ECS 170 Artificial Intelligence

PA3 Learning To Play Pong Report

Vishnu Rangiah
916562849
March 9, 2022

PART 1: Problem Representation

1. It would be presumably easier to learn from the RAM since it would offer a more accurate representation of each frame of the game. Instead, we would have to pre-process each image which would be more intensive. Also since the RAM is only 128 bytes vs the greater amount of data stored in the images our network would have to process more data.

2. The purpose of the NN is to simplify the process of choosing the optimal action for the agent by training. Instead of keeping track of a Q-Learning table, the neural network learns the best action to perform given a state by changing the weights of the neural network to favor actions that give a higher reward and output the corresponding Q values for the actions.

   The input of the NN is the screen display and the size of the input is mapped to the 32 nodes layer of the network. The NN consists of additional layers and ReLU activation functions. The output of the NN is the action space or the 6 different actions that the agent could perform.

3. The purpose of lines 48 and 57 of dqn.py is to simulate the expectation vs exploration property such that the Q learner can learn new paths given the particular epsilon value. As shown in the code, our model favors exploration in the beginning then adheres to exploitation later. The epsilon value determines how much we will favor exploration vs exploration.

4. (Programming) act function

PART 2: Making the Q-Learner Learn

1. The loss is a measurement between the Expected Q-table and the current Q-table. y - represents the expected Q values. Q(s, a, Theta) represents the current Q - table. s - state, a - action, and Theta are the model parameters. The loss is used in the gradient function to adjust the weights of the model to favor ideal actions.

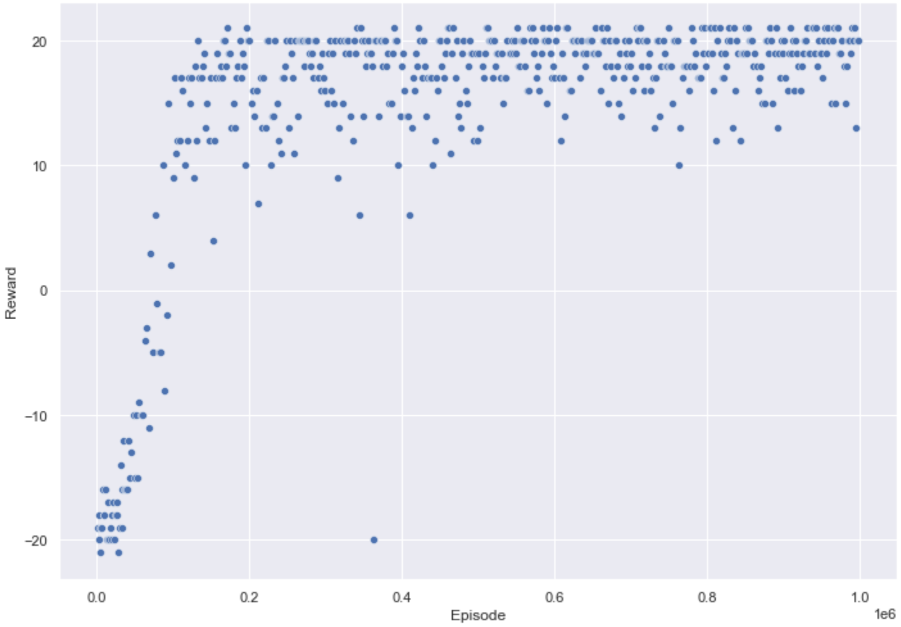
2. (Programming)

PART 3: Extend the Deep Q-Learner

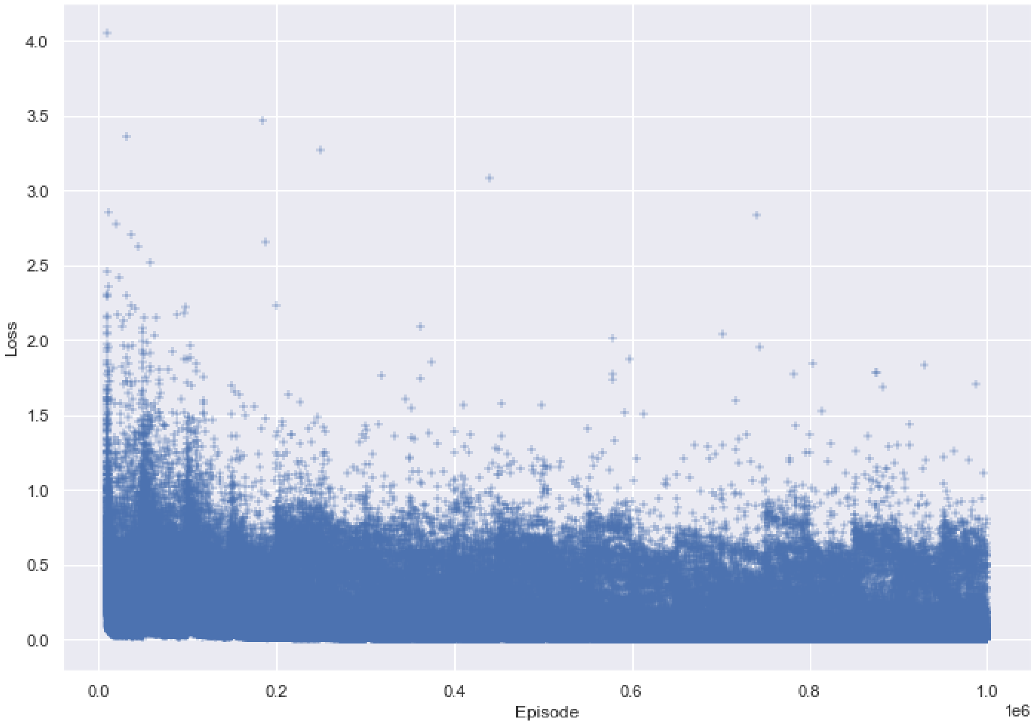
1. (Programming) Implement Replay Buffer

PART 4: Learning to Play Pong

1. Reward Graph
    a. The graph below shows that the Reward increases and reaches around 18 overtime with 1M frames.



2. Losses Graph
    a. The graph below shows that the losses from the model decrease over time with 1M frames.


4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Loss
0.0
0.2
0.4
0.6
0.8
1.0
1e6
Episode