# AU332 Quiz10

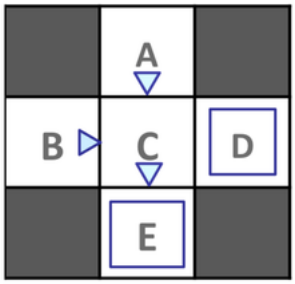
* 基本信息：

姓名： t

学号： t

*1. Model-Based RL: Grid

Input Policy π




*Assume:* γ = 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

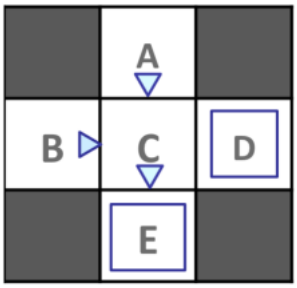
T(B, east, C)= 1

T(C, south, E)= 0.75

T(C, south, D)= 0.25

*2. Direct Evaluation

Input Policy π




*Assume:* γ = 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 are the estimates for the following quantities as obtained by direct evaluation:**

$V^{\pi}(A)=$ 8

$V^{\pi}(B)=$ -2

$V^{\pi}(C)=$ 4

$V^{\pi}(D)=$ -10

$V^{\pi}(E)=$ 10

提交