Steven Granaturov

CS-GY 6923 – Machine Learning

October 19, 2023

Project Proposal

1. What problem do you aim to tackle with? Why is it an important problem to you?

I intend to create a stock market trend predictor. It isn’t really an important problem for me, but it is a problem that I would like to explore.

1. What dataset do you plan to use? Are you going to collect data yourself? If not, is the dataset already preprocessed?

I plan on using the [Stock Market Data (NASDAQ, NYSE, S&P500) available on Kaggle](https://www.kaggle.com/datasets/paultimothymooney/stock-market-data/). Because this dataset is massive, I will limit the amount of stocks / indexes to a small amount or else the model will take a very long time to run.

1. Which ML models do you plan to use?

I plan to use a recurrent neural network by using the TensorFlow library. This allows for a layer of LSTM (Long short term memory) which I feel will be very useful for the creation of this model. I would also like to try out the linear regression model to see how it would perform.

1. Which accuracy/error measures do you plan to use?

Although I haven’t figured out a way to get the training accuracy, the testing accuracy can be measured by limiting the stock prices to a certain year, (say 2020 for example) and then plotting the predicted cost against the actual cost of the stock.

1. Are there previously documented results on the same problem? If yes, how do you plan to improve previous results?

There are previous documentations on this project and I attempt to improve previous results by playing around with the neural network layers. I am also interested to see how a linear regression model will perform.