Stock Price Trend Forecasting using Supervised Learning

Problem Statement

The aim of the project is to examine a number of different forecasting techniques to predict future stock returns based on past returns and numerical news indicators to construct a portfolio of multiple stocks in order to diversify the risk. We do this by applying supervised learning methods for stock price forecasting by interpreting the seemingly chaotic market data.

Motivation

Nowadays, as the connections between worldwide economies are tightened by globalization, external perturbations to the financial markets are no longer domestic. With evolving capital markets, more and more data is being created daily.

The intrinsic value of a company's stock is the value determined by estimating the expected future cash flows of a stock and discounting them to the present, which is known as the book value. This is distinct from the market value of the stock, that is determined by the company's stock price. This market value of a stock can deviate from the intrinsic value due to reasons unrelated to the company's fundamental operations, such as market sentiment.

The fluctuation of stock market is violent and there are many complicated financial indicators. Only few people with extensive experience and knowledge can understand the meaning of the indicators and use them to make good prediction to get fortune. Most people have to rely solely on luck to earn money from stock trading. However, the advancement in technology, provides an opportunity to gain steady fortune from stock market and also can help experts to find out the most informative indicators to make better prediction. The prediction of the market value is of paramount importance to help in maximizing the profit of stock option purchase while keeping the risk low.

Proposed Solution

1. **Preprocessing and Cleaning**

   Interpolating or recovering the missing data and removing the redundant data. This step also involves creating any useful feature from the existing ones.

2. **Feature Extraction**

   This step involves searching in the space of possible feature subsets. We then pick the subset that is optimal or near-optimal with respect to some objective function. This is done so as to avoid problems of overfitting/underfitting the dataset.

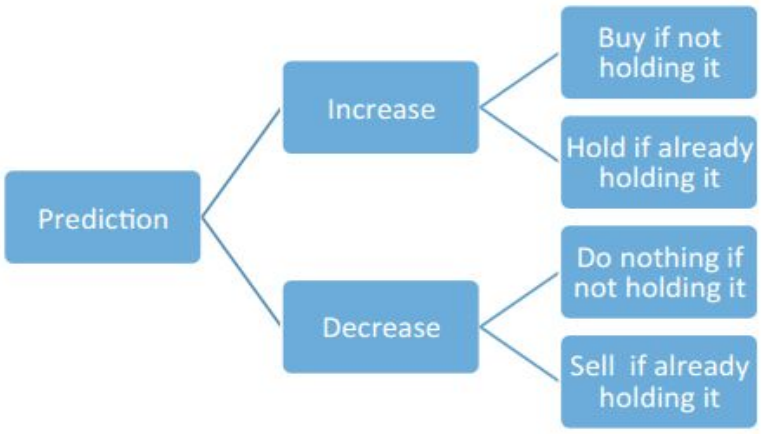
3. **Data Normalization**

   Data is needed to be normalized for better accuracy by ensuring that all features are not given excessive/low weightage.

4. **Analysis of various supervised learning methods**

   a. **Classification Methods**

      This phase would involve supervised classification methods like Support Vector Machines, Neural Networks, Naive Bayes, Ensemble classifiers (like Adaboost, Random Forest Classifiers), etc.



b. **Regression Methods**

These models would be used to get the expected numerical value of the interested stocks.This phase would involve supervised regressions methods like Linear Regressions, Support Vector Regressions, Usage of Kernel Methods, etc.

5. **Social Media Sentiment Analysis**

   Analysing the current market situation from the latest news headlines and social media platform such as Twitter to gain insights into the future of stock prices.

6. **Credit Assignment Problem**

   This step involves the assigning of appropriate weightage to different ways used for data collection.

7. **Analysis of Different Models**

   Comparison between the various methods and models implemented over the datasets.

Work Division

The work division is currently done based on a per algorithmic implementation and testing. This will be further decided as the project progresses.

Datasets

1. Nasdaq
2. QuantQuote minute resolution data files
3. Google Finance
4. Yahoo Finance

Libraries

Numpy, Scipy, Pandas, Matplotlib, Scikit-Learn, Seaborn.

References

1. http://www.nasdaq.com/
2. https://in.finance.yahoo.com
3. https://www.google.com/finance
4. https://www.princeton.edu/~erp/ERParchives/archivepdfs/R7.pdf
5. http://cs229.stanford.edu/proj2013/DaiZhang-MachineLearningInStockPriceTrendForecasting.pdf
6. http://cs229.stanford.edu/proj2012/ShenJiangZhang-StockMarketForecastingusingMachineLearningAlgorithms.pdf
7. http://cs229.stanford.edu/proj2015/009_report.pdf
8. http://quant.caltech.edu/historical-stock-data.html

Team Vortex

Saurabh Jain (201301128)

Sharvil Katariya (201301129)