

Exercise 6

Ching Anon Eladreich L. Com2021

Episode Path 1:

Step 1: $(0,0) \rightarrow D \rightarrow (1,0) [-0.1]$

Step 2: $(1,0) \rightarrow R \rightarrow (1,1) [-0.1]$

Step 3: $(1,1) \rightarrow R \rightarrow (1,2) [-0.1]$

Step 4: $(1,2) \rightarrow D \rightarrow (2,2) \text{ terminal } [+3]$

Episode Path 2:

$(0,0) \rightarrow R \rightarrow (0,1) [-0.1]$

$(0,1) \rightarrow R \rightarrow (0,2) [+1]$

$(0,2) \rightarrow D \rightarrow (1,2) [-0.1]$

$(1,2) \rightarrow D \rightarrow (2,2) \text{ terminal } [+3]$

SARSA

States	V	D	L	R
0,0	0.0	-0.05	0.0	-0.05
0,1	0.0	0.0	0.0	-0.5
0,2	0.0	0.7	0.0	0.0
1,0	0.0	0.0	0.0	-0.05
1,1	0.0	0.0	0.0	-0.05
1,2	0.0	2.25	0.0	0.0
2,0	0.0	0.0	0.0	0.0
2,1	0.0	0.0	0.0	0.0
2,2	-	-	-	-

Episode 1:

Target = $-0.1 + Q(0,0,R) = -0.1$

$Q(0,0,D) = 0 + 0.5(-0.1 - 0) = -0.05$

Target = $-0.1 + Q(1,1,R) = -0.1$

$Q(1,0,R) = 0 + 0.5(-0.1 - 0) = -0.05$

Target = $-0.1 + Q(1,2,R) = -0.1$

$Q(1,1,R) = 0 + 0.5(-0.1 - 0) = -0.05$

$Q(1,2,D) = 0 + 0.5(3 - 0) = 1.5$

Episode 2:

Target = $-0.1 + Q(0,1,R) = -0.1$

$Q(0,0,R) = 0 + 0.5(-0.1 - 0) = -0.05$

Target = $1 + Q(0,2,D) = 1$

$Q(0,1,R) = 0 + 0.5(1 - 0) = 0.5$

Target = $-0.1 + Q(1,2,D) = 1.4$

$Q(0,2,D) = 0 + 0.5(1.4 - 0) = 0.7$

$Q(1,2,D) = 1.5 + 0.5(3 - 1.5) = 2.25$

Episode Path 1:

Step 1: $(0,0) \rightarrow D \rightarrow (0,0) [-0.1]$

Step 2: $(0,0) \rightarrow R \rightarrow (0,1) [-0.1]$

Step 3: $(0,1) \rightarrow R \rightarrow (0,2) [-0.1]$

Step 4: $(0,2) \rightarrow D \rightarrow (1,2) \text{ terminal} [3]$

Episode Path 2:

$(0,0) \rightarrow R \rightarrow (0,1) [-0.1]$

$(0,1) \rightarrow R \rightarrow (0,2) [-0.1]$

$(0,2) \rightarrow D \rightarrow (1,2) [-0.1]$

$(1,2) \rightarrow D \rightarrow (2,2) \text{ terminal} [5]$

Q Learning

States	U	D	L	R
0,0	0.0	-0.05	0.0	-0.05
0,1	0.0	0.0	0.0	0.5
0,2	0.0	0.7	0.0	0.0
1,0	0.0	0.0	0.0	-0.05
1,1	0.0	0.0	0.0	-0.05
1,2	0.0	2.25	0.0	0.0
2,0	0.0	0.0	0.0	0.0
2,1	0.0	0.0	0.0	0.0
2,2	-	-	-	-

Episode 1

Target = $-0.1 + Q(0,0,R) = -0.1$

$Q(0,0,D) = 0 + 0.5(-0.1 - 0) = -0.05$

Target = $-0.1 + Q(0,1,R) = -0.1$

$Q(0,1,R) = 0 + 0.5(-0.1 - 0) = -0.05$

Target = $-0.1 + Q(0,2,D) = -0.1$

$Q(0,2,D) = 0 + 0.5(3 - 0) = 1.5$

$Q(1,2,D) = 0 + 0.5(5 - 0) = 2.5$

Episode 2

Target = $-0.1 + Q(0,0,R) = -0.1 + 0.5$

$= 0.4$

Target = $1 + Q(0,1,R) = 1$

$Q(0,1,R) = 0 + 0.5(1 - 0.5) = 0.5$

Target = $-0.1 + Q(0,2,D) = 1.4$

$Q(0,2,D) = 0 + 0.5(1.5 - 0) = 0.75$

~~$Q(1,2,D) = 0 + 0.5(2.5 - 1.5) = 1.5$~~

~~$Q(2,2,D) = 0 + 0.5(5 - 2.5) = 1.25$~~

$Q(1,2,D) = 1.5 + 0.5(3 - 1.5) = 2.25$