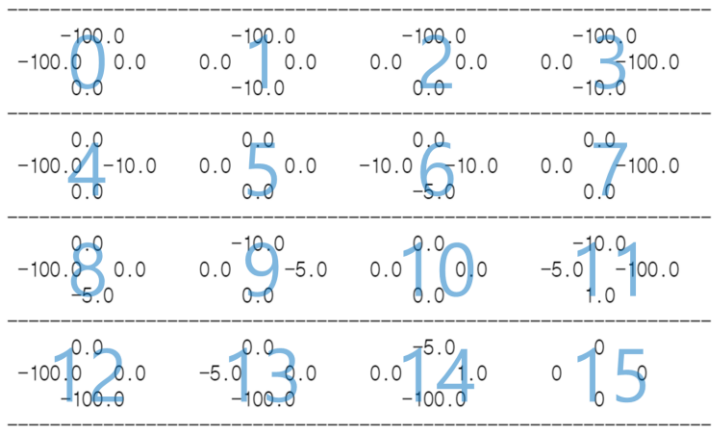
Homework #2

COSE490(02)
Reinforcement Learning

Goal: The goal of the second homework is (i) to let the student familiar with implementing an RL algorithm, and (ii) to understand the online rollout algorithm.

Tasks: Use the provided notebook file, and complete the function call "do_rollout".

- We have modified the Frozen Lake (FL). See the new MDP environment descried in matrix "P" (See line 72 in [5] of the notebook file). The transitions and rewards are deterministic. There is no hole. Instead, we consider several states that, when entering, we will get negative reward (-5 or -10). The following is the reward value for each (state, action) – state is shown in blue and action in each direction.



- We also provide the "base policy" that will be used for rollout. (See [6] of the notebook)

```
Policy:
| 00      > | 01      > | 02      > | 03      v |
| 04      v | 05      v | 06      > | 07      v |
| 08      v | 09      v | 10      v | 11      v |
| 12      > | 13      > | 14      > |           |
```

- Task 1: Complete the function "do_rollout". You should not modify the other parts. (One exception: if you want, you may modify "online_rollout" too.)
- Task 2: Make all the program run without error. We may test your code with different environments or different base policies (under which, your program has to work without error for full grade.)

Submission:

- After running the program (all the cells), print the results as a PDF file.
- Submit both your code and the PDF file through BlackBoard (BB).
- Submission is due on April 19 (before the class hour).