RL ASSIGNMENT 4 REPORT

Contributions-

1. Stuti Garg - SE22UCSE263- Section 4 step 1 and step 2; Section 7

2. Anya - SE22UCSE033- Section 4 step 3; Section 6

3. Lakshmi - SE22UCSE149 -Section 4 step 3; Section 6

Section 4 Step 1 and 2

Date: / / /

SECTION 4 (STEP 2)

MOUNTAIN CAR

(Play Game)

(1) a → Action 0 → Accelerate to the left
S → Action 1 → Do not accelerate
d → Action 2 → Accelerate to the right.

(2) noop stands for 'no operation' and defines the default action when no key is pressed. It maps to action 1 (no acceleration.)

(3) This callback function is used to track and display the total reward accumulated during an episode. It is called after each step to update the reward counter.

(4) We can slow down the game by modifying frames per second (fps) parameter in the play() function.
For example in line 31, we can change
fps = 50 $\xrightarrow{\text{change to}}$ **fps = 10**

This can reduce the speed of the game, making it easier to play manually.

Date: / / /

SECTION 4 (CONTINUATION)

(5) The initial position and velocity of the car are randomized at the start of each episode. This randomness ensures that the agent does not always start from the exact same state. It also helps in better generalization.

(6) A good convergence criterion is
(i) when the average total reward over the last 100 episodes exceeds -110.
(benchmark for solving MountainCar-v0)
(ii) or when the moving average of the rewards stabilizes.

SECTION 4 (STEP 1) Date: / / /

MOUNTAIN CAR
(Gymnasium Document)

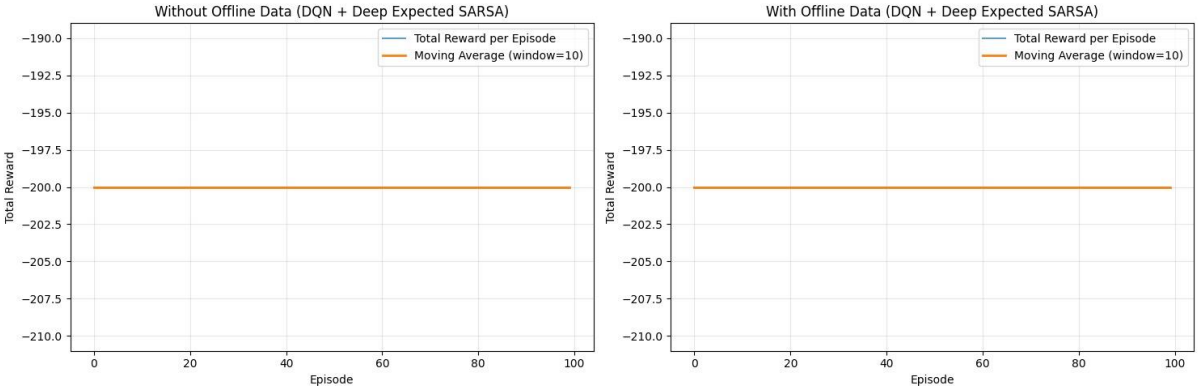
(1) There are 3 discrete actions:
0 : Accelerate to the left
1 : Don't Accelerate
2 : Accelerate to the right.

(2) No, it is not possible to accelerate to the left and to right at the same time. The action space is discrete, implying that agent can choose only one action from the 3 actions.

(3) There are 2 observations:
(i) Position of the car
(ii) Velocity of the car.

(4) There are 2 values in the observation space:
(i) Position $\in [1.2, 0.6]$
(ii) Velocity $\in [-0.07, 0.07]$

Section 6



Both models maintain a constant reward of approximately -200 across all episodes