Pseudo Exam

Gunhyeong Kim

1 Overview of Reinforcement Learning

1.1 Write the 4 Characteristics of Reinforcement Learning

1.2 Write the 4 Example of Reinforcement Learning Applications

1.3 Explain the Definition of Reward Hypothesis

1.4 What is the Sequential Decision Making? Explain about its goal.

1.5 Explain the Differences between Observation and State

1.6 Insert the collect word in the blank

At each step t the agent:

- Executes __________
- Receives __________
- Receives __________

The enviornment:

- Receives __________
- Emits __________
- Emits __________

1.7 Write the Definition of state S_t is Markov

1.8 Fully Observable Environment와 Partially Observable Environment의 차이를 수식으로 설명하시오.

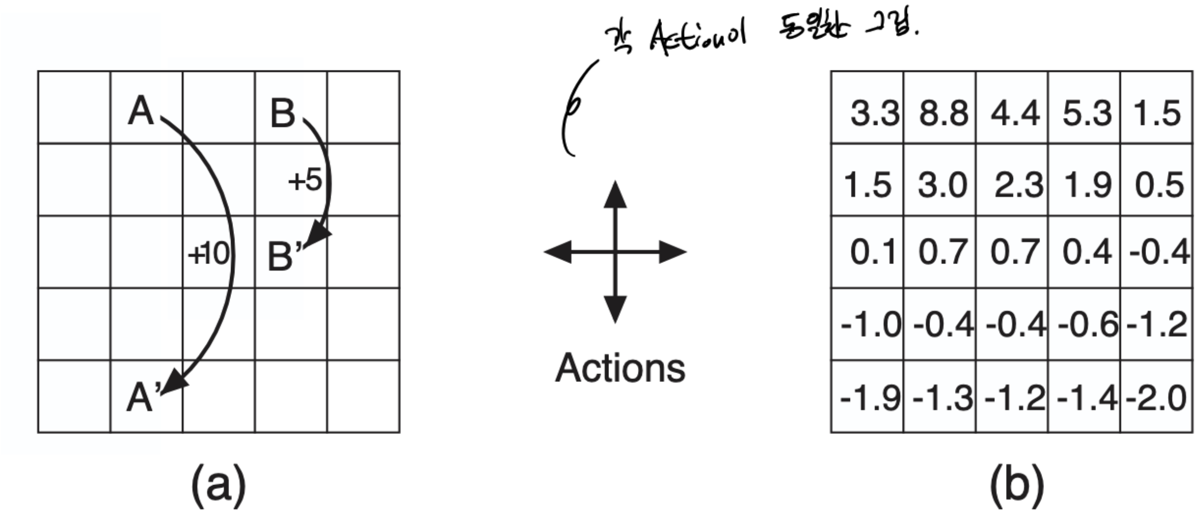
1.9 어떤 Policy π에서 state s에 대한 Value function을 수식으로 쓰시오. (discount factor γ 포함)

1.10 state s에서 state s'로의 Transition Probability를 수식으로 쓰시오. (action a 포함)

1.11 state s에서 action a를 했을 때 받는 Reward의 기대값을 수식으로 쓰시오.

1.12 Value Based와 Policy Based의 장단점에 대해 서술하시오.

1.13 다음 figure를 보고 uniform random policy의 Value function을 구하시오.

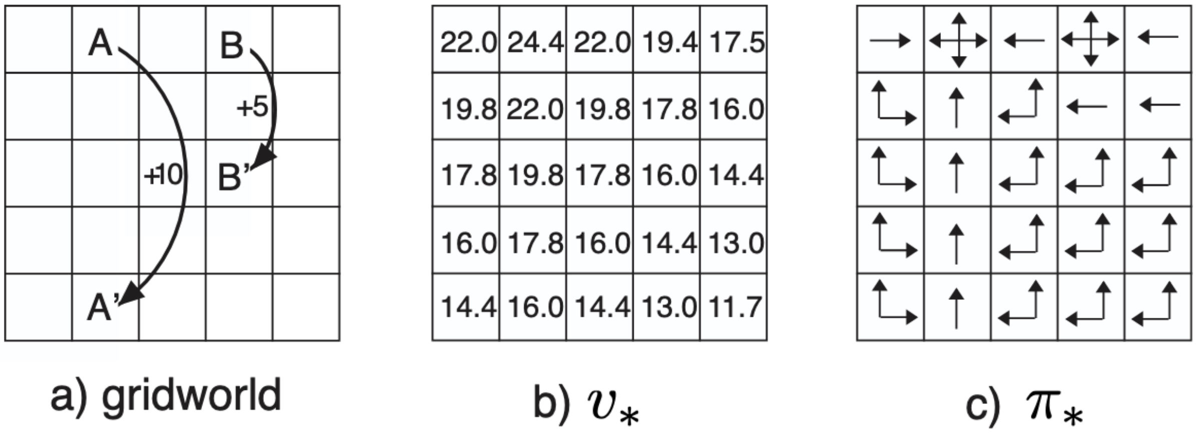


3.3	8.8	4.4	5.3	1.5
1.5	3.0	2.3	1.9	0.5
0.1	0.7	0.7	0.4	-0.4
-1.0	-0.4	-0.4	-0.6	-1.2
-1.9	-1.3	-1.2	-1.4	-2.0

각 Action이 동일한 그림.

Figure 1: Gridworld Example

1.14 다음 figure를 보고 optimal value function과 optimal policy를 구하시오.



22.0	24.4	22.0	19.4	17.5
19.8	22.0	19.8	17.8	16.0
17.8	19.8	17.8	16.0	14.4
16.0	17.8	16.0	14.4	13.0
14.4	16.0	14.4	13.0	11.7

→	↑↓←→	←	↑↓←→	←
↑→	↑	←↑	←	←
↑→	↑	←↑	←↑	←↑
↑→	↑	←↑	←↑	←↑
↑→	↑	←↑	←↑	←↑

Figure 2: Gridworld Example for Optimal Value Function and Policy

2 Markov Decision Processes

2.1 Write the Definition of Markov

2.2 Write the Definition of Markov Process

2.3 Write the Definition of Markov Reward Process

2.4 Write the Definition of Return G_t

2.5 Write the Definition of state-value function of MRP

2.6 Input the collect value in the blank

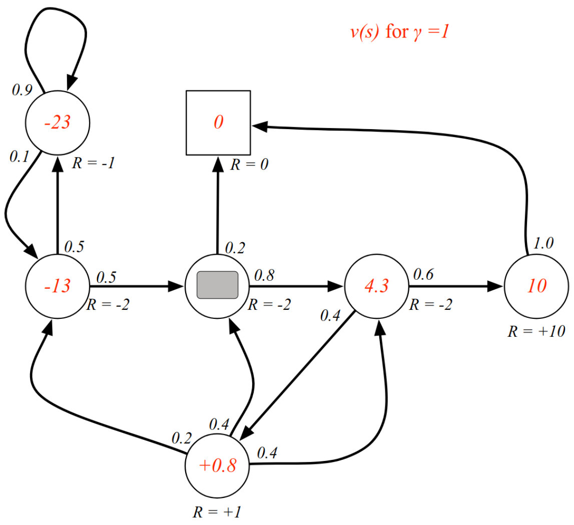


Figure 3: Markov Reward Process

2.7 Write the Bellman Equation for state-value function of MRP (and also in model based form)

2.8 Solve the Bellman Equation, and Explain why this solution is not practical in real-world applications.

2.9 Write the Definition of Markov Decision Process

2.10 Write the Definition of policy π in MDP(contains what it outputs)

2.11 Write the state-value function and action-value function in MDP under policy π

2.12 Insert the collect value in the blank

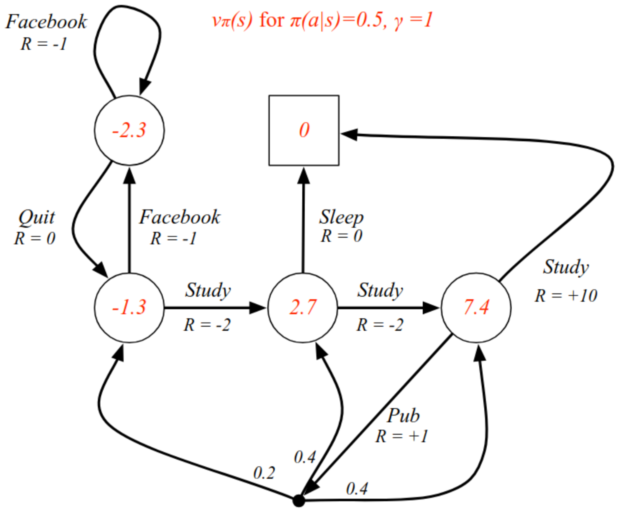


Figure 4: Markov Decision Process

2.13 Write the Bellman Expectation Equation for V^π with diagram

2.14 Write the Bellman Expectation Equation for Q^π with diagram

2.15 Write the Bellman Expectation Equation for V^π using Q^π (with diagram)

2.16 Write the Bellman Expectation Equation for Q^π using V^π (with diagram)

2.17 Write the Definition of Optimal state-value function V^* and Optimal action-value function Q^*

2.18 Write the Theorem of Optimality between π_* and π

2.19 Write the $\pi_*(a \mid s)$ by using $q_*(s, a)$

2.20 Write the Optimal values under the Actions

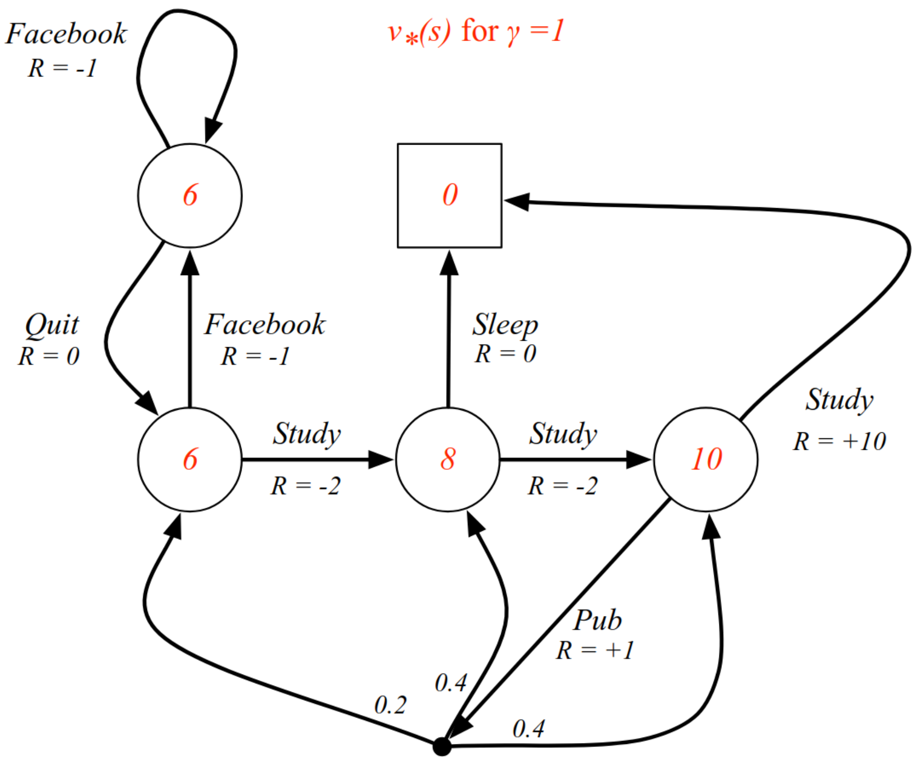


Figure 5: Optimal values under the Actions

2.21 Write the Bellman Optimality Equation for V^π with diagram

2.22 Write the Bellman Optimality Equation for Q^π with diagram

2.23 Write the Bellman Optimality Equation for V^π using Q^π (with diagram)

2.24 Write the Bellman Optimality Equation for Q^π using V^π (with diagram)

3 Planning by Dynamic Programming

3.1 What is Dynamic Programming? and What are the two properties of problems that DP can be applied to?

3.2 Explain Policy Iteration. Draw a diagram that shows the process.

3.3 Explain Value Iteration. Write down the value function update rule.

3.4 Compare and contrast Policy Iteration and Value Iteration.

3.5 What is the difference between synchronous and asynchronous dynamic programming?

3.6 Explain Contraction Mapping Theorem and why it is important in Dynamic Programming.

4 Model-Free Prediction

4.1 What is the key difference between model-based and model-free reinforcement learning?

4.2 Explain Monte-Carlo (MC) Policy Evaluation. What is the difference between first-visit and every-visit MC?

4.3 Explain Temporal-Difference (TD) Learning. Write down the TD(0) update rule.

4.4 Discuss the Bias-Variance Trade-Off between Monte-Carlo and Temporal-Difference Learning.

4.5 What is TD(λ)? Explain the role of eligibility traces.

4.6 What is the difference between bootstrapping and sampling? Describe MC, TD, and DP in terms of these concepts.

5 Model-Free Control

5.1 What is the difference between On-policy and Off-policy learning?

5.2 Explain the concept of ϵ-greedy exploration.

5.3 Explain the Sarsa algorithm. Write down the update rule for the action-value function.

5.4 Explain the Q-learning algorithm. Write down the update rule for the action-value function.

5.5 What is the key difference between Sarsa and Q-learning?

5.6 Explain Importance Sampling for Off-Policy Monte-Carlo and Off-Policy TD.