Assignment 3 – Deep Reinforcement Learning

Vishal Patil

10th Dec 2021

1.Assignment overview:

Goal of the assignment is to apply the Reinforcement learning algorithm: Q-Learning to 'teach' an agent to maximize the reward for the given dataset set and the stock trading environment. The goal of the agent is to learn the trends in the stock price and perform a series of trades over a period of time and end with a profit

2. Python Editor:

Jupyter Notebook IDE was used for all the coding implementation.

3. Data set:

Dataset on the stock price of Nividia for the last 5 years. The dataset has 1258 entries starting 10/27/2016 to 10/26/2021. The features include information such as the price at which the stock opened, the intraday high and low, the price at which the stock closed, the adjusted closing price and the volume of shares traded for the day.

4. Implementation:

4.1: Q-Learning Algorithm:

For any finite Markov decision process (FMDP), *Q*-learning finds an optimal policy in the sense of maximizing the expected value of the total reward over any and all successive steps, starting from the current state. *Q*-learning can identify an optimal action-selection policy for any given FMDP, given infinite exploration time and a partly-random policy. "Q" refers to the function that the algorithm computes – the expected rewards for an action taken in a given state.

Q-table: It is a table that is made up of rewards for a certain state-action pair. The dimensions of this table are (no. of states x possible actions). Bellman equation is used to find the optimal values that go into this Q-table at a particular step in the training phase of the model. The training phase is ran for a certain number of "episodes" till the reward values converge. In the evaluation phase, the optimal filled table is referred to decide upon an action at a particular state in order to achieve maximum reward.

4.2: Environment:

Given environment is a stock trading environment which has 4 observations: (0, 1, 2, 3) which represents 4 states of the environment. These states are: 1. prices increased, stocks held 2. prices decreased, stocks held 3. prices increased, stocks not held 4. prices decreased, stocks not held.

It has 3 actions: 1. Buy 2. Sell 3. Hold

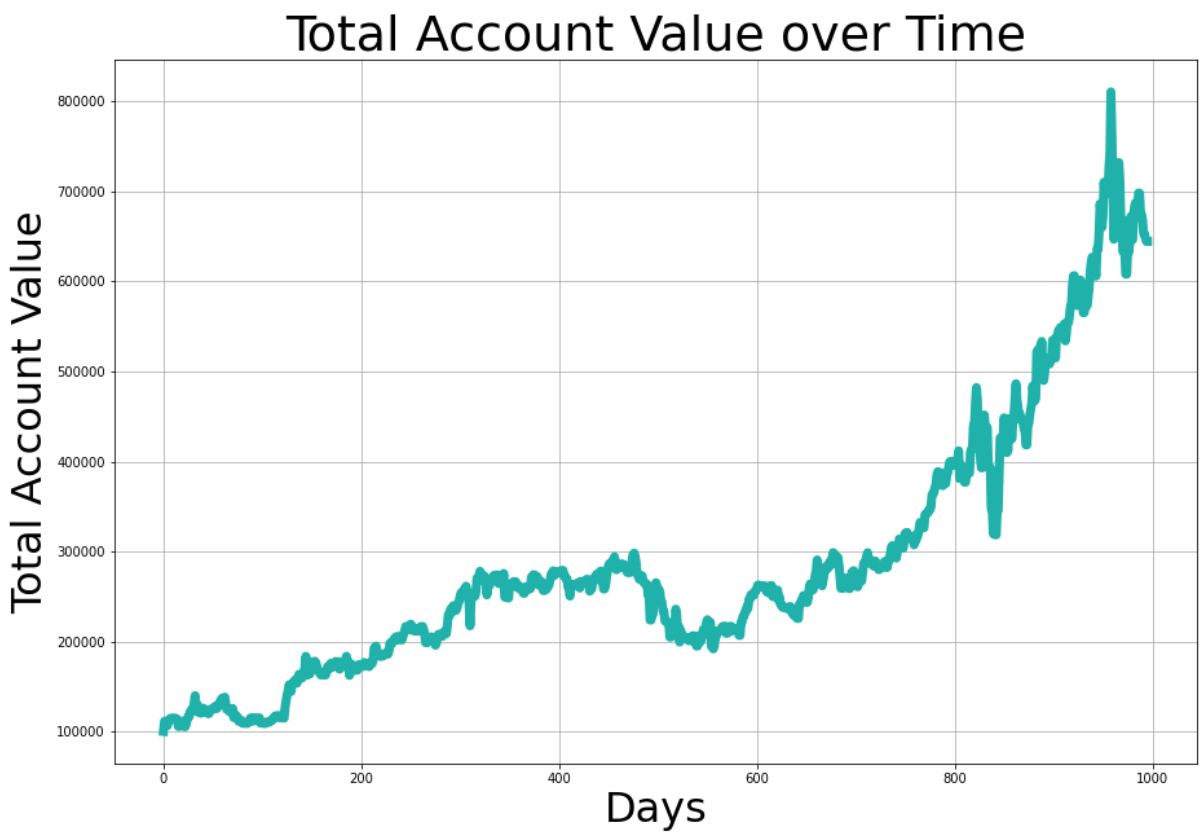
A Q-learning agent is a value-based reinforcement learning agent that trains the model to estimate the return or future rewards.

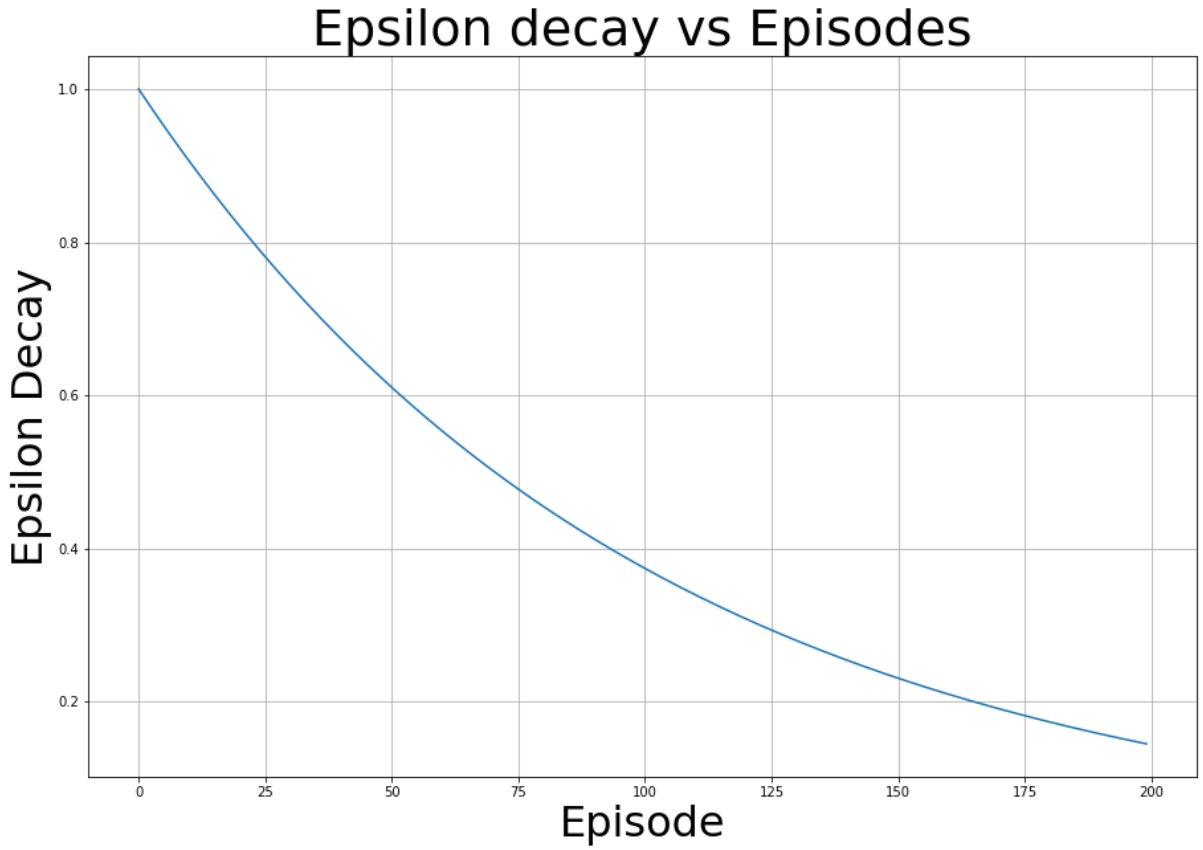
Goal of the Q-learning algorithm is to maximum the possible rewards and in the case of the given environment, make profit with the given account value (100000).

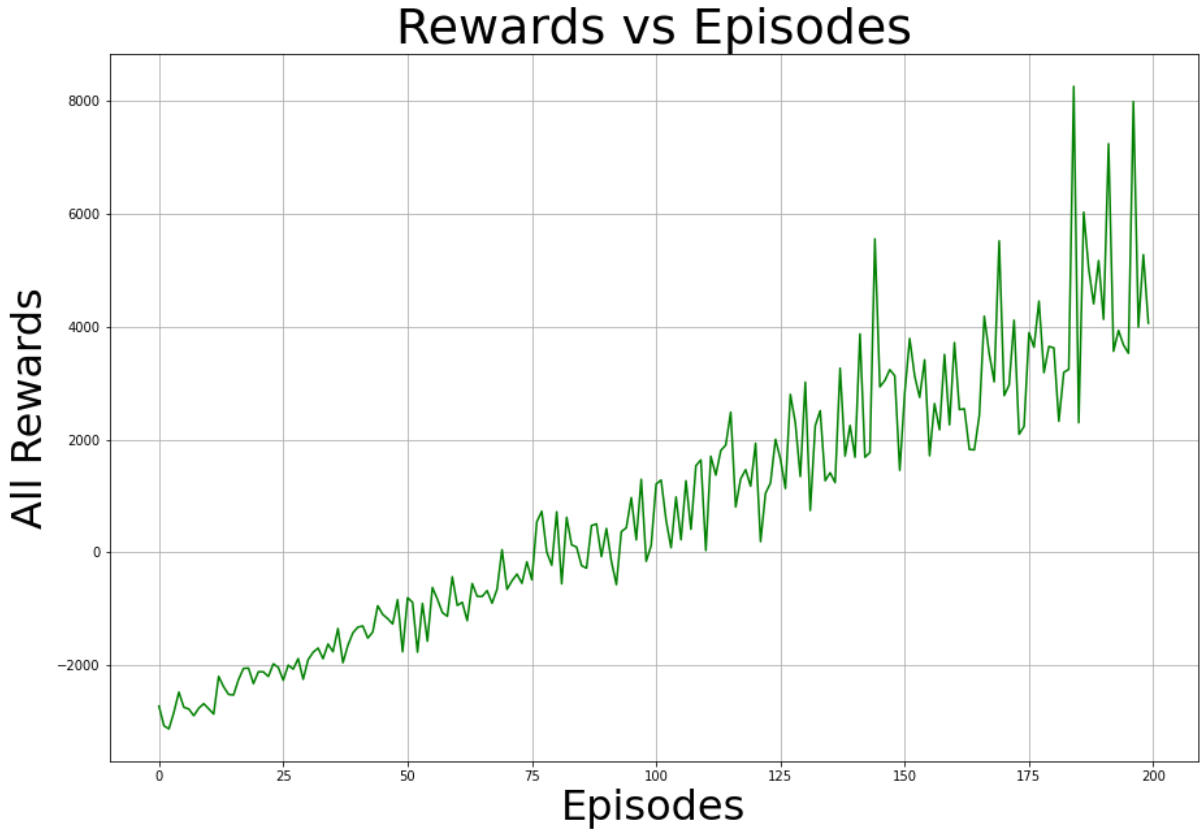
Rewards is basically the reinforcement given by the environment when the agent selects an action that is favourable to the goal and penalty is the opposite of that i.e. the environment “punishes” the agent then for making a wrong choice.

4.3 Training results:

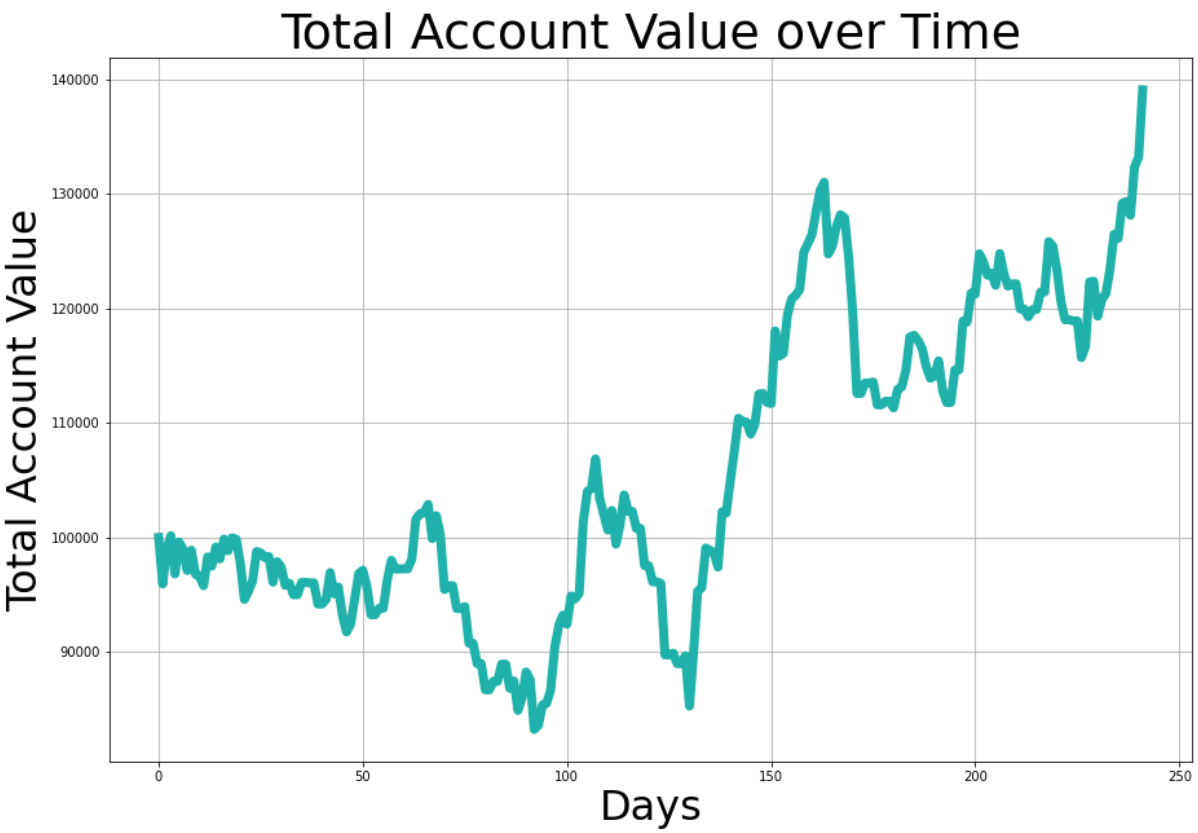
stock_trading_environment.total_account_value: 644557.8033270002







4.4 Evaluation results:



stock_trading_environment.total_account_value: `139095.30966699996`