Reinforcement Learning HW-3

Name: Li, Jieda
NetID: jlg7773

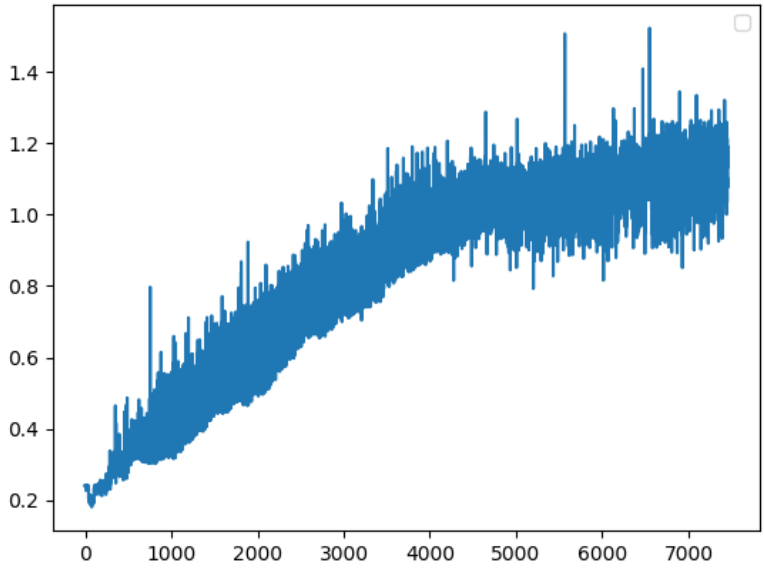
Problem 1: Breakout

Code:https://github.com/jiedali/reinforcement_learning_jieda_li/blob/master/hw3/breakout/breakout_run10.ipynb

I implemented the DQN algorithm using the typical CNN network with 3 convolutional layers, a flatten layer and 2 fully-connected layers. The final tunned parameter that gives the best results are as following: learning rate = 0.00025, RMSProp optimizer, replay buffer size of 400K, target network update frequency is 1000 optimization steps, discount rate: 0.99.

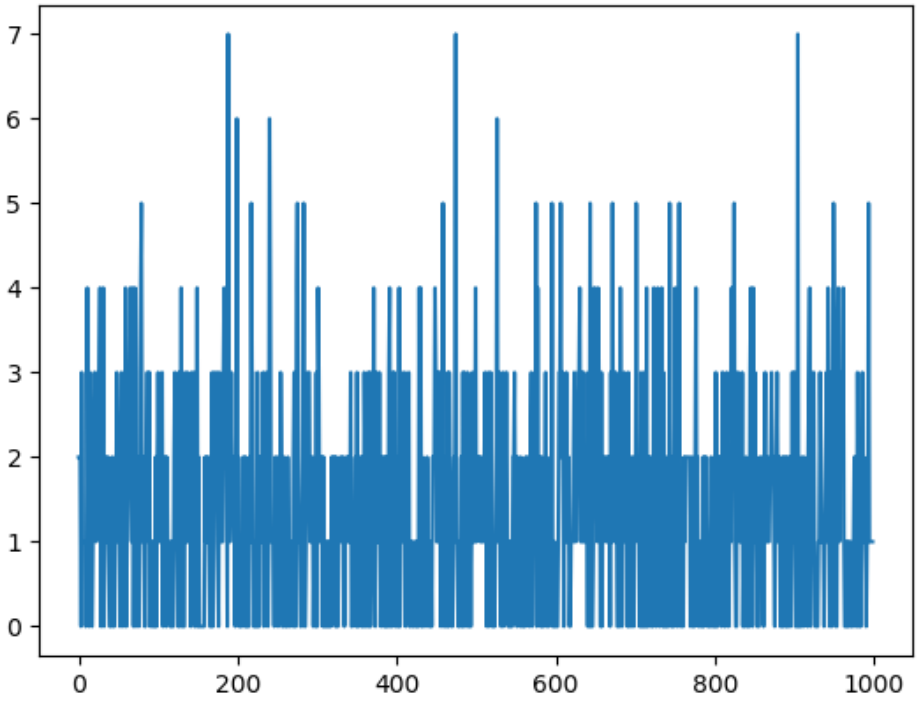
The best and longest training curve I can get is as following (~10 hours, due to server/memory issue). The best Q values I can achieve is around **1.45**

Breakout training mean max_Q



I use the trained neural network and played 1000 episode of games. The reward for each episode is plotted below: x-axis is the episode index, y-axis is un-discounted total reward of one game. Reward for each episode is between 0 and 7, the maxim single episode reward that the agent ever gets is ***7***

Breakout Agent Play 1000 games
(Maximum Single episode reward is ***7***)

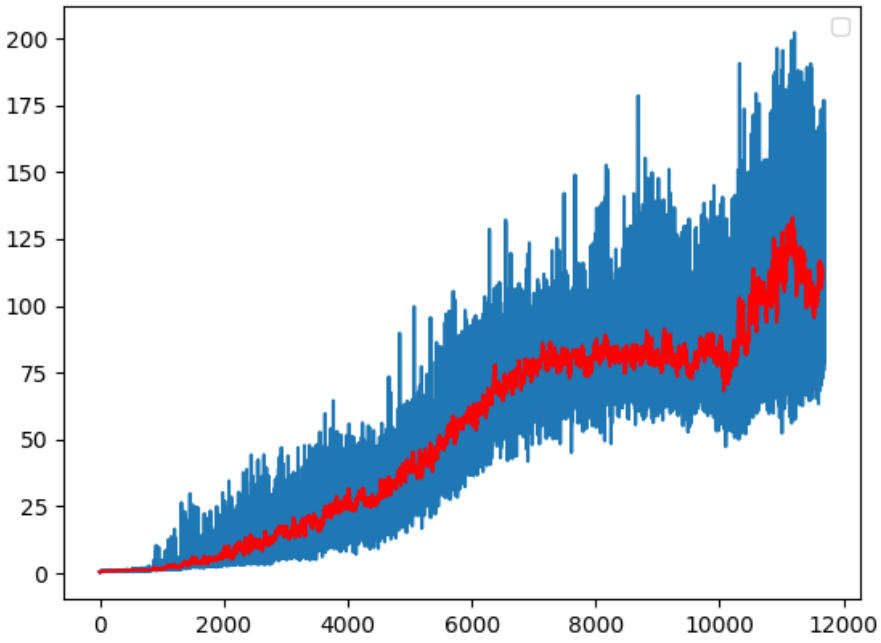


Problem 2: MsPacMan

Code:https://github.com/jiedali/reinforcement_learning_jieda_li/blob/master/hw3/MsPacman/mspacman.py

I implemented the DQN algorithm using the typical CNN network with 3 convolutional layer, a flatten layer and 2 fully-connected layer. The best and longest training curve I can obtain is as following (about 12 hours training time). X-axis is training episode; y-axis is mean max Q value. Red curve is the moving average of the mean max Q from the last 30 episode. The best Q value achieved is around **200**.

MsPacman training mean max_Q



I then use the trained DQN network and played the game for 1000 episode. Here is the undiscounted total reward for the first 250 games. The reward is in the range between 380 and 2380. The max reward agent ever gets is **4550**

MsPacman Trained Agent Play