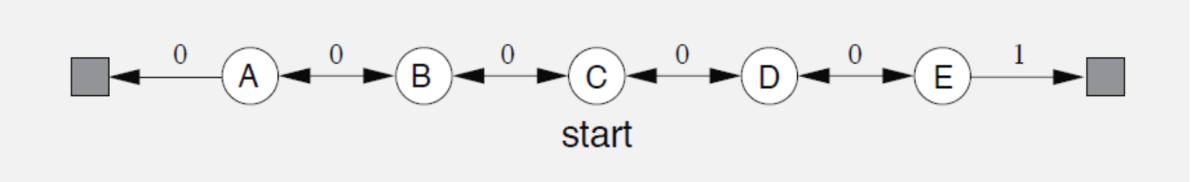
Reinforcement Learning
Mid Semester Project - 2023

Hadar Pur, Ron Azuelos

Theoretical Questions

Question 1:



We discussed the Random walk example in class.
In this Markov Reward Process (MRP), all episodes start at the centre state, C (state:3), then proceed either left or right by one state on each step, with equal probability.
Episodes terminate either on the extreme left (s0) or the extreme right (s6).
When an episode terminates on the right, a reward of +1 occurs, all other rewards are zero

The initial values are as follows:

$$v_0^{(0)} = 0 \quad v_1^{(0)} = \frac{1}{6} \quad v_2^{(0)} = \frac{2}{6} \quad v_3^{(0)} = \frac{3}{6} \quad v_4^{(0)} = \frac{4}{6} \quad v_5^{(0)} = \frac{4}{6} \quad v_6^{(0)} = 0$$

Using Dynamic Programming, run (theoretically) Policy Evaluation and show what the state's values will be at the next iteration. Show your computations in details.

Answer:

Let's define the states first:

The rewards are:

$$r_{1\to 0} = 0 \quad r_{1\to/\leftarrow 2} = 0 \quad r_{2\to/\leftarrow 3} = 0 \quad r_{3\to/\leftarrow 4} = 0 \quad r_{4\to/\leftarrow 5} = 0 \quad r_{5\to 6} = 1$$

The initial values are:

$$v_0^{(0)} = 0 \quad v_1^{(0)} = \frac{1}{6} \quad v_2^{(0)} = \frac{2}{6} \quad v_3^{(0)} = \frac{3}{6} \quad v_4^{(0)} = \frac{4}{6} \quad v_5^{(0)} = \frac{4}{6} \quad v_6^{(0)} = 0$$

We have requested to run a single iteration (first) and calculate the values for each state after these iteration.
We will use the below formula to calculate the values:

$$V(S) = \sum_a \pi(a|s) \sum_{s',r} p(s', r \mid s, a)\,[r + \gamma V(s')]$$

We will assume t the model is deterministic so $\gamma = 1$ (discount factor).
The probability to move from one state to another state is equal to $\frac{1}{2}$.

Calculate S3:

Possible transitions: $S3 \rightarrow S2,\ S3 \rightarrow S4$. Both of them have a probability of $\frac{1}{2}$ and $r = 0$.

$$V_\pi^{(1)}(S3) = (P(right \mid S3) + P(left \mid S3)) \cdot [(P_{3\rightarrow 2} \cdot (r_{3\rightarrow 2} + \gamma V'_2)) + (P_{3\rightarrow 4} \cdot (r_{3\rightarrow 4} + \gamma V'_4))]$$

$$= (\tfrac{1}{2} + \tfrac{1}{2}) \cdot [(\tfrac{1}{2} \cdot (0 + 1 \cdot \tfrac{2}{6}) + (\tfrac{1}{2} \cdot (0 + 1 \cdot \tfrac{4}{6})] = 1 \cdot (\tfrac{1}{6} + \tfrac{2}{6}) = \tfrac{1}{2}$$

Calculate S4:

Possible transitions: $S4 \rightarrow S3,\ S4 \rightarrow S5$. Both of them have a probability of $\frac{1}{2}$ and $r = 0$.

$$V_\pi^{(1)}(S4) = (P(right \mid S4) + P(left \mid S4)) \cdot [(P_{4\rightarrow 3} \cdot (r_{4\rightarrow 3} + \gamma V'_3)) + (P_{4\rightarrow 5} \cdot (r_{4\rightarrow 5} + \gamma V'_5))]$$

$$= (\tfrac{1}{2} + \tfrac{1}{2}) \cdot [(\tfrac{1}{2} \cdot (0 + 1 \cdot \tfrac{3}{6}) + (\tfrac{1}{2} \cdot (0 + 1 \cdot \tfrac{4}{6})] = 1 \cdot (\tfrac{3}{12} + \tfrac{2}{6}) = \tfrac{7}{12}$$

Calculate S5:

Possible transitions: $S5 \rightarrow S4,\ S5 \rightarrow S6$. Both of them have a probability of $\frac{1}{2}$ and $r_{5\rightarrow 4} = 0,\ r_{5\rightarrow 6} = 1$.

$$V_\pi^{(1)}(S5) = (P(right \mid S5) + P(left \mid S5)) \cdot [(P_{5\rightarrow 4} \cdot (r_{5\rightarrow 4} + \gamma V'_4)) + (P_{5\rightarrow 6} \cdot (r_{5\rightarrow 6} + \gamma V'_6))]$$

$$= (\tfrac{1}{2} + \tfrac{1}{2}) \cdot [(\tfrac{1}{2} \cdot (0 + 1 \cdot \tfrac{4}{6}) + (\tfrac{1}{2} \cdot (1 + 1 \cdot 0)] = 1 \cdot (\tfrac{2}{6} + \tfrac{1}{2}) = \tfrac{5}{6}$$

Calculate S2:

Possible transitions: $S2 \rightarrow S1,\ S2 \rightarrow S3$. Both of them have a probability of $\frac{1}{2}$ and $r = 0$.

$$V_\pi^{(1)}(S2) = (P(right \mid S2) + P(left \mid S2)) \cdot [(P_{2\rightarrow 1} \cdot (r_{2\rightarrow 1} + \gamma V'_1)) + (P_{2\rightarrow 3} \cdot (r_{2\rightarrow 3} + \gamma V'_3))]$$

$$= (\tfrac{1}{2} + \tfrac{1}{2}) \cdot [(\tfrac{1}{2} \cdot (0 + 1 \cdot \tfrac{1}{6}) + (\tfrac{1}{2} \cdot (0 + 1 \cdot \tfrac{3}{6})] = 1 \cdot (\tfrac{1}{12} + \tfrac{3}{12}) = \tfrac{4}{12} = \tfrac{1}{3}$$

Calculate S1:

Possible transitions: $S1 \rightarrow S0,\ S1 \rightarrow S2$. Both of them have a probability of $\frac{1}{2}$ and $r = 0$.

$$V_{\pi}^{(1)}(S1) = (P(right \mid S1) + P(left \mid S1)) \cdot [(P_{1\rightarrow 0} \cdot (r_{1\rightarrow 0} + \gamma V'_0)) + (P_{1\rightarrow 2} \cdot (r_{1\rightarrow 2} + \gamma V'_2))]$$

$$= (\tfrac{1}{2} + \tfrac{1}{2}) \cdot [(\tfrac{1}{2} \cdot (0 + 1 \cdot 0) + (\tfrac{1}{2} \cdot (0 + 1 \cdot \tfrac{2}{6})] = 1 \cdot (0 + \tfrac{1}{6}) = \tfrac{1}{6}$$

Calculate S0 + S6:

We don't need to calculate the values for S0 and S6 because those are termination states and we can't do any transition from them to another state, the probability to make action from those states is 0 and the calculation for both will be 0.

$$V_{\pi}^{(1)}(S0) = V_{\pi}^{(1)}(S6) = 0$$

Overall we have got that the values for all state after one iteration are:

$V_{\pi}^{(1)}(S0) = 0$	$V_{\pi}^{(1)}(S1) = \frac{1}{6}$	$V_{\pi}^{(1)}(S2) = \frac{1}{3}$	$V_{\pi}^{(1)}(S3) = \frac{1}{2}$
$V_{\pi}^{(1)}(S4) = \frac{7}{12}$	$V_{\pi}^{(1)}(S5) = \frac{5}{6}$	$V_{\pi}^{(1)}(S6) = 0$	

Question 2:

Given the following grid-world problem. The goal is to reach the upper left or lower right corners.

0	-14	-20	-22
-14	-18	-20	-20
-20	-20	S2	-14
-22	S1	-14	0

Actions: UP, DOWN, RIGHT, LEFT
Rewards: -1 for every movement (even movements that tries to move into a wall and stays in place)
Transition Model: Deterministic
Discount factor: γ = 1
Random Policy: 0.25 probability in each direction

In this environment, we were partially given the "Values" for certain states (according to the current policy). Calculate (manually) the Values (according to the current policy) for states: S1, S2, Show your computations in details.

Answer:

We need to calculate S1 and S2 under the following:

- Actions: UP, DOWN, RIGHT, LEFT
- Rewards: -1 for every movement
- Transition Model: Deterministic
- $\gamma = 1$
- Random Policy: 0.25 probability in each direction

We will use the below formula to calculate the values:

$$V(S) = \sum_a \pi(a|s) \sum_{s',r} p(s', r \mid s, a) [r + \gamma V(s')]$$

Calculate S1:

$$V(S1) = (P(right \mid S1) + P(left \mid S1) + P(up \mid S1) + P(down \mid S1)) \cdot [P_{up} \cdot (r(S1)_{up} + \gamma V'(S1)_{up}) + P_{down} \cdot (r(S1)_{down} + \gamma V'(S1)_{down}) + P_{left} \cdot (r(S1)_{left} + \gamma V'(S1)_{left}) + P_{right} \cdot (r(S1)_{right} + \gamma V'(S1)_{right})]$$

$$= (\tfrac{1}{4} + \tfrac{1}{4} + \tfrac{1}{4} + \tfrac{1}{4}) \cdot [\tfrac{1}{4} \cdot (-1 + 1 \cdot (-20)) + \tfrac{1}{4} \cdot (-1 + 1 \cdot V'(S1)_{down}) + \tfrac{1}{4} \cdot (-1 + 1 \cdot (-22)) + \tfrac{1}{4} \cdot (-1 + 1 \cdot (-14))]$$

$$= 1 \cdot (-\tfrac{21}{4} - \tfrac{1}{4} + \tfrac{1}{4}V(S1) - \tfrac{23}{4} - \tfrac{15}{4}) = -\tfrac{60}{4} + \tfrac{1}{4}V(S1) = -15 + \tfrac{1}{4}V(S1)$$

$$V(S1) = -15 + \tfrac{1}{4}V(S1) \rightarrow$$

$$\tfrac{3}{4}V(S1) = -15 \rightarrow$$

$$V(S1) = -20$$

Calculate S2:

$$V(S2) = (P(right \mid S2) + P(left \mid S2) + P(up \mid S2) + P(down \mid S2)) \cdot [P_{up} \cdot (r(S2)_{up} + \gamma V'(S2)_{up}) + P_{down} \cdot (r(S2)_{down} + \gamma V'(S2)_{down}) + P_{left} \cdot (r(S2)_{left} + \gamma V'(S2)_{left}) + P_{right} \cdot (r(S2)_{right} + \gamma V'(S2)_{right})]$$

$$= (\tfrac{1}{4} + \tfrac{1}{4} + \tfrac{1}{4} + \tfrac{1}{4}) \cdot [\tfrac{1}{4} \cdot (-1 + 1 \cdot (-20)) + \tfrac{1}{4} \cdot (-1 + 1 \cdot (-14)) + \tfrac{1}{4} \cdot (-1 + 1 \cdot (-20)) + \tfrac{1}{4} \cdot (-1 + 1 \cdot (-14))]$$

$$= 1 \cdot (-\tfrac{21}{4} - \tfrac{15}{4} - \tfrac{21}{4} - \tfrac{15}{4}) = -\tfrac{72}{4} = -18$$

Overall we have got that the values for S1 and S2 are:

$V(S1) = -20 \qquad V(S2) = -18$