

Aldrin A. Baluery

Ep 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) <sup>terminal</sup> = 3

Ep 2

Step 1 (0,0)  $\rightarrow$  R  $\rightarrow$  (0,1) = -0.1

Step 2 (0,1)  $\rightarrow$  R  $\rightarrow$  (0,2) = 1

Step 3 (0,2)  $\rightarrow$  D  $\rightarrow$  (1,2) = -0.1

Step 4 (1,2)  $\rightarrow$  D  $\rightarrow$  (2,2) = 3

(Sarsa)

States	Up	Down	Left	Right
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	3	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	-	-	-	-

Ep 1

$$J_{\text{target}} = -0.1 + Q(1,0,R) = -0.1 \quad Q(0,0,D) = 0 + 0.5(-0.1 - 0) = -0.05$$

$$J_{\text{target}} = -0.1 + Q(1,1,R) = -0.1 \quad Q(1,0,R) = 0 + 0.5(-0.1 - 0) = -0.05$$

$$J_{\text{target}} = -2.1 + Q(1,2,D) = -0.1 \quad Q(1,1,R) = 0 + 0.5(-0.1 - 0) = -0.05$$

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

Ep 2

$$J_{\text{target}} = -0.1 + Q(0,1,R) = -0.1$$

$$J_{\text{target}} = 1 + Q(0,2,D) = 1$$

$$J_{\text{target}} = -0.1 + Q(1,2,D) = 1.4$$

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



EP 1

- Step 1  $(0,0) \rightarrow D \rightarrow (1,0)$   
Step 2  $(1,0) \rightarrow R \rightarrow (1,1)$   
Step 3  $(1,1) \rightarrow R \rightarrow (1,2)$   
Step 4  $(1,2) \rightarrow D \rightarrow (2,2)$  terminal 3

EP 2

- Step 1  $(0,0) \rightarrow R \rightarrow (0,1) = -0.1$   
Step 2  $(0,1) \rightarrow R \rightarrow (0,2) = 1$   
Step 3  $(0,2) \rightarrow D \rightarrow (1,2) = -0.1$   
Step 4  $(1,2) \rightarrow D \rightarrow (2,2)$  terminal 3

(Q Learning)

States	Up	Down	Left	Right
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	—	—	—	—

EP 1

$$V_{\text{target}} = -0.1 + Q(1,0,R) = -0.1 \quad Q(0,0,D) = 0 + 0.5(-0.1 - 0) = -0.05$$

$$V_{\text{target}} = -0.1 + Q(1,1,R) = -0.1 \quad Q(1,0,R) = 0 + 0.5(-0.1 - 0) = -0.05$$

$$V_{\text{target}} = -0.1 + Q(1,2,D) = -0.1 \quad Q(1,1,R) = 0 + 0.5(-0.1 - 0) = -0.05$$

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

EP

$$V_{\text{target}} = -0.1 + Q(0,1,R) = 0 + 0.5(-0.1 - 0) = -0.05$$

$$V_{\text{target}} = 1 + Q(0,2,D) = 1 \quad Q(0,1,R) = 0 + 0.5(1 - 0) = 0.5$$

$$V_{\text{target}} = -0.1 + Q(1,2,D) = 1.4 \quad Q(0,2,D) = 0 + 0.5(1.4 - 0) = 0.7$$

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