MC3 Project 2 Report

Vedanuj Goswami
GT ID : 903126228
Georgia Tech, Atlanta, GA

December 1, 2015

1. Technical Indicators

The technical indicators used are as mentioned below :

1. **Bollinger Bands** :
$$bb\_value[t] = (price[t] - SMA[t])/(2 * stdev[t])$$
where SMA is the simple moving average with window size 5 and $stdev[t]$ is the rolling standard deviation again with a window size of 5.
```
sma = get_rolling_mean(price,window = 5)
rstd = get_rolling_std(price,window = 5)
bb_value = (price - sma)/(2*rstd)
```
2. **Momentum** :
$$momentum[t] = (price[t]/price[t - N]) - 1$$
here value of N is taken as 5.
```
momentum = price/price.shift(5) - 1
```
3. **Volatility** :
$$volatility[t] = stdev((price[t]/price[t - 1]) - 1)$$
where $stdev$ is the rolling standard deviation. Here a window size of 5 is taken for rolling standard deviation.
```
volatility = get_rolling_std((price/price.shift(1) - 1),window = 5)
```

1.1. *Training Y*

For determining the Training Y we use a 5 day change of price.

$$Y[t] = (price[t + 5]/price[t]) - 1$$

```
Y = (price.shift(-5)/price) - 1
```

1.2. *Learning*

A KNN Learner with bagging is used for learning. We use a bag size of 10 and k = 3 for this project.

2. Trading Strategy

The trading Strategy used is simple. When the predicted price increases by 5% we long the stock and hold it for 5 days. Similarly when the predicted price decreases by 5% as compared to 5 day earlier price we short the stock and hold it for 5 days. We use the assumption that we are first exiting the previous position and then entering a new position. For both Long and Short position we buy or sell 100 stocks.

3. Charts for ML4T-399

	In Sample Sine Data	**Out Sample Sine Data**
RMSE	0.0131265566018	0.0455007517506
Correlation	0.995909899496	0.950176932092
Sharpe Ratio of Fund	10.3612540378	12.1561018948
Sharpe Ratio of SPX	-0.199573937554	0.671461673758
Cumulative Return of Fund	7.51520135	3.8762989
Cumulative Return of SPX	-0.229456314437	0.110018623289
Standard Deviation of Fund	0.0065573921306	0.008313721748785
Standard Deviation of SPX	0.0219652414615	0.0113518252015
Average Daily Return of Fund	0.00427999445792	0.00636633586086
Average Daily Return of SPX	-0.0002761464963	0.00048016074666
Final Portfolio Value	85152.0135	48762.989

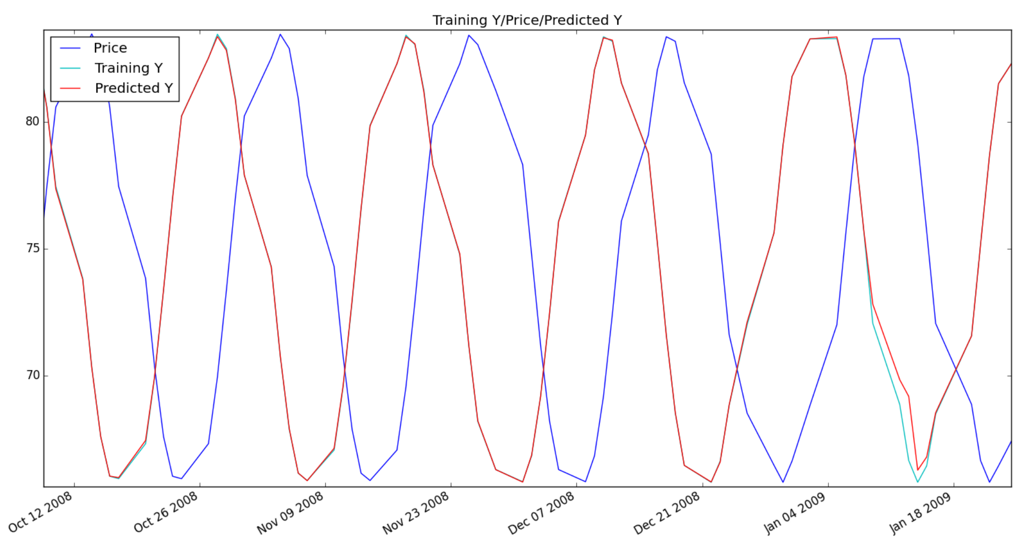

Fig. 1. **Training Y/Price/Predicted Y**: Figure illustrating Training Y values in grey, Price in blue and Predicted Y in Red.

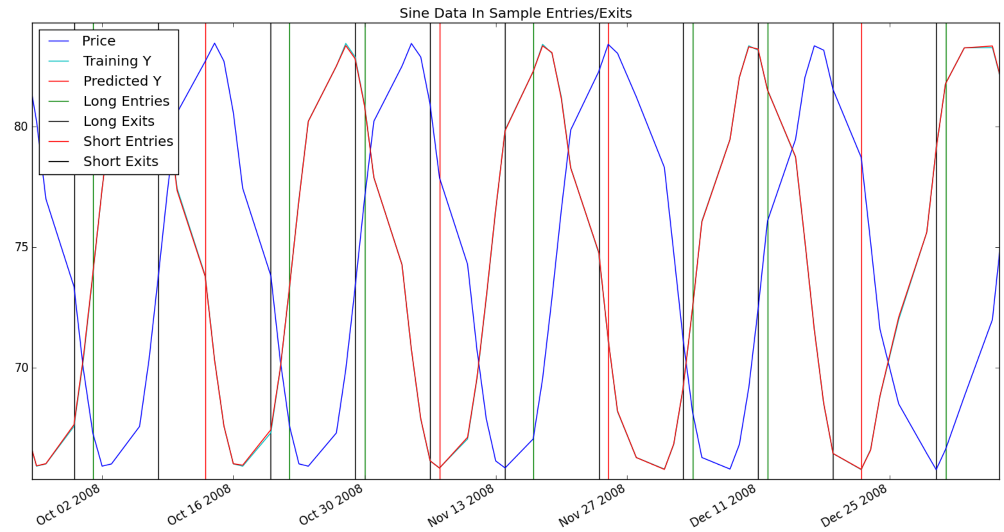

Fig. 2. **Sine Data In Sample Entries/Exits**: Figure illustrating entry and exits as vertical lines on a price chart for the in sample period 2008-2009.

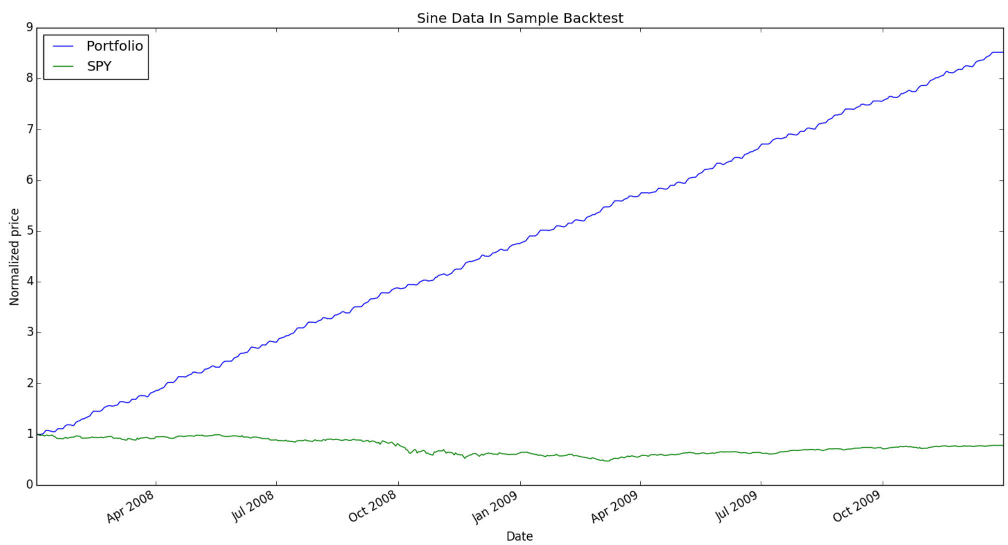

Fig. 3. **Sine Data In Sample Backtest** : Portfolio with generated orders perform substantially well with respect to SPY over the in sample period 2008-2009

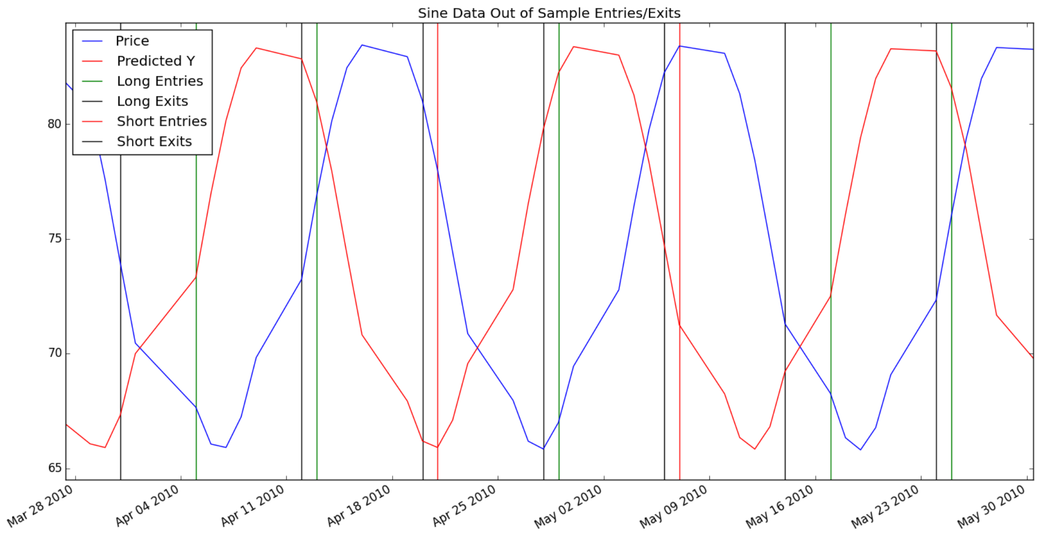

Fig. 4. **Sine Data Out of Sample Entries/Exits** : Figure illustrating entry and exits as vertical lines on a price chart for the out sample test period 2010.

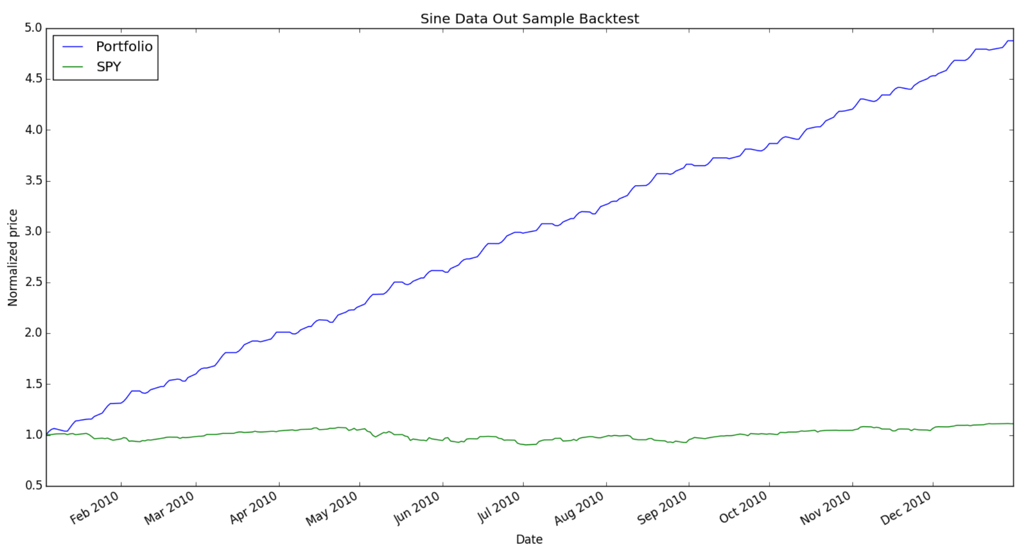

Fig. 5. **Sine Data Out of Sample Backtest** : Portfolio with generated orders over out sample predicted prices perform substantially well with respect to SPY for the out sample test period 2010.

3.1. Discussion of Results for 399 Data

We observe from the plots that KNN with bagging predicts 5 day returns outstandingly well and the orders generated from these predictions when compared against SPY shows cumulative return of more than

700%. The reason for this is because the 399 data fits properly in a sine curve. Hence KNN can predict correctly as the data to be predicted fits into the sine curve perfectly and so the average of nearest points will also lie on the sine curve. This is evident from the fact that the RMSE error is very low at 0.0131265566018 and the Correlation is very high at 0.995909899496.

For Out Sample data also the predicted change of price generates orders that outperform SPY considerably with a cumulative return of around 230%. Here also the prediction is almost accurate as is evident from the very low RMSE of 0.0455007517506 and the very high Correlation of 0.950176932092. The reason for this high performance is again that the backtest data fits properly into the sine curve that was used for training the learner and the predicted return by averaging the values of nearest points in the learner also lies on the sine curve.

4. Charts for IBM

	In Sample IBM Data	**Out Sample IBM Data**
RMSE	0.0218835055248	0.0310716180404
Correlation	0.876967106721	0.0395721327749
Sharpe Ratio of Fund	2.48998137157	0.870280849351
Sharpe Ratio of SPX	-0.199573937554	0.671461673758
Cumulative Return of Fund	0.6027	0.0632
Cumulative Return of SPX	-0.229456314437	0.110018623289
Standard Deviation of Fund	0.0.00608463880385	0.00464970469344
Standard Deviation of SPX	0.0219652414615	0.0113518252015
Average Daily Return of Fund	0.000954400438858	0.000254908623555
Average Daily Return of SPX	-0.0002761464963	0.00048016074666
Final Portfolio Value	16027.0	10632.0

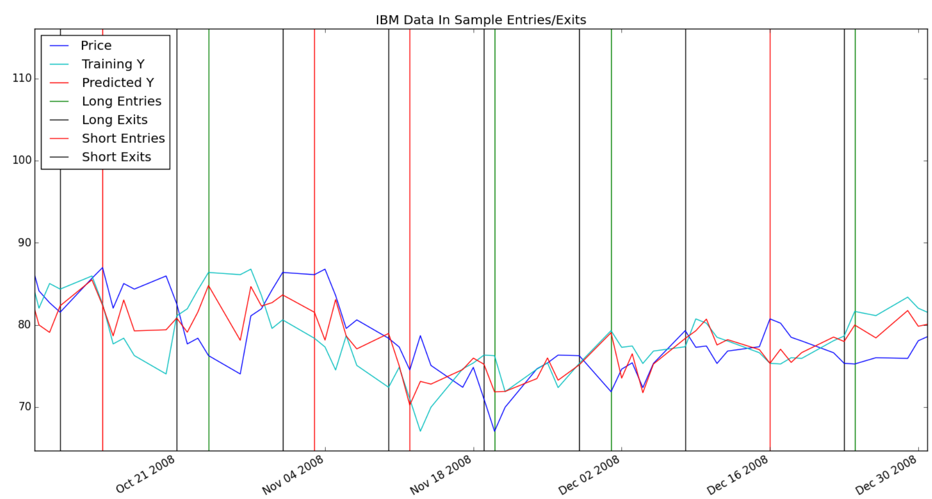

Fig. 6. **IBM Data In Sample Entries/Exits**: Plot illustrating entry and exits as vertical lines on a price chart for the in sample period 2008-2009

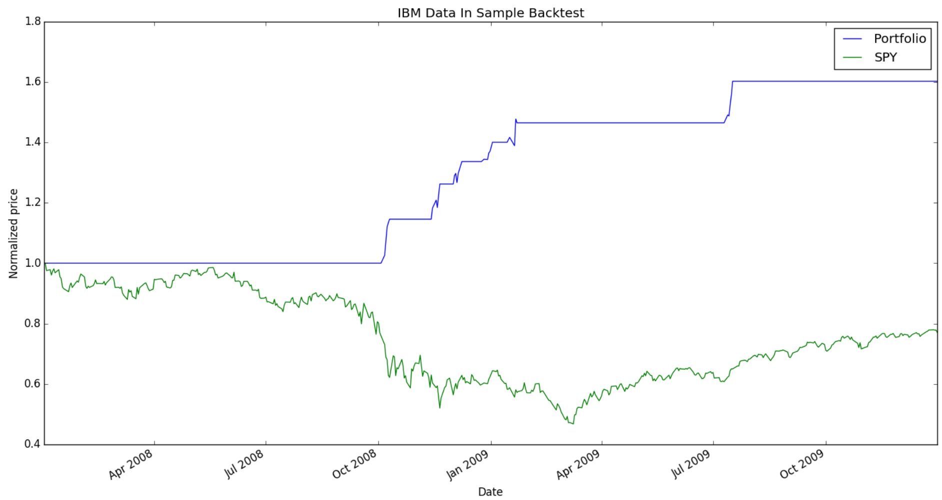

Fig. 7. **IBM Data In Sample Backtest** : Portfolio with generated orders perform just well with respect to SPY for the in sample period 2008-2009.

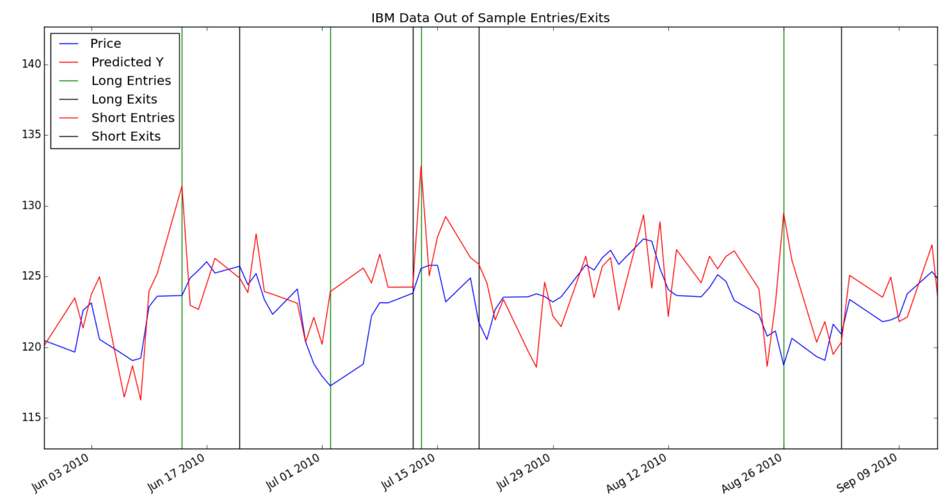

Fig. 8. **IBM Data Out of Sample Entries/Exits** : Figure illustrating entry and exits as vertical lines on a price chart for the out sample test period 2010.

4.1. Discussion of Results for IBM Data

We observe from the plots that KNN with bagging predicts 5 day returns *just well enough* for a positive return over the in sample data. The orders generated from these predictions when compared against SPY shows cumulative return of 60%. The RMSE error for in sample IBM data is 0.0218835055248 and correlation is 0.876967106721.

For Out Sample data the predicted change of price generates orders that outperforms SPY with cumulative return just touching 6%. The reason for this is that the real world stock price data is not uniform and

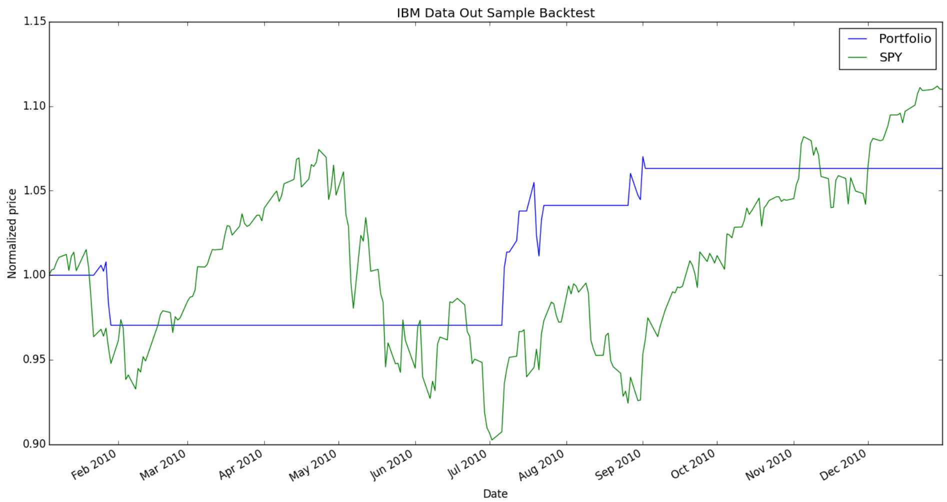

Fig. 9. **IBM Data Out of Sample Backtest** : Portfolio with generated orders over out sample predicted prices perform just marginally better with respect to SPY for the out sample test period 2010.

doesn't fit properly into a definite curve. The training and testing price ranges having no similarity. Hence making predictions on the backtest data with the help of nearest points of training data does not predict very accurately as compared to uniform sine data. Hence KNN cannot predict correctly as the learner has no nearest points when it comes to the back-testing data. The RMSE error for out sample IBM data is 0.0310716180404 and correlation is very low at 0.0395721327749.

What would you do differently? : To better predict we can use different technical indicators, and test to determine which set of technical indicators perform best for a particular stock. Also using Reinforcement Learning methods like Q Learning will help us better prediction since they find an optimal action-selection policy for any given (finite) Markov decision process. We can also try Neural Nets and genetic algorithms to check how they will perform over various stocks.