Portfolio Management

- Siddharth Chauhan (IMT2021046)
- Kushal Partani (IMT2021062)

Problem Description

Risk- Adjusted Returns

- The problem to solve is to obtain profitable returns on your investment while controlling the risk due to market volatility as well.
- Optimize the portfolio by using different currencies and different stocks and showing its performance against the well-known (baseline) strategies used in the market.

Dataset Description

Dataset

- We used the OHLCV(Open,High,Low,Close,Volume) dataset of multiple stocks and currencies.
- To make the model more robust while training we used multiple environments which samples different stocks.
- While backtesting the model we used a completely different set of stocks.

Mathematical Formulation

Action Space ; Weights

- **Weights:** Set of **n** variables corresponding to **n** stocks which determine relative percentage of current portfolio valuation **V** to be assigned to the i^{it} stock.
- Weights are chosen from a **set of predefined weights** which are formed using balls-in-boxes method.
- For instance, ways to distribute 4 balls in 2 boxes could be either [0,4],[1,3],[2,2],[3,1],[4,0], which after normalization yields [0,1],[0.25,0.75],[0.5,0.5],[0.75,0.25],[1,0].

$$0 \leq w_i \leq 1 \qquad \sum_{i=1}^{n} w_i = 1$$

Action Space ; Holdings

- **Holdings:** After getting assigned the weight w_i, holdings determine what part of the capital assigned to a stock should be invested in stock or kept liquid.
- For instance, holding value of 0.4 implies 40% of the assigned capital should be invested in the stock while keeping the rest 60% liquid.
- After assigning both weights & holdings, the shares held of each stock (s_i) can be found as follows:

$$0 \leq h_i \leq 1$$

$$s_{i,t} = \frac{h_{i,t} w_{i,t} V_{t-1}}{p_{i,t}}$$

where $p_{i,t}$ is the opening price of stock i at timestep t,

V_{t-1} is the portfolio valuation at timestep t-1

Actions; Buy / Sell?

- From our formulation of the problem, an action can be considered buy, hold or sell for the ith stock as follows:
- Buy: $s_{i,t} > s_{i,t-1}$
- Sell: $s_{i,t} < s_{i,t-1}$
- Hold: $s_{i,t} = s_{i,t-1}$

Action Space

- Our model has flexibility to choose both w_i and h_i for all stocks at each timestep.
- Discrete action space for holdings with number of possible holdings say k, (for eg k=5 implies possible holdings are [0,0.25,0.5,0.75,1]).
- Discrete action space for weights is equivalent to b balls in n boxes.

$$|A| = \binom{b+n-1}{b-1} k^n$$

Most of the models used further are trained with b=4 , n=2 and k=5, yielding a total action space size of 125.

State Space

- The financial indicators which were used were
  - RSI (Relative Strength Index)
  - MACD (Moving Average Convergence/Divergence)
  - CCI (Commodity Channel Index)
  - Slow k (Slow Stochastic Momentum Oscillator(%k)
  - ATR (Average True Range)
- The state space consists of these indicators along with OHCLV data for a particular lookback.
- **State = S_t = (X_t, w_{t-1})** where X_t is an numpy array of shape **= (indicators,num_stocks,lookBack)** and w_{t-1} is the weights of previous time step.

Rewards

- Let the portfolio value at timestep t be V_t and T is the Reward Lookback.
- The reward is

$$R_t = \sum_{n=1}^{T} log(\frac{V_t}{V_{t-n}})$$

Metrics

Metrics

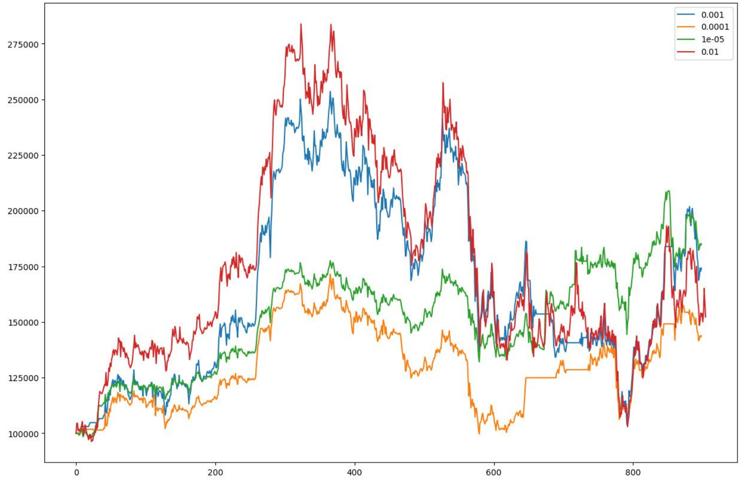
- **Max-Drawdown:** It measures the largest peak to trough loss in the portfolio.
- **Sharpe Ratio:**The Sharpe Ratio measures the risk-adjusted return of an investment or trading strategy. It quantifies the excess return generated per unit of volatility or risk.
- **Sortino Ratio:**The Sortino Ratio also measures the risk-adjusted return of an investment.It focuses only on the downside risk, ignoring the volatility associated with upside movements.
- **Returns** : Indicates the profit or loss generated by an investment over a certain period.

Algorithms

Deep Q-Network

- We used 2 policy nets and 2 target nets one to predict the weights of the assets and the other to predict the holding of the different assets.
- Loss is defined as the sum of the Smooth L1 losses of both the networks.
- To update the target net we use the vanilla soft update policy.

Hyperparameter Tuning



The learning rate was the most important hyper-parameter for fine-tuning. LR=1e-4 worked best for DQN as the model did not follow the buy/hold curve closely and learnt how to reduce drawdown.

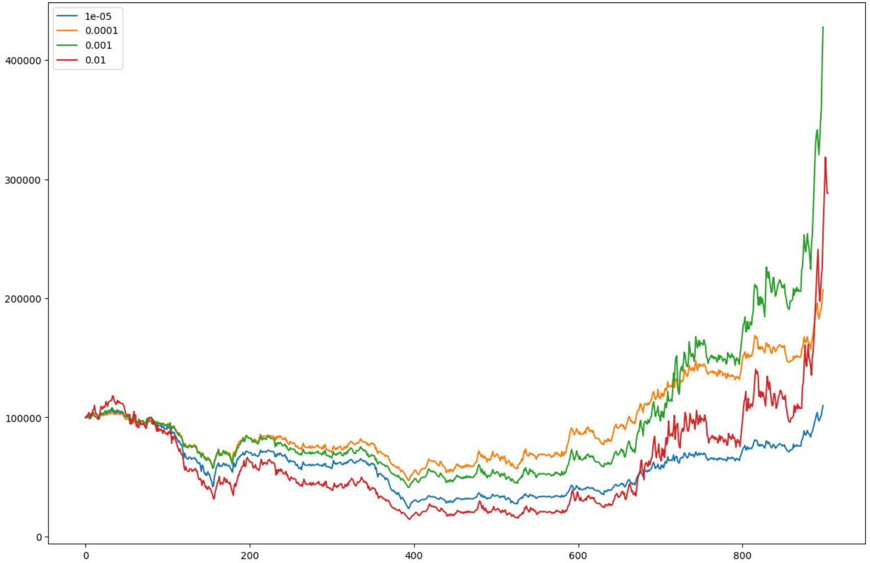
Inferences

- As we decreased the step in h the model training became slow as we approached to continuous action space.
- The DQN model learns to take only some actions and prefer to hold.
- Here there were 2 independent models for predicting the weights w_t and the holding h_t and there was no correlation between the 2 models.

Advantage Actor Critic

- In A2C we thought keeping both the models completely independent is not a good idea as the weights and holdings are having some correlation as in if the model thinks the stock is better to bet on in comparison to the other stocks then both the weight and the position should be high and vice versa.
- The combined A2C model which we used was having 3 different last layers i.e action head, weight head and the value head.All the 3 heads share the rest of the parameters.

Hyperparameter Tuning



For A2C, learning rate of 1e-3 worked best as others were unable to learn the action-reward relationship sufficiently.

Inferences

- The A2C model outperformed DQN model.
- The model was taking quite a number of actions.
- After around 700-800 episodes the model was overfitting to the stocks which we used in that particular environment.

PPO (Proximal Policy Optimization)

- We used PPO for policy optimization with architecture of the model similar to the A2C.
- The model took more number of actions than the A2C model and the training was faster than both DQN and A2C.The problem of Holding the stocks and taking not many actions which we got in DQN was solved to a point in A2C and PPO.

Baseline Strategies

Non-RL Strategies/Baseline Strategies

- Buy/Hold:Divide the initial capital equally among the stocks and wait until the last time step.
- Donchian Channel:
  - Entry Rule: Enter if a stock closes and makes a 20-day high (20-Period Donchian Channel)
  - Exit Rule: Exit if a stock closes and makes a 10-day low (10-period Donchian Channel)
- BB-based Strategy:Create 20-day BBs.
  - Entry Rule: Enter if price crosses the lower band.
  - Exit Rule : Exit if price crosses the upper band.

Back-Testing

Results; Bearish Market

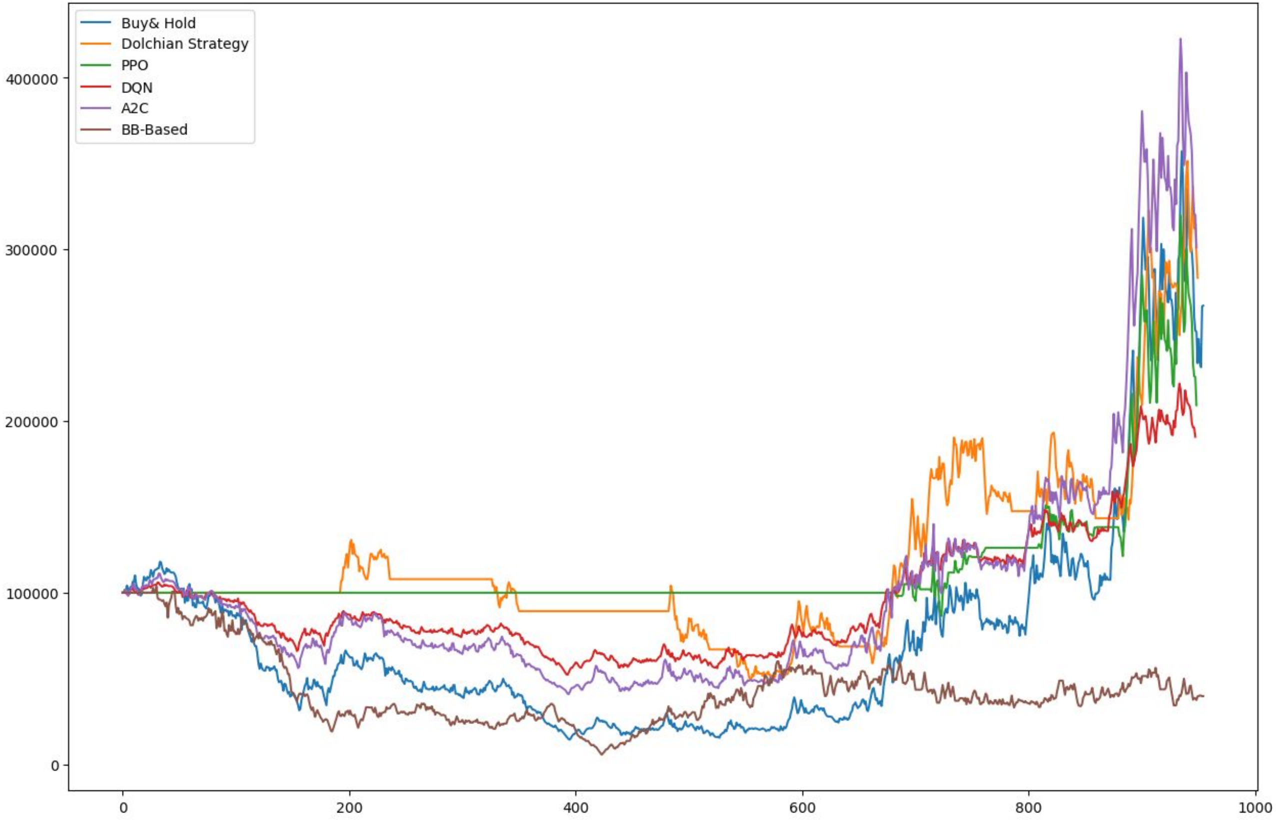
Algorithm/Strategy	Returns	Sharpe Ratio	Max-Drawdown	Sortino Ratio
DQN	93,956	0.7845	0.11860	0.6863
A2C	1,53,690	1.3588	0.69843	1.5568
PPO	1,21,671	1.2907	0.71321	1.4478
B/H	84,493	1.0559	0.67519	1.0816
Donchian Channel	1,34,050	1.3228	0.41522	1.4920
BB-based Strategy	1,16,986	1.2163	0.27975	1.3065

Performance; Bearish Market

Results; Bullish Market

Algorithm/Strategy	Returns	Sharpe Ratio	Max-Drawdown	Sortino Ratio
DQN	1,90,821	1.43235	0.50666	1.67227
A2C	3,01,177	1.50249	0.63116	1.78391
PPO	2,09,206	1.44380	0.34556	1.70087
B/H	2,67,226	1.45977	0.87686	1.71769
Donchian Channel	2,83,370	1.49911	0.62007	1.76988
BB-based Strategy	39,774	1.07971	0.94545	1.11402

Performance; Bullish Market



Inferences

- The DQN model has a very low max-drawdown and it saved almost all the money in bearish market.
- The A2C and PPO has a max-drawdown but in the end it gave almost 60-70% return in bearish market.
- The PPO learnt to hold in the period of negative drawdown and gave good returns in bullish market.
- A2C emerged as winner in both bearish and bullish markets giving moderate drawdown.

Conclusion

- We tried making new strategies for portfolio management using Reinforcement Learning and saw it outperformed the well-known baseline strategies that are in use in the financial markets.
- As we increase the number of assets in the portfolio the action space increases exponentially and the training time increased significantly.

Future Scope

Limitations

- The environment currently does not support splits in the share market price, which could further increase returns as markets become more liquid after splits.
- The environment currently does not support dividends provided by a stocks.
- The environment also does not support actions such as short (selling a stock before buying) which could also provide returns even in bearish markets.