

## Homework 3

AML 2022, Chennai Mathematical Institute

Due: Nov 15, 2022

### Open AI Gym - CartPole

Your task is to compute the optimal policy for playing the “CartPole” game from the OpenAI Gym. You are free to choose one of the reinforcement learning algorithms discussed in the lectures.

See: <https://github.com/openai/gym> on details about how to install and use this environment. See [https://github.com/openai/gym/blob/master/gym/envs/classic\\_control/cartpole.py](https://github.com/openai/gym/blob/master/gym/envs/classic_control/cartpole.py) for more details

The game is as follows:

A pole is attached by an un-actuated joint to a cart, which moves along a frictionless track. The pendulum is placed upright on the cart and the goal is to balance the pole by applying forces in the left and right direction on the cart.

See here for a short demo: <https://www.youtube.com/watch?v=T4SejVqE0X4>

### Action Space

The action takes values  $\{0, 1\}$  indicating the direction of the fixed force the cart is pushed with.

Value	Action
0	Push the cart to the left
1	Push the cart to the right

**Note:** The velocity that is reduced or increased by the applied force is not fixed and it depends on the angle the pole is pointing. The center of gravity of the pole varies the amount of energy needed to move the cart underneath it

### States

The environment gives you the following an array of 4 items with the values corresponding to the following positions and velocities:

Index	Observation	Min Value	Max Value
0	Cart Position	-4.8	4.8

1	Cart Velocity	-Infinity	Infinity
2	Pole Angle	-0.418 rad (-24 deg)	0.418 rad (24 deg)
3	Pole Angular Velocity	-Infinity	Infinity

### Starting State

At the start of an episode, all four observations are assigned a uniformly random value in  $(-0.05, 0.05)$

### Episode End

The episode ends if any one of the following occurs:

1. Termination: Pole Angle is greater than  $\pm 12$  degree
2. Termination: Cart Position is greater than  $\pm 2.4$  (center of the cart reaches the edge of the display)
3. Truncation: Episode length is greater than 500

### Rewards

1. In the classic game, the goal is to keep the pole upright for as long as possible, a reward of  $+1$  for every step taken, including the termination step, is allotted. Naturally, the goal is to remain upright for as long as possible to maximize rewards.
2. Also consider a new version of the game, where the reward is 0 for every time step. But, if the pole angle is the range  $(-12, -6)$  or  $(6, 12)$ , then there is a reward of  $+2$ .

**Task:** Your task is to determine the optimal policy for the two reward schemes. It is sufficient to output the Q-function of the optimal policy. You have to find out how to install and run Open AI Gym in Google Colab, which you can easily look up online.

### Instructions:

- You have to submit a jupyter notebook (ipynb) with all your code and outputs of the code
- If you don't include the outputs you will get partial credit
- You can work in groups of 2 or 3
- Only one member of the group should submit the assignment
- Please mention the names and roll-numbers of all group members
- You are free to build upon examples shown in class