F3

Bongseokkim

2021-02-21

차 례

Q_s_a Real

```
real_solution =np.array([
  [-5.410774, -5.107744],
  [-4.441077, -4.410774],
  [-3.666667, -3.441077],
  [-2.666667, -3.344108],
  [-1.666667, -1.666667],
  [-2.000000, -1.666667],
  [-1.000000, -1.666667],
  [ 0.000000,  0.000000]] )
```

Policy Iteration 1 -MC Control

```
num_ep = 10**5
beg_time =time.time()
q_s_a = q_s_a_init
pi = pi_50
exploration_rate = 1

for epi_i in range(1,num_ep) :
    sample_path_i = simul_path(pi, P_normal, P_speed, R_s_a)
    q_s_a = pol_eval_MC(sample_path_i, q_s_a, alpha = 0.006 )
    pi = pol_imp(pi, q_s_a, exploration_rate)
    exploration_rate *=  0.989 # exponential decay
```

```
end_time =time.time()

result_q = pd.DataFrame(q_s_a, columns =['n','s'], index= states)
result_pi = pd.DataFrame(pi, columns =['n','s'], index= states)
print("Time difference of {} sec".format(end_time- beg_time))
```

```
## Time difference of 29.693559885025024 sec
```

```
print(result_pi.T)
```

```
##     0    10   20   30   40   50   60   70
## n  0.0  0.0  0.0  1.0  0.0  1.0  1.0  1.0
## s  1.0  1.0  1.0  0.0  1.0  0.0  0.0  0.0
```

```
print(result_q.T)
```

```
##          0         10        20        30        40        50        60   70
## n -5.447228 -4.583343 -3.716340 -2.658934 -1.830896 -1.810133 -1.000000  0.0
## s -5.051796 -4.413748 -3.417744 -2.789631 -1.675030 -1.814754 -1.001935  0.0
```

```
print("MSE :",((q_s_a-real_solution)**2).mean())
```

```
## MSE : 0.053882085289518766
```

Policy Iteration 2 - TD Control (a.k.a sarsa)

```
num_ep = 10**5
beg_time =time.time()
q_s_a = q_s_a_init
pi = pi_50
exploration_rate = 1

for epi_i in range(1,num_ep) :
    s_now = 0
    while s_now != 70:
        sample_step = simul_step(pi,s_now, P_normal, P_speed, R_s_a)
        q_s_a = pol_eval_TD(sample_step, q_s_a, alpha = max(1/epi_i,0.01))
        pi = pol_imp(pi, q_s_a, epsilon=  exploration_rate)
        s_now = int(sample_step[3])
        exploration_rate *=0.9994

end_time =time.time()
result_q = pd.DataFrame(q_s_a, columns =['n','s'], index= states)
result_pi = pd.DataFrame(pi, columns =['n','s'], index= states)
print("Time difference of {} sec".format(end_time- beg_time))
```

```
## Time difference of 48.10834836959839 sec
```

```
print(result_pi.T)
```

```
##     0    10   20   30   40   50   60   70
## n  0.0  1.0  1.0  1.0  1.0  0.0  1.0  1.0
## s  1.0  0.0  0.0  0.0  0.0  1.0  0.0  0.0
```

```
print(result_q.T)
```

```
##          0        10        20        30        40        50        60   70
## n -5.454515 -4.52231 -3.621308 -2.621479 -1.634839 -1.839776 -1.000000  0.0
## s -5.315626 -4.56392 -3.711411 -2.725744 -1.764109 -1.681491 -1.450376  0.0
```

```
print("MSE :",((q_s_a-real_solution)**2).mean())
```

```
## MSE : 0.038619758567558275
```

```
"Done "
```

```
## [1] "Done "
```