RL Project Report: Banana Collection

Learning Algorithm:

The learning algorithm implemented here is the standard deep-Q learning (dqn) where neural networks are used to estimate the q values during training. The hyperparameters used include the following:

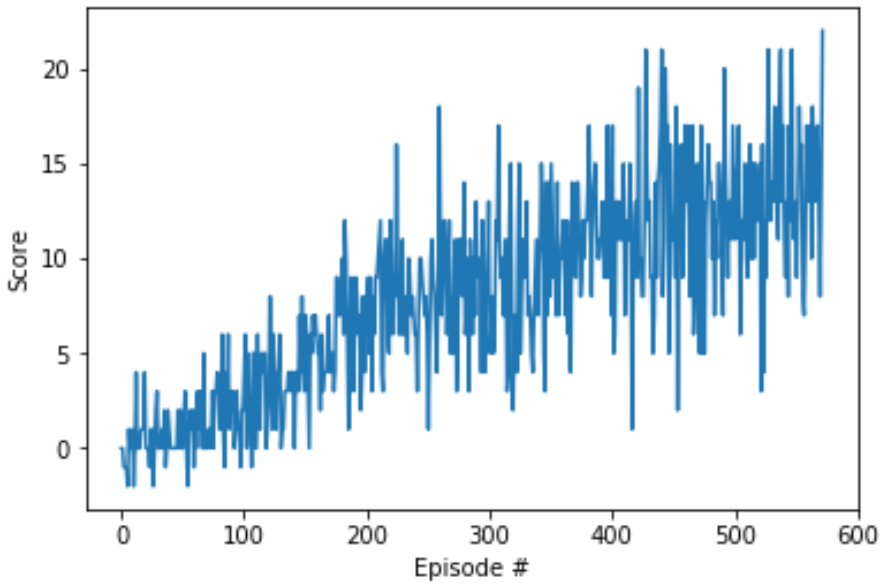
- Buffer Size: 1E5 (dictates how big the replay buffer is)
- Batch_Size: 64 (dictates how many samples to pull from replay buffer for training)
- Gamma: 0.99 (used to calculate q values for the target neural network)
- Tau: 5E-3 (used to update the target neural network's parameters)
- Learning Rate: 5E-4 (dictates how fast the neural network learns)
- Update_Every: 4 (dictates how often the target neural network is updated)
- Eps_start: initial percent value for epsilon greedy decision making
- Eps_decay: how quickly the epsilon value decreases
- Eps_end: the minimum value for epsilon greedy

The learning algorithm uses a neural network to estimate the q values and a target neural network as the target. Both have the same architecture. It contains

- Input Layer: state of the environment
- Hidden Layer 1: 64 units
- Hidden Layer 2: 64 units
- Output Layer: action size (4)

ReLu activation is used throughout the neural network.

Plot of Rewards:



Future Works:

1. Implement improvements of DQN algorithm such as double DQN, dueling DQN, and prioritized experience replay. These improvements can potentially help with the algorithm to train faster or achieve higher scores.
2. Further hyperparameter tuning for training.
3. Adjust neural network size for better generalization.