

2019 Fall EE5183 FinTech - Homework 4  
Deep learning Model: Recurrent Neural Network  
Due: Jan 4, 2020

## INSTRUCTIONS

1. In this homework, Dataset is S&P 500 daily historical data comes from [yahoo finance](#) is utilized to build regression models. The features are Date, Open, High, Low, Close, Adj Close, Volume which are common attributes for investors. We want to predict next day 'Close' value depend on historical data. RNN is usually to deal with time series data, because it can capture relation between sequence.
2. Please only use TensorFlow or Keras to build the model.
3. Name your source code that contains your *main* function as *hw4\_StudentID.py* and your report as *hw4\_StudentID.pdf*.
4. Please install the following 2 packages for this homework, *mpl\_finance* and *ta-lib*. Sometimes *ta-lib* may occur error if using pip install. You can go to this [website](#) download whl file then install as below.  
Ex: `pip3 install TA_Lib0.4.17cp36cp36mwin_amd64.whl`.
5. **You should write your own codes independently. Plagiarism is strictly prohibited.**

## PROBLEMS

### 1. Regression:

In this exercise, you will implement a RNN model for regression using *S&P.csv*. The objective in this exercise is to create and train a neural network to predict the next day 'Close' value. You need to split the data from 1994-2016 as training part and data in 2017-2018 as validation part.

- (i) (20%) Please use *S&P.csv* to plot
- a.) Candlestick chart with 2 moving average line(10days and 30days).
  - b.) KD line chart.
  - c.) Volume bar chart.

Show your image from 2018/1/1 to 2018/12/31.

(Hint: You can use *mpl\_finance* and *ta-lib* package to help you plotting these stock charts.)

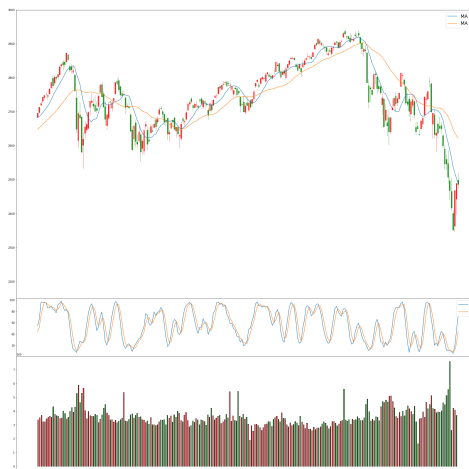


Figure 1: Example of Candlestick chart with moving average lines, KD line chart, volume bar chart.

- (ii) (15%) Please at least add 4 features from question (i) into your input which are 'Moving Average 10 days', 'Moving Average 30 days' and 'K,D from KD line chart'. And we want all features except Date to normalize on a scale of 0 to 1 by the below equation. You can also add other features you think it may help your model get better performance(e.g. If you think weekdays are important to stock price, you can add an one-hot attribute of weekday. Please talk what did you do for data preprocess?

$$z_i = \frac{x_i - \min(x)}{\max(x) - \min(x)}$$

- (iii) (10%) In RNN model, data dimension can be confused. Like the below figure, RNN has 3 dimension which can be written as  $(batch\_size, time\_step, input\_dimension)$ . In this exercise, batch size is depend on you design. As for time step, because we want to use the last 30 days to predict next day 'Close' value,  $time\_step$  equal to 30. And input\_dimension will be depend on your (ii) design. Please split data from 1994-2016 as train part and the rest as validation part.

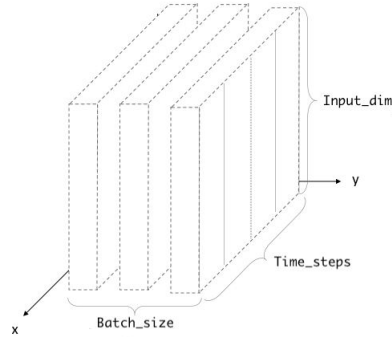


Figure 2: RNN input dimension.

- (iv) (15%) Please construct a RNN model with *SimpleRNN* cell for predicting next day 'Close' value according to Mean square error

$$MSE = \frac{1}{n} \sum (y_i - \tilde{y}_i)^2$$

Please explain how do you design your model?

- (v) (10%) Plot loss curve chart and actual and predict 'Close' value of validation part.

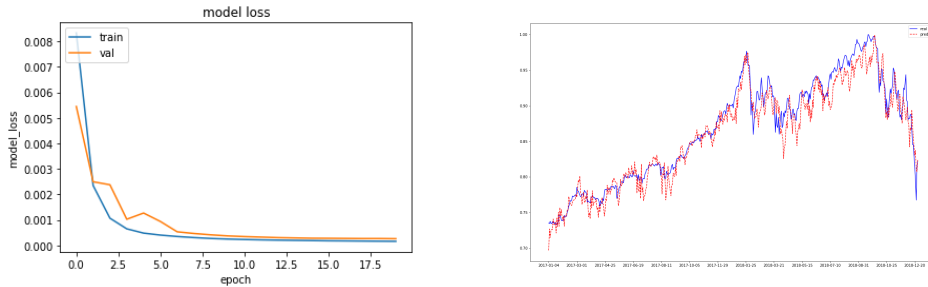


Figure 3: Example of loss curve and predict curve.

- (vi) (10%) Substitute *LSTM cell* for *SimpleRNN* and repeat (iv), (v).  
(vii) (10%) Substitute *GRU cell* for *SimpleRNN* and repeat (iv), (v).  
(viii) (10%) Discuss what do you see from (iv) to (vii)?