Assignment 2: DDPG and SAC for ATM European and American put option Hedging

Zhu Zheng, 20916267, zheng.zhu@connect.ust.hk

He Xinyi, 20914738, xhebm@connect.ust.hk

(GitHub: https://github.com/ZHUZheng1999/Reinforcement-learning-in-financial-applications/tree/main/option%20hedging)

1. Introduction

Option hedging is an effective method to reduce investment risk by using the correlation between the option and the underlying risky asset. This assignment uses Deep Deterministic Policy Gradient (DDPG) and Soft Actor Critic (SAC) to train models, aiming at finding the optimal proportion of stocks to hedge European and American put options. Two types of stocks are taken into consideration, whose returns are supposed to follow binomial distribution and GBM, respectively.

2. Problem Formulation

2.1 Binomial Model

The binomial option pricing model assumes that the stock price moves in only two directions, and that the probability and level of each upward or downward move in the stock price remains the same throughout the whole period. The parameters are calculated as follows:

$$u = e^{\sigma\sqrt{\Delta t}} \qquad d = e^{-\sigma\sqrt{\Delta t}} \qquad R = e^{r_f \Delta t} \qquad p = \frac{R-d}{u-d}$$

In this assignment, we assume that stocks can be traded in a daily basis, and there are 252 trading days in one year, so Δt is set as $1/252$ in the training models. This assignment first simulates the stock price in each step, and then deduces the binary tree of price of corresponding European option. For American option, in each period we compare the payoff of the early exercised option with the discounted price of the option, and then update the binary tree to get the initial price of American option.

2.2 GBM Model

Geometric Brownian Motion (GBM) is often used to simulate stock price movements. The stochastic process model of stock prices can be expressed as follows:

$$\frac{\Delta S}{S} = \mu\Delta t + \sigma\varepsilon\sqrt{\Delta t}$$

The above stochastic differential equation (SDE) has an explicit solution of the following form:

$$S_t = S_0 * \exp\left(\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t\right)$$

where the constants μ and σ correspond to the return and volatility of the stock price, respectively. The solution equation is used to iteratively compute the stock price of each time step in the following manner:

$$S_t = S_{t-1} * exp\left(\left(u - \frac{\sigma^2}{2}\right)t + \sigma\varepsilon\sqrt{\Delta t}\right) \text{ where } \varepsilon \sim N(0,1)$$

The price of European options is calculated according to the BS model shown below:

$$d_1 = \frac{\log(S_t/K) + \left(r + \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}} \qquad d_2 = d_1 - \sigma\sqrt{T}$$

$$P = Ke^{-rT}N(-d_2) - S_tN(-d_1)$$

Similarly in Binomial model, we compare the current payoff and the discounted price of put option at each time step to get the early exercise time for American option.

3. Methodology

3.1 Deep Deterministic Policy Gradient

DDPG is a reinforcement learning technique that combines both Q-learning and Policy Gradient. DDPG consists of actor and critic networks. The actor is a policy network that takes the state as input and outputs the exact action (continuous), instead of a probability distribution over actions. The critic is a Q-value network that takes in state and action as input and outputs the Q-value. The optimal action is taken by taking argmax over the Q-values of all actions. And for actor, the loss is simply the sum of Q-values for the states. For critic, loss is a simple TD-error where we use target networks to compute Q-value for the next state and we need to minimize it. In DDPG, we perform a "soft update" where only a fraction of main weights is transferred in the following manner:

$$\Theta_{arg}^{\mu} \leftarrow \tau\Theta_{arg}^{\mu} + (1-\tau)\theta^{\mu}$$

$$\Theta_{arg}^{Q} \leftarrow \tau\Theta_{arg}^{Q} + (1-\tau)\theta^{Q}$$

3.2 Soft Actor Critic

Soft Actor Critic is an off-policy actor-critic deep reinforcement learning algorithm based on the maximum entropy reinforcement learning framework. In this framework, the actor aims to maximize expected reward while also maximizing entropy, that is to succeed at the task while acting as randomly as possible. It consists of actor, critic, policy network. Different from DDPG, SAC combines stochastic and value learning.

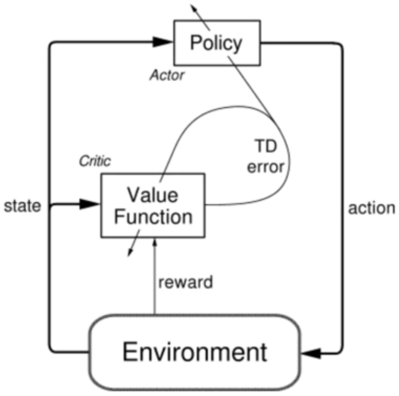

Figure 1. Process of Soft Actor Critic

4. Results

This assignment trains DDPG model $N = 18000$ episodes for European Binomial model at step $T = 60$ and $N = 20000$ episodes for American Binomial model at step $T = 80$. For GBM model, we train both DDPG and SAC models $N = 15000$ episodes for European put option at step $T = 20$, and $N = 15000$ episodes for

American put option at step $T = 60$. Due to the more information need to learn in American put option, that is the time when to early exercise, we need to train more epochs and larger time step to get the convergence of each model. For Binomial model, the state is the list of both time step and current stock price; while for GBM model, the state is the list of both time to maturity in year basis and current stock price. Two models have the same setting in action and reward. The action space is $\{0,0.01,0.02,\ldots,0.99,1\}$ with 101 uniformly elements, and the reward is negative of absolute value of the portfolio payoff by having 1 unit of long position in ATM put option and long the underlying stock. In this setting, when we can more accurately replicate the return of the put option with the stock, the reward of the model is larger.

4.1 Delta performance

For Binomial model, the delta performances are shown as below for each setting. The action of stock hedging trained in European put option is generally consistent with that of delta hedging. However, the results of American put option show that when strike price is 50 and stock price is less than 45, the model will consider early exercise by taking delta close to 0, which is in accord with the situation of early exercise in American put option in the financial market. After 20000 episodes training in DDPG algorithm, the American put option usually exercise early average at time step 54.5 of total 80 steps, and the early exercise ratio among 50 simulations is 0.48.

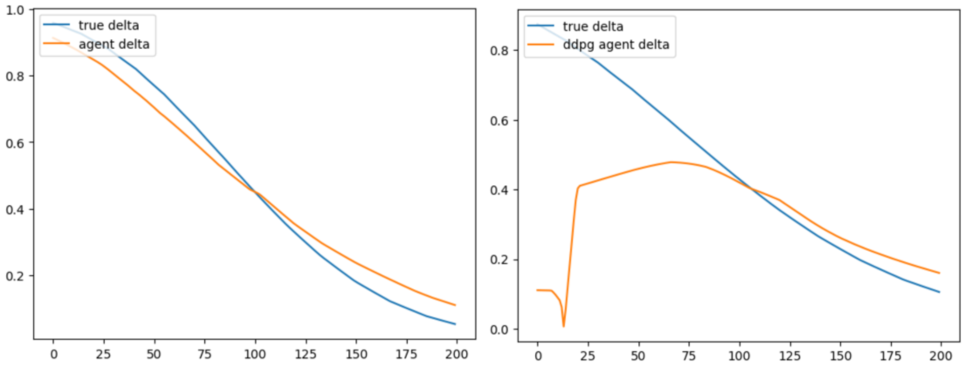

Figure 2&3. Delta Performance of Binomial Model for European and American Put Option from DDPG

According to the delta performance of GBM Model, DDPG performances better in both European and American put option, whose final action is much closer to optimal action in delta hedging, compared with SAC algorithm under the same environment setting. From the Figure 5, American put option is supposed to early exercise when stock price is less than 45 with strike of the option is 50. Based on training results after 15000 episodes in GBM model using DDPG algorithm, the American put option usually exercises early average at time step 39.3 of total 60 steps, and the early exercise ratio among 50 simulations is 0.54; while in SAC algorithm, the

American put option usually exercises early average at time step 38.3 of total 60 steps, and the early exercise ratio among 50 simulations is 0.4.

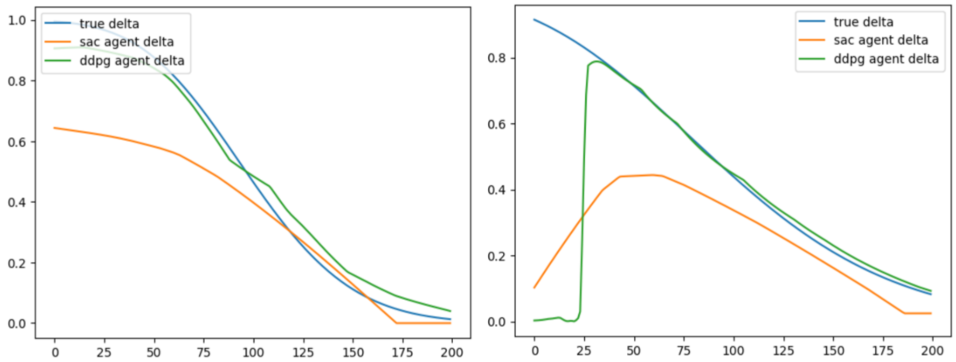

Figure 4&5. Delta Performance of GBM Model for European and American Put Option from DDPG and SAC

4.2 Convergence of Models

As can be seen from the following four figures, each model under different settings all has reach its convergence. By comparing the model under Binomial distribution stock, the hedging of European put option is more successful, that is to say, the return of European option is replicated more accurately using stocks. For GBM model, the reward of European option is generally larger in both DDPG and SAC algorithm. What's more, DDPG achieves convergence faster than SAC and has better hedging performance in both European and American put option.

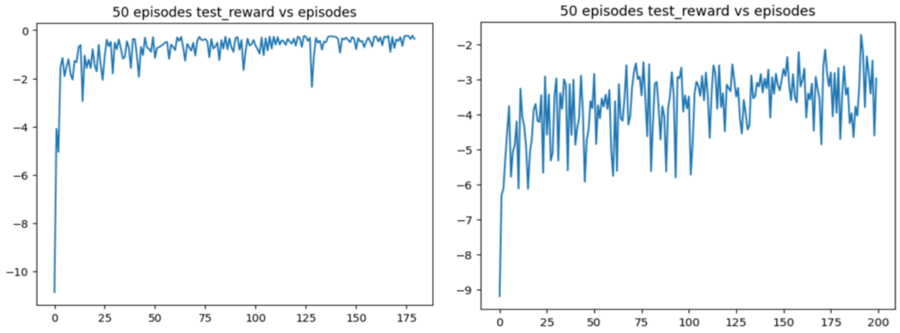

Figure 6&7. Test Reward of Binomial Model for European and American Put Option from DDPG

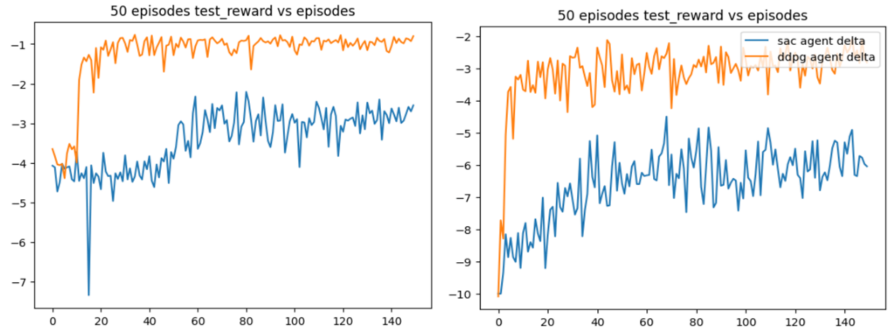

Figure 8&9. Test Reward of GBM Model for European and American Put Option from DDPG and SAC

5. Conclusion

This assignment successfully hedges European and American put option using DDPG and SAC algorithm in both Binomial stocks and GBM stocks. By comparing the results of delta performance, European is easier to be replicated using both types of stocks, and DDPG performances better than SAC in hedging put option under GBM model.

Acknowledgement

This group assignment is completed by HE Xinyi and ZHU Zheng with equal contribution. We had lots of face-to-face discussions to design the algorithm and the experiment. HE focused more on the design of the stock and option environment and ZHU focused more on the DDPG and SAC agent. In addition, HE and ZHU together finished the experiments and visualization of the results. Also, we would like to express our sincere appreciation to Professor Chak WONG and TA because of their patient help.