1 + (-2) = -3

π_{k+2} (B) = {L, R, U, D}

▷ C

q_{k+2} (C, L) = -1 + (-2) = -3
(C, R) = -1 + (-2) = -3
(C, U) = -1 + (-2) = -3
(C, D) = -1 + (-1.75) = -2.75

π_{k+2} (C) = {D}

▷ D

q_{k+2} (D, L) = -1 + (-2) = -3
(D, R) = -1 + (-2) = -3
(D, U) = -1 + (-2) = -3
(D, D) = -1 + (-2) = -3

π_{k+2} (D) = {L, R, U, D}

▷ F

π_{k+2} (F) = {D}

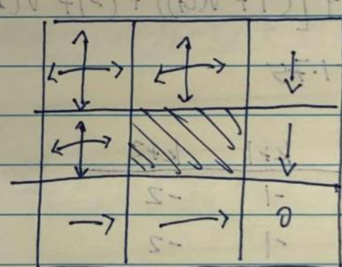
▷ G

q_{k+2} (G, L) = -1 + -2 = -3
(G, R) = -1 + ~~-2~~ -1.75 = -2.75
(G, U) = -1 + -2 = -3
(G, D) = -1 + -2 = -3

π_{k+2} (G) = ~~{L, R, U, D}~~ {Right}

▷ H

π_{k+2} (H) = {D}



NO.:
DATE:

(1.3) Value functions at ~~k+2~~ k+3

$V_{k+3}(A) = 1/4 [(-1 + (-2)) + (-1 + -2) + (-1 + -2) + (-1 + -2)]$

$V_{k+3} = -3$

$V_{k+3}(B) = 1/4 [(-1 + -2) + (-1 + -2) + (-1 + -2) + (-1 + -2)]$

$V_{k+3} = -3$

$V_{k+3}(C) = 1/4 [(-1 + -2) + (-1 + -2) + (-1 + -2) + (-1 + -1.75)]$

$V_{k+3}(C) = -2.9375$

$V_{k+3}(D) = 1/4 [(-1 + -2) + (-1 + -2) + (-1 + -2) + (-1 + -2)]$

$V_{k+3}(D) = -3$

$V_{k+3}(F) = 1/4 [(-1 + -1.75) + (-1 + -2) + (-1 + -2) + (-1 + 0)]$

$V_{k+3}(F) = -2.375$

$V_{k+3}(G) = 1/4 [(-1 + (-2)) + (-1 + \overset{-1.75}{\cancel{-2}}) + (-1 + -2) + (-1 + -2)]$

$V_{k+3}(G) = -2.9375$

$V_{k+3}(H) = 1/4 [(-1 + -2) + (-1 + -1.75) + (-1 + -1.75) + (-1 + 0)]$

$V_{k+3}(H) = -2.375$

(2.3) ~~Values for~~ Action-value function at k+3

▷ A

$q_{k+3}(A, L) = -1 + (-3)$
$(A, R) = -1 + (-3)$
$(A, U) = -1 + (-3)$
$(A, D) = -1 + (-3)$

$\pi_{k+3}(A) = \{L, R, U, D\}$

▷ B

q_{k+3}

NO.:
DATE:

1.3) Value functions at ~~k+2~~ k+3

$V_{k+3}(A) = 1/4\left[(-1+(-2)) + (-1+-2) + (-1+-2) + (-1+-2)\right]$

$V_{k+3} = -3$

$V_{k+3}(B) = 1/4\left[(-1+-2) + (-1+-2) + (-1+-2) + (-1+-2)\right]$

$V_{k+3} = -3$

$V_{k+3}(C) = 1/4\left[(-1+-2) + (-1+-2) + (-1+-2) + (-1+-1.75)\right]$

$V_{k+3}(C) = -2.9375$

$V_{k+3}(D) = 1/4\left[(-1+-2) + (-1+-2) + (-1+-2) + (-1+-2)\right]$

$V_{k+3}(D) = -3$

$V_{k+3}(F) = 1/4\left[(-1+-1.75) + (-1+-2) + (-1+-2) + (-1+0)\right]$

$V_{k+3}(F) = -2.375$

$V_{k+3}(G) = 1/4\left[(-1+(-2)) + (-1+\overset{-1.75}{\cancel{-2}}) + (-1+-2) + (-1+-2)\right]$

$V_{k+3}(G) = -2.9375$

$V_{k+3}(H) = 1/4\left[(-1+-2) + (-1+-1.75) + (-1+-1.75) + (-1+0)\right]$

$V_{k+3}(H) = -2.375$

2.3) ~~values for~~ Action-value function at k+3

▷ A

$q_{k+3}(A, L) = -1+(-3)$

$(A, R) = -1+(-3)$

$(A, U) = -1+(-3)$

$(A, D) = -1+(-3)$

$\pi_{k+3}(A) = \{L, R, U, D\}$

▷ B

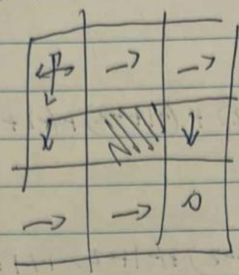
$q_{k+3}(B) = \cancel{\{L, R, U, D\}}\ \{R\}$

▷ C

$\pi_{k+3}\ C = \{D\}$

▷ D

$\pi_{k+3}(D) = \{Down\}$



▷ F

π {Down}

▷ G {Right}

▷ H {Right}

1.4

Final

NO.:
DATE:

a) compute the average expected reward for Sunny : $\boxed{r(\text{sunny}) = 1}$ //

b) compute the average expected reward for Rainy : $\boxed{r(\text{rainy}) = 2}$ //

c) Using the bellman expectation equation, solve for $V_\pi(\text{sunny})$

▷ $\boxed{V_\pi(\text{Sunny}) = 2.6666}$ //

d. $V_\pi(\text{Rainy})$:

▷ $\boxed{V_\pi(\text{Rainy}) = 3.3333}$ //

6) Consider the following gridworld MDP

A	B	C
D	E	F
G	H	[I]

- State E is a wall
- State I is the terminal state
- Entering I yields a reward of 0 and the episode ends
- Discount factor : $\gamma = 1$
- URP : 0.25
- Reward is -1 per step for every transition into a non-terminal

a) Using dynamic programming, compute the optimal state-value function $V_*(s)$ for all non-terminal states.

b) Find the optimal policy $\pi_*(s)$ for all non-terminal states.

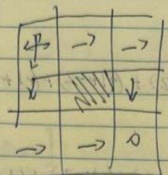
a) Initial :

0	0	0
0		0
0	0	0

	$V_k(s)$	$V_{k+1}(s)$	$V_{k+2}(s)$	k+3	k+4
A	0	-1	-2	-3	-4
B	0	-1	-2	-3	-3.984
C	0	-1	-2	-2.9375	-3.81
D	0	-1	-2	-3	-3.98
E	0				
F	0	-1	-1.75	-2.375	-2.9218
G	0	-1	-2	-2.9375	-3.8125
H	0	-1	-1.75	-2.375	-2.921
I	0				

▷ D

$\pi_{k+3}(D) = \{Down\}$



▷ F

π {Down}

▷ G {Right}

▷ H {Right}

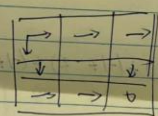
(1.4)

(A L) = -6

(R) = 4.981

(U)

(D) = 4.98



$\pi(A) = \{Down, Left\}$

Final