DS-670-Capstone: Big Data & Business Analytics

Lab 9

April 21st 2017, Madhumita D

1. **Reference:**

   Kim, J. H., Shamsuddin, A., & Lim, K. P. (2011). Stock return predictability and the adaptive markets hypothesis: Evidence from century-long US data. Journal of Empirical Finance, 18(5), 868-879.

2. **Describe the results of the paper chosen:**

Based on the regression metrics used they find evidence that return predictability fluctuates over time and is highly governed by changing market conditions. It is also found that during a market crash, no return predictability is prominently evident, mostly because of the extreme degree of associated uncertainty. The authors did not use sequential testing for understanding the market efficiency. They also state that inflation, risk-free rates and stock market volatility influence stock market predictability over a period of time.

3. **Describe the results of your paper so far:**

Applying a combination of neural networks and time series analysis, return predictability is calculated. Looking at the forecasted values we can say that the expected return of stocks in last quarter of 2014 has a negative trend while they show a positive trend in the first two quarters of 2015. However, later in third and fourth quarters in 2015 we see a similar negative trend as shown in last quarter of 2014. On the whole the averages depict a huge volatility in stock price movements. The evidence of huge volatility and how the returns predicted for Group1 outperform Group5 we can say that the publically available information has been properly reflected through our predicted returns. Our analysis addresses not only the factors driving performance of the model but also the Investability of our model with real-time data. This entails simulation of a portfolio controlling for risk and liquidity of the assets traded.

Further to understand how our analysis can be used in investing in stocks we need to perform analysis using different time periods like monthly or daily frequency to have better predictions and use the model on real time data. Also, implementing this hybrid model on various industries or sectors allows the investors to reap much higher profits by comparing the results from these industries.

4. **Describe the use of aggregation and group operations.**

The data used for the analysis consisted of 6600 files i.e. each file had information of stock prices and volume for one specific date and ticker (stock name). We also had the information on the fundamental factors which are issued by a company on a quarterly basis. Hence we used the merge function (merge())in R to merge the daily files with the quarterly information. To calculate the RSI (relative strength index) we grouped the first 100 days to compute the average gains / average loses. Finally after modelling the neural networks for the train data (15 models for 15 dates), the future date datasets are ranked using order() in R to get the stocks with higher returns. Then we have grouped these ranked datasets into 5 quantiles using head() and tail() functions. At the end average of each quantile in every quarter is calculated using mean() function. In this way we have used the concepts of aggregation and group operations to understand and get optimal results.

5. **Create a table with every operation on left column and time measurements on the right column.**

Operation	Time Measurement
Data Load	240 seconds
Merging daily files with quarterly data	60 seconds
Calculating RSI values for each file	60 seconds
Normalizing the indicators values for each file	3 seconds
Neural Networking modelling	45 minutes (each model takes about 5 minutes)
Time Series Analysis on weights	5 seconds
Predicting Returns for future quarters (using the neural network equation)	120 seconds
Ranking the datasets (using the expected returns)	5 seconds
Grouping the stocks into quantiles	5 seconds
Average of stocks in each quantile	5 seconds