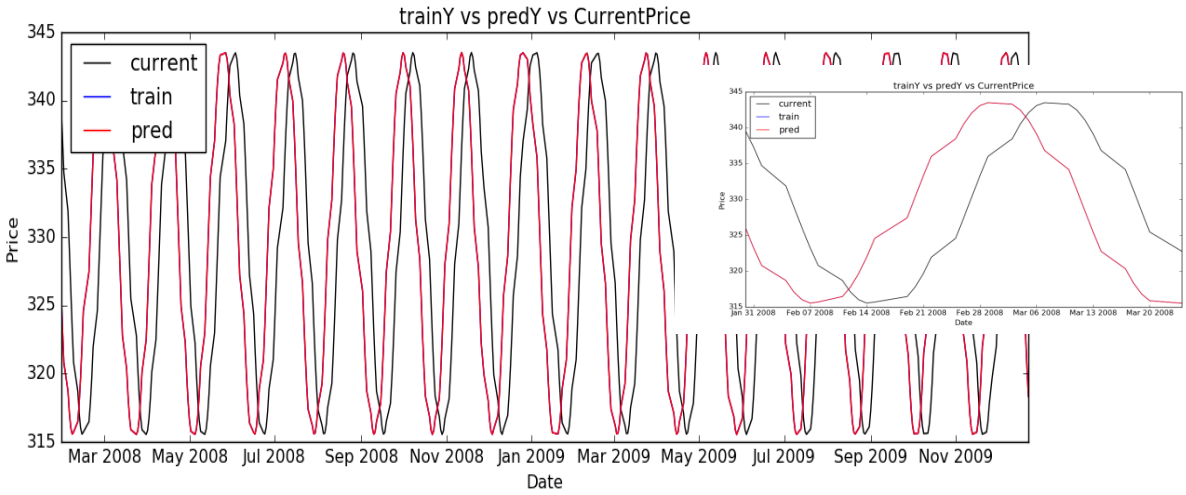
MC3-Project-2: report

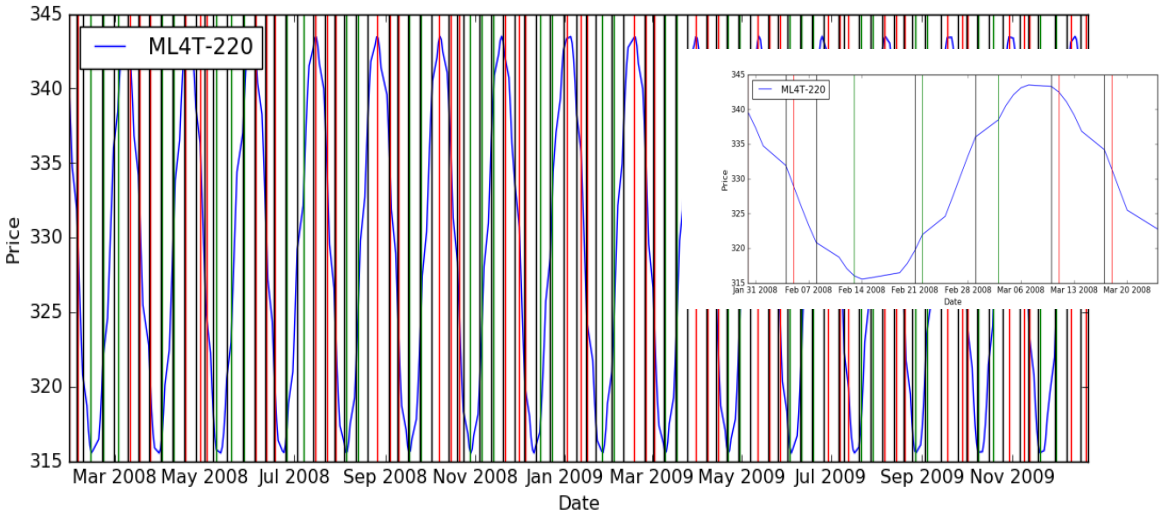
Xiaodong (Sheldon) Gu

I. Charts:

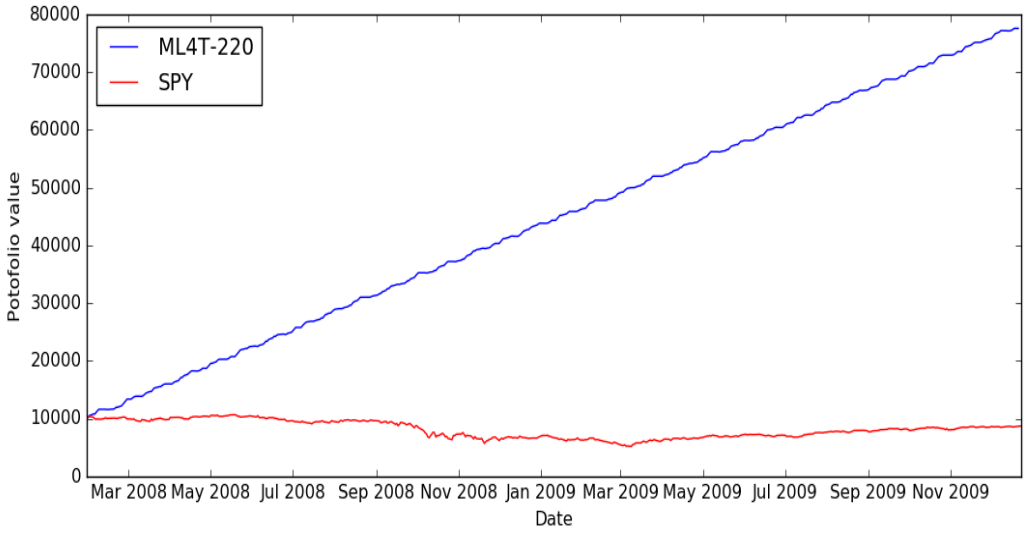
1. Training Y/Price/Predicted Y: Create a plot that illustrates your training Y values in one color, current price in another color and your model's PREDICTED Y in a third color. To help with the visualization, you should adjust your training Y and predicted Y so that it is at the same scale as the current price.



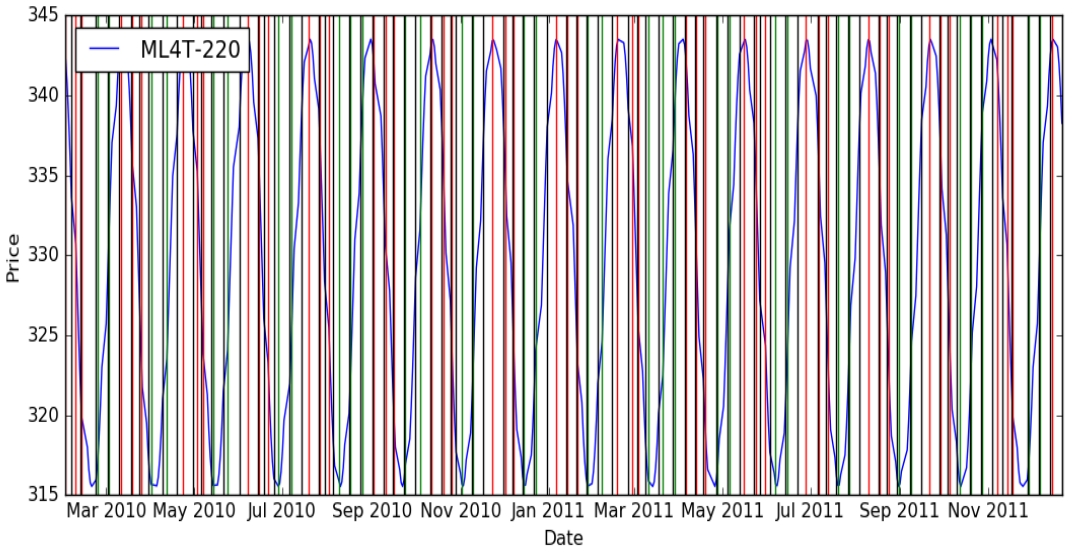
2. Sine data in-sample Entries/Exits: Create a plot that illustrates entry and exits as vertical lines on a price chart for the in sample period. Show long entries as green lines, short entries as red lines and exits as black lines. You may find it convenient to zoom in on a particular time period so this is evident.



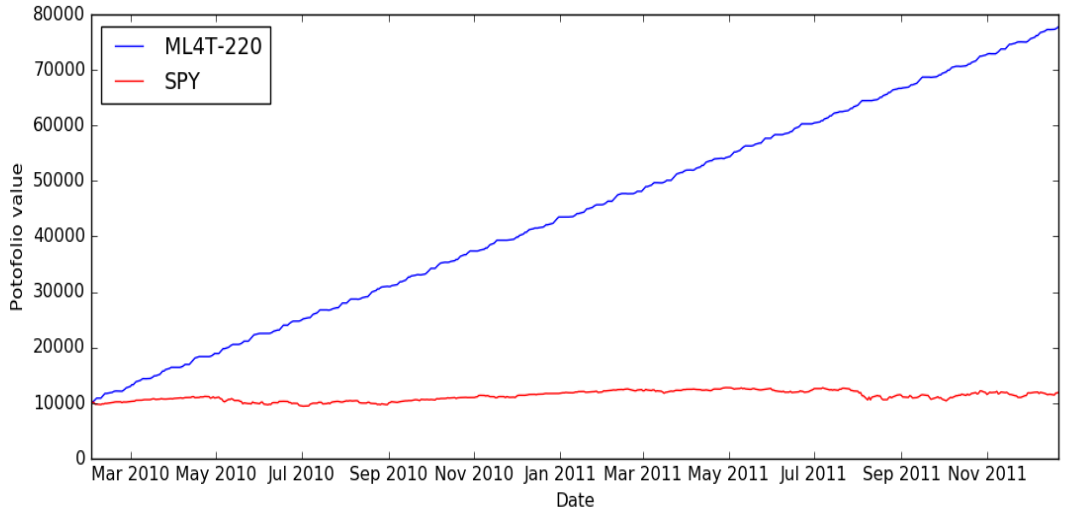
3. Sine data in-sample backtest



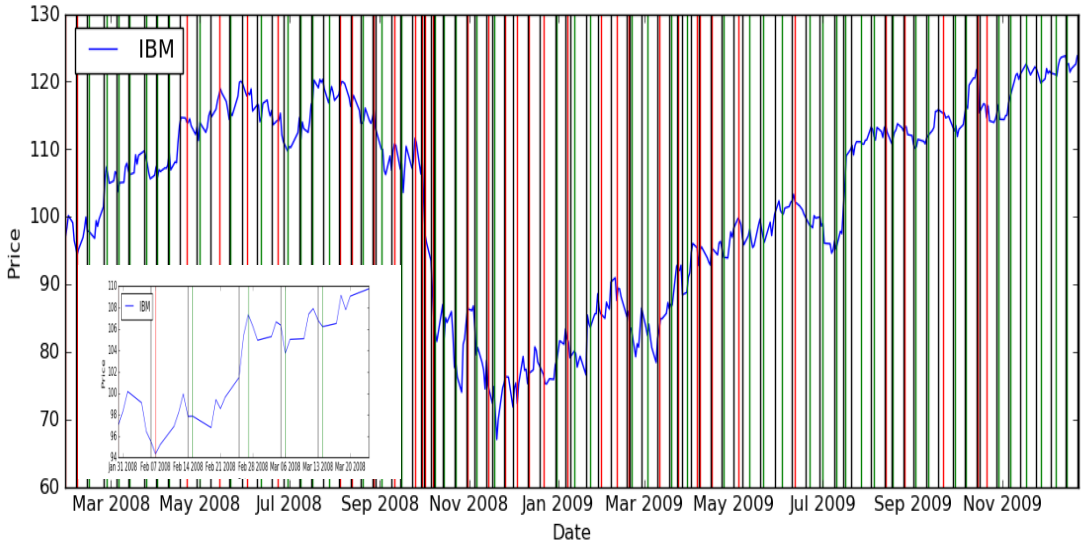
4. Sine data out-of-sample Entries/Exits: Freeze your model based on the in-sample data. Now test it for the the out-of-sample period. Plot the entry & exits, generate trades,



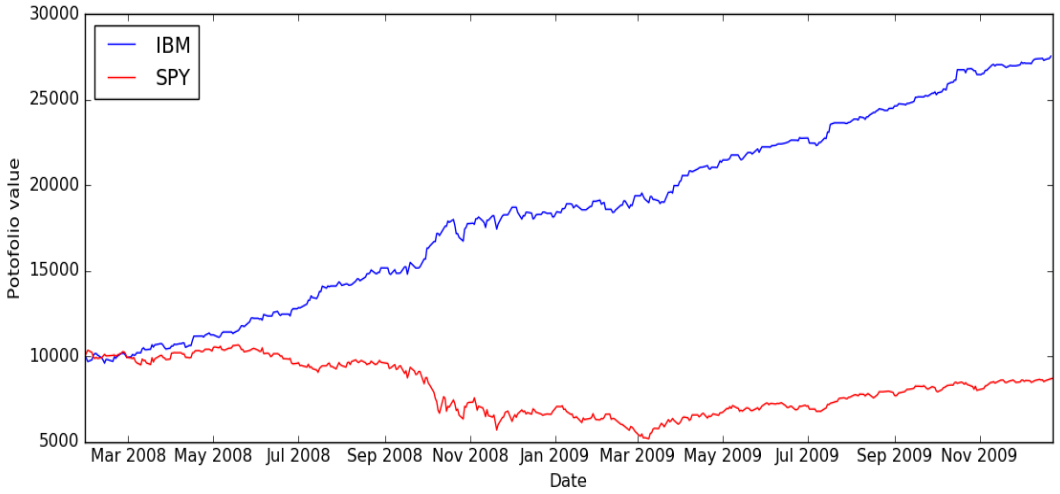
5. Sine data out-of-sample backtest.



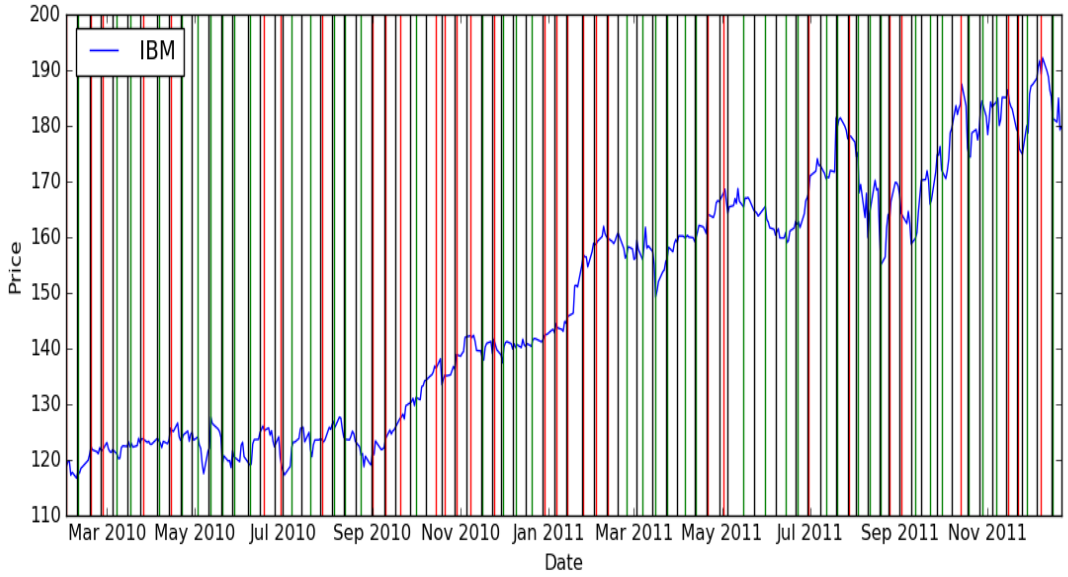
6. IBM data in-sample Entries/Exits: Create a plot that illustrates entry and exits as vertical lines on a price chart for the in sample period 2008-2009. Show long entries as green lines, short entries as red lines and exits as black lines. You may find it convenient to zoom in on a particular time period so this is evident.



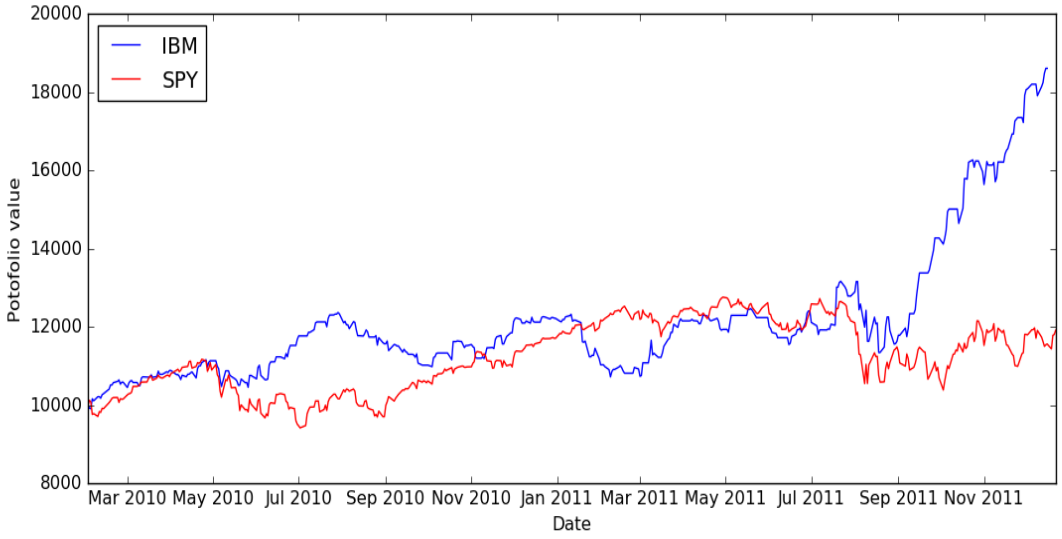
7. IBM data in-sample backtest



8. IBM data out-of-sample Entries/Exits



9. IBM data out-of-sample backtest



II. Indicators

The three technical indicators that used are: Bollinger Band value, Momentum and Volatility as Professor suggested in class with minor modifications.

Bollinger Band value: the traditional Bollinger Band refers the simple moving average plus / minus two times of standard deviations. To normalize the value in range of -1 and +1, Professor suggests a new formula:

$$bb\_value[t] = (price[t] - SMA[t]) / (2 * stdev[t])$$

The window size that I implemented in my code is 20.

Momentum: shows the trend that whether the stock goes up or goes down. The professor suggested formula is:

$$Momentum[t] = (price[t] / price[t-N]) - 1$$

However, during the implementation, I found that the momentum value is too small and need to time a factor α to fit in the range of -1 to +1.

$$Momentum[t] = ((price[t] / price[t-N]) - 1) * \alpha$$

In the training dataset, α = 1 / average (|momentum[t]|). But in the test dataset, the calculation needs information from future (peek the future). In theory, α should be generated from historical data of the specific stock, but for simplicity, I used $\alpha = 20$ in my implementation.

Volatility: is the standard deviation of daily returns. Similar as momentum value, volatility values are also small and need to be scaled up by a factor β. Theoretically, β should be generated from historical data of the specific stock, e.g. by dividing the yearly average of standard deviation. For simplicity and also not to peek the future, I used β = 100 in my implementation.

III. Trading policy

KNNLearner predicted the future 5 day return, and my trading policy is very simple and heavily relies on the accuracy of prediction.

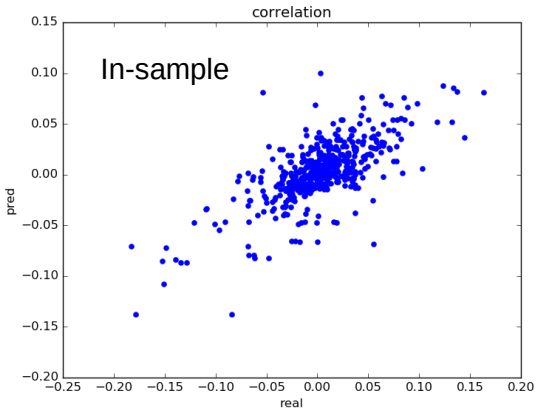
IF: future 5 days return > 1%
buy 100 share of stock
sell after 5 days or already get 2% profit

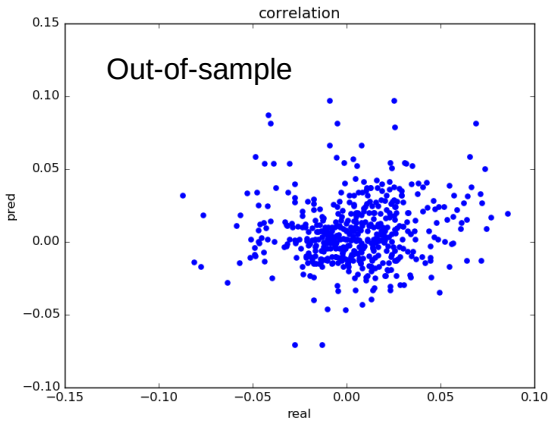
IF: future 5 days return < -1%
sell 100 share of stock
buy back after 5 days or already get 2% profit

The 2% threshold that I used reduces trading risk and increases more trading opportunities (exit earlier and trade more time).

IV. Discussion

My trading strategy works very well for in-sample sine, out-of-sample sine and in-sample IBM, as shown in these backtest charts. However, the performance for out-of-sample IBM was poor. One possible explanation is that my strategy heavily relies on the prediction of future prices. Better prediction leads to good trading performance and bad prediction leads to poor performance. I plotted the correlations between real values versus predicted values both in-sample IBM and out-of-sample IBM. The result is consistent with my hypothesis. The correlation is much better for in-sample IBM data (corr=~0.70) compared with out-of-sample IBM data (corr=~0.1).





There are a few approaches that can be used to improve my trading performance.

1, since my poor out-of-sample trading performance caused by bad future prediction, the searching for new technical indicators will help, or use more indicators (instead of using three indicators, using multiple indicators may help).

2, combine the future prediction data with other strategies like these developed in mc2-project2. Only buy/sell when these two approaches point to same directions. e.g. future prediction predicts 2% increase in 5 days, and the stock is moving from lower Bollinger band into simple moving average, thus buy the stock.