Exercise 6

1. Draw the Q. table structure

SARSA

State	Up	Down	Left	Right
0,0	-0.05	-0.05	0	0
	0	0		
0,1		0	0	0
0,2	0	0.7	0	0
1,0	0	+0.05	0	-0.05
1,1	0	0	0	-0.05
		2.25		
1,2	0		0	
2,0	0	0	0	0
2,1	0	0	0	0
2,2	0	0	0	0

Episode Path

Step 1: $(0,0) \rightarrow D \rightarrow (1,0)$ [reward -0.1]
Step 2: $(1,0) \rightarrow R \rightarrow (1,1)$ [reward -0.1]
Step 3: $(1,1) \rightarrow R \rightarrow (1,2)$ [reward -0.1]
Step 4: $(1,2) \rightarrow D \rightarrow (2,2)$ [reward $+3$]

Step 1:

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

Step 2:

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

Step 3:

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

Step 4:

Target

$$Q(4,2,D) = 0 + 0.5(3-0) = 1.5$$

Episode 2 Path

Step 1: $(0,0) \to R \to (0,1)$ reward -0.1

Step 2: $(0,1) \to R \to (0,2)$ reward $+1$

Step 3: $(0,2) \to D \to (1,2)$ reward -0.1

Step 4: $(1,2) \to D \to (2,2)$ reward $+3$

Step 1:

Target $= -0.1 + 0 = -0.1$

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

Step 2:

Target $= 1 + 0 = 1$

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

Step 3:

Target $= -0.1 + (0.5) = 0.4$

$$Q(0,2,R) = 0 + 0.5(0.4-0) = 0.7$$

Step 4:

~~Target = 5 + 0.2~~

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

Q learning

State	Up	Down	Left	Right
0,0	-0.05	-0.05	0	0
0,1	0	0	0	0
0,2	0	0.7	0	0
1,0	0	0	0	-0.05
1,1	0	0	0	-0.05
1,2	0	2.25	0	0
2,0	0	0	0	0
2,1	0	0	0	0
2,2	0	0	0	0

Episode Path

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

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

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

Step 4: $(1,2) \rightarrow D \rightarrow (2,2)$ reward   +3

Step 1

Target $= -0.1 + 0 = -0.1$

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

Step 2

Target $= -0.1 + 0 = -0.1$

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

Step 3

Target $= -0.1 + 0 = -0.1$

$Q(1,2,R) = 0 + 0.5(-0.1 - 0) = -0.05$

Step 4

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

Ep 2 Path

Step 1: $(0,0) \to R \to (0,1)$ reward -0.1
Step 2: $(0.1) \to R \to (0,2)$ reward $+1$
Step 3: $(0,2) \to D \to (1,2)$ reward -0.1
Step 4: $(1,2) \to D \to (2,2)$ reward $+3$

Step 1:

   Target $= 0.1 + 0 = -0.1$

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

Step 2:

   Target $= 1 + 0 = 1$

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

Step 3:

   Target $= -0.1 + (0.5) = 1.4$

   $Q(0,2,R) = 0 + 0.5(0.4-0) = 0.7$

Step 4:

   Target $Q_1(1,2,D) = 1.5 + 0.5(3-1.5) = 2.25$