Homework # 11. AMS 586

Name:_____SBU ID:_____

Dear all, please submit your Homework to the Brightspace, before class on Tuesday, November 28. Please include (1) Rmd code; (2) Output from Rmd; (3) Answers to all the questions asked.

LSTM & RNN Stock Price Predictions

This dataset contains Apple's (AAPL) historical stock prices. You can download the same dataset by yourself from finance.yahoo.com for free.
We will use a Long Short-Term Memory (LSTM) neural network to predict the daily close price of **AAPL** in this assignment. The data spans from 2021-03-29 to 2022-03-28.

When applying LSTM, it is very important to **normalize** the data. One widely used method is **min-max scaler**, which means we transform x to be **(x - min(x)) / (max(x) - min(x))**. However, we will tweak this method a little bit. Instead of using the min and max of that day, we use the min (low) and max (high) of the **previous** day (lagged min & max). This makes sense because we can't possibly know the min and max of today before the market closes. Therefore, we can't predict today's price if we use today's statistics.

Please review the following websites for related methods and concepts:
1. **keras**:
https://tensorflow.rstudio.com/installation/
2. **RNN and LSTM**:
https://tensorflow.rstudio.com/tutorials/beginners/basic-ml/
https://www.kaggle.com/code/rtatman/beginner-s-intro-to-rnn-s-in-r/notebook
http://datasideoflife.com/?p=1171
https://cran.r-project.org/web/packages/keras/vignettes/sequential_model.html
3. **Other Materials**:
https://www.deeplearningbook.org/
https://www.deeplearningbook.org/contents/rnn.html

1. Plot the close price vs. date to visualize the data we will analyze. Then, use the 'min-max scaler' to normalize our stock price data. Scaled_x = (x-lagged_min(x))/(lagged_max(x) – lagged_min(x)). Please report the values of the last 10 scaled close prices.

2. Divide the cleaned dataset into two parts, the last 10 prices for testing y and the rest for training (the last 10 prices in the training set will be used as testing x). Please report how many days of stock price are divided into the training set.

3. LSTM algorithm creates predictions based on the lagged values, which means we need to look back as many previous values as many points we wish to predict. Here we want to do a 10-days

forecast, so we need to base each prediction on 10 data points. (We lag the data 10 times, so that each prediction is based on 10 values, and arrange lagged values into columns) Additionally, keras LSTM expects specific tensor format of shape of a 3D array of the form [samples, timesteps, features] for predictors (X) and for response (y) values. Please create matrices for training and testing predictors and response in the 3D form, and report their dimensions by using dim().

4. Please first build the predictive model to predict 10-days stock close price using the training data and the LSTM method with only one LSTM layer with 200 hidden units, and the loss function of 'mse'. Please make predictions on the 10 observations in the testing set by using the last 10 in the training dataset and compute the Test MSE using the testing data. Scale the predicted stock price back and plot the 10-days predictions and the true stock close price in the same figure. Also, try to predict the close price of 2022-3-29.

5. Please first build the predictive model to predict 10-days stock close price using the training data and the RNN method with only one RNN layer with 200 hidden units, and the loss function of 'mse'. Please make predictions on the 10 observations in the testing set by using the last 10 in the training dataset and compute the Test MSE using the testing data. Scale the predicted stock price back and plot the 10-days predictions and the true stock close price in the same figure. Also, try to predict the close price of 2022-3-29.

6. Please compare the performance of the LSTM an RNN algorithms in Question 4 and Question 5 in 10-days forecast using the testing data.