Stock Data Prediction with Vector Grouping and Markov Modeling

Julianne Lin and Thitaree Tanprasert
April 1, 2019
Math189, Spring 2019, Harvey Mudd College

Motivation

- Explore and develop a method to predict stock trends
- Use this method to extrapolate to economic predictions affected by not just stocks but also other factors
- Apply this method to other time-series data not just stock data

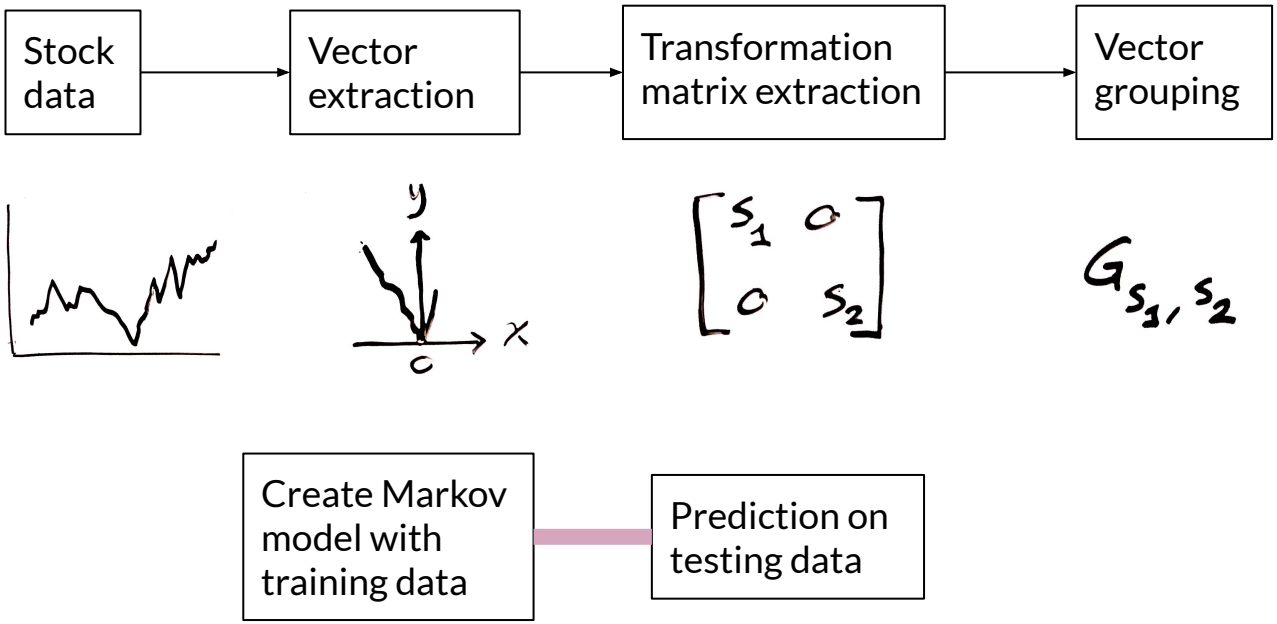
Related Work

- Techniques that other people have used to predict stock data:
  - Unsupervised learning: K-nearest neighbors, Naive Bayes classification, tree-based classification, SVM.
  - Supervised learning: time-delay, recurrent, probabilistic neural nets.
  - Hybrid approach: ridge regression + genetic algorithm

Problem Statement

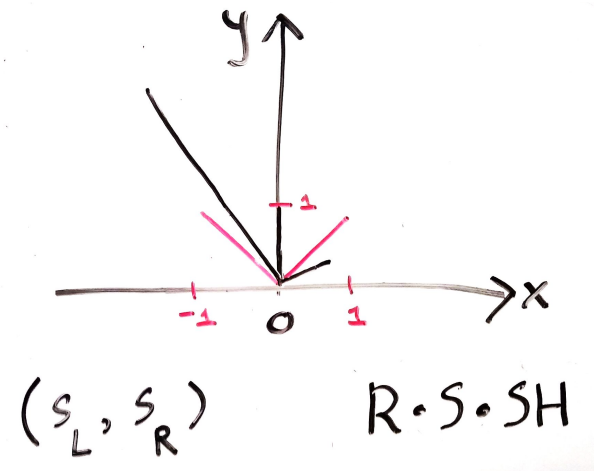
In this project, we will analyze **stock data** to extract patterns to develop a feature extraction scheme using vector transformation. More specifically, we will develop an algorithm to describe each dip and peak in the data in terms of **categories of transformation on a normalized vector.** A stretched goal is to use this technique for stock data prediction and for extracting correlations between the stock data and other economic, social, and political phenomena.

System Pipeline



Vector Grouping Algorithm

- Method 1: Left + right scaling
- Method 2: Rotation + scaling + shearing



Method 1: Left + Right Scaling

- Calculate scaling factor for left and right vectors separately, based on their euclidean distance from the dip.
- Group the scaling factor for left and right vectors separately relative to the mean scaling factor.
    - We get rid of extreme values by taking log scales and limiting the maximum and minimum values.
    - We limit the number of groups for simplicity of computation model and for visualization purpose.
- Group the vectors by looking at the pair of scaling vectors (left, right).

Method 2: Rotate + Scale + Shear

Let **H** = [-1, 1; 1, 1] and **V** = input vector.
Let **A** be a 2x2-matrix such that V = AH.
We decompose A so that A = R x S x SH, where

- **R** corresponds to rotation, defined by an angle . To rotate by angle , counterclockwise, we multiply the vector by the matrix of the form [cos, -sin; sin, cos].
- **S** corresponds to scaling, defined by a scalar s. To scale the magnitude of an original vector by s, we multiply the vector by the matrix of the form [s, 0; 0, s].
- **SH** corresponds to shearing parallel to the x-axis, defined by a scalar m. To shear a plane, we multiply the vector by the matrix of the form [1, m; 0, 1].

Markov Model Implementation

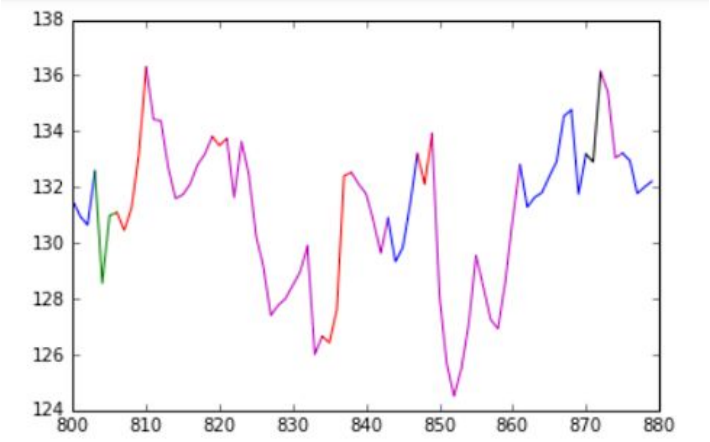
State #	0	1
0	0.3	0.7
1	0.9	0.1

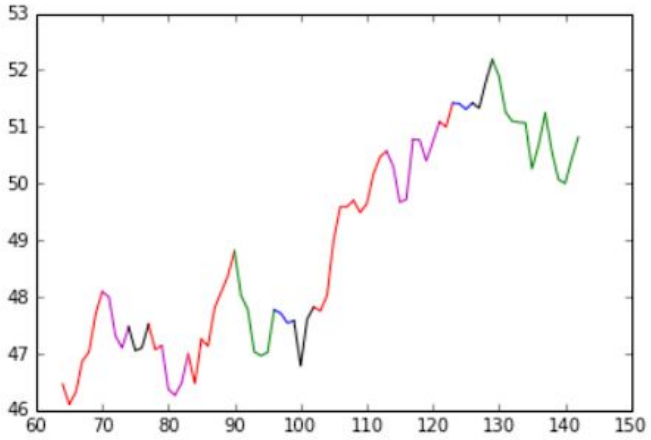
- Random next step based on the input's probability distribution of possible answers.
- The next step becomes a part of the input and is used to predict the step after that.

Dataset

- **S&P500** (https://www.kaggle.com/camnugent/sandp500)
  - 500 companies historical data over 5 years
  - Open, max, min, and close price each day
  - 619, 040 data points
- In the future, we will also use **Huge Stock Market Dataset** (https://www.kaggle.com/borismarjanovic/price-volume-data-for-all-us-stocks-etfs)
- We may also extract news to find correlation between **economic/political news** and stock price for different companies..

Data Visualization





Preliminary Result (Prediction)

- # of training data points: 119, 030 (80%)
- # of testing data points: 31, 444 (20%)
- Order of input: 3 (looks at 3 previous vectors)
  - Higher order takes much longer to generate model. Order 5 and 7 do not increase the accuracy.
- Steps of output: 3 (predict 3 steps ahead)
- Average accuracy: **~18%** (for all 3 steps)

Plan for Second-Half

- Improve prediction accuracy:
  - Developing a grouping algorithm for second method.
  - Improve grouping algorithm for scaling method.
  - Make prediction based on input of higher order.
  - Optimize the system pipeline and hyperparameters.
- Extract relevant economic events and use them to help predict the stock price.
- Find combinatorial pattern in the data based on our grouping algorithms.

Thank you for your attention.

Q&A