

Qingyang Li  
 Report: Project 8 – Strategy Learner  
 CS7646: ML4T - Spring 2019  
 April 21, 2019

## Reinforcement Learning for trading

### Introduction

In the current project, Reinforcement Learning (RL) method were used to learn trading rules from historical stock prices and apply the rules to generate orders to trade. Q-learning technique were utilized with three stock price based technical indicators, price/simple-moving-average ratio(PSR), Bollinger Bands (BB) and as input. The PSR ranges from -50% to +50% , BB indicator has a value between -1 and 1, momentum has a typical value between -0.5 and 0.5.

### Method

**Converting trading problem to a RL problem** requires to identify states, actions and rewards for the q-learning agent or investor. The actions are easy to define. The agent has three actions, *buy*, *sell* or *hold* (do nothing), to take given a state. The state can be inferred from the value of the three technical indicators using the Values to Bins table (Table 1). For example, the state where PSR = -0.5, BB indicator = 0.9, and momentum = 0.05 can be represented by 095 (e.g., state[PSR, BB, momentum] = 095). Since we can calculate the value of indicators daily, each day can be represented by the state based on the indicators. Since there is one state for each day, the reward of the states can be represented by the daily return of the day (state).

Table 1: Value to Bins

Bin	0	1	2	3	4	5	6	7	8	9
BB	< -0.8	[-0.8, -0.6)	[-0.6, -0.4)	[-0.4, -0.2)	[-0.2, 0)	[0, 0.2)	[0.2, 0.4)	[0.4, 0.6)	[0.6, 0.8)	>= 0.8
PSR	< -0.4	[-0.4, -0.3)	[-0.3, -0.2)	[-0.2, -0.1)	[-0.1, 0)	[0, 0.1)	[0.1, 0.2)	[0.2, 0.3)	[0.3, 0.4)	>= 0.4
momentum	< -0.4	[-0.4, -0.3)	[-0.3, -0.2)	[-0.2, -0.1)	[-0.1, 0)	[0, 0.1)	[0.1, 0.2)	[0.2, 0.3)	[0.3, 0.4)	>= 0.4

Thus, the trading problem can be converted to a reinforcement learning problem where state is represented by the categorized technical indicators, the actions are sell, buy or hold and the reward is the daily return rate. Now the states, actions and rewards can be used to train the Q-learning.

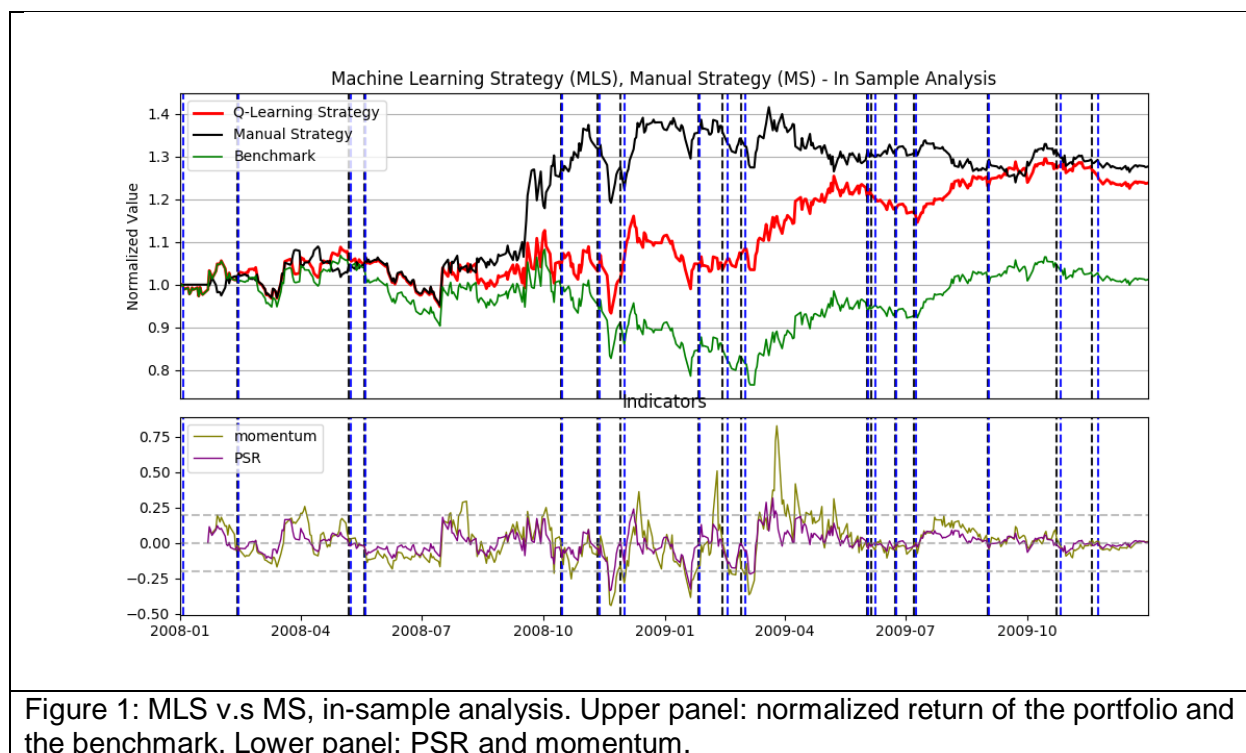
### Experiment 1

The results of the learned strategy were tested in sample using the JPM stock prices from 01/01/2008 to 12/31/2009 and out sample stock prices from 01/01/2010 and 12/31/2011, respectively. The Q-learning based strategy were compared with the manual rule based strategy(developed in Project 6, see appendix A) and a benchmark strategy (buy and hold). The net holding of the stock is limited in the range of -1000 to 1000 shares with a starting value of \$100,000. An impact of 0.005 was used when computing the portfolio value. commission = 9.95.

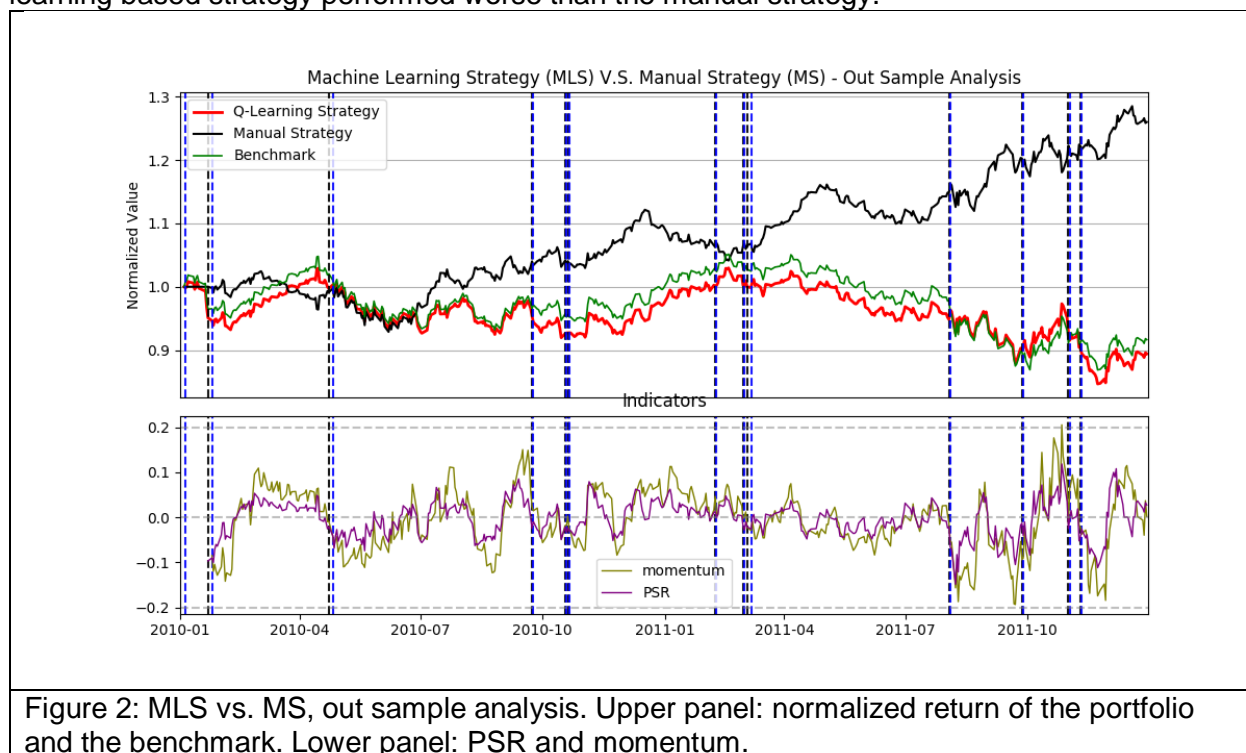
The in-sample analysis showed that the Q-learning based strategy performed better then the benchmarks but very similary to the manual strategy. The strategy has a higher cumulative return in the in-sample time period and did significantly better in returns (cumulated and average returns) than the benchmark. At the same time, the volicitities of the portfolio are compatible with the the benchmarks.

Table 2: Q-learning strategy performance

Metrics	In-sample Analysis			Out-sample Analysis		
	MLS	Manual Strategy	Benchmark	MLS	Manual Strategy	Benchmark
<b>Sharpe Ratio</b>	0.56628	0.69339	0.15702	-0.34375	1.08618	-0.25706
<b>Cumulative Return</b>	0.23807	0.27643	0.01223	-0.10591	0.26019	-0.08368
<b>Standard Deviation</b>	0.01503	0.01302	0.01704	0.00858	0.00709	0.00850
<b>Average Daily Return</b>	0.00054	0.00057	0.00017	-0.00019	0.00048	-0.00014
<b>Final Value</b>	\$ 123,807.40	\$ 127,643.15	\$ 101,017.75	\$ 89,408.55	\$ 126,019.45	\$ 91,435.75

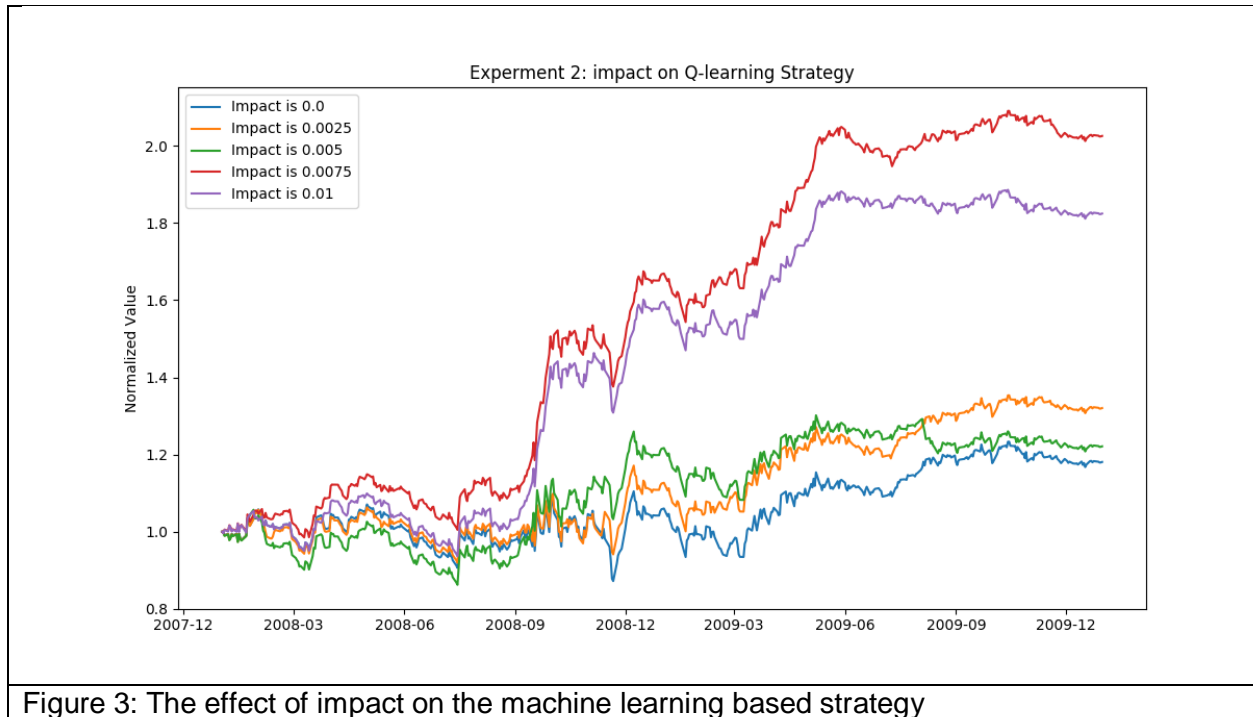


The out sample analysis, however, did not show that the Q-learning based strategy is any better than the benchmark. In fact, it is very similar to the benchmark in terms of performance. The Q-learning based strategy performed worse than the manual strategy.



## Experiment 2

The effect market impact on the Q-learning based strategy was evaluated using the normalized portfolio value. According to the results illustrated in Figure 3, we test the impact from 0 to 0.01 with a step size of 0.0025. As the impact value goes higher from 0 to 1, the return of the stock (JMP, based on in-sample data from 2008-01 to 2009-12) increases.



## Appendix A: Manual Rule-Based Strategy

For the manual strategy, I used all the indicators introduced above. The rule I chose is that when PSR is negative (price is lower than SMA) and BB indicator is smaller than -0.8 (when it is close to the lower band), and when we see a strong negative momentum, I chose to buy the stock hoping to make money since they might regress to SMA soon.

When the PSR is positive (price is higher than SMA) and BB indicator is larger than -0.8 (when it is close to the upper band), and when we see a strong positive momentum, I chose to sell the stock. This strategy should ride along when the market is positive but will do better when the market is negative.