Assignment 2

REINFORCE Algorithm for Banana Collector Game

Vedic Partap (16CS10053)

1. Introduction

The objective is to train a machine learning agent to navigate around a large square world, collecting yellow bananas and avoiding blue ones. Given this information, the agent has to learn how to best select actions.

- 0 Move Forward
- 1 Move backward
- 2 Turn Left
- 3 Turn Right

2. Learning Algorithm

REINFORCE algorithm for training the network. It is based on log_probabilities of actions and policy gradient over the rewards

Hyper Parameters:

1. `GAMMA` = 0.9
2. `LR` = 3e-3
3. `UPDATE_EVERY` = 4
4. `EPISODES` = 2500
5. `Optimizer` = Adam

3. Network Architecture

We have used a Sequential neural network with 3 layer. One input, one hidden and one output.

Input Layer: 37 neurons (State Space Size)

Hidden Layer 1: 16 neurons with relu activation

Output Layer: 4 neurons (Action Space Size) with softmax Output

4. Results

Experimentation

1. Two hidden layer: H1(37,16), H2(16,64): The performance was very poor. Stuck in the local minima.
2. Single Hidden Layer

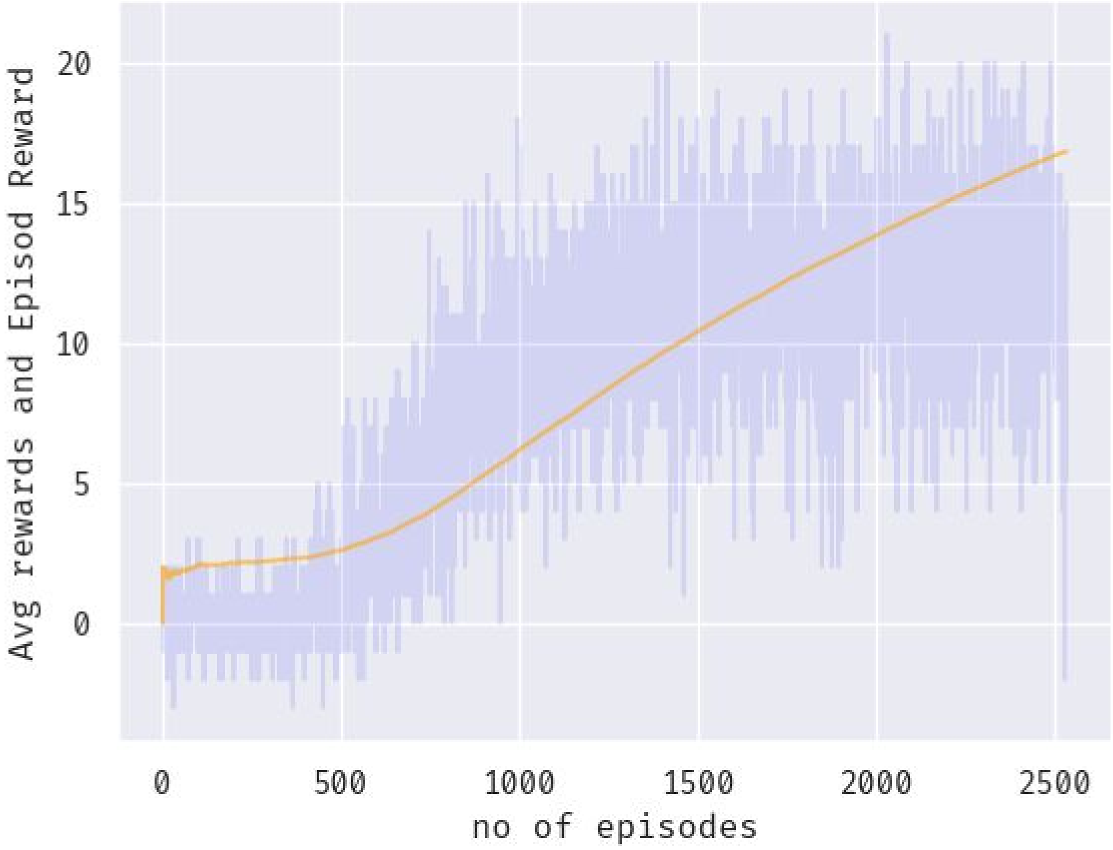

Figure 1: Number of Episodes vs Average Reward(moving average) and Episode Reward

5. Future Work

REINFORCE is much slower when compared to other state of the art algorithms. It is because being only a policy gradient it converge slower. **Improvement**: DQN trains significantly faster than REINFORCE. We may explore other state of the art algorithms too as a part of future work.