

MATH 189 Big Data: Project Proposal

Haocheng Ye

February 19th, 2019

My project focusing on using past stock price path to predict future ones. Since the market is volatile and to a certain degree random, variables that could cause a stock price to change are myriad. Through figuring out the past statistical pattern in the historical data by machine learning, insights could be drawn to indicate what factors are the major contributors to the stock price's movement. My project will approach this problem in a variety of ways.

1. Using news to predict index movements

This is originally a Kaggle challenge, which provides me with the market data (2007 to present) provided by Intrinio and new data (2007 to present) provided by Thomas Reuters.

Since this is a time series data, then we have to account for the lagging effect. In order to have stationary time series data, we can test the dataset by Dickey-Fuller test, and to use Hilbert-Huang transform. Also we are predicting whether in a given period of time, the price will be higher or lower than before. And there are two approaches to approach this problem: logistic regression, tree regression.

2. To predict future stock price of a single stock price

Data set can be accessed via Yahoo Finance. One of the most popular methods applied in this field is called "Long short-term memory", which is an RNN structure used in the field of deep learning.