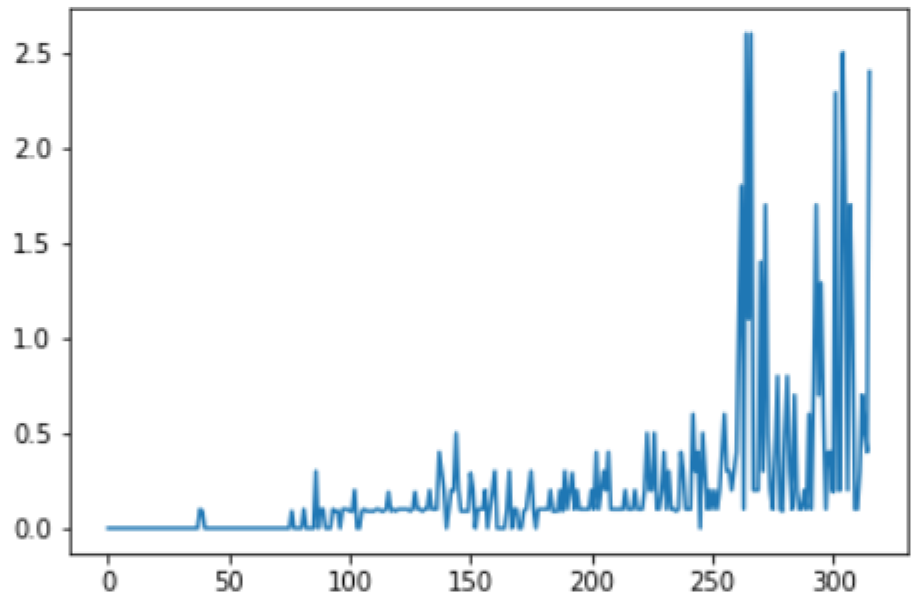
1. Model
    a. We used the DDPG algorithm for this project, slightly modified for a multi-agent implementation.
    b. The architecture for the model follows an actor critic implementation. The actor has three fully connected layers.
        i. Actor layer 1 size: 24 x 400
        ii. Actor layer 2 size: 400 x 200
        iii. Actor layer 3 size: 200 x 2
    c. The critic has three fully connected layers as well
        i. Critic layer 1 size: 24 x 400
        ii. Critic layer 2 size: 424 x 200 (concat action)
        iii. Critic layer 3 size: 200 x 1

2. Hyperparameters

    The hyperparameters used to train the model are as follows:
    BUFFER SIZE = 100000
    BATCH SIZE = 96
    TAU = 0.01
    LR_ACTOR = 1e-4
    LR_CRITIC = 1e-4
    W_DECAY = 0
    UPDATE_EVERY = 1
    GAMMA = 0.99

3. To improve the results, we can try the following approaches:
    a. Try to implement a decentralized actor-critic where they don't predict both actions at the same time.
    b. Try out other multi-agent implementations of traditional RL algorithms like PPO.
    c. Implement a prioritized experience replay.

4. Results and evaluation
    a. The environment was solved in 315 episodes
    b. Here is a plot of the rewards


2.5
2.0
1.5
1.0
0.5
0.0
0
50
100
150
200
250
300