**Description of Project and Approach:**

In this project mc3_p2, I adopted the KNN learner from mc3_p1 to predict stock prices for in and out of sample data for SINE (ML4T-220) and IBM data. This strategy is referenced as “KNN Strategy” below. I applied the KNN Learner mode to average k=3, the three closest points to average as predicted Y. In preprocessing of the data sets, I used the following technical features listed below, normalized using the in sample mean and standard deviation. Technical features were normalized for each day within the in and out of sample datasets so that values generally ranged between -1 and 1, not including outliers. The normalization formula generally applied is (technical feature - in sample mean) / (in sample standard deviation).

- Momentum
- Simple moving average (window = 5 days)
- Rolling Standard deviation or Volatility (window = 5 days)
- Bollinger Band values

The KNN Strategy generated predicted Ys, where the Ys reflected a daily predicted return. The algorithm checked the returns every 5 trading days (not including banking vacations and holidays), if the trading day had a predicted Y or daily return greater than or equal to 1.0% then it entered a long trading position 5 days period and exited the strategy 5 days later. When daily returns were greater than or equal to -1.0% then the strategy entered a short position 5 days prior and exited 5 days later. One benefit of this strategy is that it decreases the complexity of overlapping trades, which is not allowed in this assignment.

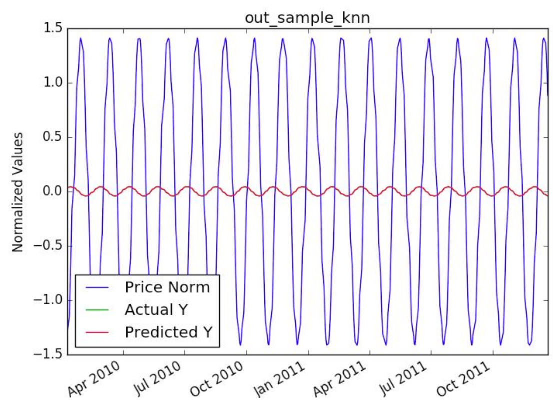
Overall KNN strategy did not perform well with IBM trading and portfolio performance. Did very well with ML4T-220 SINE data. From the entry and exit graphs of KNN Strategy with IBM prices, KNN did not accurately predict the increase and decrease within a 5 day period. I probably should have started with IBM and tuning technical features to the KNN learner since chosen technical features don’t seem to be providing enough information to KNN Learner to anticipate beneficial long and short positions. I can see these positions not being executed correctly from the exit and entry graphs and back testing.

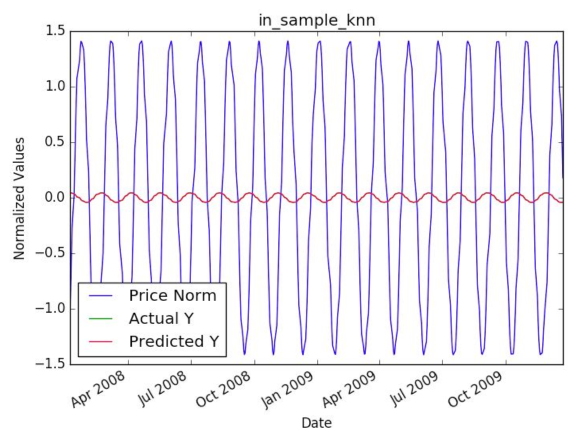
If i had more time with the project, I would center time on choosing the optimal technical features to reflect “uniqueness” in changes. Getting to uniqueness would require analyzing the correlation of technical features and choosing the optimal windows. The 5 day window I choose for simplicity may not have leverage a diverse enough set of points to help KNN model for trading.

## PART I. ML4T-200 data ("SINE")

(1) SINE graphs for in and out of sample below compare the normalized price, actual Y, and predicted Y. There are 3 lines on both of the graphs but since actual Y and predicted Y followed closely with each other, on the graph they are not distinguishable. Descriptive statistics are included for predicted Y and actual Y values to show how closed they tracked each other.

```
Actual Y stats:  count    495.000000
mean       0.000755
std        0.030305
min       -0.041824
25%       -0.029287
50%        0.000455
75%        0.030962
max        0.043655
Predicted Y stats:  count    495.000000
mean       0.000755
std        0.030305
min       -0.041823
25%       -0.029288
50%        0.000496
75%        0.030935
max        0.043654
```



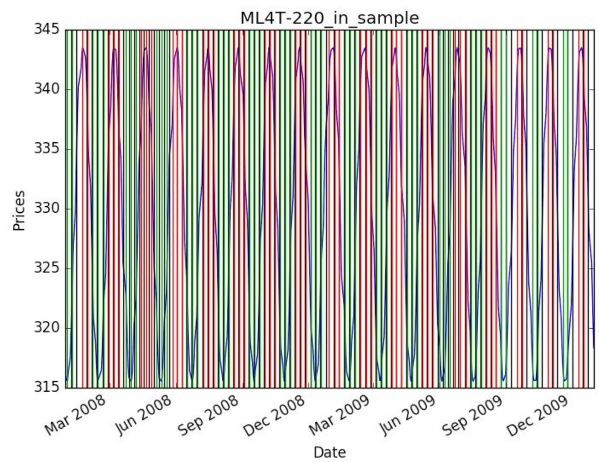
(2) Entry and Exit Graphs for ML4T-220

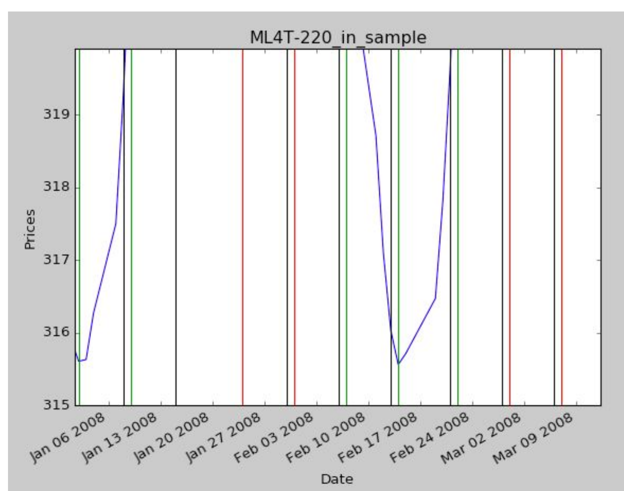
SINE was graphed with entry and exit points for short and long positions. The graphs showing all orders are difficult to read. The beginning of the time period for in sample and out of sample are zoomed in along with the first 10 orders to compare the table and graphs of orders. The trading strategy for SINE follows a similar pattern, where *green lines or entry points* are located when SINE increases and *short strategies or red lines* appear after reaching its maximum peak / starts on its decline.

**ML4T-220 Entry and Exit, Strategy Graphs**
**In and out of sample data sets**

ML4T-220 in sample Orders

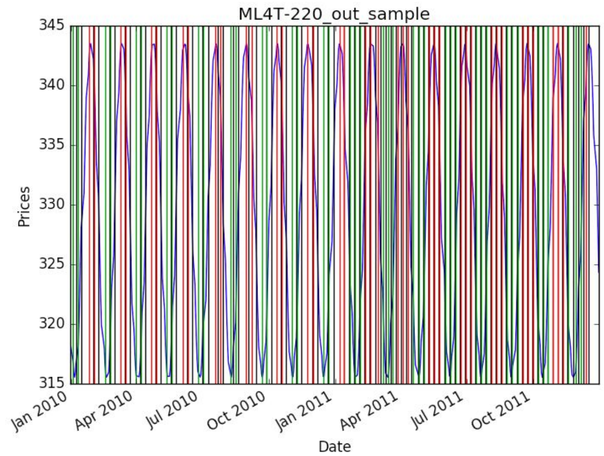
| Date | Symbol | Order | Shares | Type | Strat |
|---|---|---|---|---|---|
| 1/2/08 | ML4T-220 | BUY | 100 | Entry | Long |
| 1/8/08 | ML4T-220 | SELL | 100 | Exit | Long |
| 1/9/08 | ML4T-220 | BUY | 100 | Entry | Long |
| 1/15/08 | ML4T-220 | SELL | 100 | Exit | Long |
| 1/24/08 | ML4T-220 | SELL | 100 | Entry | Short |
| 1/30/08 | ML4T-220 | BUY | 100 | Exit | Short |
| 1/31/08 | ML4T-220 | SELL | 100 | Entry | Short |
| 2/6/08 | ML4T-220 | BUY | 100 | Exit | Short |
| 2/7/08 | ML4T-220 | BUY | 100 | Entry | Long |
| 2/13/08 | ML4T-220 | SELL | 100 | Exit | Long |

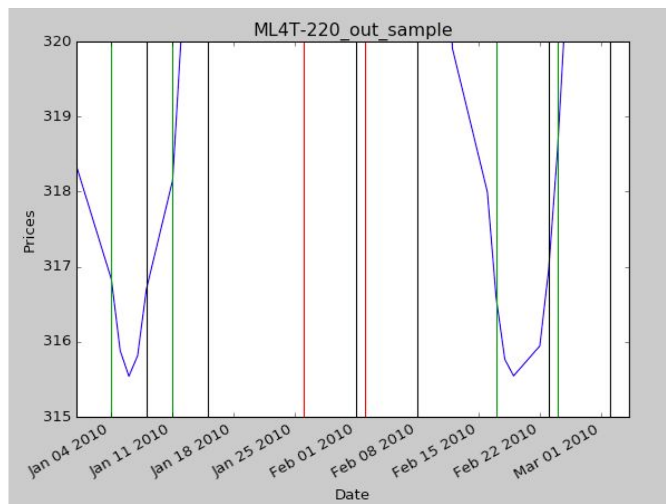


ML4T-220 out of sample Orders

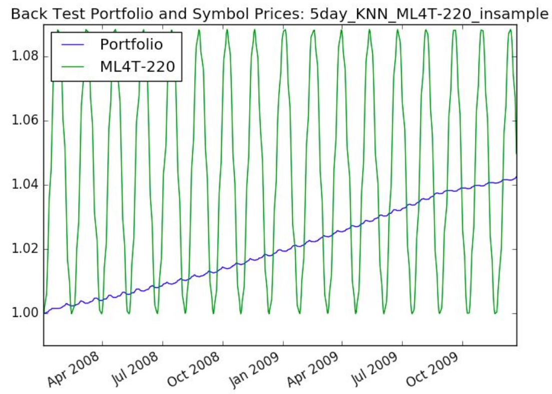
| Date | Symbol | Order | Shares | Type | Strat |
|---|---|---|---|---|---|
| 1/4/10 | ML4T-220 | BUY | 100 | Entry | Long |
| 1/8/10 | ML4T-220 | SELL | 100 | Exit | Long |
| 1/11/10 | ML4T-220 | BUY | 100 | Entry | Long |
| 1/15/10 | ML4T-220 | SELL | 100 | Exit | Long |
| 1/26/10 | ML4T-220 | SELL | 100 | Entry | Short |
| 2/1/10 | ML4T-220 | BUY | 100 | Exit | Short |
| 2/2/10 | ML4T-220 | SELL | 100 | Entry | Short |
| 2/8/10 | ML4T-220 | BUY | 100 | Exit | Short |
| 2/17/10 | ML4T-220 | BUY | 100 | Entry | Long |

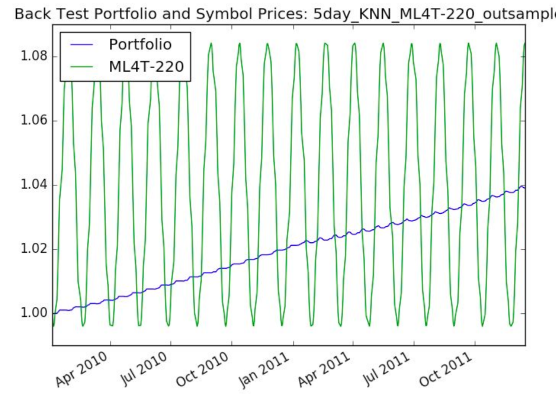



Zoomed into the beginning of the time period, troughs of the SINE

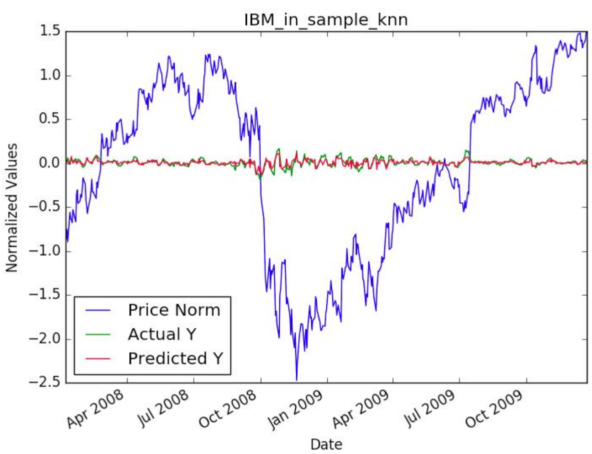
<u>(3) SINE Backtesting Graphs</u>. Show in and out of sample performance between KNN predicted trading strategies (named “Portfolio” in the back testing graphs) and ML4T-220 normalized prices. The graphs show that the KNN trading strategy performs at a consistently increasing rate. The in sample backtest ML4T-220 performance flattens out slightly at the end of the trading period, while in the out of sample backtest the rate of increase starts in the middle of the trading period.

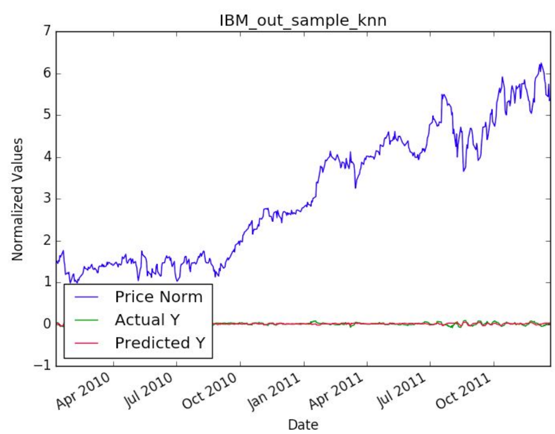



**PART II IBM data and graphs**

(1) IBM Graphs comparing IBM price, predicted Ys, and actual Ys for in and out of sample data. The Ys track pretty close to each other. However compared to SINE graphs, where Ys overlapped and predicted/actual were not distinguishable. For IBM graphs, we are able to see the differences between the green (actual) and red lines (predicted)



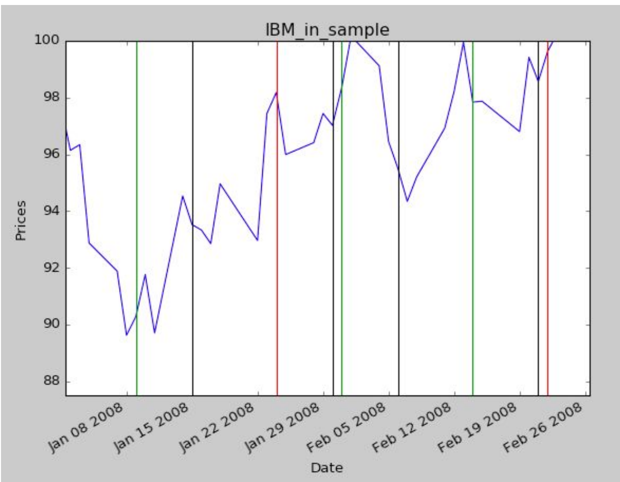



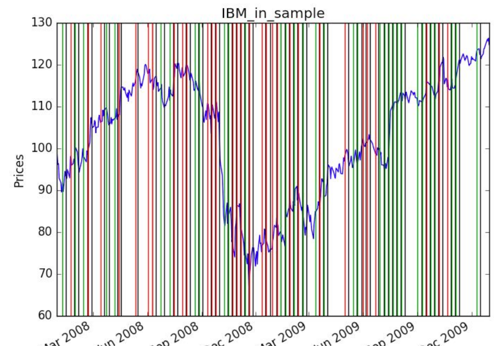

(2) IBM data Entries/Exits

It is Interesting to see regions that don't have training activity. Given the policy in place, the returns per 5 days were less than 1%. This is where I would consider re-working the policy to maximize the returns for in sample, test on out of sample, and compare results.

**IBM In Sample** trading strategy visualizations
Depicting exits and entries including sample of first 10 orders

Zoom into the first section of the IBM in_sample graph (see graph on the right).The long entries (green line) correspond with an exit (black line) and short entries (red line) correspond with an exit (black line)

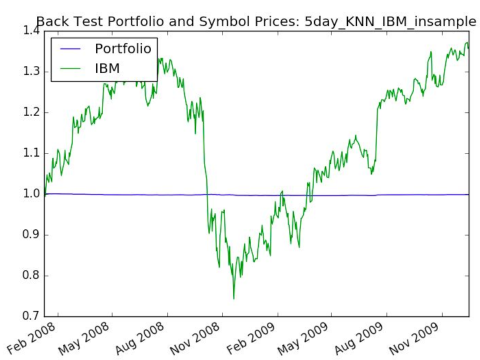



| Date | Symbol | Order | Shares | Type | Strat |
|---|---|---|---|---|---|
| 1/9/08 | IBM | BUY | 100 | Entry | Long |
| 1/15/08 | IBM | SELL | 100 | Exit | Long |
| 1/24/08 | IBM | SELL | 100 | Entry | Short |
| 1/30/08 | IBM | BUY | 100 | Exit | Short |
| 1/31/08 | IBM | BUY | 100 | Entry | Long |
| 2/6/08 | IBM | SELL | 100 | Exit | Long |
| 2/14/08 | IBM | BUY | 100 | Entry | Long |
| 2/21/08 | IBM | SELL | 100 | Exit | Long |
| 2/22/08 | IBM | SELL | 100 | Entry | Short |
| 2/28/08 | IBM | BUY | 100 | Exit | Short |

In sample comparison of Symbol ("IBM") prices and KNN predicted orders ("Fund").

The graph below shows that the return from KNN strategy is fairly flat. As compared with in sample back testing through ML4T-200 (Sine), applying the same KNN prediction policy for IBM is not a fruitful strategy.

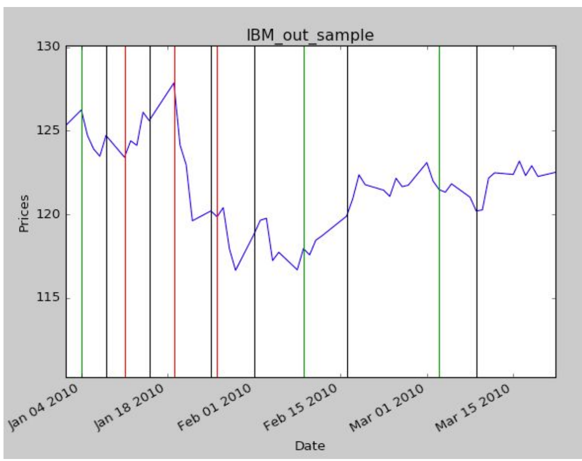



```
Date Range: 2008-01-09 to 2009-12-16*

Sharpe Ratio of Fund: -0.358430323444
```

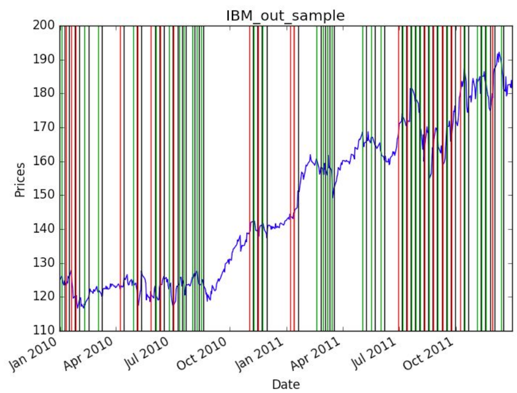
*The portfolio value is evaluate by the date range of the orders, since we are using x rolling values that restrict our ability to trade, where we can only trade x days from the first day of the period and x days from the last day of the period. As depicted below the portfolio calculations between IBM actual values and Portfolio KNN strategy are very similar. The starting values is $10K so given my strategy, this strategy lost money. If the objective was the project was to make the most money possible, i would go back and revise my strategy or policy.

```
Sharpe Ratio of Symbol: IBM    0.77374

Cumulative Return of Fund: -0.001557
Cumulative Return of Symbol: IBM    0.358773

Standard Deviation of Fund: 0.00014069046706
Standard Deviation of Symbol: IBM    0.020108

Average Daily Return of Fund: -3.17664837477e-06
Average Daily Return of Symbol: IBM    0.00098

Final Portfolio Value: 9984.0
```

**IBM Out of Sample** trading strategy visualizations
Depicting exits and entries including sample of first 10 orders

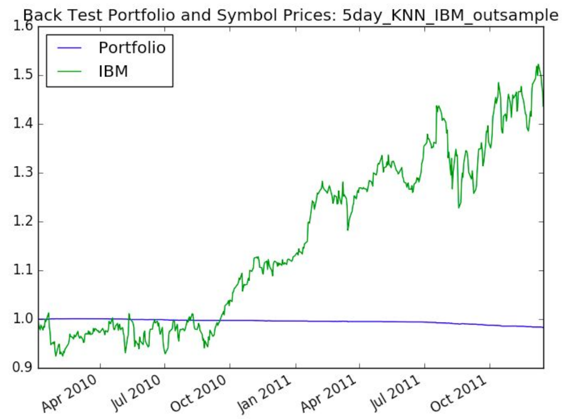



Zoom into the first section of the IBM in_sample graph (see complete graph and orders to the right).The long entries (green line) correspond with an exit (black line) and short entries (red line) correspond with an exit (black line). Applying the strategy that I applied with SINE did not do so well with real stock data. The entry points do not seem consistent with higher returns. For example around January 2011 a bunch of shorts are executed while the price of IBM goes up.




**IBM Out of Sample Orders**

| Date | Symbol | Order | Shares | Type | Strat |
|---|---|---|---|---|---|
| 1/4/10 | IBM | BUY | 100 | Entry | Long |
| 1/8/10 | IBM | SELL | 100 | Exit | Long |
| 1/11/10 | IBM | SELL | 100 | Entry | Short |
| 1/15/10 | IBM | BUY | 100 | Exit | Short |
| 1/19/10 | IBM | SELL | 100 | Entry | Short |
| 1/25/10 | IBM | BUY | 100 | Exit | Short |
| 1/26/10 | IBM | SELL | 100 | Entry | Short |
| 2/1/10 | IBM | BUY | 100 | Exit | Short |
| 2/9/10 | IBM | BUY | 100 | Entry | Long |
| 2/16/10 | IBM | SELL | 100 | Exit | Long |

| | |
|---|---|
| <br>Date Range: 2010-01-04 to 2011-12-16<br><br>Sharpe Ratio of Fund: -3.77903538457<br>Sharpe Ratio of Symbol: IBM 0.70938<br>dtype: float64<br><br>Cumulative Return of Fund: -0.017178<br>Cumulative Return of Symbol: IBM 0.436312<br>dtype: float64<br><br>Standard Deviation of Fund: 0.000147292415589<br>Standard Deviation of Symbol: IBM 0.012932<br>dtype: float64<br><br>Average Daily Return of Fund: -3.50639689164e-05<br>Average Daily Return of Symbol: IBM 0.000578<br>dtype: float64<br><br>Final Portfolio Value: 9828.0 | Out of sample comparison of Symbol ("IBM") prices and KNN predicted orders ("Fund") substantiates with the the in sample graph and backtest that the trading policy does not perform well. Unlike SINE data, IBM prices do not follow a consistent pattern in either of the in or out of sample. The 5-day KNN prediction strategy, with k=3 in KNN, failed to increase returns. In addition, the technical features applied were not able to predict distinct patterns for stock data. |