CSE 487/587 Programming Assignment #2

Due Date: March 18, 2015
Weight: 12% of your grade

Problem: Use R to implement the time-series forecast of stocks in NASDAQ

Description:
In this assignment, you are required to use R language/tools in CCR to do time-series forecast of stock price using the same data in hw#1. There are many approaches to forecasting. In this homework, you will compare three techniques, namely, Linear Regression Model, Holt-Winters Model, and ARIMA model.

Implementation:
In order to finish this task, the data for each stock will be split into two parts:

- The first part with 744 trading days is used for training
- The second part with 10 trading days is used for testing.

Evaluation:
The MAE (Mean Absolute Error) is a common measure to evaluate error in time series analysis. For each stock, the MAE is calculated as below:

$$\text{MAE}_i(\text{each day}) = \left|\text{forecastData}_i - \text{testData}_i\right|$$

$$\text{sum of MAE} = \sum_{i=1}^{10} \text{MAE}_i$$

Based on this error, you are required to find stocks with best-forecasted performance, using the three techniques as follows:

- The top 10 stocks with the minimum sum of MAE using Linear Regression Model
- The top 10 stocks with the minimum sum of MAE using Holt-Winters Model
- The top 10 stocks with the minimum sum of MAE using ARIMA Model

Plot the graph for the above three results in increasing order and save them as “lm.jpg”, “hw.jpg”, “arima.jpg”, respectively.

Data:
Use the data from hw#1 (small size only) located at:
/gpfs/courses/cse587/spring2015/data/hw2/data
/gpfs/courses/cse587/spring2015/data/hw2/stocklist.txt

The first column "Date" and the last column "Adj Close" will be used in this assignment. Other columns can be neglected.

Important Note:

- If the data file is empty, skip it, and do not assign zero value to that stock.
- If the data have less than 36 months, skip it, and do not assign zero value to that stock.

In other words, the data must have 36 months that is 754 lines of data and one line of header, which are the 755 lines in the file.

Sample Code:

Please start your work using the sample code and read the comments in the sample code carefully. This example will generate the sum of MAE of stock AAPL using ARIMA Model. Extend this example and find the top 10 minimum sum of MAE using the three models.

Bonus(1% of your grade):

Utilize MPI in your R code to improve the performance in 2 nodes.

Specific Submission Guidelines:

1. Files should be strictly organized as following structure in your own directory (**/gpfs/classes/cse587/spring2015/students/username/hw2**) and the naming of the directory should be followed exactly (case sensitive):

NOTE THAT YOU SHOULD NOT MAKE ANY CHANGES TO THE DIRECTORY AFTER THE SUBMISSION DEADLINE AS THE TIME STAMP OF THE FILES WILL BE USED FOR TIMELY SUBMISSION.

2. Your code will be evaluated using automated script in following fashion.

```
[username@rush:~] sbatch SLURM-R
```

3. Please compress:

- R code files
- Three jpg images

into one file(tar or zip) and submit to UBLearn.

Grading Criteria

- Program correctness (for each technique): 25%
- Report: 25%