Identifying Trading Opportunities using Unsupervised Learning
About Us

- MS Data Science at Columbia University
- Capstone Project: Identifying Trading Opportunities
- Mentors
  - Naftali Cohen
  - Zhen Zeng
  - Srijan Sood
Motivation

Technical Analysis

- Identify investment opportunities using price alone
- Subjective (particularly chart patterns)
- “Price is all that matters!”
Motivation

Technical Analysis

- Identify investment opportunities using price alone
- Subjective (particularly chart patterns)
- “Price is all that matters!”

Question: Can we objectively identify meaningful **multiscale** patterns in financial time-series data using unsupervised machine learning?
Motivation

Technical Analysis

- Identify investment opportunities using price alone
- Subjective (particularly chart patterns)
- “Price is all that matters!”

Question: Can we objectively identify meaningful *multiscale* patterns in financial time-series data using unsupervised machine learning?
Multiscale patterns in time-series

- Stock price:
  - Long timescale for context
  - Short timescale for current behavior
- Multiple harmonics for sine waves
● Murphy, John. **Technical Analysis of the Financial Markets.** Penguin, 1999


Key findings

- Time-series separable into clusters using unsupervised methods
- Simple harmonic functions best characterize the data
- Time, sector, profitability did not add predictive power
Key findings

- time-series separable into clusters using unsupervised methods
- simple harmonic functions best characterize the data
- time, sector, profitability did not add predictive power

Areas to build upon

- Preprocessing techniques
- Clustering algorithms
- Cluster quality
- Multiscale pattern evaluation
Outline

- Workflow
- Data generation
- Clustering pipelines
- Multiscale pattern evaluation
- Conclusion, next steps
Workflow of the project

Data Collection
- Real S&P 500 data
- Synthetic Data
  - Generate artificial \textit{multiscale} data
- Scaling & Random Sampling

Feature Extraction & Clustering Pipeline
- \textbf{Multiscale Feature Extraction}
  - Autoencoders, Discrete Cosine & Fourier Transform, Padded Sampling, Perceptually Important Points, Skip-sampling, ...

- \textbf{Cluster Check and Parameter Optimization}
  - Elbow Method KMeans, DTW & Euclidean metrics, Silhouette Score

- \textbf{Combine Best methods}
  - DCT + autoencoders, DCT + skip & padded sampling

Multiscale Evaluation: Evaluate all pipelines on real & synthetic data
- \textbf{PIP Permutation Entropy}
  - Which patterns are clustered together?
- \textbf{Conditional Distribution}
  - Which patterns are predictive of future behavior?
Data Generation
Synthetic data

How can we simulate multiscale data?

Discrete Cosine Transform (DCT):

Decomposition of a signal into a sum of long scale, short scale & noise patterns.

We generate patterns by creating DCT coefficients for each scale.
Pipelines
Pipeline: Important Concepts

- DCT & Fourier Transform
  - Smoother

- Dynamic Time Warping
  - Similarity measurement between 2 time series
  - Can capture similarity in patterns when time series are out of sync
Pipeline: Clustering & Optimizing Parameters

K-Means: Often used on time series to discover the existing patterns within each signal.

Optimizing k:
- Silhouette Score accounts for the intra and inter-cluster distance.
- The higher the score, the better; the less clusters the better.
- The Elbow Method enables us to find a balance between the two.
DCT & Skipped: Adding 2 long scale dimensions to each time series
Main takeaway:
The clusters are capturing **harmonic trends** and the times series are **evenly spread** among the clusters.
Autoencoders: Neural Networks to extract features of the long scale time series
Pipeline: DCT & Autoencoders

Main takeaway:
Results tend to be independent of the autoencoder architecture: CNN, LSTM or single linear layer
Multiscale Evaluation
This is a **complicated** graph!

How do we know if the clusters capture **multiscale** patterns?
Multiscale Evaluation: Long-term Scale

Overall increasing trend in 60 days
Multiscale Evaluation: Short-term Scale

They should **NOT** belong in the same cluster
Multiscale Evaluation: Perceptually Important Point (PIP)

- A method for dimensionality reduction of time series
- Extract the most important points from a human observer perspective
- Use PIP to evaluate multiscale patterns
Multiscale Evaluation: Permutation Entropy

- A method for time series complexity measurement
- Downsampling 3 or more points using PIP
- Classify motifs using a sliding window
Multiscale Evaluation: Permutation Entropy

$x = [6, 9, 11, 12, 8, 13, 5], n = 3, \tau = 1$

Second sliding window: increasing motif (1,2,3)
Third sliding window: decreasing motif \((2,3,1)\)
Multiscale Evaluation: Cluster: Skipped-Values

Same long-term scale, What about short-term scale?
Multiscale Evaluation: Permutation entropy: Skipped Values

- First graph in the 3-PIP histogram of all ts
- Second graph is the 5-PIP point sliding window of length 3

Different in the short-term scale
Conditional Probability Analyses

- insights on predictive power of clusters
- how short term (or future) trends relate to long term (or past) trends

Steps:

- Split all time series chunks at fixed points to obtain two different sets
- Cluster independently
- Probability of cluster assignment in 2nd set conditioned on cluster assigned in 1st set

\[ P(\text{cluster}_{2\text{nd set}} | \text{cluster}_{1\text{st set}} = c) \]
Conditional Probability Analyses

Long scale/Past

Short scale/Future

60 days

20 days

6

7

Conditional probability distribution
Conditional Probability Analyses: k-means

Past time-series clusters
(4 shown here)

Populations of future timeseries clusters
Conditional Probability Analyses: Autoencoders

Past time-series clusters (4 shown here)

Populations of future timeseries clusters
Conditional Probability Analyses: Evaluation

Normalized shannon entropy

\[ \eta(X) = \frac{H}{H_{max}} = -\sum_{i=1}^{n} \frac{p(x_i) \log_b(p(x_i))}{\log_b(n)} \]

\( n = \text{number of clusters} \)

Mean normalized shannon entropy across clusters obtained using different pipelines on 80 days time series with split as \( X/Y \)

<table>
<thead>
<tr>
<th>Pipeline</th>
<th>40/40</th>
<th>50/30</th>
<th>60/20</th>
</tr>
</thead>
<tbody>
<tr>
<td>( k)-means</td>
<td>0.978</td>
<td>0.975</td>
<td>0.986</td>
</tr>
<tr>
<td>Autoencoders</td>
<td>0.962</td>
<td>0.949</td>
<td>0.978</td>
</tr>
</tbody>
</table>
Conclusion & Next Steps
Conclusions & Next steps

Key Findings
- Our research opens new directions for multiscale analysis and evaluation in stock prices
- Ensembling different methods leads to more compact clusters
- Results varied across methods and clusters

Next Steps
- Fine-tuning existing pipelines to make sure that all clusters capture multiscale properties
- Introducing Profitability:
  - Which clusters are capturing profitable patterns?
  - Can we use clustering to predict future stock behavior?
Thank you!