

CSC 215-01 Artificial Intelligence (Fall 2020)

Mini-Project 2: Time Series Forecasting using NN, LSTM and CNN

Due at 4:00 pm, Wednesday, October 14, 2020

Peer Review: class time, Wednesday, October 14, 2020

1. Problem Formulation

Time series forecasting is an important area of AI. It is important because there are so many prediction problems that involve a time component. This time component makes time series problems more difficult to handle. In this project, you practice with time series data to predict stock price.

This project is threefold:

- Task 1: Predict [Close] of a day based on the last 7 days' data [Open, High, Low, Volume, Close] using a **full-connected neural network 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 has $7 * 5 = 35$ input features and 1 output feature.

- Task 2: Do the same as Task 1 but use a **LSTM model**. Report the RMSE of the model. Show the “regression lift chart” of your test data.

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

- Task 3: Do the same as Task 1 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 viewed 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=1Bp4_UvWcfZLK2fwiLCpyDcUO9PL9i-Z5

Find the file CSC215_P2_Stock_Price.csv. This file has seven columns as follows:

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

Remove date and adj_close columns first since we do not need them.

3. Requirements

- Split data to training and testing. Use training data to train your models and evaluate the model quality using test data

- Do feature normalization. **Notice that you should never normalize the output feature when training any regression models.** Otherwise, the RMSE of the regression model will be also normalized.

Hint: Create two separate copies of the “close” column and use one as output feature while the other as part of the input

- You must use EarlyStopping and ModelCheckpoint when training neural networks, LSTM, and CNN using Tensorflow.
- Tuning the following hyperparameters to see how they affect performance
 - **Activation:** relu, sigmoid, tanh
 - **Number of layers and neuron count in each layer**
 - **Optimizer:** adam and sgd
 - **Kernel number and kernel size** (for CNN only)
 - **Number of LSTM layers and neuron count in each layer** (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.

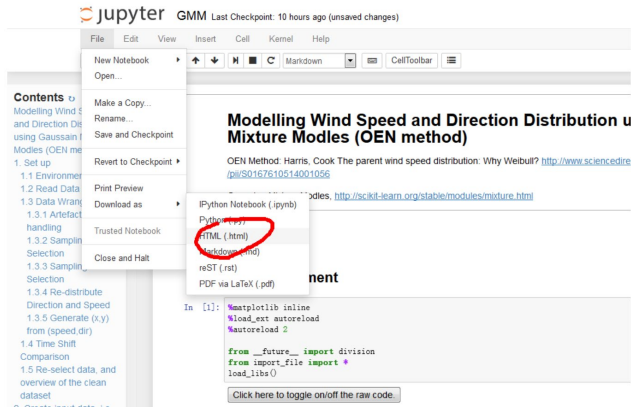
Use [the evaluation form on Canvas](#) as a checklist to make sure your work meet all the requirements.

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) The **HTML version of your notebook** that includes all your source code.** Go to “File” and then “Download as”. Click “HTML” to convert the notebook to HTML.



5 pts will be deducted for the incorrect file format.

- (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.

In the section “Additional Features”, you describe and claim credit for additional features.

To submit your notebook and the report, go to Canvas “Assignments” and use “Project X (submit your code and report here)”. Use the evaluation form on Canvas as a checklist to make sure your work meet all the requirements.

- (3) **Link to your video presentation shared to the discussion board.** Each team has **five minutes** to demo your work. Failure to submit the video presentation will result in **zero** point for the project. The following is how you should allocate your time:

- Model design (1 minute)
- Findings/results (1 minute)
- Task division (1 minute)
- Challenges encountered and what you have learned from the project (1 minutes)
- Additional features (1 minute)

To submit the link to your video presentation, go to Canvas “Discussions” and use “Post Your Presentation for Project X Here”. Share your link by replying directly to my main discussion post.

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

4:00 pm, Wednesday, October 14, 2020

NO late submissions will be accepted.

7. Possible Additional Features (5 pts per feature, 10 pts at most)

- 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 most accurate model?
- Can you use LSTM to predict the stock prices for a continuous future time period (e.g., the prices in the next five days)? Show the true prices and predicted prices in the same chart.

Hint: train a multi-output regression model.

- Can you use Keras layer wrappers to create an even more complicated layer? For example a bidirectional LSTM?

https://keras.io/api/layers/recurrent_layers/

- In the shared Google drive folder, you can find the stock price for the following companies.
 - Royal Dutch Shell
 - Apple
 - Google
 - JPMorgan

Based on your observations from the project, build a good model for stock price prediction. Show RMSE and regression lift chart.

- Go to Yahoo! Finance. Then you can find the stock price historical data for more companies.

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

8. Peer Review

During the class on the due day, please review and comment on the presentations from other teams by replying to their posts. It is a great chance for you to learn from other people's work. Please be nice, and provide constructive, specific feedbacks. You will become a better, more effective learner when you found yourself in a community of active learners!