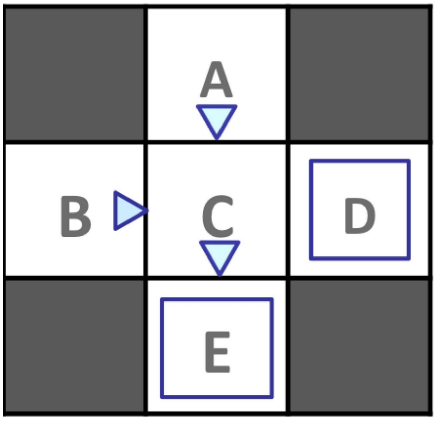
hw5_rl_q1_model_based_rl_grid

Question 1: Model-Based RL: Grid

0.0/4.0 points (graded)

Input Policy π



Assume: $\gamma = 1$

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 model would be learned from the above observed episodes?

T(A, south, C) =

1

Answer: 1

The action south is taken twice from state A, and both times results in state C. $\frac{2}{2} = 1$

T(B, east, C) =

1

Answer: 1

The action east is taken twice from state B, and both times results in state C. $\frac{2}{2} = 1$

T(C, south, E) =

0.75

Answer: .75

The action south is taken four times from state C, and results in state E three times. $\frac{3}{4} = .75$

T(C, south, D) =

0.25

Answer: .25

T(C, south, D): The action south is taken four times from state C, and results in state D one time. $\frac{1}{4} = .25$

Submit

Answers are displayed within the problem