

MC3-Project-2

Xiaobing Li
Xli605@gatech.edu

Overview

In this project I trained a KNN regression learner without bagging on ML4T-399 and IBM data from 2008 to 2009 and tried to predict their future 5 day returns.

Methods

There were 5 technical features used in the model: Bollinger bands, Momentum (N=5days), Momentum (N=10days), SMA (N=5days) and SMA (N=10days).

These features were normalized so their output values are ranging from -1.0 to 1.0. This will help avoid the situation where one feature overwhelms the results. The normalization formulae are shown below:

Normalized Bollinger bands: $bb_value[t] = (price[t] - SMA[t,20]) / (2 * stdev[t])$

Normalized Momentum: $momentum[t] = (price[t] / price[t-N]) - 1$ (N=5 for 5 days momentum and N=10 for 10 days momentum)

Normalized SMA: $sma[t,N] = SMA[t,N] / price[t] - 1$ (N=5 for 5days SMA and N=10 for 10 days SMA)

The values to-be-predicted are future 5 day returns, which are defined as:

$$Y[t] = (price[t+5] / price[t]) - 1.0$$

For a given dataset, I took the price data from 2007-12-31 to 2009-12-31, then calculated five TA features as my training set.

Ideally I should test different k values to get an optimized one, however, based on my conclusion from the last project, bagging can effectively remove the overfitting and reduce the MSE, so I just set k=3 and with bags=20.

Results

ML4T-399 Dataset

I first did the training and backtesting with ML4T-399 Dataset.

Training results Figure 1 shows the training result from the dataset ML4T-399. The training Y values (future-5-day returns) are shown in Blue, prices in Green and predicted Y from my model in Red. For clarification, only a two-month period is shown.

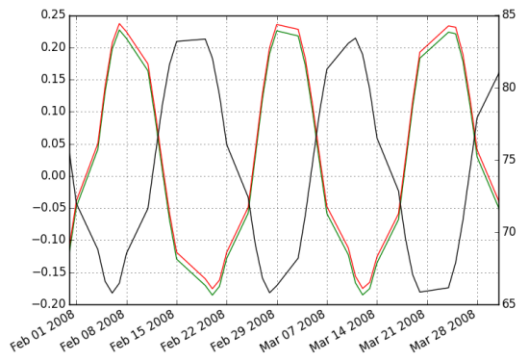
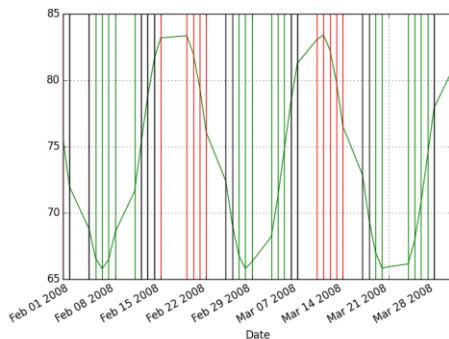


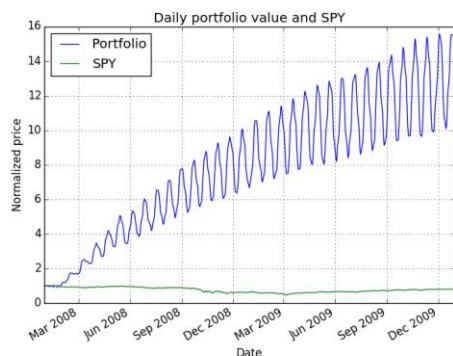
Figure 1. Actual Y values are in Green, Predicted Y values are in Red and the original prices are in Black. Because Predicted Y values are perfectly matched with the actual Y values, the green line cannot be seen. Therefore in the right panel, I shifted the red line slightly upward 0.01.

Trading policy Based on the training result, I choose a trading policy as below: when the learner predicts the price will go up more than 10%, I will buy and hold for 5 days and decide to sell if the learner predicts less than 10% increase at the fifth day; when the learner predicts a more than 10% drop, I will short and hold until the learner predicts a less than 10% drop.

Figure 2 illustrates entry and exits as vertical lines based on this trading policy. Long entries are shown as green lines, short entries as red lines and exits as black lines.



Orders are generated through my market simulator and Figure 3 is the returns for this backtest. It is doing VERY well for this training set at the time period 2008-2009. The summary of the backtest is listed in Table 1. (i did something wrong here. have to double-check)



Data Range: 2007-12-31 to 2009-12-31

Sharpe Ratio of Fund: 1.8095237921

Sharpe Ratio of SPY: -0.149575888341

Cumulative Return of Fund: 12.411505

Cumulative Return of SPY: -0.201395139514

Standard Deviation of Fund: 0.0599476848508

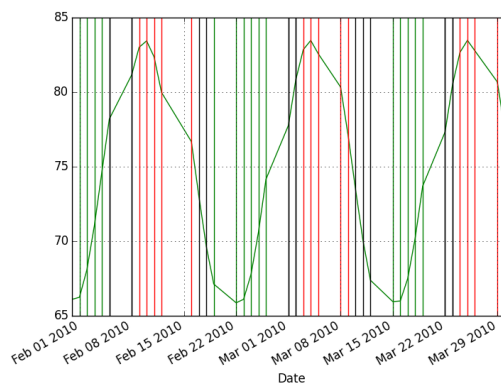
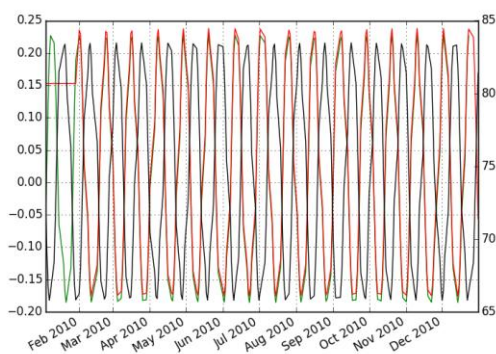
Standard Deviation of SPY: 0.0219136847778

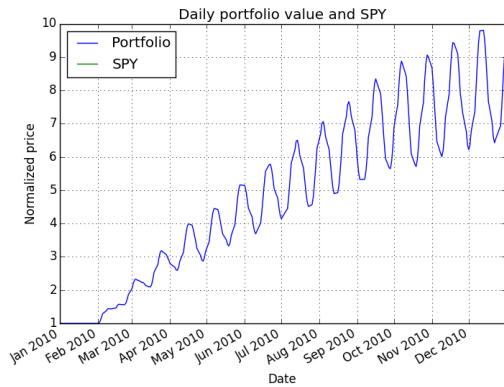
Average Daily Return of Fund: 0.00683339369838

Average Daily Return of SPY: -0.000206479400499

Final Portfolio Value: 134115.05

Testing results Next I applied the trained model to the data on the time period 2010. Figure 4 plots the stock price, actual 5-day returns and predicted returns in Green, Blue and Red, respectively. And Figure 5 shows the entries and exits. Finally, a backtest result is shown in Figure 6 and Table 2.





Data Range: 2010-01-01 to 2010-12-31

Sharpe Ratio of Fund: 2.92148060788

Sharpe Ratio of SPY: 0.758702117743

Cumulative Return of Fund: 8.45679036

Cumulative Return of SPY: 0.126812598134

Standard Deviation of Fund: 0.0545801169815

Standard Deviation of SPY: 0.0112823505018

Average Daily Return of Fund: 0.0100447053023

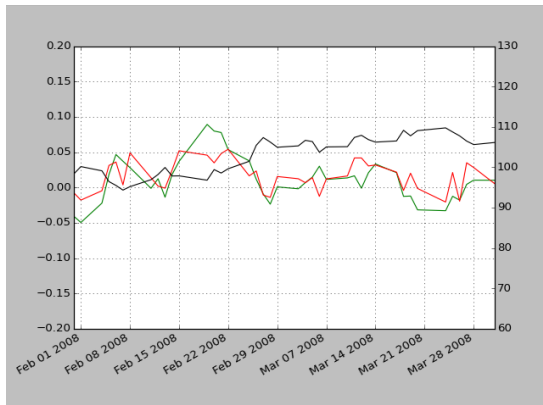
Average Daily Return of SPY: 0.000539225737952

Final Portfolio Value: 94567.9036

IBM Dataset

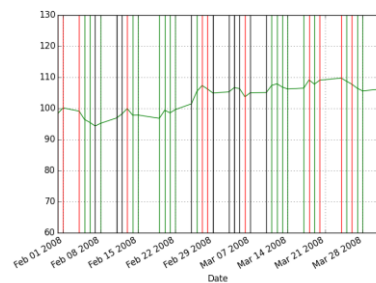
Next I did the training and testing with the IBM dataset. All the parameters and colors used in this analysis are identical to the ones used in the analysis to the ML4T-399 Dataset.

Training results Figure 7 illustrates the training Y values (future-5-day returns) in Blue, prices in Green and predicted Y from my model in Red. From this chart we can see the knn learner performed so-so. For clarification, only a two-month period is shown.

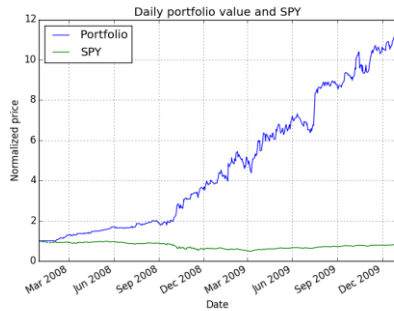


Trading policy Based on the training result, I choose a trading policy as below: when the learner predicts the price will go up more than 1%, I will buy and hold for 5 days until the learner predicts a less than 1% return; when the learner predicts a more than 1% drop, I will sell and hold until the learner predicts a less than 1% drop.

Figure 8 illustrates entry and exits as vertical lines based on this trading policy. Long entries are shown as green lines, short entries as red lines and exits as black lines



Orders are generated through my market simulator and Figure 9 is the returns for this backtest. It is doing VERY well for this training set at the time period 2008-2009. The summary of the backtest is listed in Table 3.



Data Range: 2007-12-31 to 2009-12-31

Sharpe Ratio of Fund: 2.9333947998

Sharpe Ratio of SPY: -0.149575888341

Cumulative Return of Fund: 10.0706

Cumulative Return of SPY: -0.201395139514

Standard Deviation of Fund: 0.0272777175168

Standard Deviation of SPY: 0.0219136847778

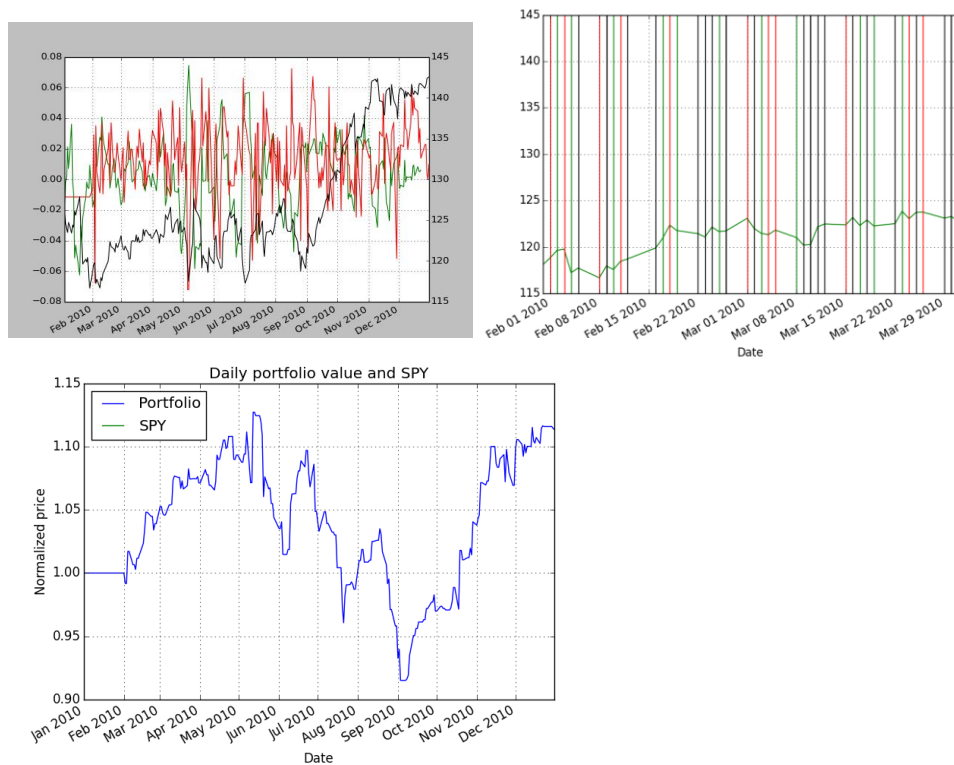
Average Daily Return of Fund: 0.00504055403717

Average Daily Return of SPY: -0.000206479400499

Final Portfolio Value: 110706.0

Testing results

Next I applied the trained model to the data on the time period 2010. Figure 10 plots the stock price, actual 5-day returns and predicted returns in Green, Blue and Red, respectively. And Figure 11 shows the entries and exits. Finally, a backtest result is shown in Figure 12 and Table 4.



Data Range: 2010-01-01 to 2010-12-31

Sharpe Ratio of Fund: 0.69329221214

Sharpe Ratio of SPY: 0.758702117743

Cumulative Return of Fund: 0.1135

Cumulative Return of SPY: 0.126812598134

Standard Deviation of Fund: 0.0106535749575

Standard Deviation of SPY: 0.0112823505018

Average Daily Return of Fund: 0.000465276820654

Average Daily Return of SPY: 0.000539225737952

Final Portfolio Value: 11135.0

Discussion

The knn model with bagging is working very well with the data for ML4T-399 but not IBM. Because ML4T-399 repeats the same pattern but IBM is not.