

Machine Learning for Trading: Strategy Learner

September 6, 2021

0.1 Introduction

In this project we will apply the Q learning method to the trading problem. First it is necessary to frame the trading problem in terms of a learner to maximize the rewards for its actions.

We use the historical stock closing price data in a fixed time as training data. And we will use three indicators to form states and LONG (to a maximum of 1000 shares), CASH (do nothing), SHORT (to a maximum of 1000 shares shorted) as 3 actions. The reward is the daily return (next trading day vs today) as a result of an action taken today and the goal is to maximize the cumulative return.

The structure of our strategy learner is arranged as below:

1. select several technical features, and compute their values for the training data
2. discretize the values of the features (so the states are discrete)
3. instantiate a Q-learner
4. for each day in the training data: compute the reward of the last action, 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 (LONG, CASH, SHORT) and update portfolio value (dyna-Q is implemented)
5. repeat the above loop multiple times until cumulative return stops improving

0.2 Indicators

We used three indicators, which are:

1. Price/SMA(PSMA): A simple moving average (SMA) is an arithmetic moving average calculated by adding the closing price of the security for a number of time periods and then dividing this total by the number of time periods.

$$SMA_t^{(n)} = \frac{1}{n} \sum_{i=0}^{n-1} Price_{t-i}$$

PSMA is the quotient of prices divided by simple moving average:

$$PSMA_t = Price_t / SMA_t$$

2. Bollinger Bands (bb): Bollinger Bands are a highly popular technical analysis technique. It is plotted two standard deviations away from a simple moving average.

$$\text{upper-band} = \text{rolling-mean} + 2 * \sigma(\text{prices})$$

$$\text{lower-band} = \text{rolling-mean} - 2 * \sigma(\text{prices})$$

$$\text{bollinger band} = (\text{price} - \text{lower-band}) / (\text{upper-band} - \text{lower-band})$$

3. Momentum: Momentum of a stock over n days is the return of a certain day compared to n days before.

$$\text{momentum (n day)} = (\text{price} / \text{price before n days}) - 1$$

To implement the Q-learner method, we need to further discretize the indicators, after calculating three indicators, which we call I_1, I_2, I_3 , we then separate the values of each indicator across all trading days being used into 10 bins. Therefore each indicator for a date will be given an index value from 0-10, corresponding to a bin it belongs to. The discretized state of a trading day is then given by

$$I_1 * 100 + I_2 * 10 + I_3$$

Therefore for 3 indicator each having 10 possible values, we have 1000 states in total for the Q-learner.

0.3 Experiment

The trading details are: AAPL

1. in-sample period: 1st Jan 1, 2008 - 31 Dec, 2009
2. out-of-sample/testing period: Jan 1, 2010 - 31 Dec 2011
3. starting cash: \$100,000
4. Allowable positions: 1000 shares long, 1000 shares short, 0 shares
5. impact: 0.005, commission: 9.95
6. Q-learner parameters: learning rate $\alpha = 0.2$, discount rate $\gamma = 0.9$, random action rate $rar = 0.5$, random action decay rate $radr = 0.99$, dyna 0.

The outcome

The experiment's result is shown in figure 1.

The result shows that the Strategy Learner performs greatly in terms of culmulative return.

