John Efren V. Gannaban

	up	down	left	right
0, 0		-0.05		-0.05
0, 1				0.5
0, 2		0.7		
1, 0				-0.05
1, 1				-0.05
1, 2		1.5 → 2.25		
2, 0				
2, 1				
2, 2				

episode 1

step 1, (0,0) → DOWN [reward -0.1]
step 2, (1,0) → RIGHT [reward -0.1]
step 3, (1,1) → RIGHT [reward -0.1]
step 4, (1,2) → DOWN [reward +3]
[Terminal]

① Target = -0.1 + 0 = [-0.1]
Q(0,0,D) = 0 + 0.5 (-0.1 - 0) = [-0.05]
② Target = -0.1 + 0 = [-0.1]
Q(1,0,R) = 0 + 0.5 (-0.1 - 0) = [-0.05]
③ Target = -0.1 + 0 = [-0.1]
Q(1,1,R) = 0 + 0.5 (-0.1 - 0) = [-0.05]
④ Target = 3 + 0 = [3]
Q(1,2,D) = 0 + 0.5 (3 - 0) = [1.5]

episode 2

step 1, (0,0) → RIGHT [reward -0.1]
step 2, (0,1) → RIGHT [reward 1]
step 3, (0,2) → DOWN [reward -0.1]
step 4, (1,2) → DOWN [reward 3]

① Target = -0.1 + 0 = [-0.1]
Q(0,0,R) = 0 + 0.5 (-0.1 - 0) = [-0.05]
② Target = 1 + 0 = [1]
Q(0,1,R) = 0 + 0.5 (1 - 0) = [0.5]
③ Target = -0.1 + 0.5 = [0.4]
Q(0,2,D) = 0 + 0.5 (1.4 - 0) = [0.7]
④ Target = 3 + 0 = [3]
Q(1,2,D) = 1.5 + 0.5 (3 - 1.5) = [2.25]

Q-learning

	up	down	left	right
0,0		-0.05		-0.05
0,1				0.5
0,2		-0.05 -> 0.7		
1,0				-0.05
1,1				
1,2		1.5 -> 2.25		
2,0				
2,1				
2,2				

episode 1

step 1 $q(0,0,D) = 0 + 0.5\,[-0.1 + 0 - 0] = -0.05$

step 2 $q(1,0,R) = 0 + 0.5\,[0.1 + 0 - 0] = -0.05$

step 3 $q(1,1,R) = 0 + 0.5\,[-0.1 + 0 - 0] = 0.05$

step 4 $q(1,2,D) = 0 + 0.5\,[3 + 0 - 0] = 1.5$

episode 2

step 1 $q(0,0,R) = 0 + 0.5\,[-0.1 + 0 - 0] = -0.05$

step 2 $q(0,1,R) = 0 + 0.5\,[1 + 0 - 0] = 0.5$

step 3 $q(0,2,D) = 0 + 0.5\,[-0.1 + 1.5 - 0] = 0.7$

step 4 $q(1,2,D) = 1.5 + 0.5\,[3 + 0 - 1.5] = 2.25$

4-)

1. no difference (same q-values)
2. SARSA
3. Q-Learning