Assignment 2 Report

Introduction:

The Objective of this project isto use Reinforcement Learning technique to predict the optimal action (out of Buy, Sell, Hold) to take at a given time for some stock of a company.

This is achieved by creating a Agent that trains on data and creates an optimal policy that can effectively take optimal decision. This is done by the use of Q-learning technique and Neural Networks, hence the method is known as Deep Q Learning.

Implementation:

1. **Deep Q Network Class:**
   This consists of functions associated with the Deep Q Network. This is used to mainly form two Q networks: namely Evaluation and Target Q networks.

   The Init function initialises the NN using Pytorch Framework and sets the parameter values. The NN uses two fully connected layers (except the input layer), uses Adam as optimiser and MSE as the error matrix.

   The forward function evaluates the Q value for various actions when given a state as input.

2. **Agent Class:**
   This mainly deals with all the functions associated with the agent.

   The Init function mainly initialises the agent with all the parameters involved, and also the two Q networks mentioned above, it also initialises the replay memory which is used to store the experiences of the agent, (These experiences will be combined in batch and uses to train the model)

   The store transition function stores the transitions while the choose Action function takes an action according to the epsilon greedy strategy.

The learn function function takes a batch of experiences from the replay memory, uses the Q networks and Bellman equation on the State-Action pairs and updates the Model.

3. **Episode:**
   In each episode we parse through the dataset once.

4. **State:**
   State at time t is defines by the data of previous n days.

5. **Reward:**
   The Agent gets reward based on the net value of the profile previous day and the value that day.

Working:

The main components involved are Evaluation Q network, Target Q network , Replay memory.
While training for each episode we iterate through the states, in each state we take an action, based on the action we update the parameters like total money stock bought etc.

We calculate all the values of the experience tuple consisting of State, Action, next_state, reward and update it into the replay memory, using the above defined functions. We further train the model on the NNs by selecting a batch of experiences from the replay memory again by using the above mentioned functions.

The Agent initially has epsilon value 1 which decays to epsilon min gradually, Implying that initially it will prefers to choose actions based on exploration but gradually it will exploit its learnings.

This helps in forming Optimal Policy based on which we evaluate the model on the test Dataset.