Machine Learning Engineer Nanodegree

Capstone Proposal

Simran Kaur

April 16th, 2020

Domain Background

Investment firms, hedge funds and even individuals have been using financial models to better understand market behaviour and make profitable investments and trades. A wealth of information is available in the form of historical stock prices and company performance data, suitable for machine learning algorithms to process.
The uncertainty in the stock market refrain people from investing in stocks. Thus, there is a need to accurately predict the stock market which can be used in a real-life scenario. [1]
DeepAR is an LSTM neural network that can be used to forecast time-series data, accounting for trends and seasonality of the time series in order for the network to learn and give accurate forecasts.[2]
This project will utilize RNN, or LSTM, an RNN architecture, to predict stock prices.

Problem Statement

The aim is to predict the future closing value of a certain stock using historic data.
The solution should include the steps in the order:
1. Obtaining the data from yahoo finance of a particular stock.
2. Data Preprocessing.
3. Using Keras to build RNN with LSTM.
4. Training the Neural Network.
5. Applying an optimizer (Adam or ADAgrad). [3]
6. Making Predictions.

Datasets and Inputs

The dataset for this project will be obtained from Yahoo Finance. The stock price data will be acquired using pandas-datareader API.

For a stock, there are 4 value particular features namely, **Open, High, Low, and Close.** In addition to those, there is also a **Volume** column which contains the volume of stock and **Date**. This dataset into 60% train, 20% validation, and 20% test.
Automated feature extraction can be done using restricted Boltzmann machines.

Solution Statement

I will use the Keras library along with Tensorflow to build the framework. After training the network, choosing the type of optimizer will be important. Using the Dropouts methods for preventing overfitting is also to be added in the framework. To evaluate the effectiveness of the method, I will use the root mean square error (RMSE) and mean absolute percentage error (MAPE) metrics.[4]

Benchmark Model

The Benchmark model will utilize Linear Regression. The predicted values from the benchmark model and the actual price of the stock will determine the performance of the model.

Evaluation Metrics

To evaluate the effectiveness of our methods root mean square error (RMSE) and mean absolute percentage error (MAPE) metrics between the actual price and the predicted price will be used.

Project Design

Data Preprocessing will be an important part of this project. 1. Data Discretization, 2. Normalizing - Data transformation, Data Cleaning will play a major role in determining the predicted output.
In the next phase, Feature Extraction, the chosen features will be fed to the network in the next step for training.
To build my model I will use Keras library with Tensorflow to build an LSTM.

References

1. Stock Market Prediction Using Machine Learning Algorithms
2. Introducing Amazon Forecast and a Look into the Future of Time Series Prediction
3. An overview of gradient descent optimization algorithms
4. Machine Learning Techniques applied to Stock Price Prediction