Encoding Stock Charts Patterns for Technical Analysis

Thitaree Tanprasert (Mint) & Julianne Lin

MATH189 SP2019

Background

- Stock data is very visual. Traders can look at them and forecast future price with a process called **technical analysis**.
- Experts can predict trends based on specific patterns:

Cup & Handle

Double-bottom

Image source: https://www.investopedia.com/university/charts/

Problem Statement

We aim to develop a **tool** that **detects human-recognizable patterns** in **stock charts** data.

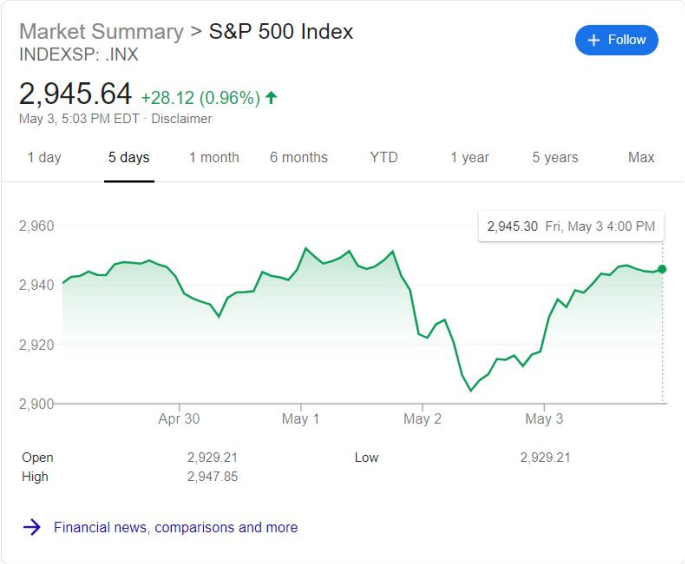
More specifically, we aim to show **a proof of concept** that by transforming time-series data into sequence of **V-shaped vectors**, we can detect **cup-and-handle** and **double-bottom** patterns based on **characteristics** of the vectors.

Related Work

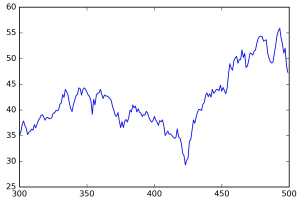
- Almost all approaches are based on neural network
  - Need annotated and labeled data, which are not made available
  - Their preprocessing methods could be implemented for our recognition algorithm
  - Explored models:
    - **Self-organizing maps**: update pattern templates with exponential average
    - **Multi-layered perceptrons:** one hidden layer, each output neuron corresponds to probability of the data matching a pattern
    - **Recurrent neural net**
- All approaches need human pattern recognition, either for labeling the data or for initializing pattern template

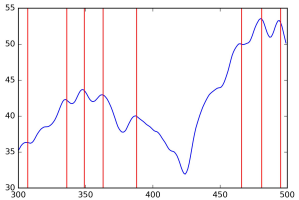
Dataset

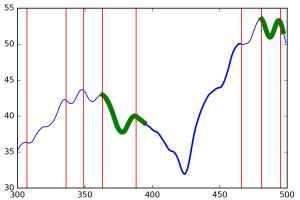
- **S&P500** (https://www.kaggle.com/camnugent/sandp500)
  - 500 companies historical data over 5 years
  - Open, max, min, and close price each day
  - 619, 040 data points



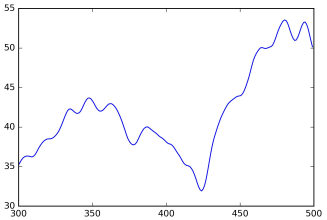
System Pipeline

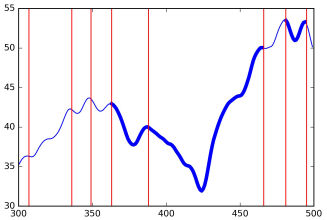






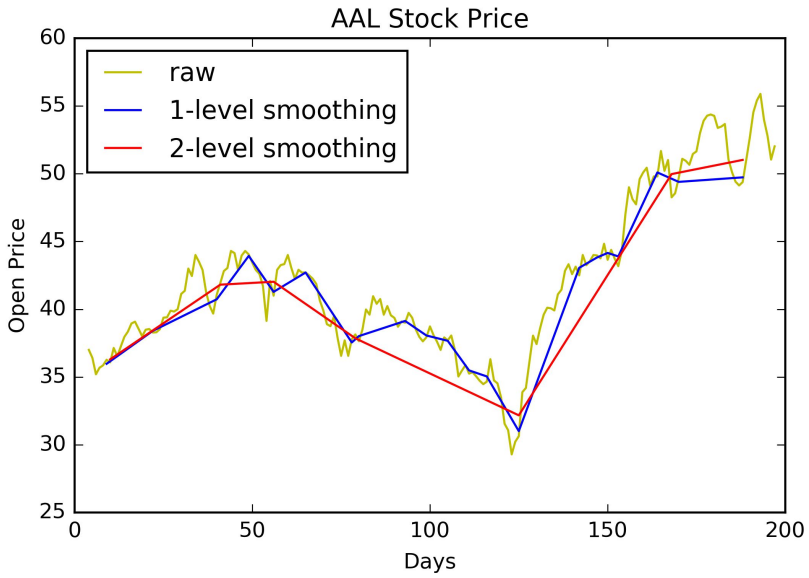
Stock data → Moving-average smoothing → V-shaped vector extraction → Identify starting vector candidates → Pattern detection





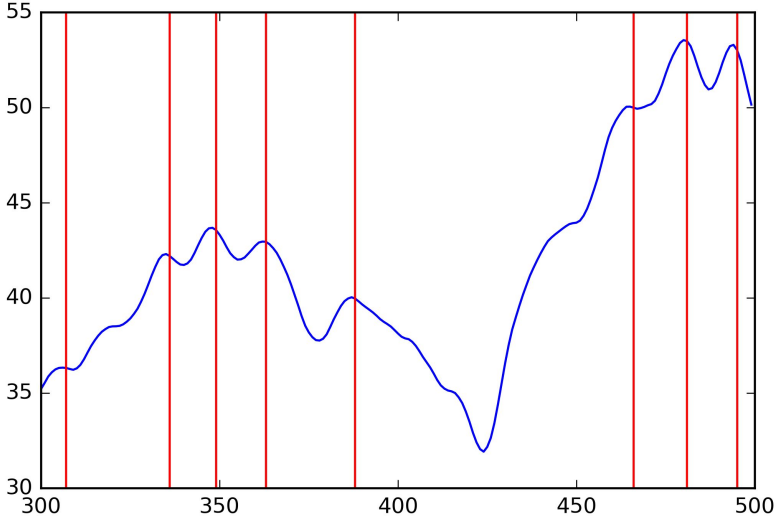
Moving-Average Smoothing

- Double 1-D convolution with averaging filter
- The size of the moving window is the mean of the length of V in the previous level of smoothing



V-shaped Vectors Extraction

- Identify all local maxima in the smoothened data
- For each V-shaped vector, calculate the rotation, scaling, and shearing parameters for each vector



Vector Representation & Transformation Matrix

$$V = \begin{bmatrix} x_m - x_l & x_r - x_m \\ y_l - y_m & y_r - y_m \end{bmatrix},$$

where (x_l, y_l) = timestamp and corresponding value of the left end of the vector

(x_m, y_m) = timestamp and corresponding of the local minimum

(x_r, y_r) = timestamp and corresponding of the right end of the vector

$$A = VH^{-1} = \frac{1}{2}\begin{bmatrix} -V_{11} + V_{12} & V_{11} + V_{12} \\ -V_{21} + V_{22} & V_{21} + V_{22} \end{bmatrix}$$

Vector Parameters Calculation

$$R = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \qquad S = \frac{1}{2}\begin{bmatrix} -s_l + s_r & s_l + s_r \\ -s_l + s_r & s_l + s_r \end{bmatrix} \qquad SH = \begin{bmatrix} 1 & m \\ 0 & 1 \end{bmatrix}$$

$$\begin{aligned} A &= R * S * SH \\ &= \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} * \begin{bmatrix} -s_l + s_r & s_l + s_r \\ -s_l + s_r & s_l + s_r \end{bmatrix} * \begin{bmatrix} 1 & m \\ 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} s_l \cos\theta & m s_l \cos\theta - s_r \sin\theta \\ s_l \sin\theta & m s_l \sin\theta + s_r \cos\theta \end{bmatrix} \end{aligned}$$

Vector Parameters Calculation

$$
\begin{aligned}
\theta &= \arctan(A_{21}/A_{11}) \\
s_l &= A_{11}/\cos\theta \\
s_r &= A_{22} - A_{12}\sin\theta \\
m &= (A_{12} + s_r \sin\theta)/s_l \cos\theta
\end{aligned}
$$

Starting vector candidates identification

- Find V-shaped vectors in which the left-antenna scaling factor is bigger than the “average” in that window of data
  - Only look for big left antenna because we’re currently only identifying cup & handle and double-bottom patterns.
- The average could be mean or median

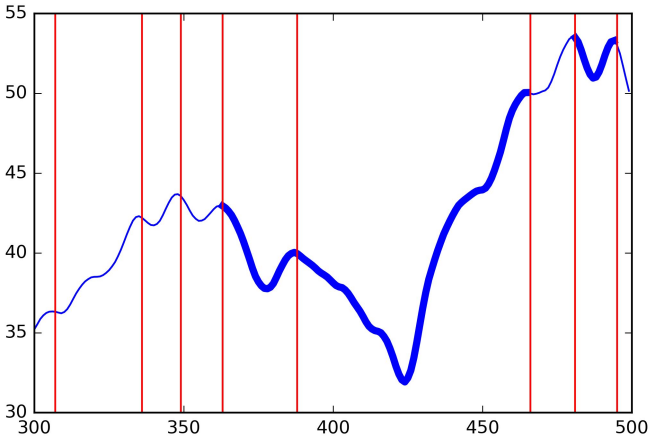
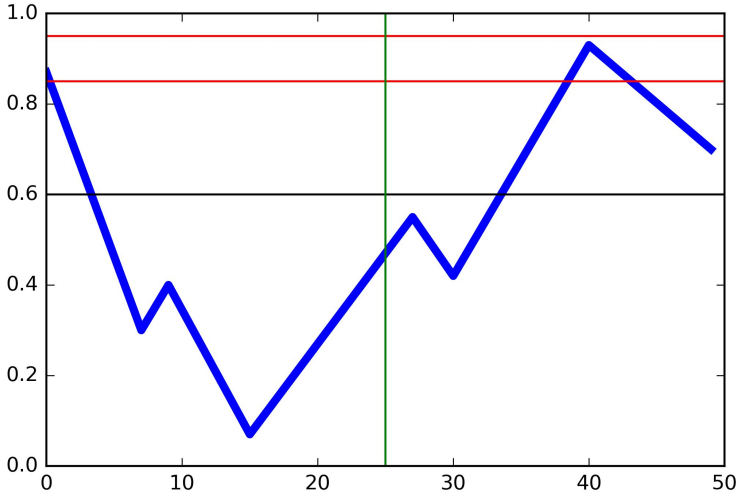


Image source: https://www.investopedia.com/university/charts/

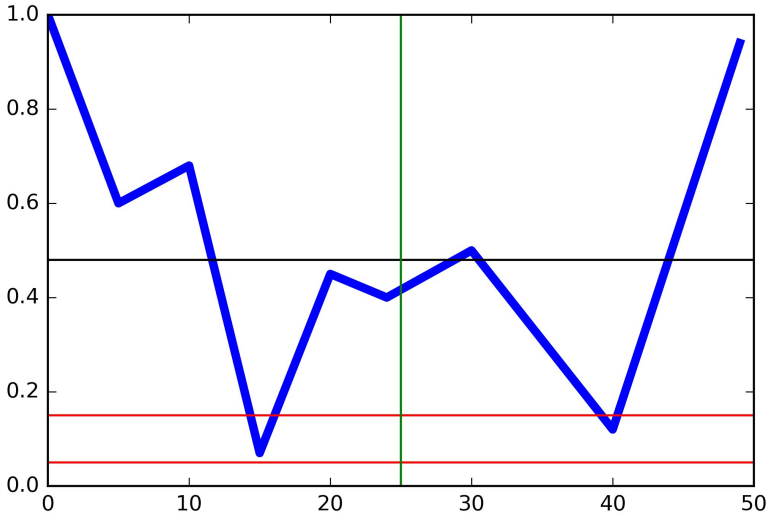
Pattern Encoding: Cup & Handle

- A big V shape follows by a fall
  - The 2 ends of the V must have similar heights under a preset threshold.
  - The 2 ends of V must be far apart by at least a preset length.
  - The big V can make up of many Vs.
  - The big V cannot have a local maximum inside higher than a preset threshold.



Pattern Encoding: Double-Bottom

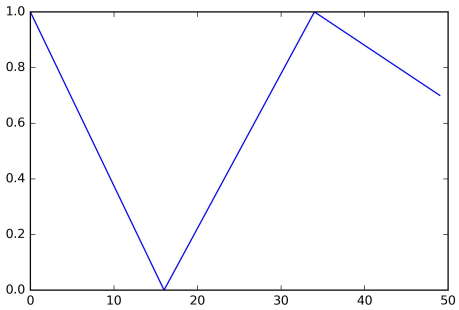
- Two big Vs
  - There must be 2 dips that are a lot lower than others between the 2 ends.
  - The 2 dips must differ by only under a threshold.
  - The right end has to be higher than a preset percentage of the left end.
  - Each big V can make up of many Vs.
  - The middle peak must be higher than a preset percentage of the left end.

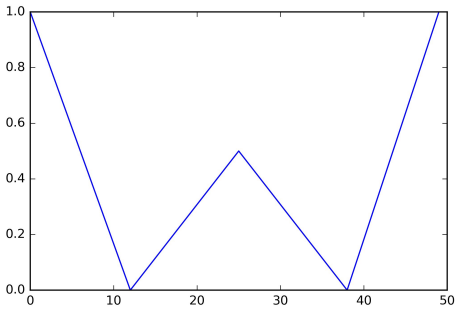


Baseline System

Modified from self-organizing map method

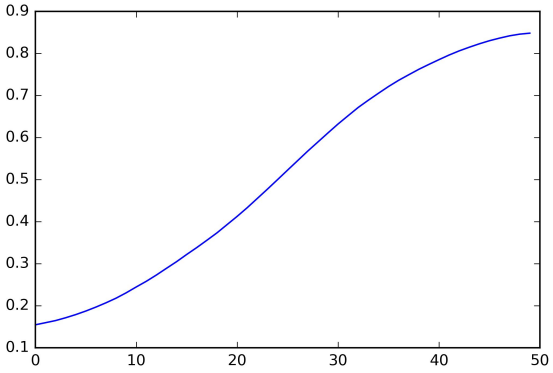
- Initialize the template for each pattern
- Identify a pattern that matches more to the input window of data with low sum-of-square error (under a preset threshold)
- Update the template by averaging it with data input



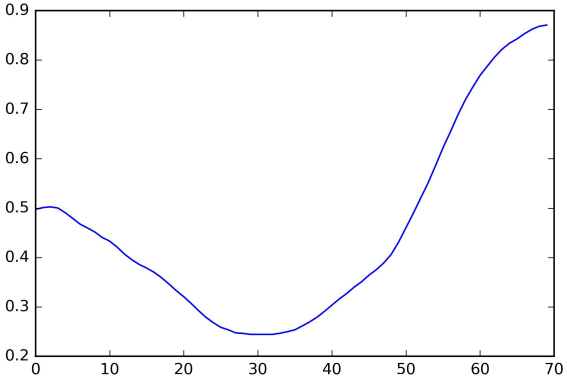


Baseline Result

- Final templates after training
- It is clear that this method is not good for identifying patterns in test data

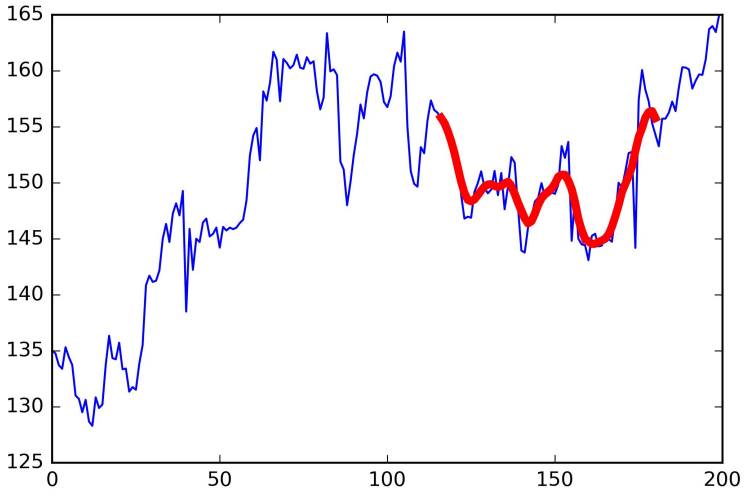


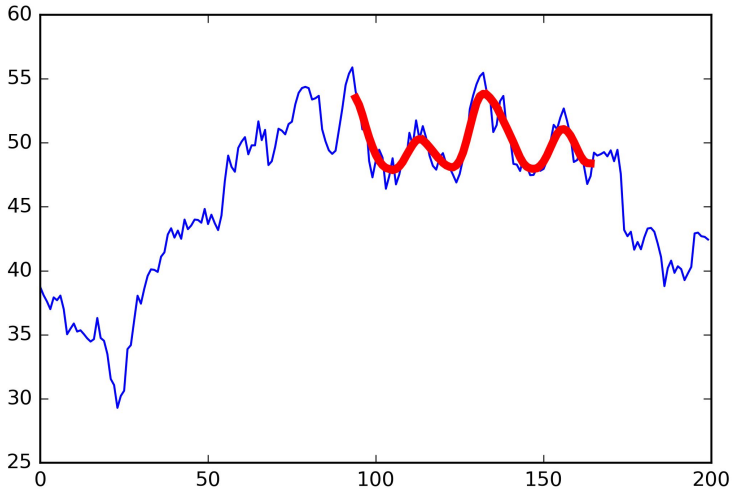
Cup & Handle



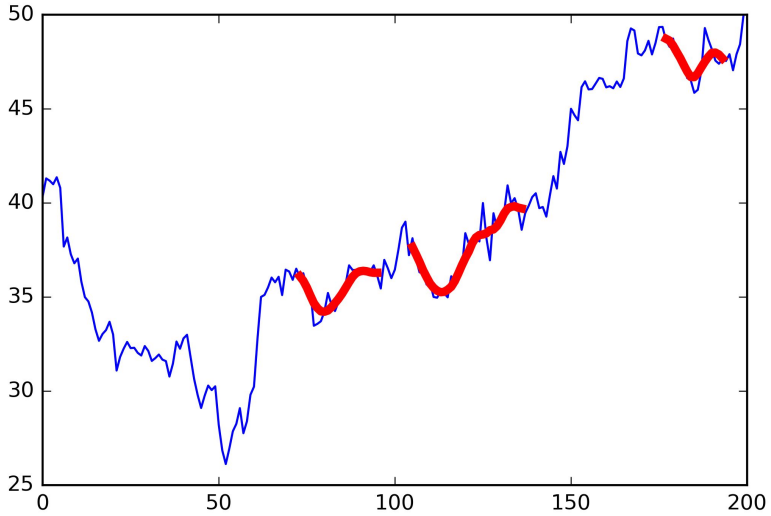
Double-bottom

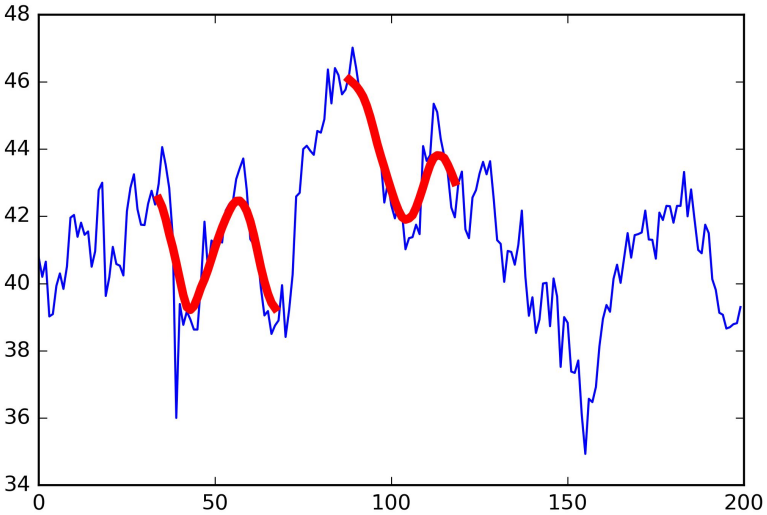
Our Result: Cup & Handle





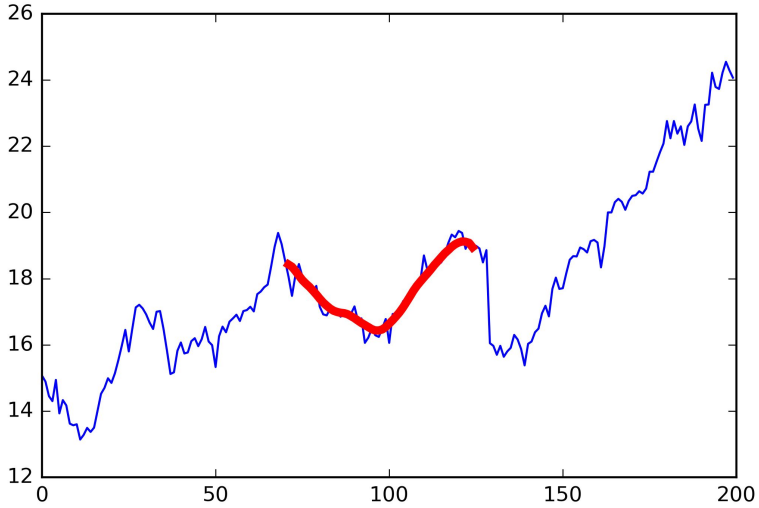
Our Result: Cup & Handle

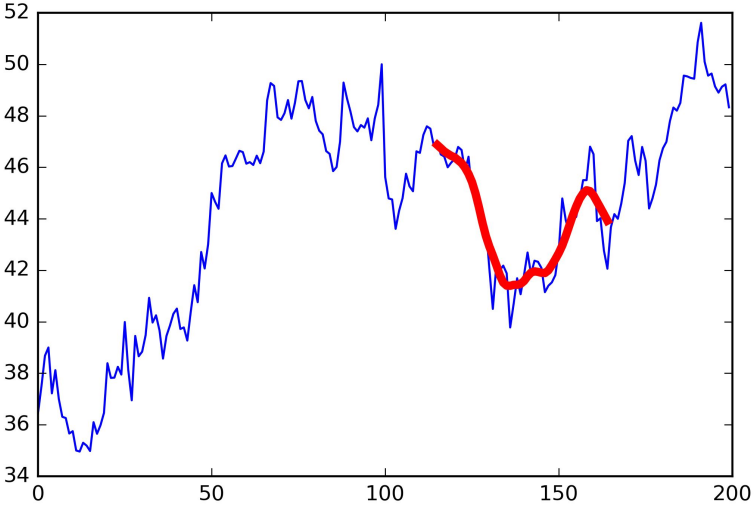




Decrease allowable height for peaks between ends

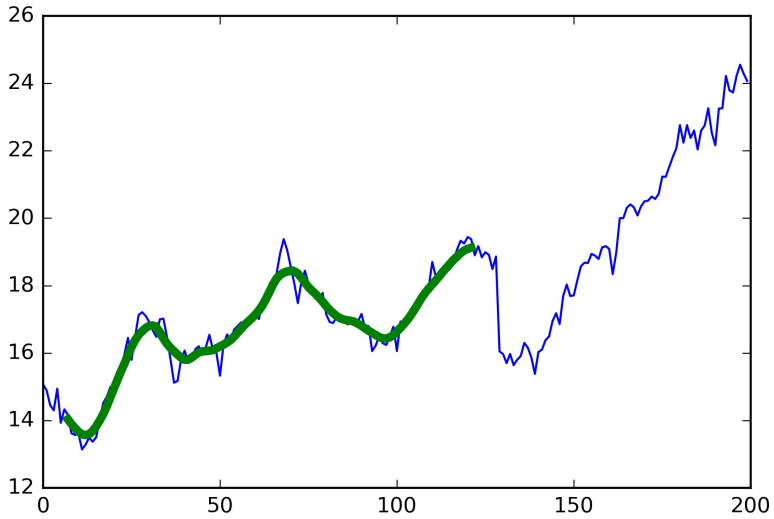
Our Result: Cup & Handle

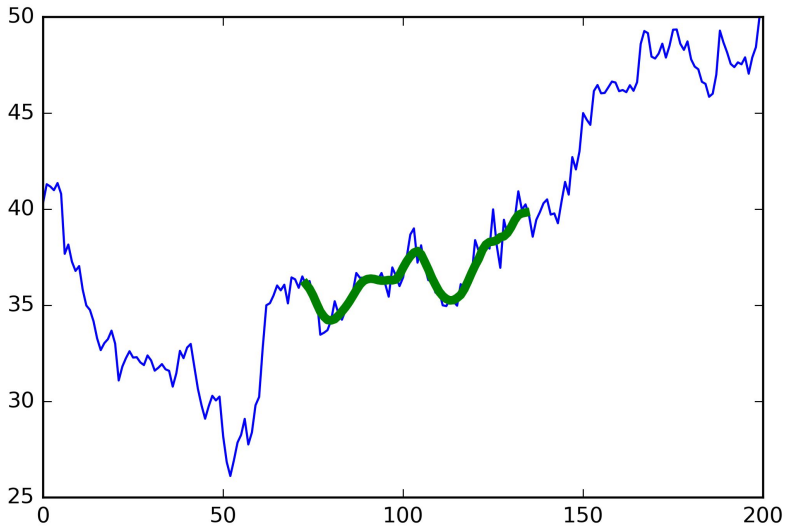




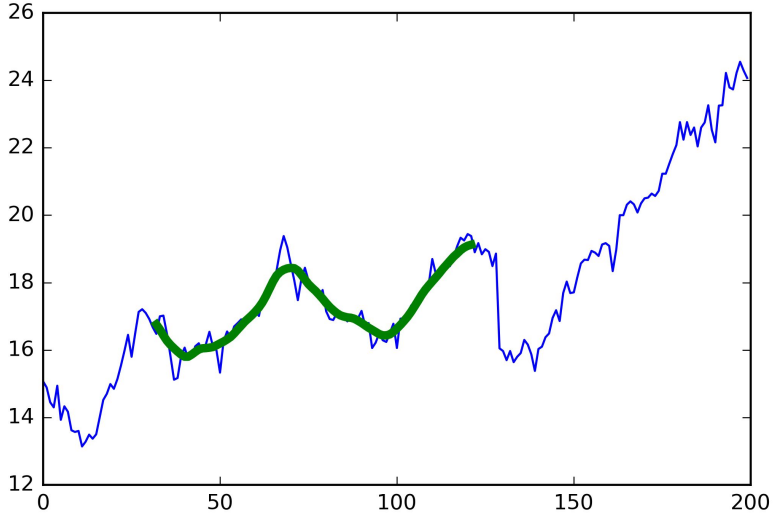
Decrease allowable difference between 2 ends

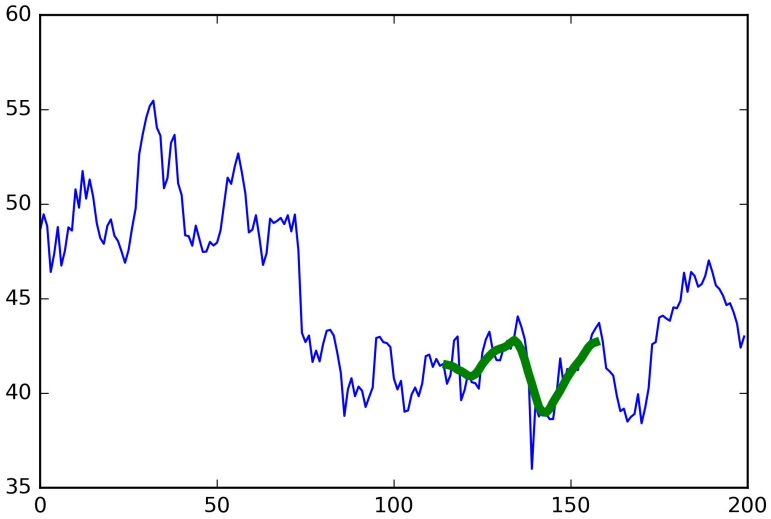
Our Result: Double-Bottom





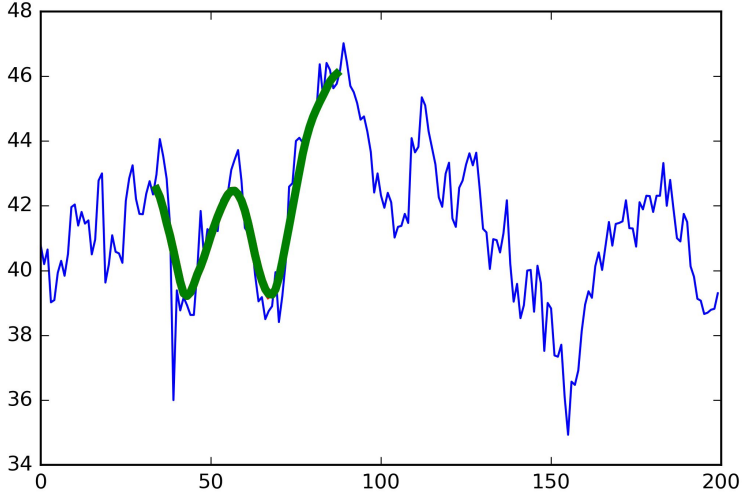
Our Result: Double-Bottom

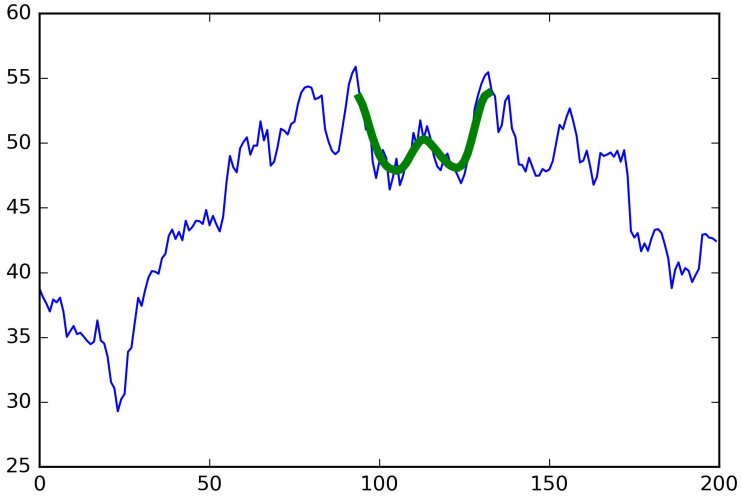




Constrain allowable height of the middle peak

Our Result: Double-Bottom

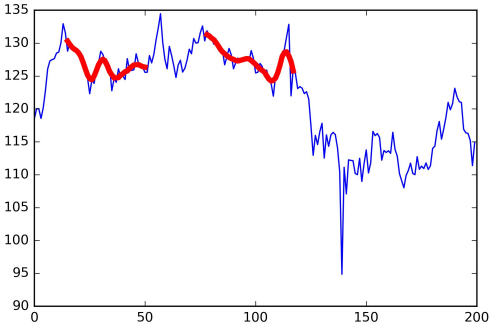


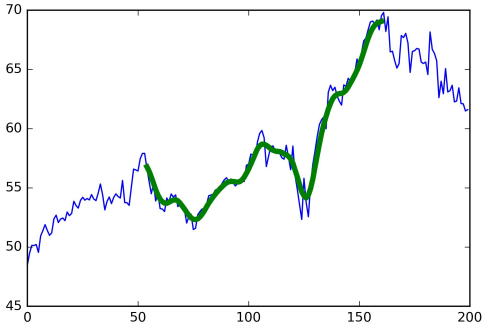


Decrease allowable difference between the 2 bottoms (dips)

Future Work

- With more domain-specific knowledge on stock pattern analysis, we can make the pattern hardcoding more robust and more accurate.
  - Use angle information to make patterns more specific
  - Refine our criteria for identifying starting vectors





Future Work

- Our system has the potential to extend to other patterns, since the encoding of each pattern is intuitive and is based on the same data preprocessing method.

Triangle

Flag & Pennant

Image source: https://www.investopedia.com/university/charts/

Future Work

- This system could be implemented for other time-series data to extract human-recognizable patterns:
  - Rainfall/precipitation data
  - Electricity consumption
  - Sales figures
  - Etc.

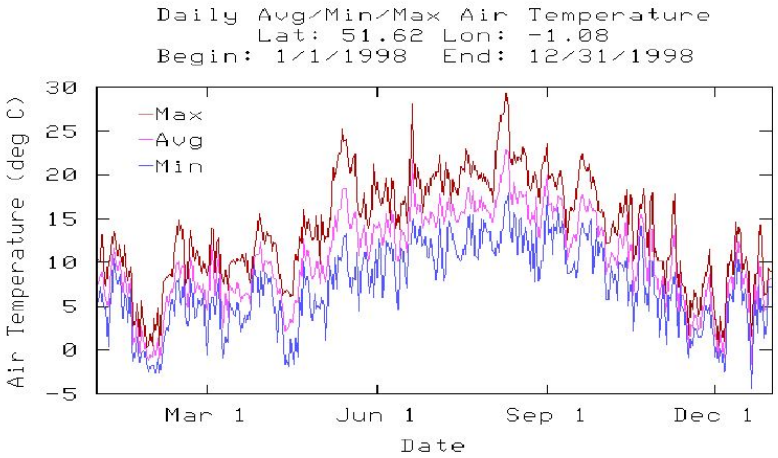


Image Source: NASA's Surface Meteorology and Solar Energy Web Portal (Release 6.0) - Scientific Figure on ResearchGate. Available from: https://www.researchgate.net/figure/Daily-averaged-temperature-time-series-data_fig2_268328855 [accessed 5 May, 2019]

Conclusion

- We developed a tool that, given 4 parameters of user's specification, recognizes basic stock charts patterns.
- The tool does not need training data or data annotation, only domain-specific knowledge on the general structure of the patterns.
- The tool works, effectively, for detecting cup & handle and double-bottom patterns. It can be easily extended to include other patterns as well.
- This shows that transforming data into V-shaped vectors make pattern recognition more intuitive and more robust to negligible oscillation.

Thank you for your attention

Questions?