Machine Learning Engineer Nanodegree

Capstone Proposal:

Chaitanya Potnis

Stock Price Trend Prediction Using LSTM.

December 8 2018

Proposal:

Domain Background:

Predicting stock prices is always being research topic in the market and there are not even more than 100 research paper on stock predictions. As we all know Stock prices of a company depends on various features. If we are able get some features and develop a algorithm or model to predict stock prices, it won't be accurate as the real stock prices. Therefore with some limited feature we can predict the stock price trend (is it an upwards trend or an downwards trend). As we know stocks price are time series data and there are very few algorithms or model that handle the time series data well, therefore I selected a state of the art model that is an LSTM(Long Short Term Memory) which is an Recurrent Neural Network and it performs well on time series data. Presently with all the computation power and data we can build an powerful LSTM model to predict the stock price trend.

Stocks of a company are fuel that runs the whole organization. As we know that stocks are highly volatile in nature and first rule of finance is based on the Brownian motion which states that stock prices have random behavior and are continuously changing over the time and are driven by Brownian motion process. There are many factors affecting the stock of a particular company, like Company A's stock price depend on their rival Company B. If Company B launches a Product in market and it has a huge success so, it will affect the stock prices of their rival Company A and visa versa. In this scenario it's just one company but in reality a company's stock may be affected by it's own decision and products as well as the products and decision of their subsidiary, dependent companies and rivals. Therefore it makes prediction of future stock prices more difficult.
As most of the research work in stock price prediction is done using financial concepts like Monte Carlo technique and by using technical indicators like Rate of change, momentum, relative index strength, average direction index etc. But in this project I am using LSTM model which is a recurrent neural network to predict the future stock prices trend.

Problem Statement :

As discussed earlier there are various ways to find future stock prices by using financial concepts like Monte Carlo technique and by using technical indicators like Rate of change, momentum, relative index strength, average direction index etc. But in this project I will be using the power of Deep learning to find the future stock prices. First we will train the LSTM model with prices (consider Open price) which is a time series data. And try to predict the next day price. As mentioned earlier it not possible to predict the accurate price, we will convert the problem from regression to classification, to predict the trend that is whether the stock price will increase (upward trend) or it will decrease (downward trend) which makes this problem a binary classification problem. This information itself can help many organizations and individuals in decision making.

Dataset and Inputs:

In this project is use Google Stock Data to predict the trend in the company's stock prices. The stock data can be retrieved from pandas datareader package source being yahoo finance.
Features in data are : Open, High, Low, Close, Adj Close, Vloume.

	Open	High	Low	Close	Adj Close	Volume
Date						
2017-01-03	778.809998	789.630005	775.799988	786.140015	786.140015	1657300
2017-01-04	788.359985	791.340027	783.159973	786.900024	786.900024	1073000
2017-01-05	786.080017	794.479980	785.020020	794.020020	794.020020	1335200
2017-01-06	795.260010	807.900024	792.203979	806.150024	806.150024	1640200
2017-01-09	806.400024	809.966003	802.830017	806.650024	806.650024	1272400

First 5 Rows of Google Dataset.

Solution Statement:

Data analysis is a key step in problem solving and deep learning. First, we need to perform Exploratory dataset analysis to find patterns ,correlations, calculating Moving averages, Daily Returns in data. Next step is problem solving here the problem is to predict the future stock prices trend.

As the stock data is random and volatile, it will be difficult to predict the "Exact stock Price " but by using LSTM and feeding it previous 60 time steps, the predicted value obtained can much closer to the real stock price and can capture the maximum trend
LSTM is a version of Recurrent Neural Network which doesn't face Vanishing Gradient Problem and can handle the long time series data very well as compared to Traditional RNN.
LSTM also can learn patterns in stock prices which will be help for Stock Trend Prediction.
Therefore this will be a perfect model for the our Time Series Data(Stock Prices).

Benchmark Model:

As this a binary classification problem that is I am trying to predict whether next day stock prices goes up(upward trend) or it goes down(downward trend). We can Randomly predict the stock price trend for a particular day as it will have 50-50 % probability.

Evaluation Metrics:

Using the Actual Prices we can find the Actual Trend and in similar way the predict trend can be calculated by Predicted prices.

Algorithm used for calculating the trend:

```
if (Next Day Open Price > Previous Day open Price):
        Trend =  +1
else:
        Trend  = -1
```

+1 Denotes Upwards Trend
-1 Denotes Downwards Trend

Then the we can calculate the accuracy from predicted trend and actual trend.

Project Design:

- Programming Language : Python
- Libraries : Numpy, Pandas, Matplotlib, Seaborn, Keras
- Workflow
    - Data Gathering:
        Using Pandas Datareader and yahoo finance data will be gathered.

    - Exploratory Data Analysis:
        EDA gives better understanding of the data and problem.
        Plot the stocks price
        Calculating moving average and daily return and plotting them

    - Data Preprocessing:
        There is lot of pre-processing required for the data and we need previous 60 days stock data(that is 3 financial months) to predict the next day stock price. Pre-processing includes changing the dimensions, creating special data-frames which include **T-60** (previous 60 time steps) features for **T'**th time step and this row will be used to predict the **T+1** Time step data

    - Benchmark Model:
        A simple Naïve baseline model which will help to evaluate the final LSTM model. Baseline model will be a random choice of upwards trend or a downwards trend (with Probability of 50 -50%)

    - Designing and Developing the LSTM model:
        Designing the LSTM model with Kears (Tensorflow as Backend)

    - Training the LSTM Model:
        Train the model on a n number of epochs where n is number of epoch eher model doesn't overfit or underfit.

- Testing the LSTM model:
  Test the model a test set which is kept aside during training

- Evaluating the Model:
  Calculate the accuracy from the Actual Trend and Predicted Trend