Report

Name: Arpit Patel
GT ID: 902891506

1.0 METHODOLOGY

In this project, we develop a strategy learner based on the Qlearning approach, to help us make informed decisions for trading stocks. In this project, the learner will be focused on trading a single stock, but similar approach can be extended to train on multiple stocks and develop portfolios. We will be training our learner based on Q Learning and will develop a policy based on available market data of the stock. Further, we will test our learner based on future data of the stock and compare its performance with the buy and hold strategy. To test the robustness of our leaner, we will use both highly oscillating stocks and stocks with low volatility and compare its performance.

Technical Indicators

For this project we will be using 4 different types of indicators – Momentum,Ballinger Bands,Moving Average Convergence Divergence[MACD],Relative Strength Index(RSI)

Momentum[Window Size=**5 days**] : "Momentum" in general refers to prices continuing to trend. The momentum indicators shows trend by remaining positive while an uptrend is sustained, or negative while a downtrend is sustained.

Ballinger Bands [**Percentage**][Window Size=**20 days**] : Used for system building and pattern recognition. Used for identification of opportunities arising from relative extremes in volatility and trend identification

MACD[Window Size=**12 days,26 days**] : MACD series can indicate changes in the trend of a stock. It can reveal subtle shifts in the stock's trend.

RSI[Window Size=**5 days**] : RSI compares the magnitude of recent gains and losses over a specified time period to measure speed and change of price movements of a security. It is used to attempt to identify overbought or oversold conditions in the trading of an asset.

Pre-Processing

During the process ,we will be tracking our current holdings of the stock using the variable ‘Holding State’. To simplify the problem, we will be allowing only 3 possible states {0,1,2} → to you own{-200,0,200} shares. Similarly the actions will be limited to {0,1,2,3,4} → buy {-400,-200,0,200,400} shares.

To form our states of the Q learner, will be discretizing the data into bins based on a histogram of the data. After turning the continuous technical indicators into discrete values(corresponding bins), we will definite our state as

State[dt] =Momentum[dt]*Ballinger[dt]*MACD[dt]*RSI[dt]*(HoldingState+1)-1

Environment

The environment function performs the following tasks

1 . Assign Rewards [for taking an action]:

Rewards = DailyReturn – FixedTransactionCost-MarketImpact

2. Disallow not possible trades:

Allow the trade only if we have the allowable bank balance and the action doesnt result in a new holding state greater than 200 shares of lesser than -200 shares

3. Track the Portfolio Value and assign new State

Training

The psuedo algorithm for the training process is :

- Compute the technical indicator values for the training data
- Discretize the values of the features
- Instantiate a Q-learner
- For a epoch: list of training days
  - Compute the current state (including holding)
  - Query the learner with the current state and reward to get an action
  - Implement the action the learner returned, and update portfolio value
- Repeat the above loop multiple times(for multiple epochs) until cumulative return stops improving beyond a **Stopping Threshold** .

Testing

The psuedo algorithm for the training process is :

- For each day in testing data
  - Compute the current state
  - Query the learner with current state to get an action

    - Implement the action
- Return the list of actions taken

2.0 RESULTS

To understand the robustness of the learning algorithm, we will be varying the hyperparamters and evaluating its performance.

EXPERIMENT 1: Vary the Number of Bins

We will be varying the bin size and observing how it changes the cumulative return on different stocks stocks. Using InSample cumulative return as a benchmark we will see how the learner performs.

Bins	Insample[CR]			
	Stock 1	Stock 2:	Stock 3	Stock 4:
2	1.3523890566	0.3502044	2.722	0.023
3	1.0575	-0.2286	3.4314	-0.73148
4	0.9267	-0.0676	2.6598	-0.0985
5	1.0566	-0.1975	3.8579	0.064
6	1.30385	0.1531	3.5034	0.026
7	1.342	0.2315	3.677	-0.041

We realize that depending the type of stock,stock behavior(volatility,trend,etc,etc) andlength of data the bin size that gives the maximum cumulative return could be different. However, as a general idea, more number of bins(assuming you have enough data) can predict cumulative return better

EXPERIMENT 2: Stopping Threshold

Stopping Threshold	Insample[CR]			
	Stock 1	Stock 2:	Stock 3	Stock 4:
0.2	-0.4926	-0.3997	2.426	-0.04003
0.15	1.405	0.08343	3.13	0.05984
0.1	1.44808	0.2728	2.9728	0.01944

As we reduce the stopping threshold, the in sample accuracy improves