Lecture D3. Dynamic Programming

Baek, Jong min

2021-01-18

차 례

Exercise 1

How would you generalize this game with arbitrary value of m_1 (minimum increment), m_2 (maximum increment), and N (the winning number)?

Answer.
t is your turn number, you don't need to care opponent's turn number.
first is you(t = 1), next turn is opponent(don't care), second is you(t = 2), op, third is you(t = 3)

$N_t = m_1 + (t-1) * (m_1 + m_2)$
optimal number is $N_1 = 1,\ N_2 = 4, \cdots, N_{10} = 28, N_{11} = 31$

Exercise 2

Two players are to play a game. The two players take turns to call out integers. The rules are as follows. Describe A's winning strategy.
- A must call out an integer between 4 and 8, inclusive.
- B must call out a number by adding A's last number and an integer between 5 and 9, inclusive.
- A must call out a number by adding B's last number and an integer between 2 and 6, inclusive.
- Keep playing until the number larger than or equal to 100 is called by the winner of this game.

Exercise 3

There is only finite number of deterministic stationary policy. How many is it?
Answer.
$|A|^{|S|}$

Exercise 4

Formulate the first example in this lecture note using the terminology including state, action, reward, policy, transition. Describe the optimal policy using the terminology as well.

state = $\{1, 2, 3, \cdots, 30, 31\}$
action = $\{a_1, a_2\}$
reward = $R(29, a_2) = 1,\ R(30, a_1) = 1$, all other $R(s, a) = 0$
optimal policy =
transition = $P^a_{ss'} = P(S_{t+1} = S' \mid S_t = s, A_t = a) = 1$
$s' = s + 1, if(a = a1)$
$s' = s + 2, if(a = a2)$

Exercise 5

From the first example,
- Assume that your opponent increments by 1 with prob. 0.5 and by 2 with prob. 0.5.
- Assume that the winning number is 10 instead of 31.
- Your opponent played first and she called out 1.
- Your current a policy π_0 is that
If the current state s $\leq$ 5 then increment by 2.
If the current state s > 5 then increment by 1.
Evaluate V^{π_0}

D3.Rmd

```
"Hello"
```

```
## [1] "Hello"
```