Reinforcement Learning
Exercise 5 - Solution

Jonathan Schnitzler - st166934
Eric Choquet - st160996

May 21, 2024

Task 1) - Random Walk

In the Random walk example from lecture 5 slide 12 the value is a prediction of the probability of terminating on the right side of the chain. Let us first recall the update rule for $TD(0)$:

$$V(S_t) \leftarrow V(S_t) + \alpha \left[R_{t+1} + \gamma V(S_{t+1}) - V(S_t)\right] \tag{1}$$

When only the first state $V(A)$ is updated, the sampled Random walk is immediately terminated at the left. Then the equation abouve evaluates to

$$V(A) \leftarrow V(A) + 0.1 \left[0 + 0 - V(A)\right] = V(A) - 0.1V(A) = 0.5 - 0.1 \cdot 0.5 = 0.45 \tag{2}$$

Otherwise, also the state of another value would have been changed, e.g. if first the Random walk would have gone to the right, then $V(B)$ would have been updated as well.