

2022-2023 # Group 04	AI & Deep Learning	Chenghao Lyu Daniel Stulberg Huf Kerryan Macé
-------------------------	--------------------	---

Using long short-term memory (LSTM) recurrent neural network to predict stock prices

Resources

- YouTube video introducing the topic of stock pricing using LSTM:

<https://youtu.be/QIUxPv5PJ0Y>

- Academic paper about predicting stock market index using LSTM:

<https://www.sciencedirect.com/science/article/pii/S2666827022000378>

- Some general explanations about the LSTM RNN topic can be found in these links:

<https://towardsdatascience.com/lstm-networks-a-detailed-explanation-8fae6aefc7f9>

<https://colah.github.io/posts/2015-08-Understanding-LSTMs/>

Goals

- Getting to know the basic functioning of the LSTM technique.
- Getting know how data is prepared and structured to be used in a LSTM network.
- Discovering a little more about how does stock market prediction work at a basic level (yet an important tool for quantitative trading).
- Introducing to students the tools:
 - **Pandas**: Python library for data analysis, especially useful for handling big datasets and any sort of time series.
 - **TensorFlow**: Open-source machine learning platform created by Google which can be used in Python as an alternative to PyTorch. Although the training time of TensorFlow is generally higher compared to PyTorch, its memory usage is lower and it is more appropriate to use when custom features are needed in the neural network.

Main ideas

We will try to explore the use of a long short-term memory network, which is a type of recurrent neural network used to learn order dependence in sequence prediction problems, therefore being a powerful tool for predicting stock prices because of its dependency of previous prices.

We will focus on stocks from some of the biggest companies of the S&P 500 index such as Apple, Meta and Amazon, and we will try to predict their evolution based on the last 5 years of operations.

Finally, we intend to explain the results achieved at the end of the model validation, and also discuss about the limitations of the methods used (Were the results correctly back tested? What is the accuracy of the model for this week's results of the stock market?) and its potential uses (Can we anticipate potential economic crisis, booms or bubbles using this model?).

Applicative context

RNNs, especially LSTMs, can be very useful when working with sequences of words and paragraphs in the field of Natural Language Processing (NLP). Other applications of LSTMs include, but are not limited to:

- Robot control
- Speech recognition
- Music composition
- Medical care predictions
- **Time series predictions**

It is this last case that we are mainly going to cover in our applied course. Nowadays, RNNs and LSTMs are being tested on time series forecasting problems, although the results are still not satisfactory. During our course, we will try to investigate the reasons why the results regarding time series prediction of stock prices are still reasonably poor, in a way that we can only capture upward and downward trends, but not solid variations of future prices.

Tools

In order to follow our applied course, the respective tools are required:

- **yFinance**
- **Numpy**
- **Matplotlib**
- **Pandas**
- **Scikit-Learn**
- **Tensorflow**

Data

In this notebook, data will be gathered from *yFinance*, which is an open-source Python library used to download market data from Yahoo! Finance's API without any cost. Data will be obtained in the form of a time series containing the open price, highest price, lowest price, close price, adjusted close price and volume of the daily stocks over the last 5 years.

Organization

Task 1

- Description: Data acquisition, preprocessing and preparation
- Outcome: Data should be gathered and properly adjusted to prepare the inputs of the LSTM network.
- Deadline: 09/01/2023

Task 2

- Description: Model development
- Outcome: The parameters for the LSTM network and the best strategies for its operation should be well defined until the deadline.
- Deadline: 13/01/2023

Task 3

- Description: Visualization tools
- Outcome: Finish the functions to plot the time series and the results of the training and validation sets in an appropriate way (easy to understand and to pick the message).
- Deadline: 16/01/2023

Task 4

- Description: MCQ and do-it-yourself coding of the notebook
- Outcome: Finish MCQ questions and let some cells of the notebook unfinished for the students to complete them.
- Deadline: 18/01/2023