Problem 2

You are running a 3 mile race. Every 10 minutes, you must decide whether to walk or run for the next 10 minutes based on your current distance from the start. You represent the milestones from the starting point, namely $0, 1, 2$ or 3 miles from the starting point, as states, and consider a Markov Decision Process with the following transition probability and rewards.

If you walk, you will advance 1 mile over the next 10 minutes. If you run, you have a 50% chance to advance 1 mile, and a 50% chance to advance 2 miles over the next 10 minutes. You want to finish the race, but running is tiring and takes effort. You will receive a reward of 10 for finishing the race (ending up in state 3). However, every time you run, you get additional negative "reward" of -1. You decide to use a Markov Decision Process with $\gamma = 0.5$ to determine what action you should take from each state.

The full table of transition probabilities and rewards is shown below:

s	a	s'	$T(s,a,s')$	$R(s,a,s')$
0	WALK	1	1.0	0
1	WALK	2	1.0	0
2	WALK	3	1.0	10
0	RUN	1	0.5	-1
0	RUN	2	0.5	-1
1	RUN	2	0.5	-1
1	RUN	3	0.5	$+9$
2	RUN	3	1.0	$+9$

2. (1)

1.0/1 point (graded)

Suppose we initialize $Q_0(s,a) = 0$ for all $s \in \{0,1,2\}$ and all $a \in \{\text{WALK}, \text{RUN}\}$. We assume that the values in state $s = 3$ are always zero for any action. Evaluate the Q-values $Q_1(s,a)$ after exactly one Q-value iteration.

a	$s=0$	$s=1$	$s=2$
WALK	a	b	c
RUN	d	e	f

Please enter the values of a, b, c, d, e, f in that order as a comma-separated list. (For example, if $a = 1, b = 2, c = 3, d = 4, e = 5,$and $f = 6$, enter `[1,2,3,4,5,6]`).

Note: Please double check your answer to this problem. Later problems will depend on your answers here.

$[a,b,c,d,e,f] =$ [0, 0, 10, -1, 4, 9] ✓ **Answer:** [0,0,10,-1,4,9]

Solution:

Since $Q_0(s,a) = 0$, $Q_1(s,a) = \sum_{s'} P(s'|s,a)\, R(s,a,s')$ As a result, the resulting table would be:

a	$s=0$	$s=1$	$s=2$
WALK	0	0	10
RUN	-1	4	9

Submit You have used 1 of 3 attempts

ℹ Answers are displayed within the problem

2. (2)

3/3 points (graded)
What is the ideal policy derived from $Q_1(s, a)$?

$\pi_1^*(s = 0) =$

● WALK

○ RUN

✓

$\pi_1^*(s = 1) =$

○ WALK

● RUN

✓

$\pi_1^*(s = 2) =$

● WALK

○ RUN

✓

Solution:

The ideal policy is the action a that maximizes $Q_1(s, a)$ for each state s.

Submit You have used 1 of 3 attempts

ℹ Answers are displayed within the problem

2. (3)

3/3 points (graded)
What is the value of $V_1(s)$ using the values of $Q_1(s, a)$ calculated above?

$V_1(s = 0) =$

0 ✓ **Answer:** 0

$V_1(s = 1) =$

4 ✓ **Answer:** 4

$V_1(s = 2) =$

10 ✓ **Answer:** 10

Solution:

$V_1(s) = \max_a Q_1(s, a)$

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

2. (4)

1/1 point (graded)

Consider now iterating Q-values one more time to obtain $Q_2(s, a)$. Here we are only interested in what happens at $s = 0$. What is the smallest value $\gamma_{\min}$ of the discount factor $0 \leq \gamma \leq 1$ for which the action derived from $Q_2(0, a)$ would suggest that we RUN? (Keep at least 4 decimal place accuracy in your answer.)

$\gamma_{\min} =$ 1/3 ✓ **Answer:** 1/3

Solution:

$$\begin{aligned} Q_2(0, \mathrm{WALK}) &= 0 + \gamma V_1(1) = \gamma \cdot 4 \\ Q_2(0, \mathrm{RUN}) &= 0.5(-1 + \gamma V_1(1)) + 0.5(-1 + \gamma V_1(2)) = -1 + \gamma \cdot 7 \end{aligned}$$

Hence $Q_2(0, \mathrm{RUN}) \geq Q_2(0, \mathrm{WALK})$ if $\gamma \geq 1/3$.

Submit You have used 1 of 3 attempts

Answers are displayed within the problem

Error and Bug Reports/Technical Issues

Show Discussion

Topic: Final exam (1 week):Final Exam / Problem 2