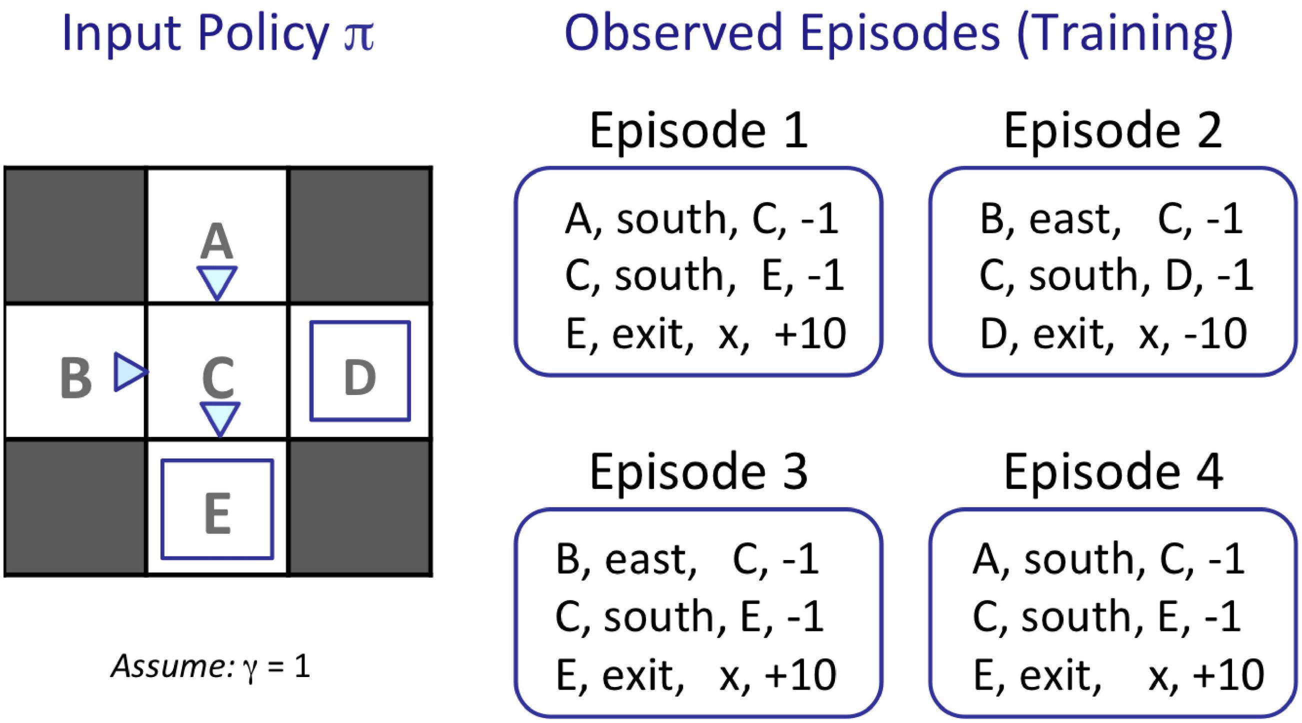
hw5_rl_q3_direct_evaluation

Question 3: Direct Evaluation

0.0/10.0 points (graded)



Observed Episodes (Training)

Episode 1

A, south, C, -1
C, south, E, -1
E, exit, x, +10

Episode 2

B, east, C, -1
C, south, D, -1
D, exit, x, -10

Episode 3

B, east, C, -1
C, south, E, -1
E, exit, x, +10

Episode 4

A, south, C, -1
C, south, E, -1
E, exit, x, +10

What are the estimates for the following quantities as obtained by direct evaluation:

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

8

Answer: 8

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

-2

Answer: -2

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

4

Answer: 4

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

-10

Answer: -10

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

100

Answer: 10

The estimated value of $\hat{V}^{\pi}(s)$ is equal to the average value achieved starting from that state.

$\hat{V}^{\pi}(A)$: Episodes 1 and 4 start from state A and both result in a utility of 8. $\frac{8+8}{2} = 8$

$\hat{V}^{\pi}(B)$: Episodes 2 and 3 start from state B. Episode 2 results in -12, while episode 3 results in 8. $\frac{8-12}{2} = -2$

$\hat{V}^{\pi}(C)$: State C is visited in every episode. The remaining rewards from C in episodes 1, 3, and 4 total 9, while the remaining rewards in episode 2 total -11. $\frac{9+9+9-11}{4} = 4$

$\hat{V}^{\pi}(D)$: State D is only visited in episode 2 and has a remaining utility of -10.

$\hat{V}^{\pi}(E)$: State E is visited in episodes 1, 3, and 4 and has a remaining utility of 10 in each state. $\frac{10+10+10}{3} = 10$

Submit

Answers are displayed within the problem