PROJECT REPORT

Topic - Navigation Using Deep Reinforcement Learning

Name – Pranav Sivadas Menon

Environment – Banana Collector Environment from Unity Machine Learning Agents toolkit.

Goal – Train an agent to navigate and collect as many yellow bananas as possible while avoiding blue bananas in a large, square world.

Reward structure - . A reward of +1 is provided for collecting a yellow banana, and a reward of -1 is provided for collecting a blue banana.

State space - The state space has 37 dimensions and contains the agent's velocity, along with ray-based perception of objects around agent's forward direction.

Action space - The action space is discrete and the agent can take 4 possible actions:

- **0** - move forward.
- **1** - move backward
- **2** - turn left.
- **3** - turn right.

Stopping Criterion - The task is episodic and is considered solved when the agent gets an average score of +13 over 100 consecutive episodes.

In this project I trained 3 different models:

- Double Deep Q Network with replay buffer and fixed Q target
- Double Deep Q Network with priority experience replay and fixed Q target
- Double Deep Q Network with experience replay , Dueling Q networks and fixed Q target

To learn more about these topics please refer the following papers:

- Deep RL
- Deep RL with Double Q Learning
- Deep RL with Prioritized Experience Replay
- Deep RL with Dueling Networks

The neural networks in all the aforementioned model consisted of the following architecture:

Input – vector 37 X 1 given by environment

2 hidden layers.

First fully connected layer had 128 units and used Relu nonlinearity.

Second fully connected layer had 64 units and also used Relu nonlinearity.

Output – fully connected layer with single neuron for each action

Hyperparameters chosen:

General:

lr = 5e-4

batch_size = 64

capacity = int(1e5)

tau = 1e-3

gamma = 0.99

UPDATE_EVERY = 4

For PER

alpha = 0.6

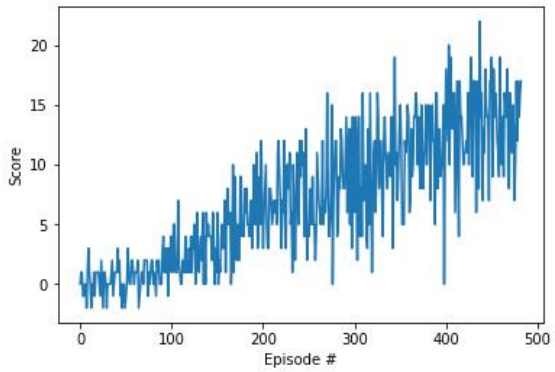
beta = 0.4

beta_update = 0.001

e = 0.01

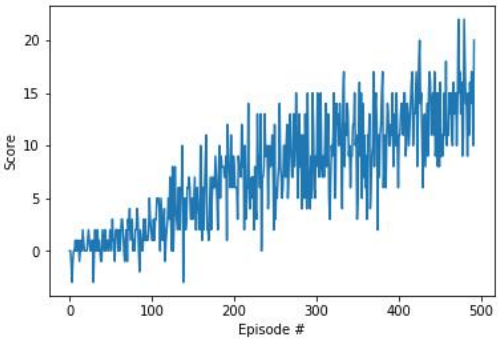
Model 1 - Double Deep Q Network with replay buffer and fixed Q target

(Best Performance)



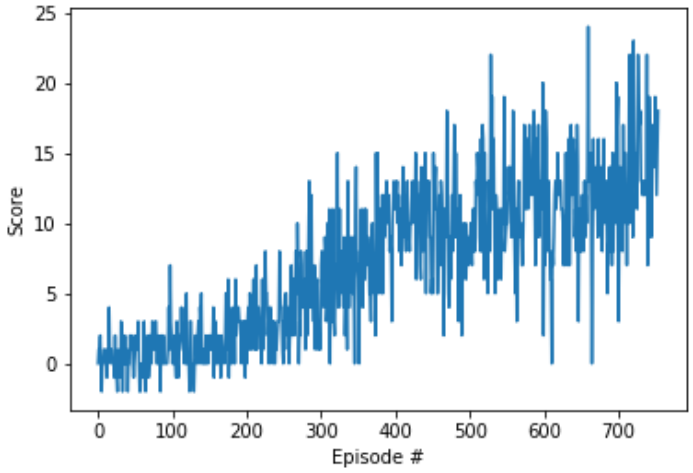
Environment solved in 383 episodes.

Model 2 - Double Deep Q Network with experience replay , Dueling Q networks and fixed Q target



Episode was solved in 393 episodes

Model 3 - Double Deep Q Network with priority experience replay and fixed Q target



Environment was solved in 654 episodes

Future Work

- Test Noisy networks on environment
- Test Rainbow method on environment
- Experiment with hyperparameters