

Amrita Vishwa Vidyapeetham
Amrita School of Computing, Bangalore
Department of Computer Science and Engineering

21AIE311 - Reinforcement Learning

Lab Worksheet - 3

**Bandit Walk (BW), Bandit Slippery Walk (BSW) and Frozen Lake (FL)
Environment Setup**

Exercise

1. Write a simple python code to model the Bandit Walk (BW) environment using python Dictionary data structure. Note: Action 0-Left and 1-Right.

I SPEAK PYTHON
The bandit walk (BW) MDP

```
P = {
    0: {
        0: [(1.0, 0, 0.0, True)],
        1: [(1.0, 0, 0.0, True)],
    },
    1: {
        0: [(1.0, 2, 0.0, True)],
        1: [(1.0, 2, 1.0, True)],
    },
    2: {
        0: [(1.0, 2, 0.0, True)],
        1: [(1.0, 2, 0.0, True)],
    }
}
```

(1) The outer dictionary keys are the states.

(2) The inner dictionary keys are the actions.

(3) The value of the inner dictionary is a list with all possible transitions for that state-action pair.

(4) The transition tuples have four values: the probability of that transition, the next state, the reward, and a flag indicating whether the next state is terminal.

(5) You can also load the MDP this way

```
import gym, gym_walk
P = gym.make('Banditwalk-v0').env.P
```

- Write a simple python code to model the Bandit Slippery Walk (BSW) environment using python Dictionary data structure.

```

P = {
    0: {
        0: [(1.0, 0, 0.0, True)],
        1: [(1.0, 0, 0.0, True)]
    },
    1: {
        0: [(0.8, 0, 0.0, True), (0.2, 2, 1.0, True)],
        1: [(0.8, 2, 1.0, True), (0.2, 0, 0.0, True)]
    },
    2: {
        0: [(1.0, 2, 0.0, True)],
        1: [(1.0, 2, 0.0, True)]
    }
}

import gym, gym_walk
P = gym.make('BanditSlipperyWalk-v0').env.P

```

Assignment

- Write a simple python code to model the Frozen Lake (FL) environment using python Dictionary data structure.
- Write a simple python code to model the Walk Three environment using python Dictionary data structure.

Walk Three Environment Properties

- Deterministic environment
- 3 non-terminal states, 2 terminal states
- only reward is at the right-most cell in the walk
- episodic environment, the agent terminates at the left- or right-most cell
- agent starts in state 2 (middle of the walk) T-1-2-3-T
- actions left (0) or right (1)

0 (Hole, Terminal)	1	2	3	4 (Goal, Terminal)
--------------------	---	---	---	--------------------