D1_Exercises

Kwon do yun

2021-01-13

차 례

Recap (P. 10)

```
import numpy as np

def soda_simul(this_state):
    n=np.random.random()

    if this_state=='c':
        if n<=0.7:
            next_state='c'
        else:
            next_state='p'
    else:
        if n<=0.5:
            next_state='c'
        else:
            next_state='p'

    return next_state

def cost_eval(path):
    cost_one_path=path.count('c')*1.5+path.count('p')*1
    return cost_one_path

MC_N=10000
spending_records=np.zeros((MC_N,))

for i in range(MC_N):
    path='c'

    for t in range(10):
        this_state=path[-1]
        next_state=soda_simul(this_state)
        path+=next_state

    spending_records[i]=cost_eval(path)
```

```
print(spending_records)
```

```
## [14.  14.  14.5 ... 14.5 14.  16. ]
```

MC simulation for estimating state-value function (P. 11)

```
episode_i = 0
cum_sum_G_i = 0
num_episode = 10000
H = 10 #time_horizon
while episode_i < num_episode:
  path = 's'
  for time in range(H):
    this_state = path[-1]
    next_state = soda_simul(this_state)
    path = path+next_state

  G_i = cost_eval(path)

  cum_sum_G_i = cum_sum_G_i + G_i

  episode_i +=1

V_t = cum_sum_G_i / num_episode

print(V_t)
```

```
## 13.04495
```

For general t, Exercise (P. 17)

$$V_t(s) = \mathbb{E}[G_t|S_t = s] \tag{1}$$
$$= \mathbb{E}[\sum_{i=t}^{H-1} r_i|S_t = s] \tag{2}$$
$$= \mathbb{E}[r_t + r_{t+1} + ... + r_{H-1}|S_t = s] \tag{3}$$
$$= \mathbb{E}[r_t|S_t = s] + \mathbb{E}[r_{t+1} + ... + r_{H-1}|S_t = s] \tag{4}$$
$$= R(s) + \mathbb{E}[G_{t+1}|S_{t+1} = s] \tag{5}$$
$$= R(s) + \sum_{s' \in S} P_{ss'}\mathbb{E}[G_{t+1}|S_{t+1} = s'] \tag{6}$$
$$= R(s) + \sum_{s' \in S} P_{ss'}V_{t+1}(s') \tag{7}$$

P. 20

```
import numpy as np
P = np.array([0.7,0.3,0.5,0.5]).reshape(2,2)
R = np.array([1.5,1.0]).reshape(2,1)
H = 10
v_t1 = np.array([0,0]).reshape(2,1)
t = H-1

while (t>=0):
    v_t = R+ np.dot(P,v_t1)
    t = t-1
    v_t1 = v_t
print(v_t)
```

```
## [[13.35937498]
##  [12.73437504]]
```

```
while (t>=0):
    v_t = R+P*v_t1
    t = t-1
print('Backward induction',v_t)
```

```
## Backward induction [[13.35937498]
##  [12.73437504]]
```

```
"D1_Exercises"
```