Haonan Hu

4/29/2021

2863545

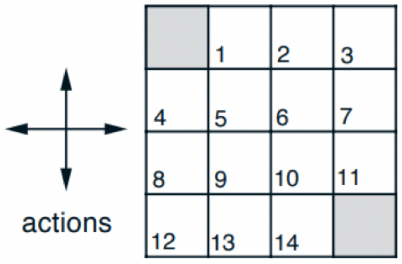
In-Class Problem

Assume:

- Monte Carlo Learning algorithm
- Same Gridworld as in example to the right.
- Random-walk policy:
  - Start robot at random location in grid.
  - Robot moves randomly around grid until it encounters a terminal state.
- Let γ = 0.9
- The sequence of states and rewards for episode 1 is:
  {(12,-1),(8,-1),(9,-1),(8,-1),(4,-1),(t,0)}

1. For First-Visit calculate the following:
   a) N(s)
   b) S(s)
   c) V(s)
2. For Every-Visit calculate the following:
   a) N(s)
   b) S(s)
   c) V(s)

$$G(s) = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \ldots + \gamma^{k-1} r_k$$



r = 0, if s = t
r = -1, if s ≠ t

1. a. N(s)=

1			
1			
1	**1**		
1			

1.b.

K	S	R	G(s)
1	12	-1	-4.0951
2	8	-1	-3.439
3	9	-1	-2.71
4	8	-1	-1.9
5	4	-1	-1
6	t	-1	0

S(s)

0			
-1			
-3.439	-2.71		
-4.0951			

1.c. V(s)

0			
-1			
-3.439	-2.71		
-4.0951			

2. a N(s)

1			
1			
2	**1**		
1			

2.b

K	S	R	G(s)
1	**12**	**-1**	**-4.0951**
2	**8**	**-1**	**-3.439**
3	**9**	**-1**	**-2.71**
4	**8**	**-1**	**-1.9**
5	**4**	**-1**	**-1**
6	**t**	**-1**	**0**

S(s)

0			
-1			
-5.339	**-2.71**		
-4.0951			

2.c

V(s)

0			
-1			
-2.6695	**-2.71**		
-4.0951			