LAB #11

Markov Decision Process

(10 pts)

Due by Nov 6th Mon 4pm

This is continuation of Lab #10

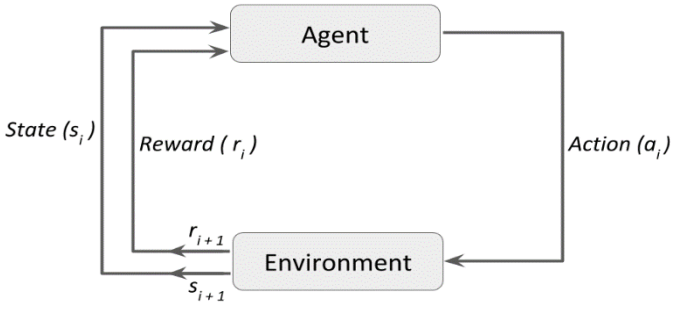
We will continue to explore the RL concept to better understand its core processes.

You will work with the same partner from last lab. Per your **six (6)** different environments, apply the Bellman equation and provide a visual illustration using fictitious values per state (refer to the Dr.Choi's lecture on Bellman's equation).

You and your partner will need to decide what discount rate based on what logic, and calculate the V (Q) value (refer to the Dr.Choi's lecture on Bellman's equation).

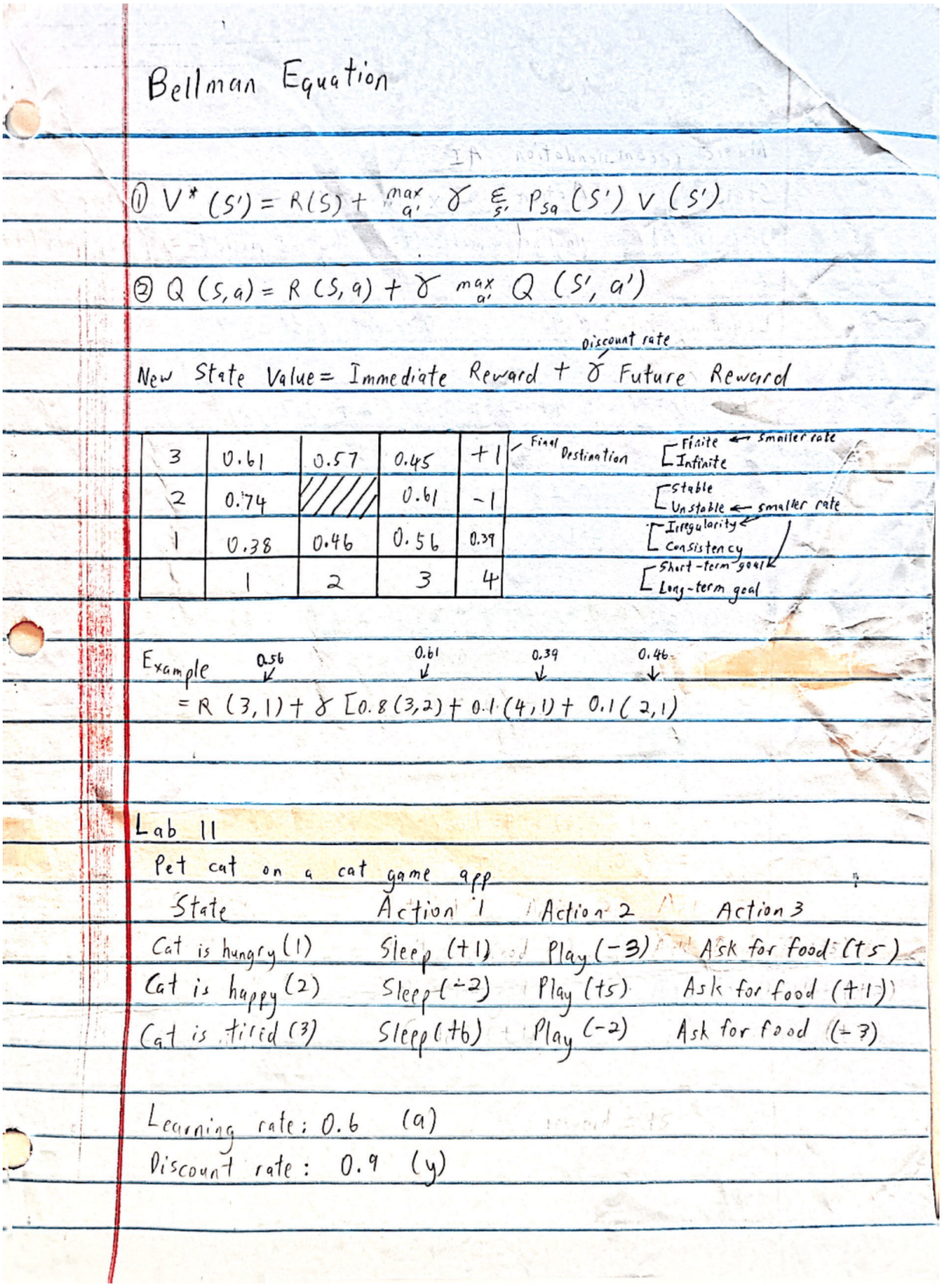
Provide your actual equations and calculations for each environment which will yield the final Q table and direction of the agent.

Complete your work using MS Word or PPT and submit via Canvas.



Agent:	A hypothetical entity which performs actions in an environment to gain some reward.
Action (a):	All the possible moves that the agent can take.
Environment (e):	A scenario the agent has to face.
State (s):	Current situation returned by the environment.
Reward (R):	An immediate return sent back from the environment to evaluate the last action by the agent.

Name: Zhen Rui Ng, Vannarith Om



Bellman Equation

① $V^*(s') = R(s) + \max_{a'} \gamma \sum_{s'} P_{sa}(s') V(s')$

② $Q(s,a) = R(s,a) + \gamma \max_{a'} Q(s', a')$

New State Value = Immediate Reward + γ (discount rate) Future Reward

3	0.61	0.57	0.45	+1	← Final Destination
2	0.74	////	0.61	-1	
1	0.38	0.46	0.56	0.39	
	1	2	3	4	

- Finite ← smaller rate
- Infinite
- Stable
- Unstable ← smaller rate
- Irregularity
- Consistency
- Short-term goal
- Long-term goal

Example (0.56, 0.61, 0.39, 0.46)

$= R(3,1) + \gamma [0.8(3,2) + 0.1(4,1) + 0.1(2,1)$

Lab 11

Pet cat on a cat game app

State	Action 1	Action 2	Action 3
Cat is hungry (1)	Sleep (+1)	Play (-3)	Ask for food (+5)
Cat is happy (2)	Sleep (-2)	Play (+5)	Ask for food (+1)
Cat is tired (3)	Sleep (+6)	Play (-2)	Ask for food (-3)

Learning rate: 0.6 (a)

Discount rate: 0.9 (y)

For state a: Cat is hungry and ask for food

$Q(s,a) = R \cdot (s,a) + \gamma \max_{a'} Q(s', a')$

$Q(1,3) = (1-0.6) \cdot 5 + 0.6 \cdot (5 + 0.9 \cdot 6)$

$= 0.4 \cdot 5 + 0.6 \cdot (10.4)$

$= 2 + 0.6 \cdot 10.4$

$= 8.24$

For state a: Cat is hungry and play

$Q(1,2) = (1-0.6) \cdot -3 + 0.6 \cdot (-3 + 0.9 \cdot 6)$

$= 0.4 \cdot -3 + 0.6 \cdot (2.4)$

$= -2.6 + 1.44$

$= -1.16$

For stat a: Cat is hungry and sleep

$Q(1,1) = (1-0.6) \cdot 1 + 0.6 \cdot (1 + 0.9 \cdot 6)$

$= 0.4 \cdot 1 + 0.6 \cdot (6.4)$

$= 0.4 + 3.84$

$= 1.536$

Q (1,3) has a higher Q-value, so the agent is more likely to choose that action to maximize reward

Online multiplayer game

State	Action 1	Action 2	Action 3
Player move around	kill enemy (+3)	Do nothing (-2)	Help teammates (+2)

Learning rate : 0.5 (a) Discount rate = 0.8

$$Q(1,1) = (1-0.5)\cdot 3 + 0.5\cdot(3 + 0.8\cdot 5)$$
$$= 0.5\cdot 3 + 0.5\cdot(7)$$
$$= 1.5 + 3.5$$
$$= 5$$

$$Q(1,2) = (1-0.5)\cdot -2 + 0.5\cdot(-2 + 0.8\cdot 5)$$
$$= 0.5\cdot -2 + 0.5\cdot 2$$
$$= -1 + 1$$
$$= 0$$

$$Q(1,3) = (1-0.5)\cdot 2 + 0.5\cdot(2 + 0.8\cdot 5)$$
$$= 0.5\cdot 2 + 0.5\cdot(6)$$
$$= 1 + 3$$
$$= 4$$

Music recommendation AI

State	Action 1	Action 2	Action 3
User is sad	play happy music (+2)	Play sad music (-2)	play blues (+4)

Learning rate : 0.9 Discount rate : 0.7

$$Q(1,1) = (1-0.9) \cdot 2 + 0.7 \cdot (2 + 0.7 \cdot 9)$$
$$= 0.1 \cdot 2 + 0.7 \cdot 8.3$$
$$= 0.2 + 5.81$$
$$= 5.81$$

$$Q(1,2) = (1-0.9) \cdot -2 + 0.7 \cdot (-2 + 0.7 \cdot 9)$$
$$= 0.1 \cdot -2 + 0.7 \cdot 4.3$$
$$= -0.2 + 3.01$$
$$= 2.81$$

$$Q(1,3) = (1-0.9) \cdot 4 + 0.7 \cdot (4 + 0.7 \cdot 9)$$
$$= 0.1 \cdot 4 + 0.7 \cdot 10.3$$
$$= 0.4 + 7.21$$
$$= 7.61$$

Music Concert

State: Teenages deciding weather or not to clean the dishes.

Action 1: Clean (+1)

Action 2: Not Clean (-1)

Action 3: Attend the music concert (+10)

Learning rate: .6

Discount rate: .5

$$Q(1,1) = ((1-.6)(1)) + (.5)(1 + .5(.6)) = \boxed{1.05}$$

$$Q(1,2) = ((1-.6)(-1)) + (.5)(-1 + .5(.6)) = \boxed{-.75}$$

$$Q(1,3) = ((1-.6)(10)) + (.5)(10 + .5(.6)) = \boxed{9.15}$$

Arachnid Superhero

State: Spidey has to decide quickly what he is going to do..

Action 1: Save both (+10)

Action 2: Save only one (+5)

Action 3: Don't save either of them (-10)

Learning rate: .4 Discount rate: .3

$Q(1,1) = ((1-.4)(10)) + ((.3)(10 + .3(.4)) = \boxed{9.04}$

$Q(1,2) = ((1-.4)(5)) + ((.3)(5 + .3(.4)) = \boxed{4.54}$

$Q(1,3) = ((1--.4)(-10)) + ((.3)(-10 + .3(.4)) = \boxed{-8.96}$

Alex, The Lion

State: Alex is scared and about to try something in the cage

Action 1: Imitate the lion figure (+4)

Action 2: Perform a dance move (+5)

Action 3: Hide from the audience (-2)

Learning rate: .7 Discount rate: .5

$Q(1,1) = ((1-.7)(4)) + ((.5)(4 + (.5(.7)))) = \boxed{3.38}$

$Q(1,2) = ((1-.7)(5)) + ((.5)(5 + (.5(.7)))) = \boxed{4.18}$

$Q(1,3) = ((1-.7)(-2)) + ((.5)(-2 + (.5(.7)))) = \boxed{-1.43}$