Machine Learning Engineer Nanodegree

Capstone Proposal

Mostafa ElGhandour

March, 31st, 2019

Proposal

Domain Background

Predicting how the stock market will perform is one of the most difficult things to do. There are so many factors involved in the prediction – physical factors vs. physiological, rational and irrational behavior, etc. All these aspects combine to make share prices volatile and very difficult to predict with a high degree of accuracy, which makes it a very project to implement.

https://www.analyticsvidhya.com/blog/2018/10/predicting-stock-price-machine-learningnd-deep-learning-techniques-python/

Problem Statement

Stock market analysis is divided into two parts – Fundamental Analysis and Technical Analysis.

Fundamental Analysis: involves analyzing the company's future profitability on the basis of its current business environment and financial performance.

Technical Analysis: on the other hand, includes reading the charts and using statistical figures to identify the trends in the stock market.

our focus will be on the technical analysis part.

The problem shall be a regression problem and not a classification problem.

The inputs will be some features of a specific item (date, location, etc.) and I shall get a predicted price of the item.

Datasets and Inputs

I will be using data sets from https://www.quandl.com/ to predict gold price in London.

Most of the features I will be using are the date, time and location of the stock items.

I can not specify right now how much of the data I will be using as this will depend on how fast my model trains.

Splitting the data would be a critical point for sure, as I will have the past data as my training data and as I progress in time I will use the future data for evaluation and testing.

Solution Statement

I will be trying a few models (naive Bayes would be a good choice as well) to predict the stock prices and using the data sets from the following website I will be able to have some test and evaluation data that I can evaluate.

Benchmark Model

There are a lot of benchmark models out there with linear regression algorithms that can be used to validate my results relative to the real stock price predicted data.

Evaluation metric

The evaluation metric would be comparing the predicted number to the real stock price number at a certain time, squaring this difference I get the square error and using it as my error function/ evaluation metric.

Project Design

1-i will be pre-processing the dataset to achieve usable raw data, as I'm not sure of what I will be doing to the data but I'm sure of a couple of things: a- I will try to normalize the data as much as possible b- any outliers will be removed to avoid any biasing.

2-i will choose a suitable machine learning model (possibly more than one and ensemble method would be great for this of course).

3-i will train the data using a part of my dataset.

4-i will test the data using another test data of my dataset.

5-i will calculate the accuracy.

6-i will tune my model parameters and calculate the model's accuracy till i achieve the highest accuracy i can achieve.

7-i will try the model on a real stock data if applicable.