Home Work #1

Ali BaniAsad 401209244

October 29, 2023

Question 1

Suppose $\gamma = 0.5$ and the following sequence of rewards is received:

t	R_t	G_t
1	1	?
2	2	?
3	6	?
4	3	?
5	2	0

We can work backward to compute the returns G_t:

$$
\begin{aligned}
&G_5 = 0 \\
&G_4 = R_4 + \gamma G_5 = 3 + 0.5 \cdot 0 = 3 \\
&G_3 = R_3 + \gamma G_4 = 6 + 0.5 \cdot 3 = 7.5 \\
&G_2 = R_2 + \gamma G_3 = 2 + 0.5 \cdot 7.5 = 6.25 \\
&G_1 = R_1 + \gamma G_2 = 1 + 0.5 \cdot 6.25 = 4.125 \\
&G_0 = R_0 + \gamma G_1 = 0 + 0.5 \cdot 4.125 = 2.0625
\end{aligned}
$$

Therefore, the returns are:

t	R_t	G_t
1	1	4.125
2	2	6.25
3	6	7.5
4	3	3
5	2	0

Question 2

$$
\begin{aligned}
v_\pi(s) &= \sum_a \pi(a|s) \sum_{s',r} p(s',r|s,a)[r+\gamma v_\pi(s')] \\
&= \frac{1}{4} \sum_{s',r} p(s',r|s,a)[r+\gamma v_\pi(s')] \\
&= \frac{1}{4}(0+\gamma v_\pi(A)) + \frac{1}{4}(0+\gamma v_\pi(B)) + \frac{1}{4}(0+\gamma v_\pi(C)) + \frac{1}{4}(0+\gamma v_\pi(D)) \\
&= \frac{1}{4}\gamma(v_\pi(A)+v_\pi(B)+v_\pi(C)+v_\pi(D)) \\
&= \frac{1}{4}\gamma \sum_{s'} v_\pi(s') \\
&= \frac{1}{4}\gamma(0.7+2.3+0.4-0.4) \\
&= \frac{1}{4}\gamma(3) = 0.7
\end{aligned}
$$

Contents