Strategy Learner Project Report

Part 1:

Step 1:
Get the indicators' value of training data. Combine them with regard to date. Get the training matrix X.

Step 2:
Classify y based on N day return. Classify the example as a +1 or "LONG" if the N day return exceeds a certain value(YBUY). Classify the example as a -1 or "SHORT" if the N day return is below a certain value (YSELL). In all other cases the sample should be classified as a 0 or "CASH." Get the training matrix Y.

Step 3:
Put the training matrix X and Y into out machine learning model the train the bag learner model.

Step 4:
Get the indicators' value of testing data. Combine them with regard to date. Query it on the model we built before and get the predication of classification Y.

My indicators are Simple Moving Average, Simple Moving Standard Deviation, Bollinger Bands and Momentum.
Since it is a classification problem, I use discretization on Y and normalization on all indicators (X).

Part 2:

I use the in sample data from January 1, 2008 to December 31 2009 with start value 100000. For the machine learning strategy, I choose bag learner with number of bags = 20. And the basic learner for bag learners is random tree with leaf_size = 5. Here is the plot for the portfolio of both strategies along with the benchmark using in sample data.

Normalized Portfolio of Benchmark, Manual Strategy, Machine Learning Strategy
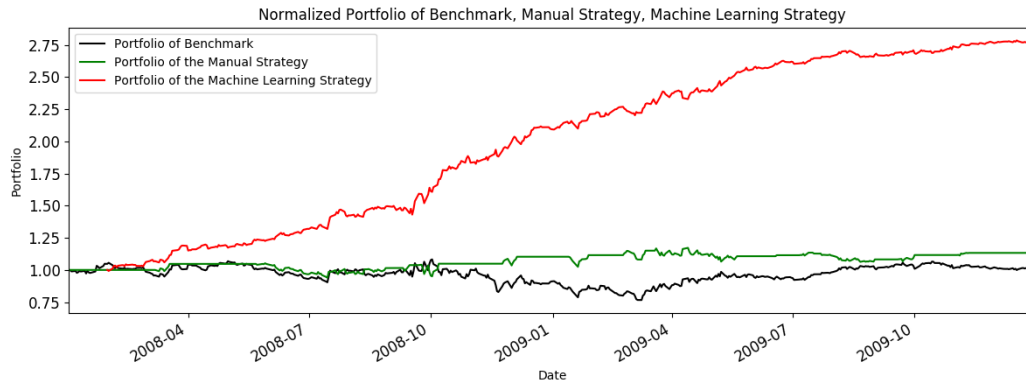
As we see from the plot, it is pretty obvious that the manual strategy only performs a little bit better than benchmark. But the machine learning algorithm works much more better than both manual strategy and benchmark. I also calculate some statistical features which determines the outstanding performance of machine learning algorithm:
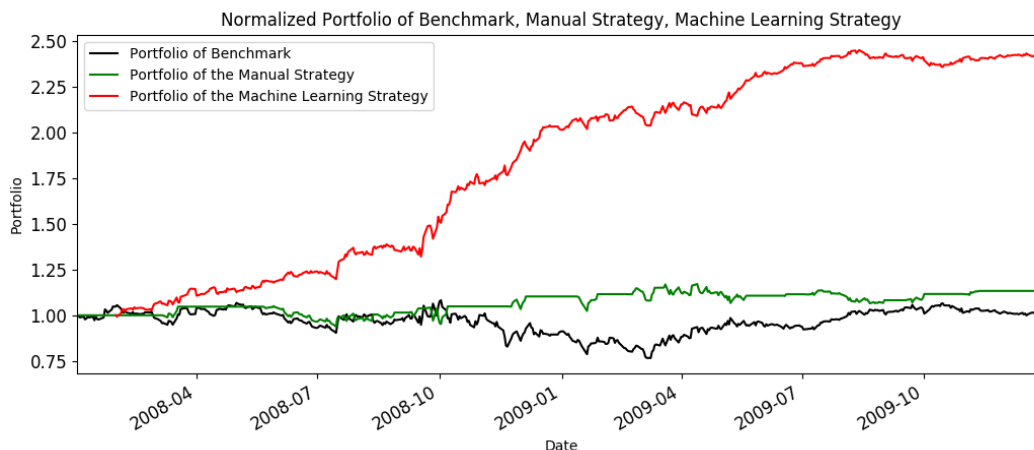
	Benchmark	Manual Strategy	Machine Learning Strategy
Cumulative return	0.0123	0.1326	1.7722
Stdev of daily returns	0.017004	0.000307	0.009336
Mean of daily returns	0.000168	0.010934	0.002152
Sharpe Ratio	0.156918	0.445043	3.659414
Final Value of portfolio	101230.0	113260.0	277220.0

I do not think the exact value of portfolio in machine learning strategy stays same every time running this algorithm since we use random trees and the split attribute varies each time. But I do expect the much better performance of machine learning algorithm than benchmark and manual strategy, since with machine learning algorithm we can predict the future more accurately to some degree.

Part 3:

I guess that with the increasing of the value of impact, people tends to decrease trading frequency. So that each of their trading behavior has more significant changes on portfolio values which means portfolio values become more unstable.

On this experiment, I use the in sample data from January 1, 2008 to December 31 2009 with start value 100000 and I set the impact value = 0.005. For the machine learning strategy, I choose bag learner with number of bags = 20. And the basic learner for bag learners is random tree with leaf_size = 5. All same as part 2 except the impact value. Here is the plot for the portfolio of both strategies along with the benchmark using in sample data. We can find that some parts of the red line become more steep than in part 2, which means the portfolio earned is more unstable than without impact value.

I also calculate some statistical features which proves my hypothesis:

	Benchmark	Manual Strategy	Machine Learning Strategy
Cumulative return	0.0123	0.133263	1.42043501949
Stdev of daily returns	0.017004	0.000308	0.010143
Mean of daily returns	0.000168	0.01092	0.001879
Sharpe Ratio	0.156918	0.447082	2.940855
Final Value of portfolio	101230.0	113326.3	242043.501949

We can find that the standard deviation of daily returns is larger than those in part 2.