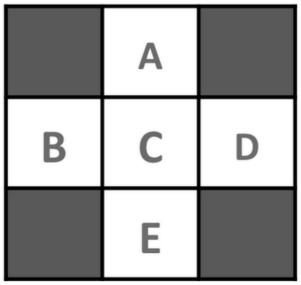
Quiz 4: TD Learning

Quiz 4: TD Learning

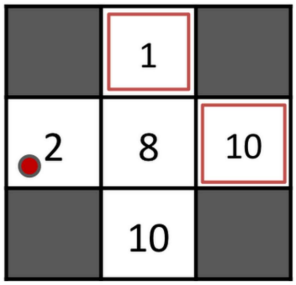
5/5 points (ungraded)

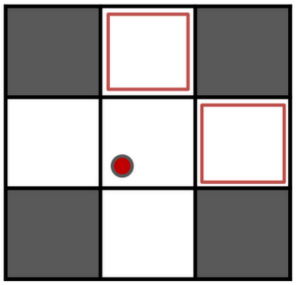
Consider the gridworld shown below. The left panel shows the name of each state A through E. The middle panel shows the current estimate of the value function V^{π} for each state. A transition is observed, that takes the agent from state B through taking action east into state C, and the agent receives a reward of -2. Assuming $\gamma = 1, \alpha = \frac{1}{2}$, what are the value estimates after the TD learning update? (note: the value will change for one of the states only)

States



Observed Transition: B, east, C, -2





Assume: γ = 1, α = 1/2

$$V^{\pi}(s) \leftarrow (1-\alpha)V^{\pi}(s) + \alpha\left[R(s,\pi(s),s') + \gamma V^{\pi}(s')\right]$$

$\hat{V}^{\pi}(A) =$

1 ✔

$\hat{V}^{\pi}(B) =$

4 ✔

$\hat{V}^{\pi}(C) =$

8 ✔

$\hat{V}^{\pi}(D) =$

10 ✔

$\hat{V}^{\pi}(E) =$

10 ✔

Submit

✔ Correct (5/5 points)