Cartpole and LunarLander Environments

Sudhir Yarram
University at Buffalo
UB No:- 50305566
sudhirya@buffalo.edu

Vamsi Krishna Velivela
University at Buffalo
UB No:- 50318486
vvelivel@buffalo.edu

1 Introduction

Deep Q-Network is one of the effective ways to train the reinforcement learning agent. It is highly effective when there are too many states and too many actions. It predicts the action on unseen state by the experience gained from the seen state using the Deep learning techniques. We implemented DQN on Cartpole and LunarLander.

$$Q(s, a, w) \approx Q^{\pi}(s, a)$$

Replay in DQN :- with increase in size of memory, the learning experience of the agent increases.

Target Network in DQN :- Moving target is one of the biggest concern in DQN. The results fluctuate rapidly with moving target. Hence the concept of Target Network gives better results.

$$\mathcal{L}(w) = \mathbb{E}_{s,a,r,s' \sim \mathcal{D}} \left[\left(r + \gamma \max_{a'} Q(s', a', w) - Q(s, a, w) \right)^2 \right]$$

2 Cartpole

Cartpole has two actions 0 to push left and 1 to push right. It gets a reward of 1 for every step taken, including the termination step. The threshold is 475.
Objective :- The objective is to make the pole stand for as long as possible.

Num	Observation	Min	Max
0	Cart Position	-2.4	2.4
1	Cart Velocity	-Inf	Inf
2	Pole Angle	~ -41.8°	~ 41.8°
3	Pole Velocity At Tip	-Inf	Inf

Figure 1: Catrpole Observation

3 LunarLander

Landing pad is always at coordinates (0,0). Coordinates are the first two numbers in state vector. Reward for moving from the top of the screen to landing pad and zero speed is about 100..140 points. If lander moves away from landing pad it loses reward back. Episode finishes if the lander crashes or comes to rest, receiving additional -100 or +100 points. Each leg ground contact is +10. Firing main engine is -0.3 points each frame.
Objective :- The main Objective is to land on the launch pad in between the flags.

4 Results

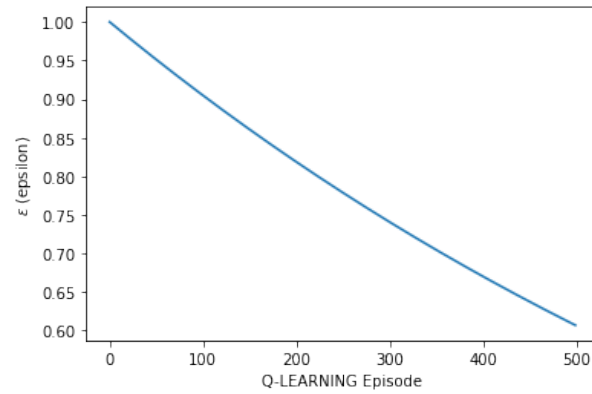We have taken same epsilon decay algorithm. The minimum epsilon is 0.1
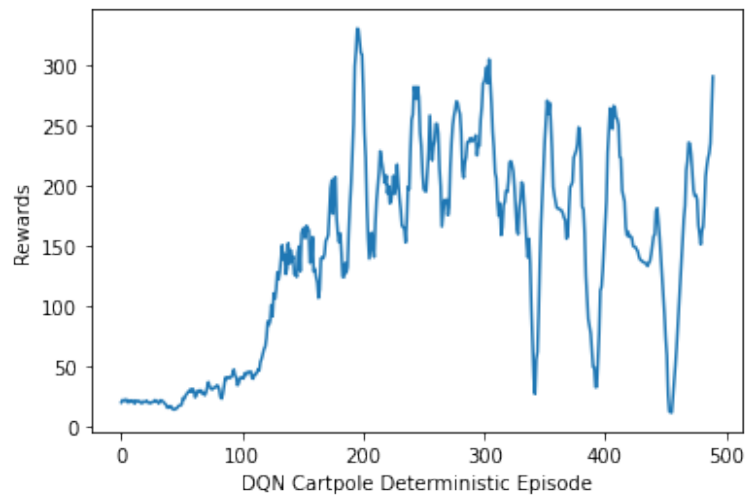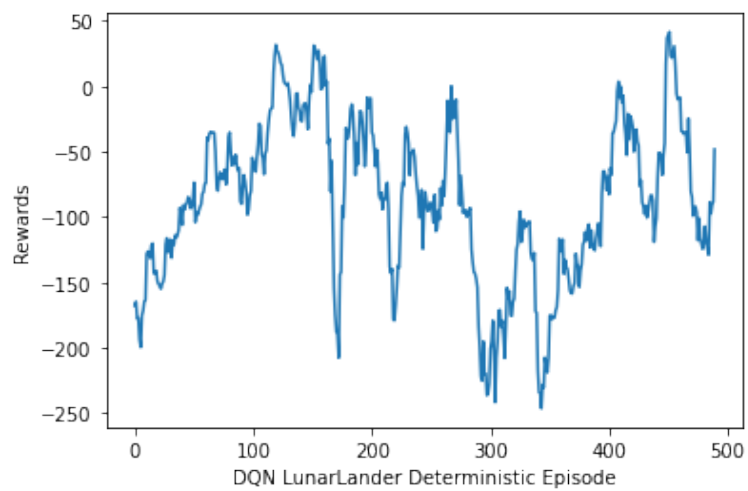

Figure 2: Epsilon decay for all Environments

Figure 3: Cartpole DQN

Figure 4: LunarLanding DQN