CSC 215-01 Artificial Intelligence (Fall 2018)

Mini-Project 4: Stock Price Prediction using Neural Network, LSTM, and CNN

Due at 4:00 pm, Monday, November 5, 2018

Demo Session: class time, Monday, November 5, 2018

1. Problem Formulation

This project is threefold:

- Task 1: Use the daily [Open, High, Low, Volume] to predict [Close] on that day using **a fully-connected neural network**. Use the first 70% of the records for training and the remaining 30% of the records for test. Report the RMSE of the model. Show the "regression lift chart" of your test data.

- Task 2: Predict [Close] of a day based on the last 7 days' data [Open, High, Low, Volume, Close] using a **LSTM model**. In other words, we want to predict the price in the green cell using all the numbers in the red cell. Use the first 70% of the available records for training and the remaining 30% of the available records for test. Report the RMSE of the model. Show the "regression lift chart" of your test data.

  Hint: Each record in X can be viewed as a sequence of 7 vectors, each vector with 5 dimensions.

- Task 3: Do the same as Task 2 but use a **CNN model**. Report the RMSE of the model. Show the "regression lift chart" of your test data.

  Hint: The red cell can be considered as an image that has 7 pixels, each pixel with 5 channels.

	A	B	C	D	E
1	# http://finance.yahoo.com/quote/GOOG/history?ltr=1				
2	# Open	High	Low	Volume	Close
3	828.66	833.45	828.35	1247700	831.66
4	823.02	828.07	821.655	1597800	828.07
5	819.93	824.4	818.98	1281700	824.16
6	819.36	823	818.47	1304000	818.98
7	819	823	816	1053600	820.45
8	816	820.959	815.49	1198100	819.24
9	811.7	815.25	809.78	1129100	813.67
0	809.51	810.66	804.54	989700	809.56
1	807	811.84	803.19	1155300	808.38
2	803.99	810.5	801.78	1235200	806.97

2. Dataset

https://drive.google.com/open?id=1epxIFUlSWj9p11frJPqHq2hWY03EUQxh

The data has seven columns as follows:

Date, Open, High, Low, Close, Adj_Close, Volume

Remove date and adj_close columns first since you don't need them.

3. Additional Requirements

- You are required to split data for training and testing. Use training data to train your models and evaluate the model quality using test data
- Do feature normalization.
- You must use EarlyStopping and ModelCheckpoint when training neural networks, LSTM, and CNN using Tensorflow.

- Add dropout layer(s) to see how it changes RMSE.
- Tuning the following hyperparameters to see how they affect performance
    - **Activation:** relu, sigmoid, tanh
    - **Layers and neuron counts**
    - **Optimizer:** adam, sgd, rmsprop, and others
    - **Kernel number and kernel size** (for CNN only)
    - **LSTM neuron count** (for LSTM only)

4. Grading breakdown

You may feel this project is described with some certain degree of vagueness, which is left on purpose. In other words, **creativity is strongly encouraged**. Your grade for this project will be based on the soundness of your design, the novelty of your work, and the effort you put into the project.

Use the evaluation form on Canvas as a checklist to make sure your work meet all the requirements.

Implementation	**70 pts**
Your report	**15 pts**
In-class defense	**10 pts**
Additional features (novelty)	**5 pts**

5. Teaming:

Students must work in teams of 2 people. Think clearly about who will do what on the project. Normally people in the same group will receive the same grade. However, the instructor reserve the right to assign different grades to team members depending on their contributions. So you should choose partner carefully!

6. Deliverables:

(1) All your source code in Python Jupyter notebook.

(2) Your report in PDF format, with your name, your id, course title, assignment id, and due date on the first page. As for length, I would expect a report with more than one page. Your report should include the following sections (but not limited to):

- **Problem Statement**
- **Methodology**
- **Experimental Results and Analysis**
- **Task Division and Project Reflection**
- **Additional Features.**

In the section "**Task Division and Project Reflection**", describe the following:

- who is responsible for which part,
- challenges your group encountered and how you solved them
- and what you have learned from the project as a team.

10 pts will be deducted for missing the section of task division and project reflection.

All the files must be submitted **by team leader** on Canvas before

4:00 pm, Monday, November 5, 2018

NO late submissions will be accepted.

7. In-class defense:

Each team member must defend your work during the scheduled defense session. Each team have **five minutes** to defend your work in class. In the defense, briefly describe **the basic steps** you took to finish this project by referring audience to the data/code/figures in your Jupyter notebook.

If you implement **additional features (novelty)**, please do mention them to receive credit for novelty.

Failure to show up in defense session will result in **zero** point for the project.

8. Additional Features

Some possible ways:

- In the project, you predict [Close] of a day based on the last 7 days' data. Can you find the best N value (number of the days we should consider in the past) that yields the best model?
- Go to Yahoo! Finance. Download the stock price historical data for some companies. For example:

https://finance.yahoo.com/quote/GOOG/history?p=GOOG&.tsrc=fin-srch

Based on your observations from the project, build a good model for stock price prediction.

Show RMSE and regression lift chart.

- Can you use LSTM or other RNN models to predict the stock prices for a particular company for a continuous time period (e.g., the prices in the next five days)? Show the true prices and predicted prices in the same chart.