Ningmu Zou
nzou3@gatech.edu

In order to make this report clear and easy to grade, I follow the rubric step by step.

Indicators (up to 20% potential deductions):

Is each indicator described in sufficient detail that someone else could reproduce it? (-5% for each if not)

In this part, I used four indicators: Price/SMA ratio, Bollinger Bands percentage, 15 days momentum and Accumulation/distribution index. Here is how I calculated each indicator:

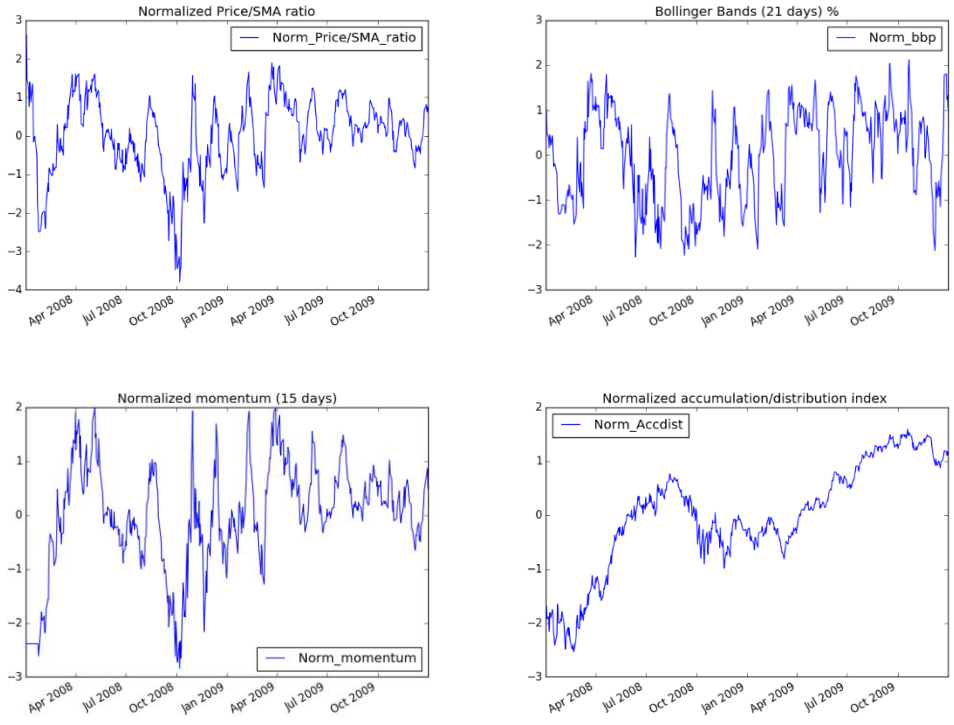
Price/SMA ratio: Price is “Adj Close” column in the data; SMA is 21 days rolling average, with all non-value being backfilled.

Bollinger Bands percentage: (Price - SMA + 2*STD/ (4*STD), in which STD is the 21 days rolling standard deviation, with all non-value being backfilled.

15 days momentum: Price(t) / Price(t-15) – 1, with all non-value being backfilled.

Accumulation/distribution index: CLV(t) * Volume(t), in which Close Location Value (CLV) was calculated by $\frac{(close-low)-(high-close)}{high-low}$ in each day, Volume is “Volume” column in the data.

Is there a chart for each indicator that properly illustrates its operation? (-5% for each if not)



Is at least one indicator different from those provided by the instructor's code (i.e., another indicator that is not SMA, Bollinger Bands or RSI) (-10% if not)

Both 15 days momentum and Accumulation/distribution index are different from instructor’s code.

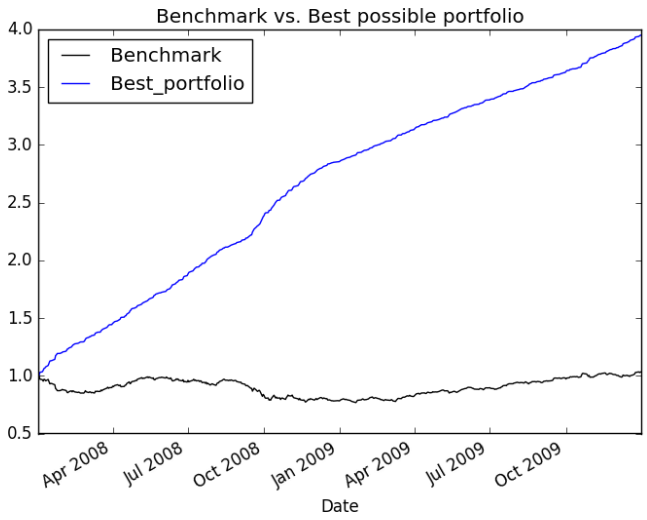
Does the submitted code indicators.py properly reflect the indicators provided in the report (-20% if not)

Please see “indicators.py”.

Ningmu Zou
nzou3@gatech.edu

Best possible (up to 5% potential deductions):

Is the chart correct (dates and equity curve) (-5% for if not)



Is the reported performance correct within 5% (-1% for each item if not)

Cumulative return of the benchmark is 0.03164

Cumulative return of the best possible portfolio is 2.94948

Standard deviation of daily returns of benchmark is 0.00872321348875

Standard deviation of daily returns of best portfolio is 0.00321528788052

Mean of daily returns of benchmark is 9.98709603011e-05

Mean of daily returns of portfolio is 0.00272878701461

Manual rule-based trader (up to 20% deductions):

Is the trading strategy described with clarity and in sufficient detail that someone else could reproduce it? (-10%)

In this task, I applied the rule with indicators of both “AAPL” and “SPY”. The detailed strategy is that:

If Price/SMA_ratio(“AAPL”) > 1.3 and Price/SMA_ratio(“SPY”) <= 0.65 and Bollinger Bands percentage(“SPY”) < 0.8: the stock is overbought, DO SELL;

If Price/SMA_ratio(“AAPL”) <0.95 and Price/SMA_ratio(“SPY”) >= 0.95 and Bollinger Bands percentage(“SPY”) >=0 : the stock is oversold, DO BUY;

Otherwise, DO NOTHING.

The detailed reason of this strategy is that Price/SMA ratio typically represents the overbought situation of a stock. If the ratio of a stock is too high while the ratio of SPY is not that high, it often means that this stock is overbought or overestimated, which would bring a possible opportunity for selling this stock. On the other hand, if the ratio of a stock is too low while the ratio of SPY is

not that low, it often means that this stock is oversold or underestimated, which would bring a possible opportunity for buying this stock. Bollinger Bands percentage of SPY represents the whole trend of the market. Here, I confined it in a certain range so that the volatility of the market is controllable. I would prefer to choose a conservative strategy rather than an aggressive one. Here, I gave up the momentum indicator and accumulation/distribution index indicator. I found the momentum indicator is highly correlated with SMA, which is understandable if we consider the calculation process of these two indicators. The accumulation/distribution index is a volume-based indicator designed to measure the cumulative flow of money into and out of a security. Generally, a low negative number combined with high volume reflects strong selling pressure that pushes the indicator lower. I checked and tweaked this indicator and found that this volume based indicator is not that important for this stock during the in-sample period.

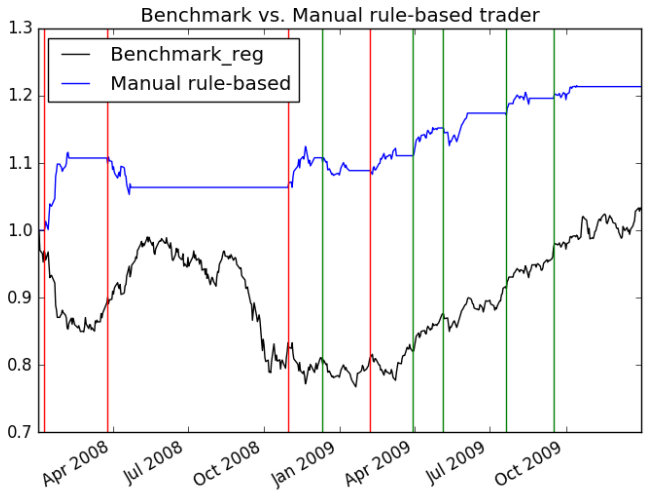
Does the provided chart include:

Historic value of benchmark normalized to 1.0 with black line (-5% if not) See below

Historic value of portfolio normalized to 1.0 with blue line (-10% if not) See below

Are the appropriate date ranges covered? (-5% if not) See below

Are vertical lines included to indicate entries (-10% if not)



Does the submitted code rule_based.py properly reflect the strategy provided in the report? (-20% if not)

Please see “rule_based.py”.

Does the manual trading system provide higher cumulative return than the benchmark over the in-sample time period? (-5% if not)

Yes.

ML-based trader (up to 30% deductions):

Is the ML strategy described with clarity and in sufficient detail that someone else could reproduce it? (-10%) See below

Are modifications/tweaks to the basic decision tree learner fully described (-10%) See below

Does the methodology utilize a classification-based learner? (-30%)

Here, the first thing to do is to change the RTlearner from a regression learner to a classification learner. The change I made is based on a "voter" model: in previous regression learner, initialization of Y-value in a "leaf" is based on the mean Y-value of training data:

```
leaf = np.array([-1, np.mean(dataY), np.nan, np.nan])
```

In order to change it to classification learner, I replace the code with

```
leaf = np.array([-1, max(dataY.tolist(),key=dataY.tolist().count), np.nan, np.nan])
```

Instead of using the average value of training Y data, I used the most frequent value of training Y data. Specifically, in this question, it depends on the occurrence frequency of -1, +1, and 0. This change can fulfill the requirement of classification learner.

Using the same method, I also changes the BagLearner. Instead of using the average result from each training, I changed into a "vote" mode to make sure the classification result is from the most frequent value of Y:

```
for i in range(0,points.shape[0]):
    Y_return[i] = Counter(Y[i, :]).most_common(1)[0][0]
```

The other details are described below:

I used 4 indicators as I described before. YBUY and YSELL values are set as 0.06 and -0.06, which can be changed but the difference is not apparent. According to the YBUY and YSELL value, all 21 days return data can be converted into +1, -1, and 0, as training Y data. The other parameters are very similar as previous project. In this task, the leaf size was set as 5 and the bag size of BagLearner was set as 10.

Does the provided chart include:

Historic value of benchmark normalized to 1.0 with black line (-5% if not)See below

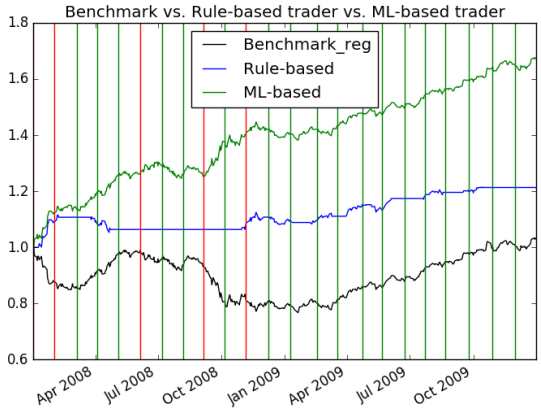
Historic value of rule-based portfolio normalized to 1.0 with blue line (-5% if not)See below

Historic value of ML-based portfolio normalized to 1.0 with green line (-10% if not)See below

Are the appropriate date ranges covered? (-5% if not)See below

Are vertical lines included to indicate entry (-10% if not)See below

Ningmu Zou
nzou3@gatech.edu



Does the submitted code ML_based.py properly reflect the strategy provided in the report? (-30% if not)

Please see "ML_based.py".

Does the ML trading system provide 1.5x higher cumulative return or than the benchmark over the in-sample time period? (-5% if not)

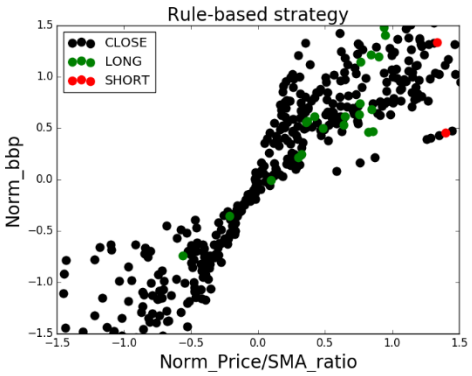
Yes

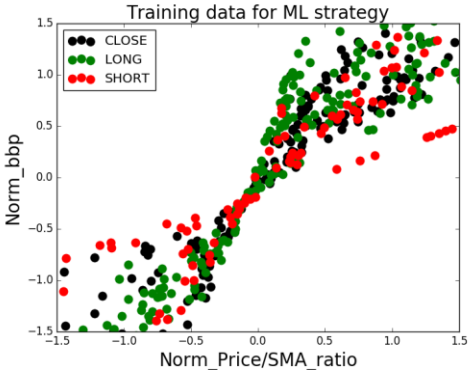
Data visualization (up to 15% deductions):

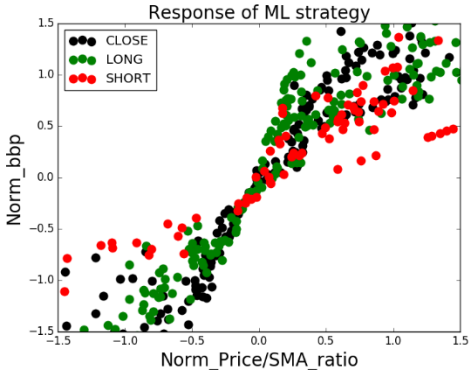
Is the X data reported in all three charts the same? (-5% if not) See below

Is the X data standardized? (-5% if not) See below

Is the Y data in the train and query plots similar (-5% if not)



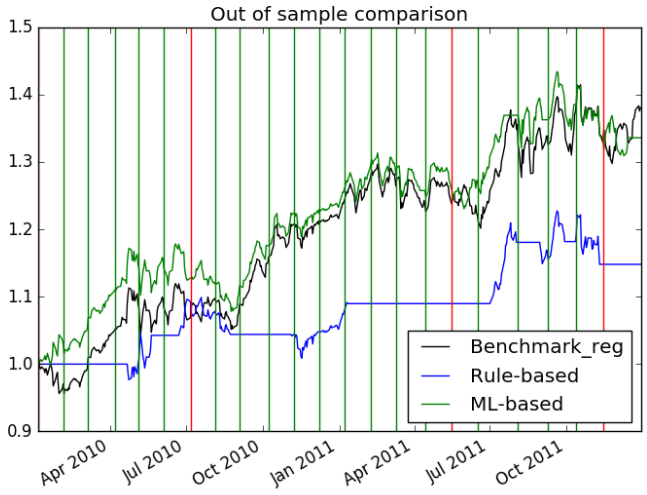




Ningmu Zou
nzou3@gatech.edu

Comparative analysis (up to 10% deductions):

Is the appropriate chart provided (-5% for each missing element, up to a maximum of -10%)



Is there a table that reports in-sample and out-of-sample data for the baseline (just the stock), rule-based, and ML-based strategies? (-5% for each missing element)

	Start Date	End Date	Sharpe Ratio	Cumulative Return	Standard Deviation	Average Daily Return
Benchmark	'2008-01-02'	'2009-12-31'	0.182	0.0316	0.00874	0.0001
Rule-based	'2008-01-02'	'2009-12-31'	1.57	0.213	0.00397	0.000392
ML-based	'2008-01-02'	'2009-12-31'	2.83	0.675	0.00583	0.00104
Benchmark	'2010-01-04'	'2011-12-30'	1.26	0.38	0.00855	0.000678
Rule-based	'2010-01-04'	'2011-12-30'	0.822	0.148	0.00561	0.00029
ML-based	'2010-01-04'	'2011-12-30'	1.21	0.336	0.00794	6.08E-04

Are differences between the in-sample and out-of-sample performances appropriately explained (-5%)

The difference between the in-sample data and out-of-sample data is from several aspects. First, for the benchmark result, the in-sample data shows a decrease and then increase trend, with the cumulative return as 0.0316, which is much lower than out-of-sample cumulative return as 0.38. This means that a fund manager needs more skills to gain profit during the in-sample period rather than the out-of-sample. Second, because I used one indicator of SPY, the correlation between SPY and AAPL would also affect the result. That is why my rule-based strategy for out-of-sample data works not as good as benchmark. We can consider the out-of-sample period as a bull market. In this situation, one of the good options is keeping a stock for a long time, which is similar as benchmark's behavior. Third, for the machine learning-based strategy, in sample period has 0.675 cumulative return while out sample period only has 0.336 cumulative return, similar as benchmark. This result is also expected. Although the overfitting issue in this problem is not apparent, the out of sample's high benchmark return makes the ML-based method not that exciting.