Assignment 2 – avv1004

Imagine that you are looking for the perfect rental apartment. You have clearly formulated your preferences in terms of the size, location, condition, and price of the apartment. Using your preferences, you are able to assign a score (such as a real number in the open interval (0-1)) to each apartment that you see and you would like to find one with the maximal score. The challenge is that the market is very competitive and if you want to get the particular apartment, you have to make an offer right after you see it. There is also no way of predicting the score that you'd assign to apartments that will come on the market later. You want to make a decision after seeing 5 apartments.

Formulate this problem as an MDP and compute the optimal policy and value function. Use either a computer or a pen and paper. Be careful when defining state and action sets. They should be small enough to solve the problem by hand.

Reading: See the secretary problem on Wikipedia, and Section 3.4.3 in [MDP] which describes the problem and its formulation.

Solution :

⃝ - States

● - Actions

2 states :

t = 3 t = 4

1. ① Best apartment seen so far -----accept------> X ①



1. ⓪ Didn’t see the best so far reject 1/4

●---3/4--->⓪

Actions :

1. Accept
2. Reject

Rewards :

1, if Best

0, Otherwise

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  | ● |  |

If 5th is not better, will get ‘0’

S = Ꝭ = Set of States = {0,1}



a = {a,r} (actions, rejections)

Pt = (st1,a1,st+1), where Pt is Probability,

Pt (o,r,1) = 1/(t+1)

Pt (1,r,1) = 1/(t+1)

Rt (st,at)

Rt (st,at1,st+1) , as going to another state

Scenario : Rt (0,a) = 0, where 0 = not best, a = accept, 0 = reward.

Finite Horizon

Example : R3 (1,a) = 3/5

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1/5 | 1/5 | 1/5 |  |  |

3/5 2/5

Required optimal policy :

History, dep. not possible

Methods : 1) Try them all

2) Dynamic Programming/ Value Iteration. (Backward Induction)

a S1 = 0.5

S0  (where only value by policy matters and not the policy)

r S2 = 0.3

t = 0 t = 1 t = 2 (1/3) (1/(t+1)) t = 3

① v(1) = 1/2 ① V2a = (2/3 , 0) ①

(1) = reject V2r = (1/3,1/3)

Va = 1/3 V2 = (2/3,1/3) V3 = (1,0)

Vr = (½(2/3)+ ½(1/3))

= 1/2 (2/3)(t/(t+1))

⓪ 2(1) = accept ⓪

1(1) = reject 2(0) = reject

--Generally, proof goes by backward induction

Number of apartments to be visited, N = 5

Range of Score, S = 0minimum to 10maximum

Optimal Policy Derivation :

