Strategy Learner Project

ZHENG FU (zfu66@gatech.edu)

Part 1: Describe the steps you took to frame the trading problem as a learning problem for your learner.

I. What are your indicators?

In this project I selected Simple Moving Average (SMA), Bollinger Band® and Momentum as my technical indicators to build my strategy learner, and they were also applied in the manual strategy as well.

1. Simple Moving Average (SMA)

In this project, SMA is defined as the average of the daily adjusted close price values of the security over a specific number of periods. Here a specific number of periods are 20 days and 50 days.

2. Bollinger Band®

In this project Bollinger Band® is a set of lines plotted *n* standard deviations away from a *m*-days simple moving average of the security's adjusted close price. Here *m* was set to 20 and *n* was set to 2. And the Bollinger Band® feature at given time point *t* was defined as:

$$\text{Bollinger Band® feature} = \frac{Adjusted\ Close\ Price(t) - 20days\ SMA(t)}{2\ \times 20days\ Stdev\ of\ Adjusted\ Close\ Price(t)}$$

3. Momentum

In this project Momentum was defined as:

Momentum = [(Current adjusted close price of the security - Adjusted close price of the security at *m*-days ago) / Current adjusted close price of the security] - 1

Here *m* was set to 20.

II. Learning.

In order to frame the trading problem as a learning problem for my learner, the key step is to generate the training and testing data sets and train the learner. In this project the training data set (X_train) has 7 columns corresponding to the following 7 features:

1) Price minus upper Bollinger Band®,
2) Price minus lower Bollinger Band®,
3) 20 days SMA,
4) 50 days SMA,
5) 50 days SMA minus 20 days SMA,
6) Momentum and 7) Bollinger Band® feature.

Each row of the training data set represents the dates of in-sample time period, and their labels (Y_train) were set to 0 (HOLD) initially. Here 40-days return was used to determine the labels of each trading day:

1). If (40-days Later Price) / (Current Date Price) no less than (1.025 + impact factor), the label of that trading day was set to 1 (BUY).

2). If (40-days Later Price) / (Current Date Price) no more than (0.975 - impact factor), the label of that trading day was set to -1 (SOLD).

The testing data set (X_test) has the same seven features as the training data set, and its rows represent the dates of out-sample time period. Both X_train and Y_train will be introduced to the strategy learner to train the model. Then the X_test will be feed to the trained learner and predict the labels of its rows (Y_test). Using the rules mentioned in **"Part II – Manual strategy – Making orders"** section (see page 4 for more details), the Y_test could be converted to a trading table followed by portfolio value calculation.

III. Did you adjust the data in any way (discretization, standardization)?

The "Prices" dataframe was forward filled then backward filled so as to eliminate the missing price values. The price values were normalized to 1.0 at the start date, so the potential biases introduced by the differences between in-sample and out-sample data could be reduced to the maximum extend.

Part 2: Experimental 1

I. Assumption.

My assumption is the performance of learning strategy will be better than manual strategy and benchmark in terms of Sharpe Ratio and Cumulative Return.

II. Experiment details.

All trades were made on security symbol JPM. The in-sample period is 01-01-2008 to 12-31-2009 with starting cash = $100,000. The commission cost and impact factor are both set to zero. The allowable positions are: 1000 shares long, 1000 shares short, 0 shares.

1. Learning strategy.

For learning strategy, I employed the bag leaner with number of bags = 20. The boost function of bag learner was nullified. I selected random tree learner as the basic learner of bag learner with leaf size = 5. Random seed was set to 19683. X_train, Y_train, X_test and Y_test data sets were generated as described in **Part 1 – Learning**. Using the rules mentioned in **"Part II – Manual strategy – Making orders"** section (see page 4 for more details), the Y_test could be converted to a trading table followed by portfolio value calculation.

2. Manual strategy.

For manual rule-based trader, I applied three technical indicators to determine the trading signal of the system.

A. SMA.

(a). Buy signal (Golden Cross): The security's 20-days SMA is no less than its 50-days SMA at current date, and the security's 20-days SMA is no more than its 50-days SMA at previous date.

(b). Sell signal (Death Cross): If the security's 20-days SMA is no more than its 50-days SMA at current date, and the security's 20-days SMA is no less than its 50-days SMA at previous date.

B. Bollinger Band®.

(a). Buy signal: The security's adjusted close price is no less than its 20-days lower Bollinger band at current date, and the security's adjusted close price is no more than its 20-days lower Bollinger band at previous date.

(b). Sell signal: The security's adjusted close price is no more than its 20-days upper Bollinger band at current date, and the security's adjusted close price is no less than its 20-days upper Bollinger band at previous date.

C. Momentum.

(a). Buy signal: the current day's momentum is more than zero and the previous days' momentum is less than zero.

(b). Sell signal: the current day's momentum is less than zero and the previous days' momentum is more than zero.

D. Making orders.

If at least one of the three buy signals mentioned above appears, and no signal confliction occurs, the system buy signal is set to TRUE and the trader will make an order to long shares of the security.

If at least one of the three sell signals mentioned above appears, and no signal confliction occurs, the system sell signal is set to TRUE and the trader will make an order to short shares of the security.

If both system BUY and SELL signals are FALSE, the trader just does nothing and holds the positions. Meanwhile the manual rule-based trader was constrained on shares holding (Allowable positions are: 1000 shares long, 1000 shares short, 0 shares), so the following scenarios should be considered while making orders:

a). Positions are zero shares:
If the system buy signal is TRUE, the trader will make an order to long 1000 shares of the security. If the system sell signal is TRUE, the trader will make an order to short 1000 shares of the security.

b). Positions are 1000 shares long:
If the system buy signal is TRUE, the trader will do nothing. If the system sell signal is TRUE, the trader will make an order to short 2000 shares of the security.

c). Positions are 1000 shares short:
If the system buy signal is TRUE, the trader will make an order to long 2000 shares of the security. If the system sell signal is TRUE, the trader will do nothing.

3. Benchmark strategy.

The trader will buy and hold 1000 share of the security throughout the in-sample period.

III. Results

The results of Experiment 1 were summarized in Figure 1 and Table 1. We could see that learning strategy has higher Sharpe Ratio as well as higher Cumulative Return comparing with manual strategy and benchmark strategy, respectively. This observation is in accordance with my assumption. Note that the manual strategy makes 67 trades during the in-sample period whereas the learning strategy only makes twice. Considering the commission fee is not zero in actual trades, the learning strategy may also have better performance than manual strategy in real world.

IV. Would you expect this relative result every time with in-sample data?

I would not expect this relative result every time with in-sample data. Since the learning strategy was based on Random Tree learner, each time the Random Tree learner will random split the features and generate different results with in-sample data. However for facilitating TAs to repeat my figures and tables, I set up random seed = 19683 for Random Forest learner and every time the learning strategy will produce same relative result.

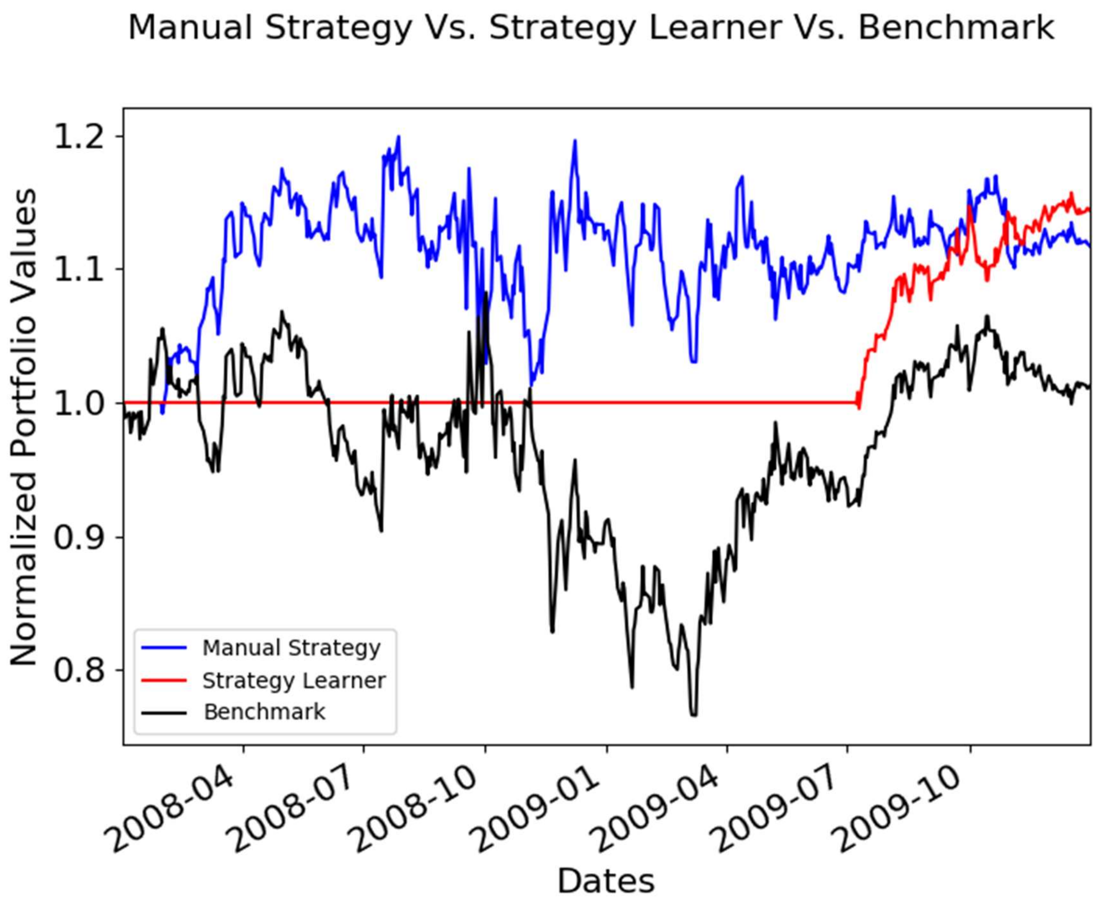

Figure 1. Comparing learning strategy, manual rule-based strategy and benchmark strategy over in-sample time period at Experiment 1.

	Learning Strategy	Manual Strategy	Benchmark Strategy
Sharpe Ratio	1.124	0.357	0.157
Cumulative Return	14.36%	11.69%	1.23%
Average Daily Returns	2.737E-04	3.202E-04	1.680E-04
Standard Deviation of Daily Returns	3.864E-03	1.422E-02	1.700E-03
Trading Times	2	67	2
First Trading Date	07-07-2009	01-30-2008	01-02-2008

Table 1. Statistics of portfolio values using learning strategy, manual rule-based strategy and benchmark strategy over in-sample time period at Experiment 1.

Part 3: Experimental 2

I. Assumption.

My assumption is increasing the value of impact factor will reduce the trading frequency. It also rises the costs of trading hereby the Sharpe Ratio and Cumulative Return will decrease accordingly.

II. Experiment details.

All trades were made on security symbol JPM. The in-sample period is 01-01-2008 to 12-31-2009 with starting cash = $100,000. The commission cost was set to zero. The allowable positions are: 1000 shares long, 1000 shares short, 0 shares.

1. Learning strategy.

For learning strategy, I employed the bag leaner with number of bags = 20. The boost function of bag learner was nullified. I selected random tree learner as the basic learner of bag learner with leaf size = 5. Random seed was set to 19683. Then I generated four learning strategies with impact factor = 0.0, 0.005, 0.05 and 0.5, respectively. X_train, Y_train, X_test and Y_test data sets were generated as described in **Part 1 – Learning**. Using the rules mentioned in **"Part II – Manual strategy – Making orders"** section (see page 4 for more details), the Y_test could be converted to a trading table followed by portfolio value calculation.

2. Benchmark strategy.

The trader will buy and hold 1000 share of the security throughout the in-sample period.

III. Results

The results of Experiment 2 were summarized in Figure 2 and Table 2. We could see that learning strategy with impact factor = 0.005 has lower Sharpe Ratio and lower Cumulative Return comparing with learning strategy whose impact factor is 0.0. Moreover, if the impact factor is no less than 0.05, no trading occurs. Both observations support my hypothesis. In order to eliminate the impacts of random seed on this experiment, I reset the random seed = 19583 and redo the experiment. From Figure 3 and Table 3 we could see the results using different random seeds are similar and both of them support my hypothesis.

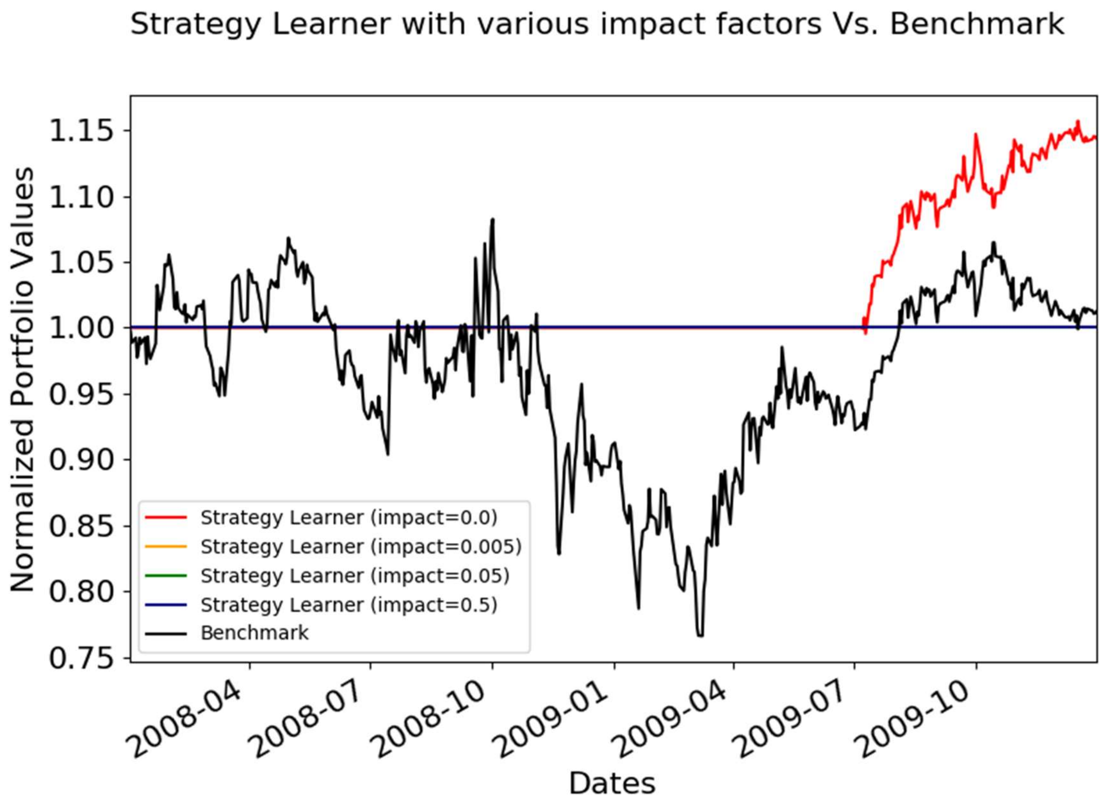

Figure 2. Comparing learning strategy with various impact factors and benchmark strategy over in-sample time period at Experiment 2, random seed = 19683.

	Learning Strategy (Impact Factor = 0.0)	Learning Strategy (Impact Factor = 0.005)	Learning Strategy (Impact Factor = 0.05)	Learning Strategy (Impact Factor = 0.5)	Benchmark Strategy
Sharpe Ratio	1.124	0	0	0	0.157
Cumulative Return	14.36%	0%	0%	0%	1.23%
Average Daily Returns	2.737E-04	0	0	0	1.680E-04
Standard Deviation of Daily Returns	3.864E-03	0	0	0	1.700E-03
Trading Times	2	0	0	0	2
First Trading Date	07-07-2009	No trading	No trading	No trading	01-02-2008

Table 2. Statistics of portfolio values using learning strategy with various impact factors and benchmark strategy over in-sample time period at Experiment 2, random seed = 19683.

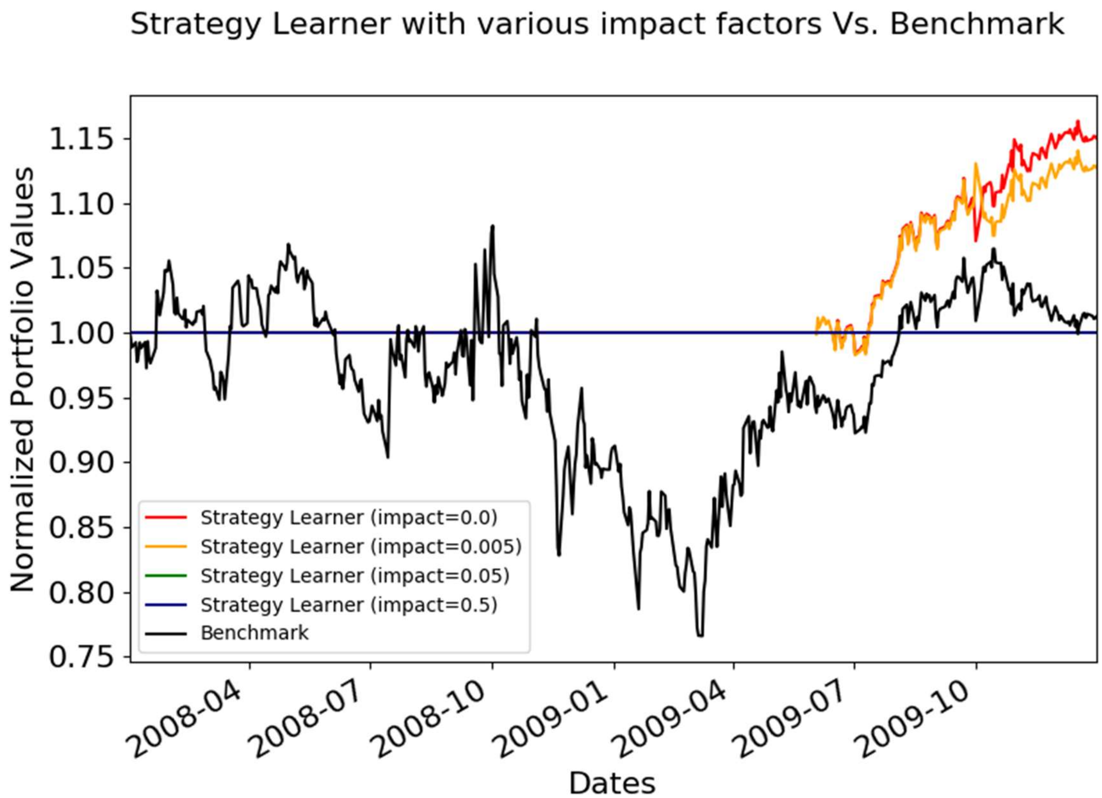

Figure 3. Comparing learning strategy with various impact factors and benchmark strategy over in-sample time period at Experiment 1, random seed = 19583.

	Learning Strategy (Impact Factor = 0.0)	Learning Strategy (Impact Factor = 0.005)	Learning Strategy (Impact Factor = 0.05)	Learning Strategy (Impact Factor = 0.5)	Benchmark Strategy
Sharpe Ratio	1.090	0.916	0	0	0.157
Cumulative Return	14.99%	12.70%	0%	0%	1.23%
Average Daily Returns	2.858E-04	2.464E-04	0	0	1.680E-04
Standard Deviation of Daily Returns	4.162E-03	4.269E-03	0	0	1.700E-03
Trading Times	2	2	0	0	2
First Trading Date	06-15-2009	06-03-2009	No trading	No trading	01-02-2008

Table 3. Statistics of portfolio values using learning strategy with various impact factors and benchmark strategy over in-sample time period at Experiment 2, random seed = 19583.