STA3140-01 Deep Learning
Homework 6
DUE Monday, June 3

Copying homework solutions from others lead to 0 score. No late submission is allowed. Your solution should contain both code and corresponding explanations for the answer. Submit your HW (pdf form of report, code file) through LearnUs.

1. In this exercise, we will build a LSTM model to estimate the daily confirmed cases of COVID-19. You should solve the problem based on `KoreaCovid.csv` file that I uploaded at LearnUs.

   (a) The `KoreaCovid.csv` data set includes the number of confirmed, recovered and deaths from January 22, 2020 to October 17, 2020. Draw the time series plot of the confirmed, recovered and deaths and explain the pattern of the plot. Divide the dataset into training (first 200 days) and test (remaining 70 days) sets (20 points).

   (b) Using training data, construct an LSTM model to estimate the confirmed cases. Let the model uses the previous 30 days' confirmed cases. Then compare the predicted and the true confirmed cases in the test data (20 points).

   (c) Using training data, construct an LSTM model to estimate the confirmed cases. Let the model uses the previous 30 days' confirmed and recovered cases. Then compare the predicted and the true confirmed cases in the test data (20 points).

   (d) Using training data, construct an LSTM model to estimate the confirmed cases. Let the model uses the previous 30 days' confirmed, recovered and death cases. Then compare the predicted and the true confirmed cases in the test data (20 points).

   (e) Compare the prediction accuracy between LSTM models in (b), (c), (d). Which one is the best? Explain the reason why (20 points).