Optimising Stock Trading Strategy With Reinforcement Learning

~ Anweasha Saha

Problem Statement:
The goal is to find an investment strategy on a single exchanged traded fund (SPY) that would maximize our cumulative wealth.

Data Acquisition and Understanding:

Data source:
The stock market performance of the last ten years has been considered here. The reinforcement learning trading strategies have been tested on the S&P 500 from January 4, 2010 to December 30, 2020, using the index fund SPY as the trading instrument.
The data has been extracted from Yahoo Finance, exporting the following features: date, open price, and close price.

Data Preparation:
One of the primary purposes of data preparation is to ensure that raw data being readied for data processing and analysis is accurate and consistent.
After data collection, the next steps are data cleaning, data analysis, data visualization and Exploratory Data Analysis (EDA). In the next phase, the dataset will be split into train and test data.

Modeling:

Methods:
In this project, we explore and compare the potential of three RL (Reinforcement Learning) algorithms:

- discretized Q-learning,
- Hill Climbing, and
- Deep Q-learning,

to optimize stock trading strategy and maximize our cumulative wealth.
As input, we define two versions of the Markov Decision Process (MDP), one with discretized state space for Q-learning and Hill Climbing and one with a continuous state space for Deep Q-learning. Hill Climbing is a policy-iteration based algorithm, whereas discrete Q-learning and Deep Q-learning are both value-iteration based algorithms.

For discretized RL techniques (tabular Q-learning and Hill Climbing), our MDP will be as follows:

- State s: set of states (u,t). In total we have 4 possible states (u=1, u=0, t=1, t=0).
- Action a: set of actions on the SPY stock. Available actions include: buying, holding or selling.
- Reward r: the cumulative wealth

For Deep Q-learning method, our MDP will be given as:

- State s: set of states (r,t). In total we have 4 possible states (r=-1.0, r=1.0, t=1, t=0).
- Action a: same as the MDP given above
- Reward r: the change in wealth

We train these models using the training data we have at hand.

Deployment:

In order to start using a model for practical decision-making, it needs to be effectively deployed into production. If we cannot get practical insights from our model, then the impact of the model is severely limited. It is one of the last stages in the cycle.

We conducted many different experiments with the three different models against the test data. Our primary metric for measuring model performance was the ending portfolio value on the validation set, as this gives us a direct way to compare each of our policies against other well-known trading strategies.

Conclusion:

We find that Hill Climbing is the most stable method compared to the value-search-based methods: Q-learning and Deep Q-learning. Hill Climbing manages to find a policy that outperforms Q-learning and Deep Q-learning. This may be due to the fact that Q-learning and Deep Q-learning are value-search-based approaches that highly depend upon our estimation of the value function and state space definition.