Machine Learning Engineer Nanodegree

Capstone Proposal

Lisa Hatch
June 19, 2017

Proposal

Domain Background

Individuals and Organizations invest in stock markets as a means to increase financial wealth. To maximize the return on their investments in the stock markets, the resources are time, money, existing stock shares and information about market performance.

Problem Statement

Machine Learning models when applied to resource components may be used to analyze a stock's data to generate via the model a prediction of a stock's future performance. Prediction of the future stock performance is a regression problem. The regression target will be the stock's adjusted close price for the targeted prediction date.

Datasets and Inputs

Daily stock market data will be retrieved from Quandl's free API for WIKI data (see quandle.com). The dataset includes data for both the New York Stock Exchange (NYSE) and NASDAQ.

The model inputs will be a stock symbol, a past date range of at least a month's duration and 30 days previous for model training data. The return values will be predictions for 7 days, 14 days and 28 days after the training data end date.

Solution Statement

Machine Learning regression models will be explored and analyzed to identify a optimal and efficient model for stock price prediction.

The solution will be a Python program requesting the necessary input parameters, acquiring the appropriate data from the Quandl API, modeling the data, then displaying the price prediction.

Benchmark Model

The benchmark model will be a Linear Regression model applied to a single stock's past data using daily features of opening price, highest price, trading volume, and adjusted closing price to predict stock price in the future.

Evaluation Metrics

The solution model will be evaluated based on it's performance in comparison to the benchmark model. The predicted stock price from the solution model must be closer to the adjusted close price of the stock for the prediction date than the prediction from the benchmark model.

Project Design

Data Acquisition

Review Quandl documentation on getting stock data. Research for examples using Python or REST API for acquiring data. Identify any limitations for frequency for allowed pull of stock data or amount of data pulled regarding date ranges.

Create prototype notebook pulling stock data to verify API documentation or examples.

Data Review

Acquire stock data for example stock symbol and date range. Use pandas or matplotlib graphing to explore the data features.

Model Selection

Identify possible regression models, excluding baseline Linear Regression approach. Create exploratory notebook(s) applying models to acquired stock data. Examine the predicted stock prices and compare to actual adjusted close for the prediction date.