Machine Learning Engineer Nanodegree

Capstone Project Proposal

"Machine Learning Approach to

Stock Price Prediction using LSTM"

Omar Ali

November 28,2018

Domain Background

Stock price prediction is a popular topic throughout the world. The stock market has always been a goldmine of adventures for mathematicians and statisticians. They keep trying to find patterns such that the behavior of stock market and how it can be predicted, however the huge amount of data and buy/sell decisions carried out every day makes it almost impossible to be analyzed manually. However, in the recent few years the emergence of new organized data, high computational power and machine learning algorithms has given us the ability to use computer for predicting the behavior of stock market. In this project I would try machine learning to predict the stock prices and their behavior.

Problem Statement

The project aims to create a machine learning model that can predict stock price by using historical information as a timeseries data. The task is to build a stock price predictor that takes daily trading data over a certain date range as input, and outputs projected estimates for given query dates. The inputs will contain multiple metrics, such as opening price (Open), highest price the stock traded at (High), lowest stock price of the day (Low) and the closing stock price of the day.

Datasets and Inputs

In this project, I will use the American Airline Stock prices in the S&P 500 as inputs and target. For each stock, the data will contain the four metrics mentioned in the problem statement. The Closing stock will be the target, and the rest variables will be our inputs. Our data will feature the stock prices for American Airline company for the last 20

years. We will get our data from Kaggle and Alpha vantage by using an api key and then saving the data in a file.

Solution Statement

In this project, I would like to use recurrent neural network to solve the problem. I will use Long-short term memory network as the model, the closing stock price as the target and all the historical time series data of the target stock itself as inputs. I will be using several hyperparameters to optimize the model. The prediction will be one-step-ahead stock price of the target stock.

Benchmark Model

The benchmark model for this project would be a linear regression. Including lagged features of the dependent variable and other exogenous features in a linear regression is called autoregressive model with exogenous inputs, which is proved to be very successful in statistical time series modeling. This benchmark will use exact the same input as our LSTM network model and provide a benchmark performance for the LSTM.

Evaluation Metrics

This is a regression problem, so the metrics of choice would be R-square and root-mean-squared-error. R-square can provide how much variation in the dependent variable can be explained by the variation in the independent variables. Root-mean-squared-error can provide what is the average deviation of the prediction from the true value, and it can be compared with the mean of the true value to see whether the deviation is large or small.

Project Design

After loading the data, I will spilt into training and testing sets. I will use the mid-price calculated by taking the average of the highest and lowest recorded prices on a day, Now I can split the training data and test data. The training data will be the first 11,000 data points of the time series and rest will be test data. and perform data formalizing so they all can be in the region of 0and 1. In normalizing, different time periods of data have different value ranges, we normalize the data by splitting the full series into windows. If

we don't do this, the earlier data will be close to 0 and will not add much value to the learning process. Here I choose a window size of 2500.After that we can begin with our averaging method and our benchmark model (Linear Regression) so we can compare with our LSTM model after that. In the LSTM, the network will have three layers each with hidden states of 200-200-150 We can then measure the performance of our model by using R-square and RMSE.

Acknowledgement

Sreelekshmy Selvin, Vinayakumar R, Gopalakrishnan E.A, Vijay Krishna Menon, Soman K.P "STOCK PRICE PREDICTION USING LSTM,RNNAND CNN-SLIDING WINDOW MODEL", 2017 International Conference on Advances in Computing, Communications and informatics (ICACCI)