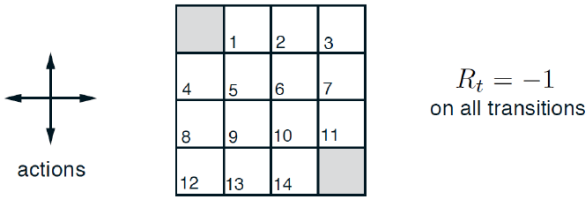
지능제어 1차 숙제 (2021. 03. 18)

학번: 이름:

다음 문제에 대해서 value function을 선형 연립방정식으로 만들어서 푸는 프로그램을 작성하시오.

Example 4.1 Consider the 4×4 gridworld shown below.



The nonterminal states are $\mathcal{S} = \{1, 2, \ldots, 14\}$. There are four actions possible in each state, $\mathcal{A} = \{\texttt{up}, \texttt{down}, \texttt{right}, \texttt{left}\}$, which deterministically cause the corresponding state transitions, except that actions that would take the agent off the grid in fact leave the state unchanged. Thus, for instance, $p(6, -1|5, \texttt{right}) = 1$, $p(7, -1|7, \texttt{right}) = 1$, and $p(10, r|5, \texttt{right}) = 0$ for all $r \in \mathcal{R}$. This is an undiscounted, episodic task. The reward is -1 on all transitions until the terminal state is reached. The terminal state is shaded in the figure (although it is shown in two places, it is formally one state). The expected reward function is thus $r(s, a, s') = -1$ for all states s, s' and actions a. Suppose the agent follows the equiprobable random policy (all actions equally likely). The left side of Figure 4.1 shows the sequence of value functions $\{v_k\}$ computed by iterative policy evaluation. The final estimate is in fact v_π, which in this case gives for each state the negation of the expected number of steps from that state until termination. ■

```
def evaluate_state_value_by_matrix_inversion(env, discount=1.0):
    WIDTH, HEIGHT = env.size()

    # Reward matrix R
    R = np.zeros((WIDTH, HEIGHT))
    for i in range(WIDTH):
        for j in range(HEIGHT):
            expected_reward = 0
            for action in ACTIONS:
                (next_i, next_j), reward = env.interaction([i, j], ACTIONS[action])
                expected_reward += ACTION_PROB*reward
            R[i, j] = expected_reward
    R = R.reshape((-1,1))
    R = R[1:-1,:]


    # Transition matrix T
    #
    #
    #
    # Write your code here
    #
    #
    #
```

3월 18일 강의에서 배부된 코드

"3_0_policy_evaluation_matrix_students.ipynb"

를 기반으로 "Write your cod here" 부분의 빈칸을 채우고 위의 코드(코드 부분 스크린샷)와 결과를 word 파일로 작성하고 pdf로 변환하여 4월 1일 오전 9시까지 제출하세요.