Title: Long Short-Term Memory: A Recurrent Neural Network to Determine the Direction of Commodity Futures Prices

Team: Tim Decilveo

Summary:
Commodity futures are agreements to buy or sell a predetermined amount of a commodity at a specific price on a particular date in the future. These products are traded on various exchanges worldwide and are used to hedge or protect an investment position or speculate on how a price may move over time. Futures contracts are standardized for quality and quantity of a commodity to facilitate ease of trading on an exchange.

The Commitment of Traders (COT) Report is published by the Commodity Futures Trading Commission (CFTC), helping the public understand the market dynamics in the commodity futures markets. The COT reports are based on position data supplied by reporting firms (futures commission merchants - - FCMs - -, clearing members, foreign brokers, and exchanges). The reports break down the open interest of futures and options on futures markets in which 20 or more traders hold positions equal to or above the reporting levels established by the CFTC. The reports provide a positioning of the market participants to determine how each group of investors views the markets.

Historically, investors in financial markets have used past data to decide whether to buy or sell a specific security, and recurrent neural networks offer market participants a way to apply this thinking systematically. Specifically, Long Short-Term Memory Recurrent Neural Networks (LSTM RNN) allow investors to utilize feedback connections to process data sequences, or time series data, which is the standard data set used in the financial markets. LSTMs can be valuable models for financial markets as there may be lags of unknown time between significant data points within a timer series. LSTMs can process data and maintain information about prior data points without knowing if the specific data point is relevant a priori.

This paper will provide a novel way to implement an LSTM model to predict the direction of commodity market prices using historical price action and COT reports data[1]. I will implement experiments along different historical time windows, determine the overfitting/underfitting of data, measure the mean-squared/root-mean-squared error, and determine classification.

Resources:
Code: I will be implementing my code with Python, using packages from PyTorch, scikit-learn, Numpy, Pandas, Matplotlib, Seaborn, Quandl, and others. I may also use AWS Forecast to develop a baseline model.

Data: the data for this paper will come from two sources. Nasdaq Data Link's API will be used to retrieve the COT report data. The commodity futures price data has been acquired from Commodity Systems Inc. The futures prices have been back adjusted to provide a continuous time series that eliminates the gaps between expiring and newly active contracts. The idea of back adjusted contracts involves concatenating

[1] As I am still learning about LSTM models, I am unsure of the exact metrics I will use to evaluate my results. The standard measurements of loss/cost and others will be used, but I may also implement a classification algorithm to help determine if the predictions are up/down (0 or 1) as I don't anticipate being able to predict exact prices implementing this methodology for the first time.

historical contracts of a given commodity into the past while adjusting to smooth transitions between delivery months. The commodities that I will evaluate are the Australian Dollar, Soybean Oil, British Pound, Corn, Cocoa, Canadian Dollar, WTI-Crude Oil, Cotton, Euro, U.S. Dollar Index, Euro Dollar, Feeder Cattle, 5-Year T-Note, Gold, HG Copper, Heating Oil, Japanese Yen, Coffee, Lumber, Live Cattle, Lean Hogs, Mexican Peso, Nasdaq 100, New Zealand Dollar, Natural Gas, Oats, Orange Juice, Palladium, Platinum, Rough Rice, Soybeans, Sugar #11, Swiss Franc, Silver, Soybean Meal, S&P 500, 2-Year T-Note, 10-Year T-Bonds, 30-Year T-Bonds, Wheat. As each security was introduced to the exchanges individually, the date each contract started can vary. The earliest start date in the time series data is January 1, 1946.

Research Resources & Papers:

1. https://arxiv.org/pdf/2106.12747.pdf
2. https://www.ai.uga.edu/sites/default/files/inline-files/ernest_foster.pdf
3. https://jfds.pm-research.com/content/early/2021/06/10/jfds.2021.1.065
4. https://en.m.wikipedia.org/wiki/J%C3%BCrgen_Schmidhuber
5. https://en.wikipedia.org/wiki/Long_short-term_memory
6. https://people.idsia.ch//~juergen/rnn.html
7. http://www.felixgers.de/papers/phd.pdf
8. https://www.jmlr.org/papers/volume3/gers02a/gers02a.pdf
9. http://www.overcomplete.net/papers/nn2012.pdf
10. http://etd.uwc.ac.za/xmlui/handle/11394/249
11. https://towardsdatascience.com/lstm-networks-a-detailed-explanation-8fae6aefc7f9
12. http://christianherta.de/lehre/dataScience/machineLearning/neuralNetworks/LSTM.php
13. https://www.frontiersin.org/articles/10.3389/fnins.2020.00584/full
14. https://www.bioinf.jku.at/publications/older/2604.pdf
15. https://www.ai.uga.edu/sites/default/files/inline-files/ernest_foster.pdf
16. https://machinelearningmastery.com/gentle-introduction-long-short-term-memory-networks-experts/