Identifying Trading Opportunities using Unsupervised Learning

About Us

Hritik Jain

Romane Goldmuntz

Amaury Sudrie

Vy Tran

Kassie Papasotiriou

Maxime Tchibozo

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

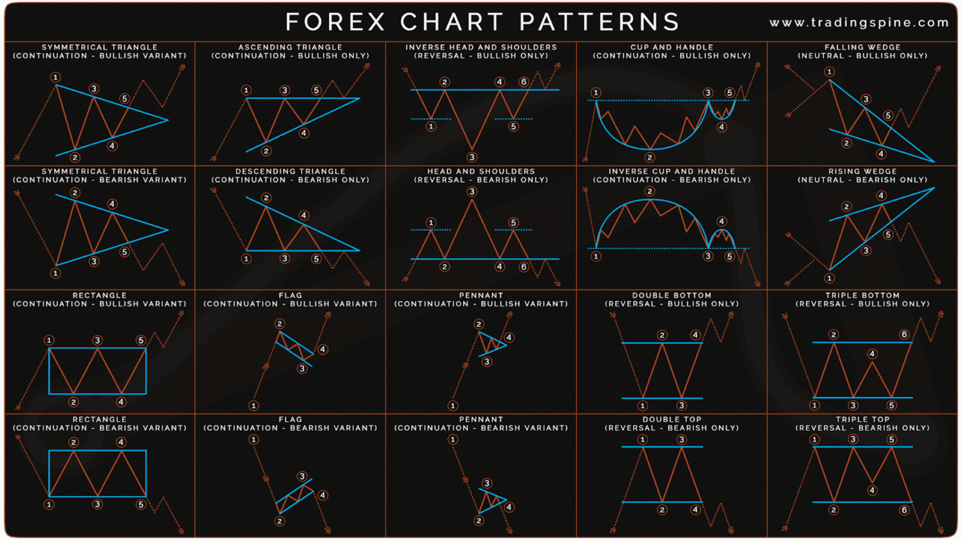
Motivation

Technical Analysis

- Identify investment opportunities using price alone

- Subjective (particularly chart patterns)

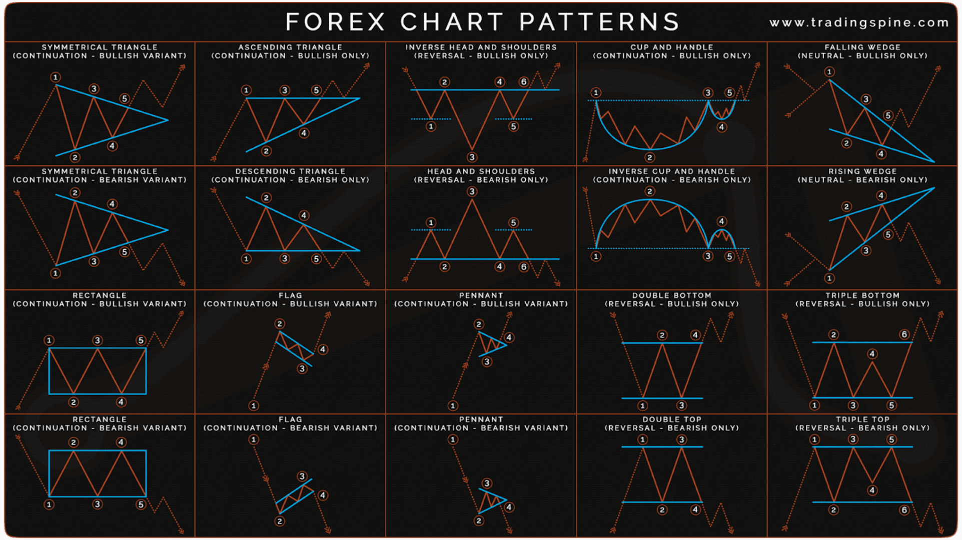
- “Price is all that matters!”



Motivation

Technical Analysis

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

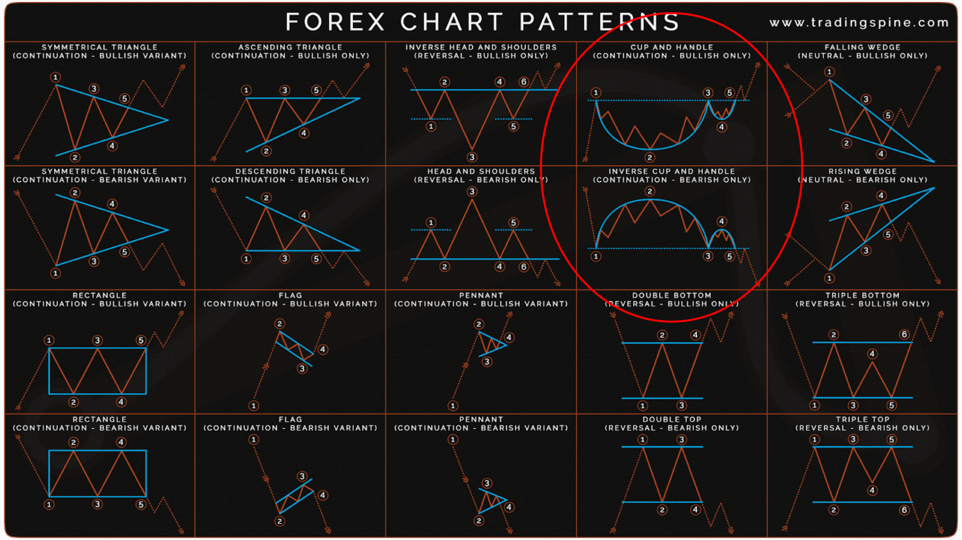


Question: Can we objectively identify meaningful ***multiscale*** patterns in financial time-series data using unsupervised machine learning?

Motivation

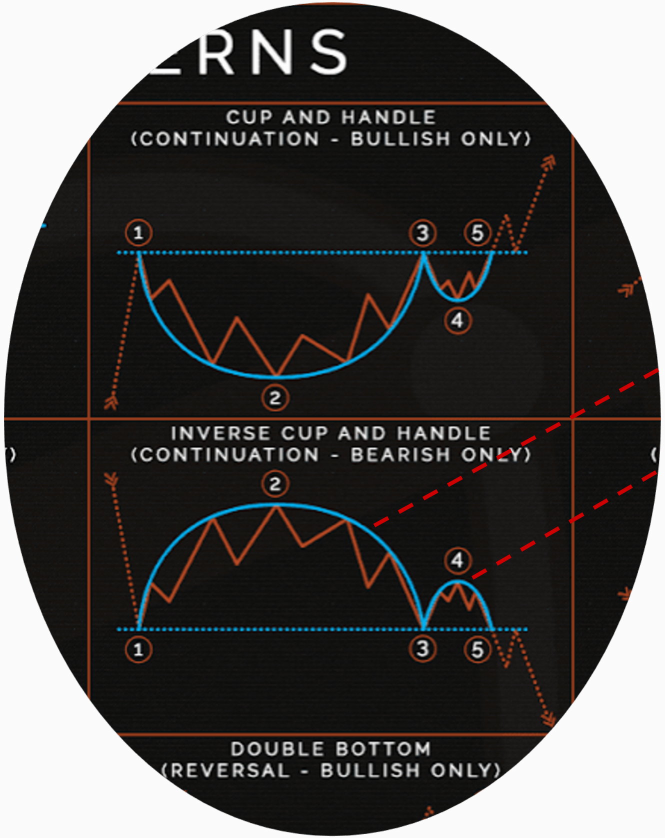
Technical Analysis

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

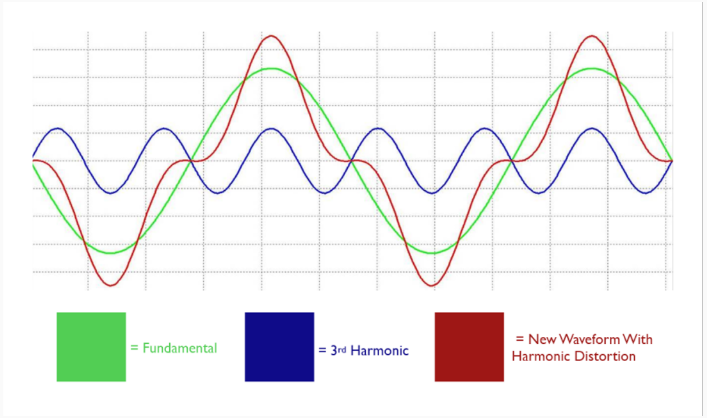


Question: Can we objectively identify meaningful ***multiscale*** patterns in financial time-series data using unsupervised machine learning?

Multiscale patterns in time-series



- Stock price:
  - Long timescale for context
  - Short timescale for current behavior
- Multiple harmonics for sine waves



Literature

- Murphy, John. **Technical Analysis of the Financial Markets.** Penguin, 1999

- Lo, Andrew W., Mamaysky, Harry and Wang, Jiang. **Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation**. *The Journal of Finance* (2000)

- Leigh, William, et al. **Stock market trading rule discovery using technical charting heuristics.** *Expert Systems with Applications* 23.2 (2002)

- Wang, Jar-Long, and Shu-Hui Chan. **Stock market trading rule discovery using pattern recognition and technical analysis.** *Expert Systems with Applications* 33.2 (2007)

Previous work: Lawrence Huang, AI Research Intern, Summer 2020

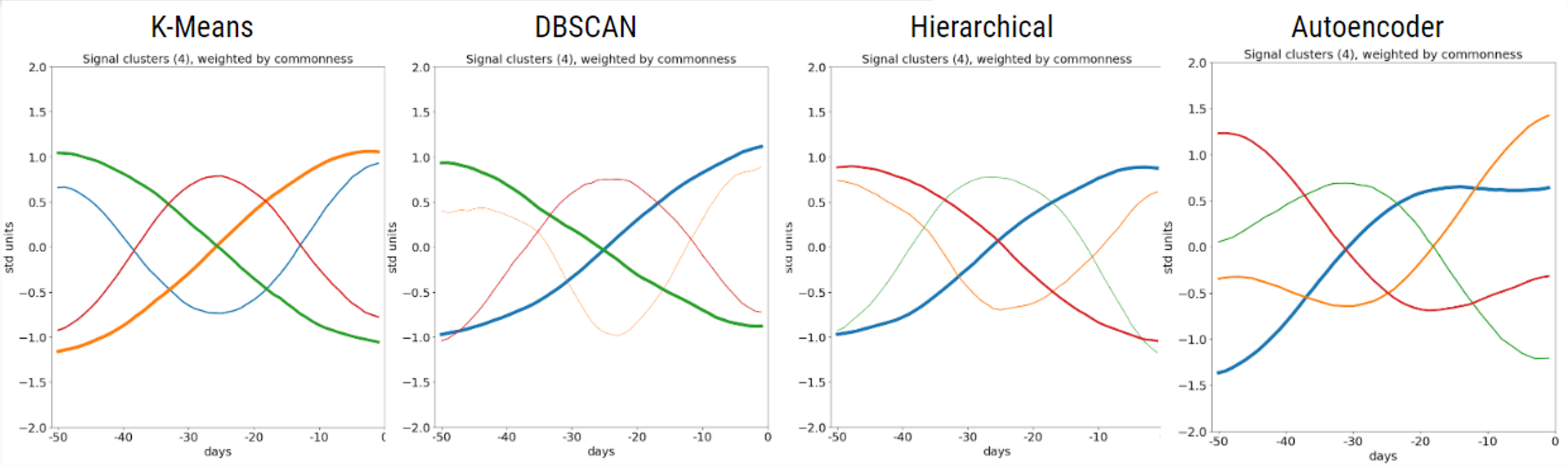


Figure 8: Average time series of clusters using four different clustering methods.

Source: ***Searching for Patterns in Daily Stock Data: First Steps Towards Data-Driven Technical Analysis*** *By Lawrence Huang, AI Research Intern, Summer 2020*

- 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

Previous work: Lawrence Huang, AI Research Intern, Summer 2020

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 **multiscale** data

Scaling & Random Sampling

Feature Extraction & Clustering Pipeline

Multiscale Feature Extraction

Autoencoders, Discrete Cosine & Fourier Transform, Padded Sampling, Perceptually Important Points, Skip-sampling, ...

Cluster Check and Parameter Optimization

Elbow Method KMeans,
DTW & Euclidean metrics
Silhouette Score

Combine Best methods

DCT + autoencoders,
DCT + skip & padded sampling

Multiscale Evaluation: Evaluate all pipelines on real & synthetic data

PIP Permutation Entropy

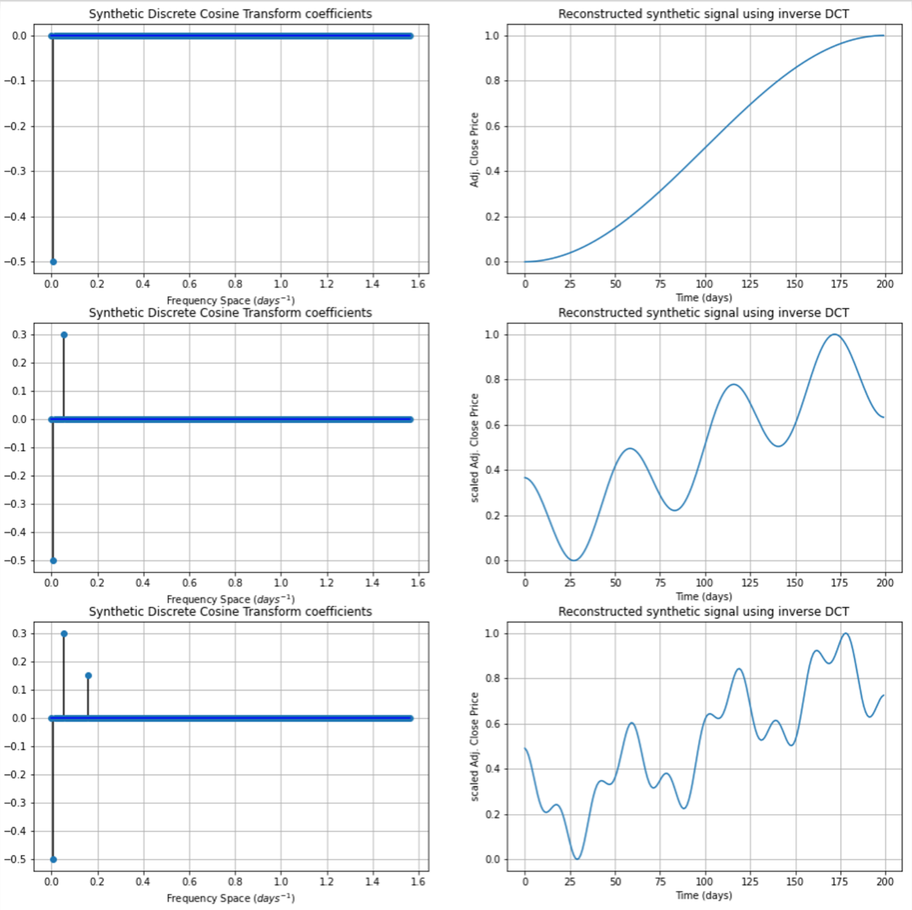
Which patterns are clustered together?

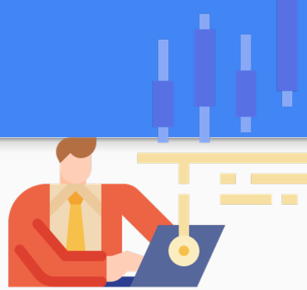
Conditional Distribution

Which patterns are predictive of future behavior?

Data Generation

Data Generation





<u>Synthetic data</u>

How can we simulate multiscale data?

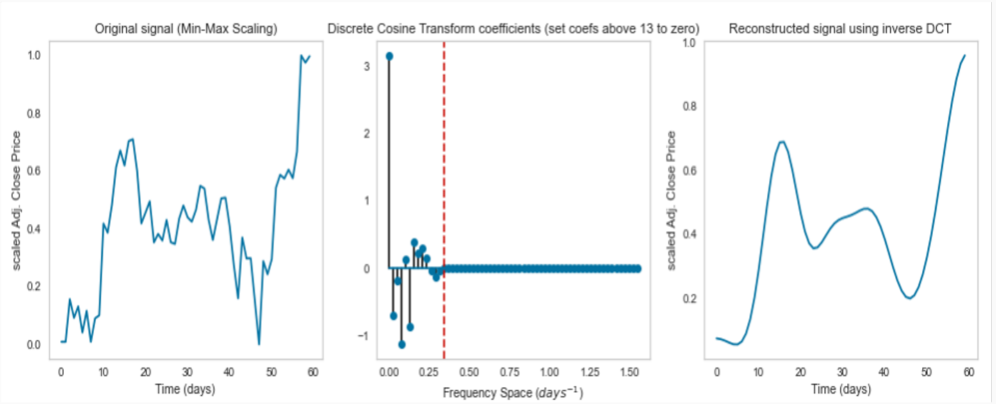
Discrete Cosine Transform (DCT):

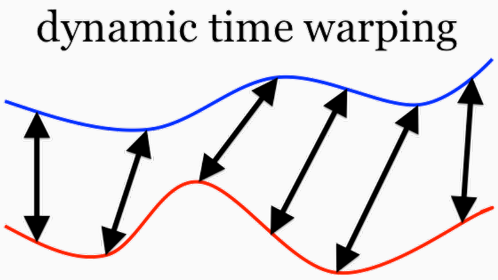
Decomposition of a signal into a sum of long scale, short scale & noise patterns.

We generate patterns by creating DCT coefficients for each scale.

Pipelines

Pipeline : Important Concepts

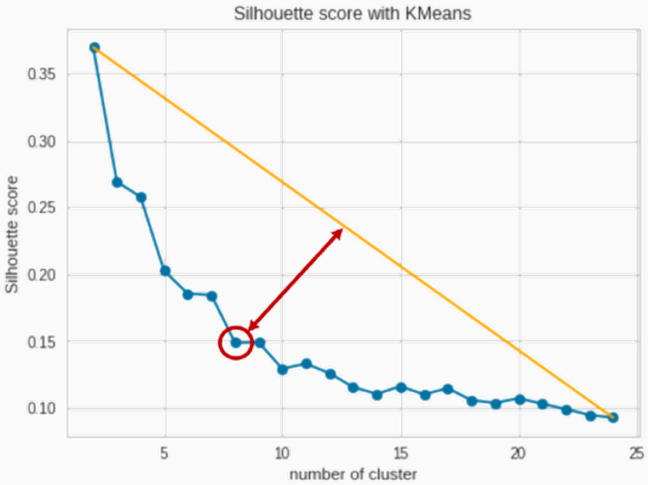




- **DCT & Fourier Transform**
  - **Smoother**

- **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

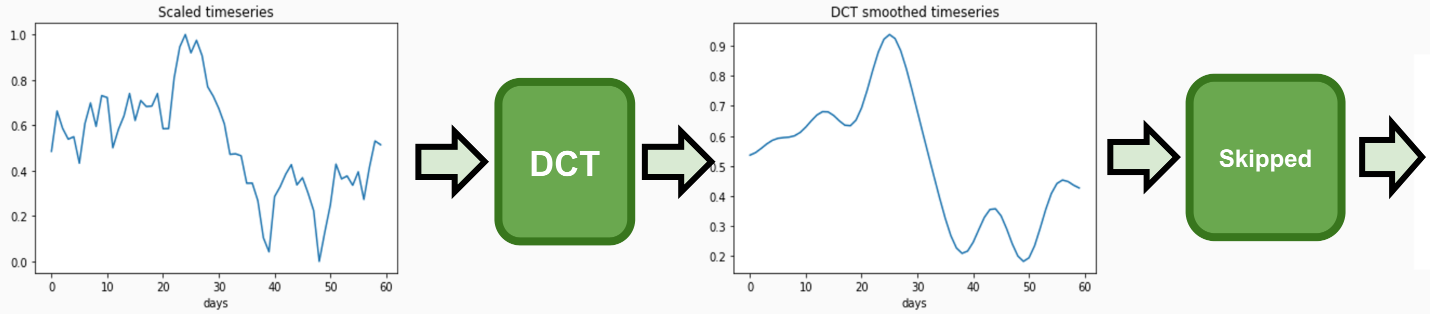


K-Means: Often used on time series to discover the existing patterns within each signal.

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.**

Pipeline: DCT & Skipped



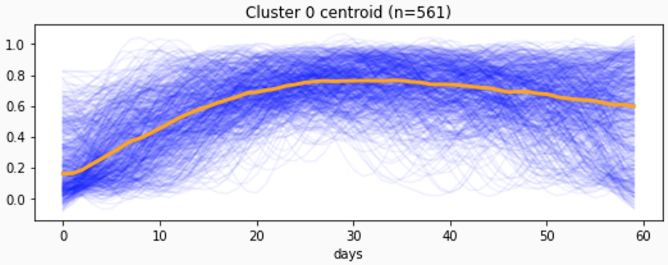
$$\begin{pmatrix} 0.54 & 0.55 & 0.56 & 0.57 & \dots & 0.43 \\ 0.54 & 0 & 0.56 & 0 & \dots & 0 \\ 0.54 & 0 & 0 & 0.57 & \dots & 0 \end{pmatrix}$$

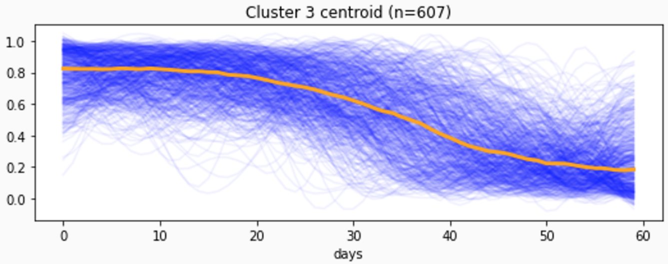
DCT & Skipped : Adding 2 long scale dimensions to each time series

Pipeline: DCT & Skipped

$$\begin{pmatrix} 0.54 & 0.55 & 0.56 & 0.57 & \ldots & 0.43 \\ 0.54 & 0 & 0.56 & 0 & \ldots & 0 \\ 0.54 & 0 & 0 & 0.57 & \ldots & 0 \end{pmatrix}$$

K-Means Clustering DTW

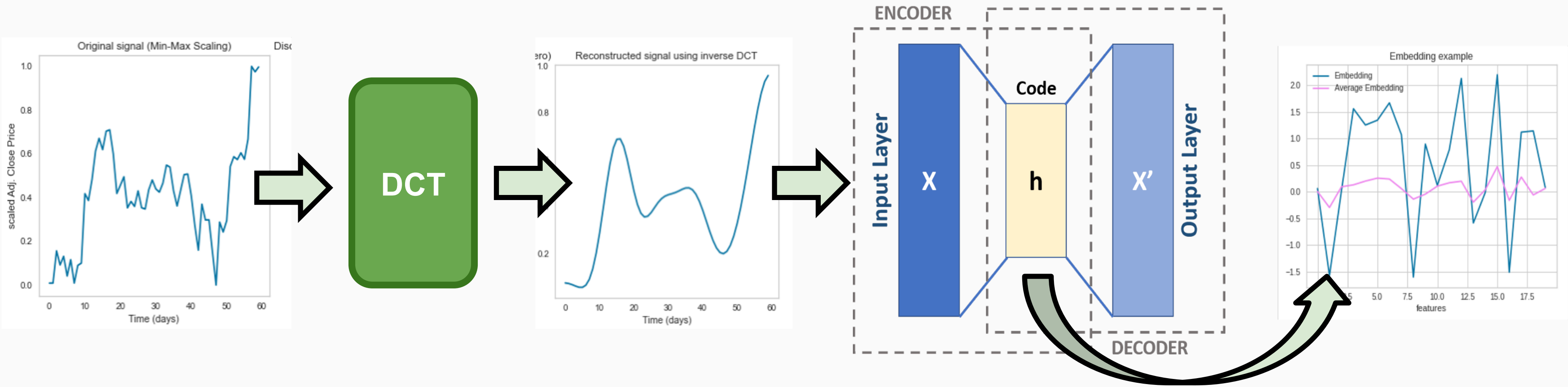




Main takeaway:

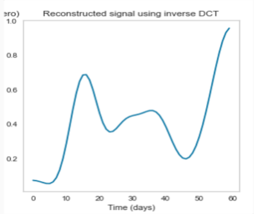
The clusters are capturing harmonic trends and the times series are evenly spread among the clusters

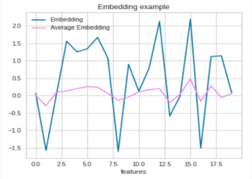
Pipeline: DCT & Autoencoders



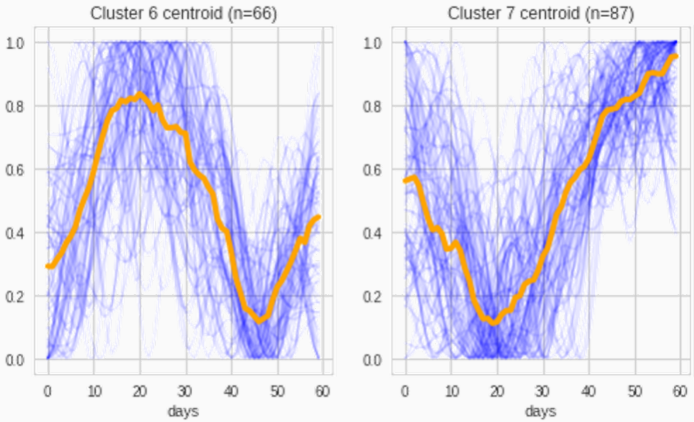
Autoencoders: Neural Networks to extract features of the long scale time series

Pipeline: DCT & Autoencoders





K-Means clustering

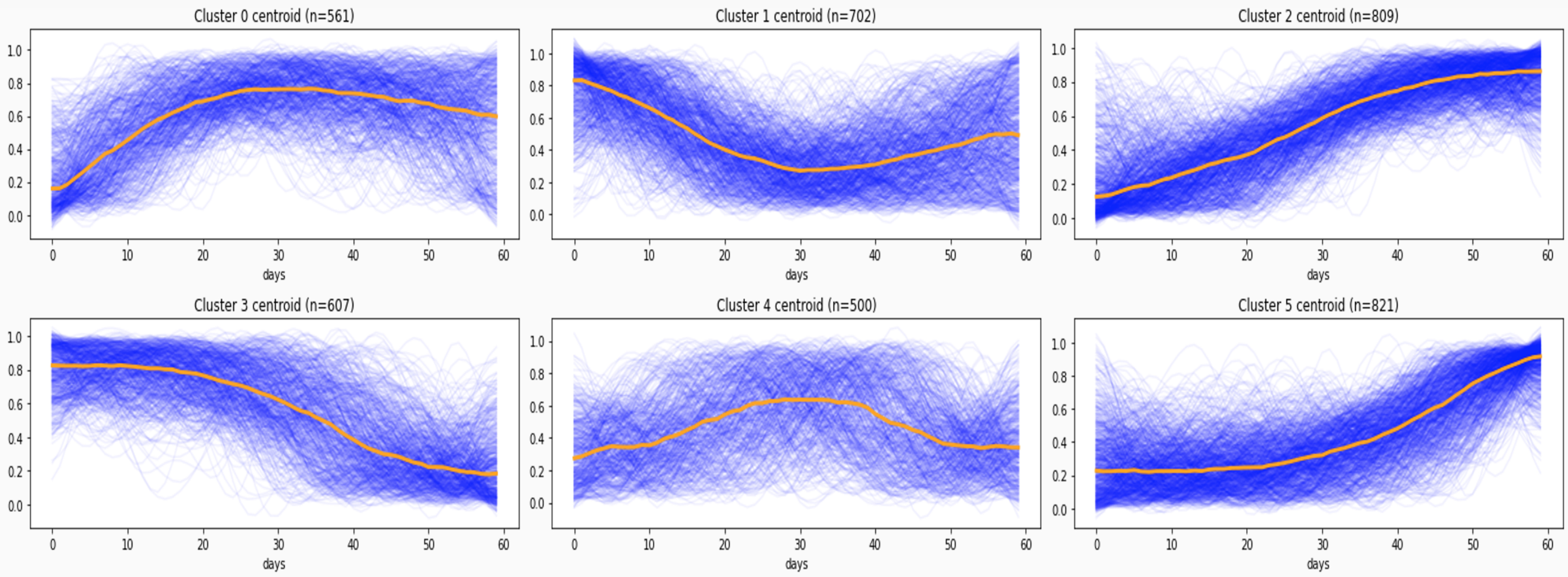


Main takeaway:

Results tend to be independent of the autoencoder architecture: CNN, LSTM or single linear layer

Multiscale Evaluation

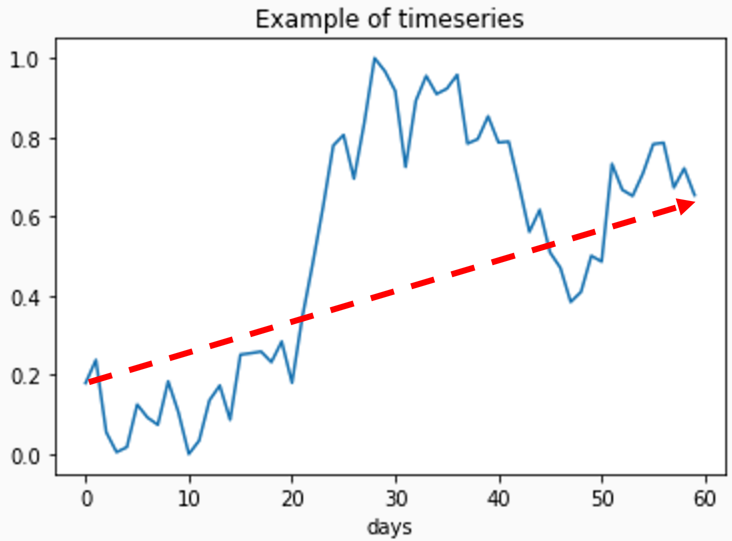
Multiscale Evaluation

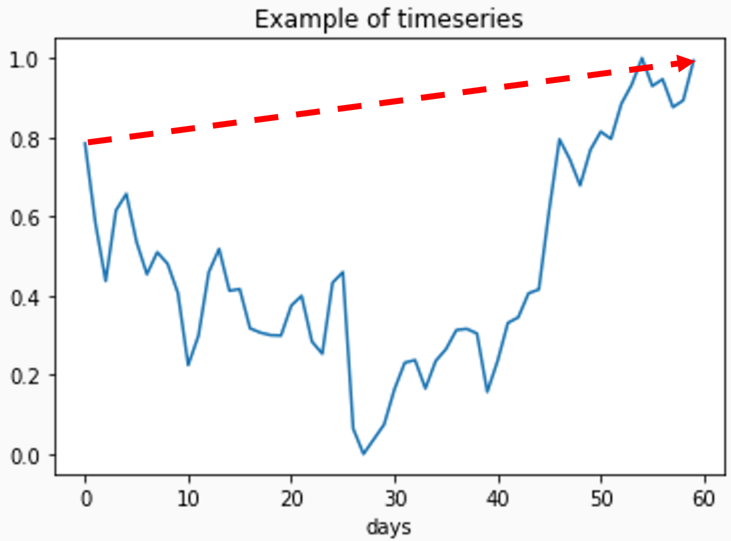


This is a complicated graph!

How do we know if the clusters capture multiscale patterns?

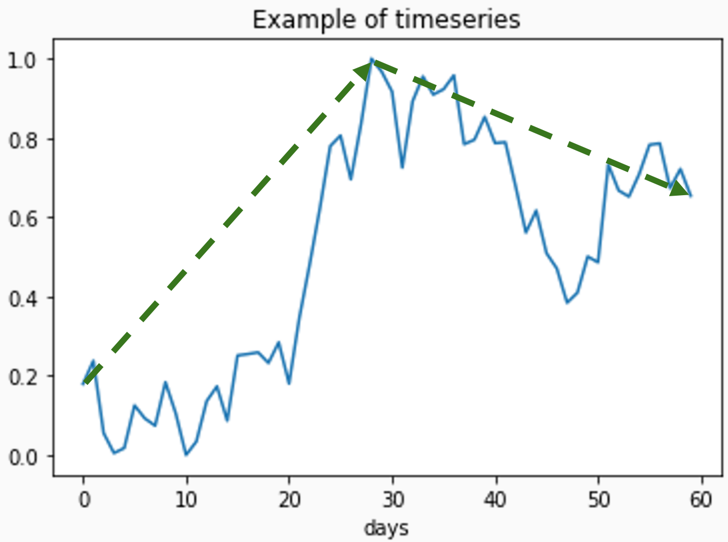
Multiscale Evaluation: Long-term Scale



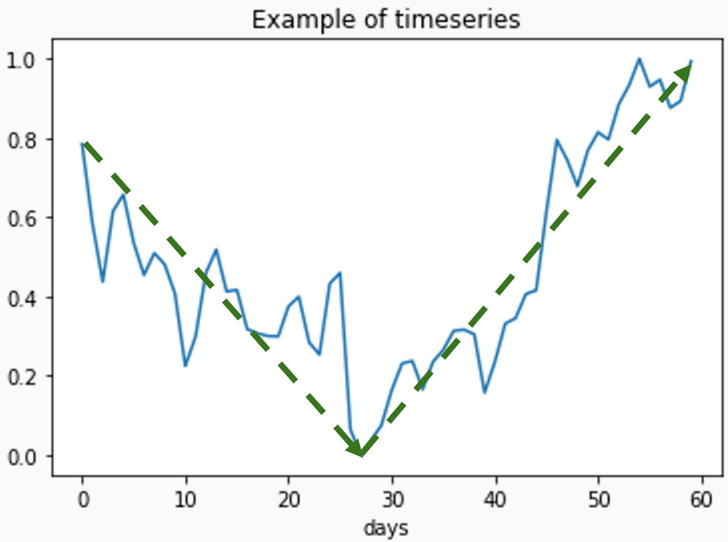


Overall increasing trend in 60 days

Multiscale Evaluation: Short-term Scale



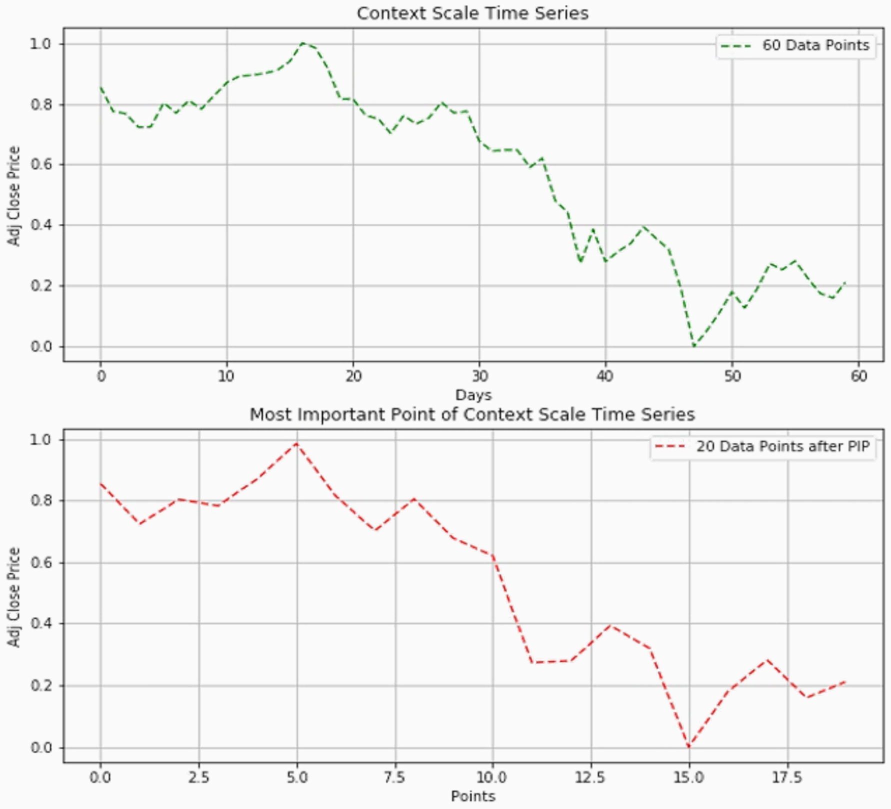
Up then Down



Down then Up

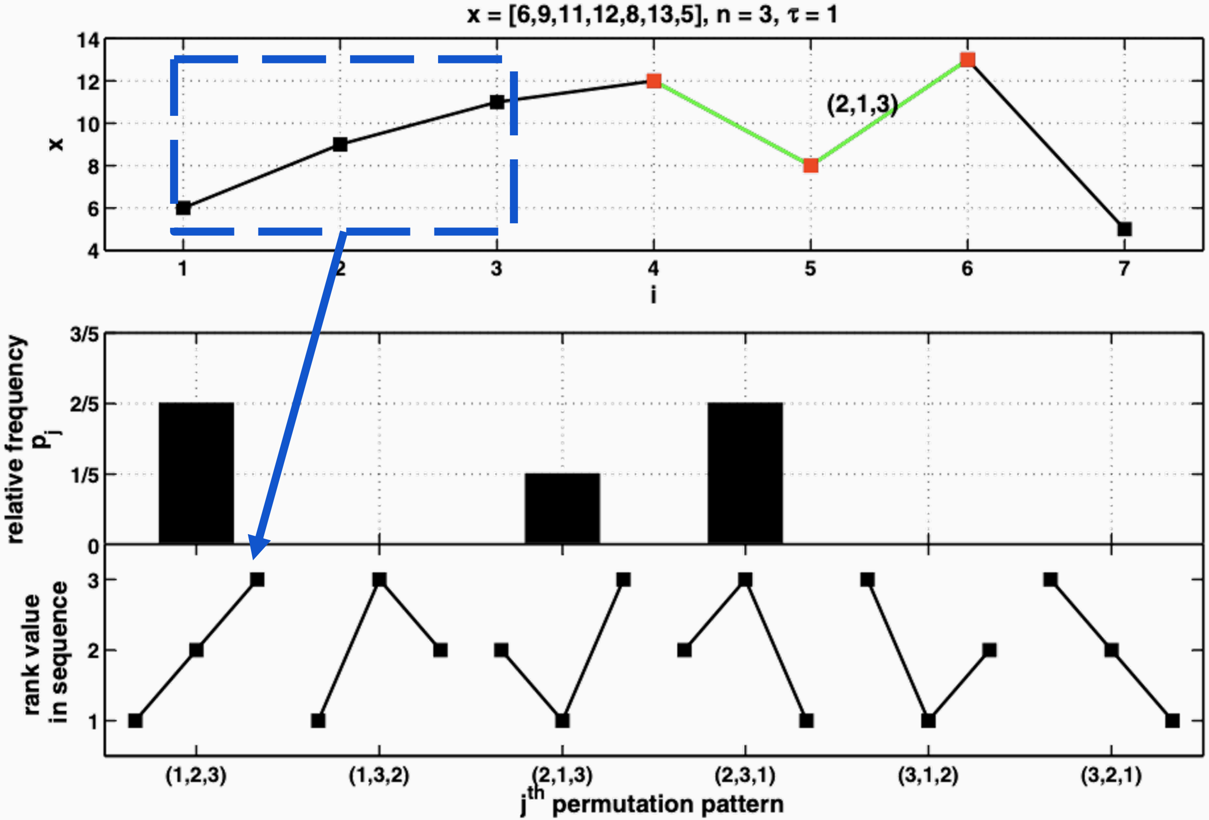
They should NOT belong in the same cluster

Multiscale Evaluation: Perceptually Important Point (PIP)



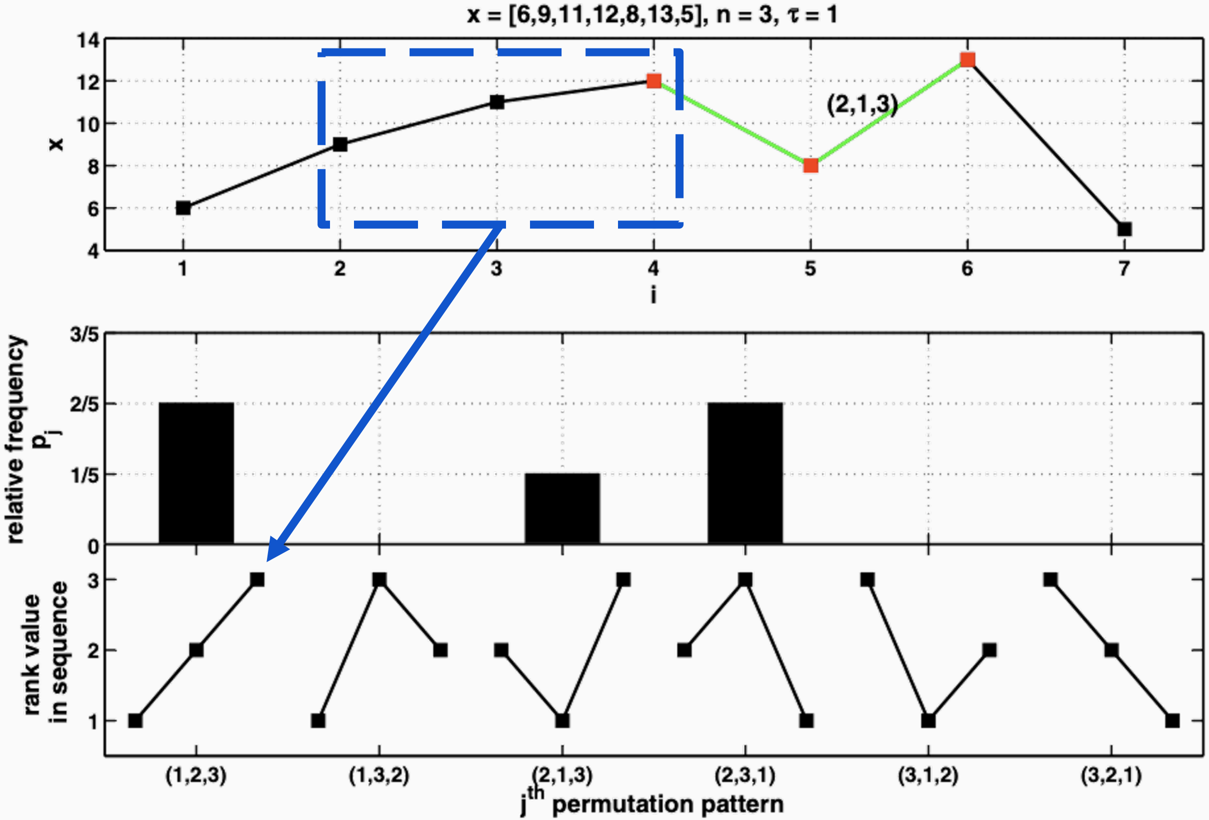
- **A method for dimensionality reduction of time series**
- **Extract the most important points from a human observer perspective**
- **Use PIP to evaluate multiscale patterns**

Multiscale Evaluation: Permutation Entropy



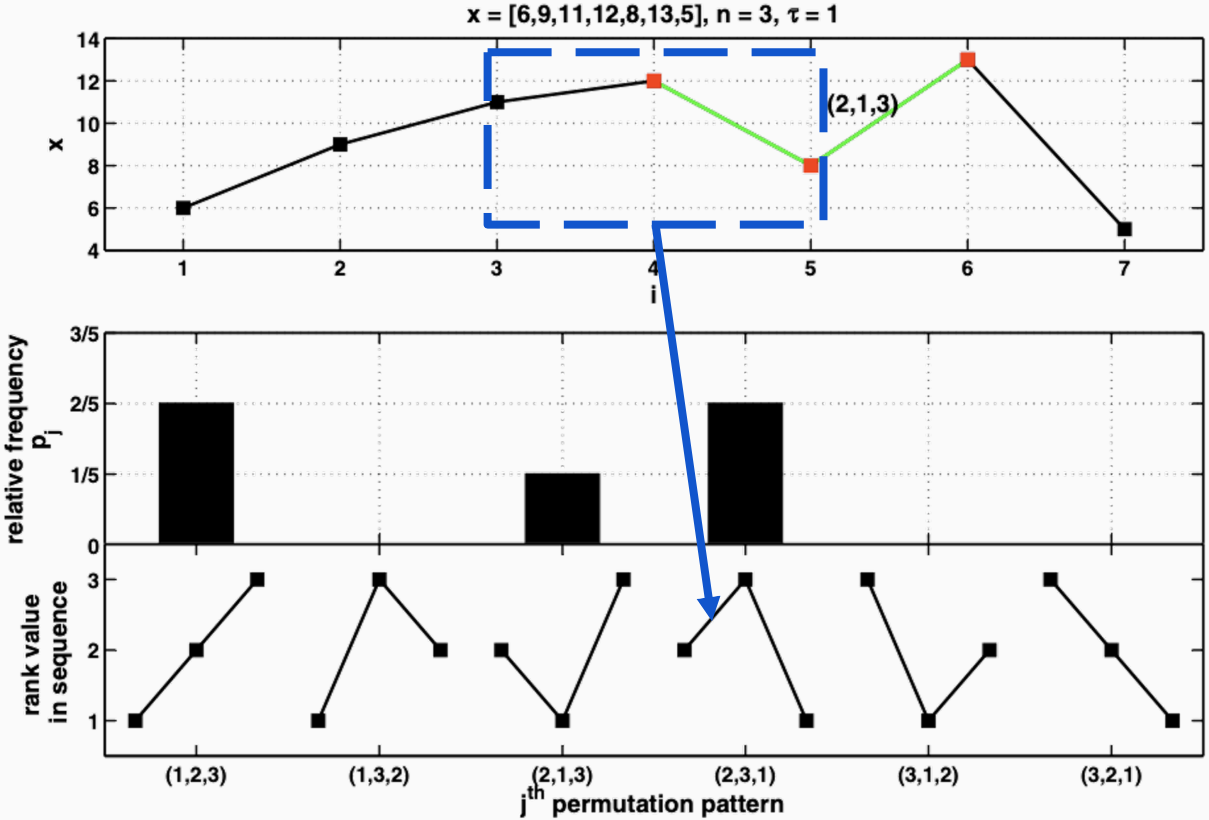
- **A method for time series complexity measurement**
- **Downsampling 3 or more points using PIP**
- **Classify motifs using a sliding window**

Multiscale Evaluation: Permutation Entropy



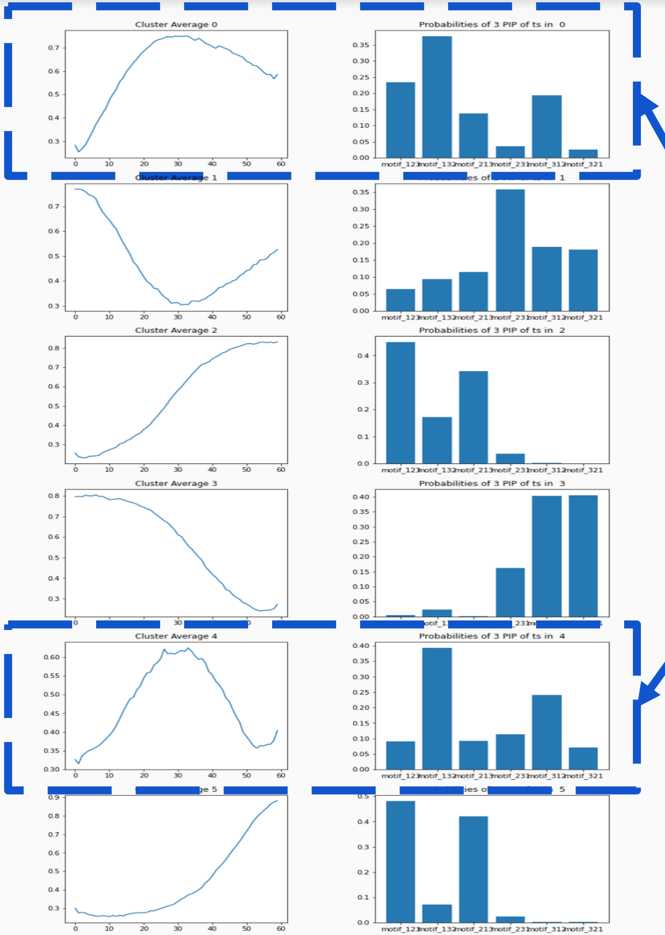
Second sliding window:
increasing motif (1,2,3)

Multiscale Evaluation: Permutation Entropy



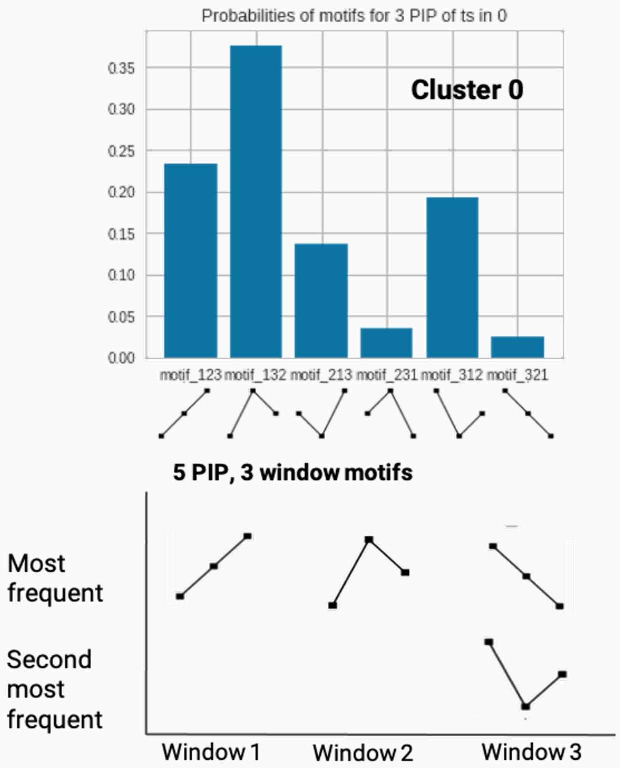
Third sliding window:
decreasing motif
(2,3,1)

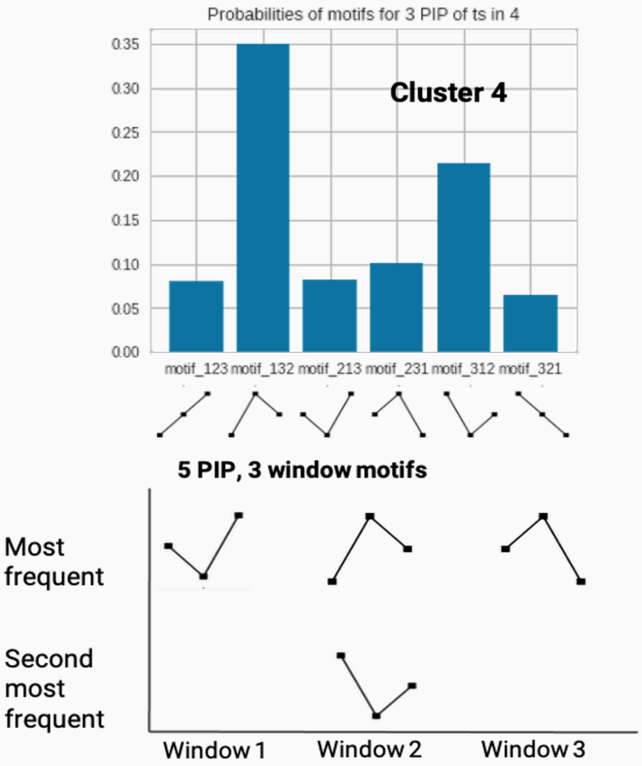
Multiscale Evaluation: Cluster: Skipped-Values



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

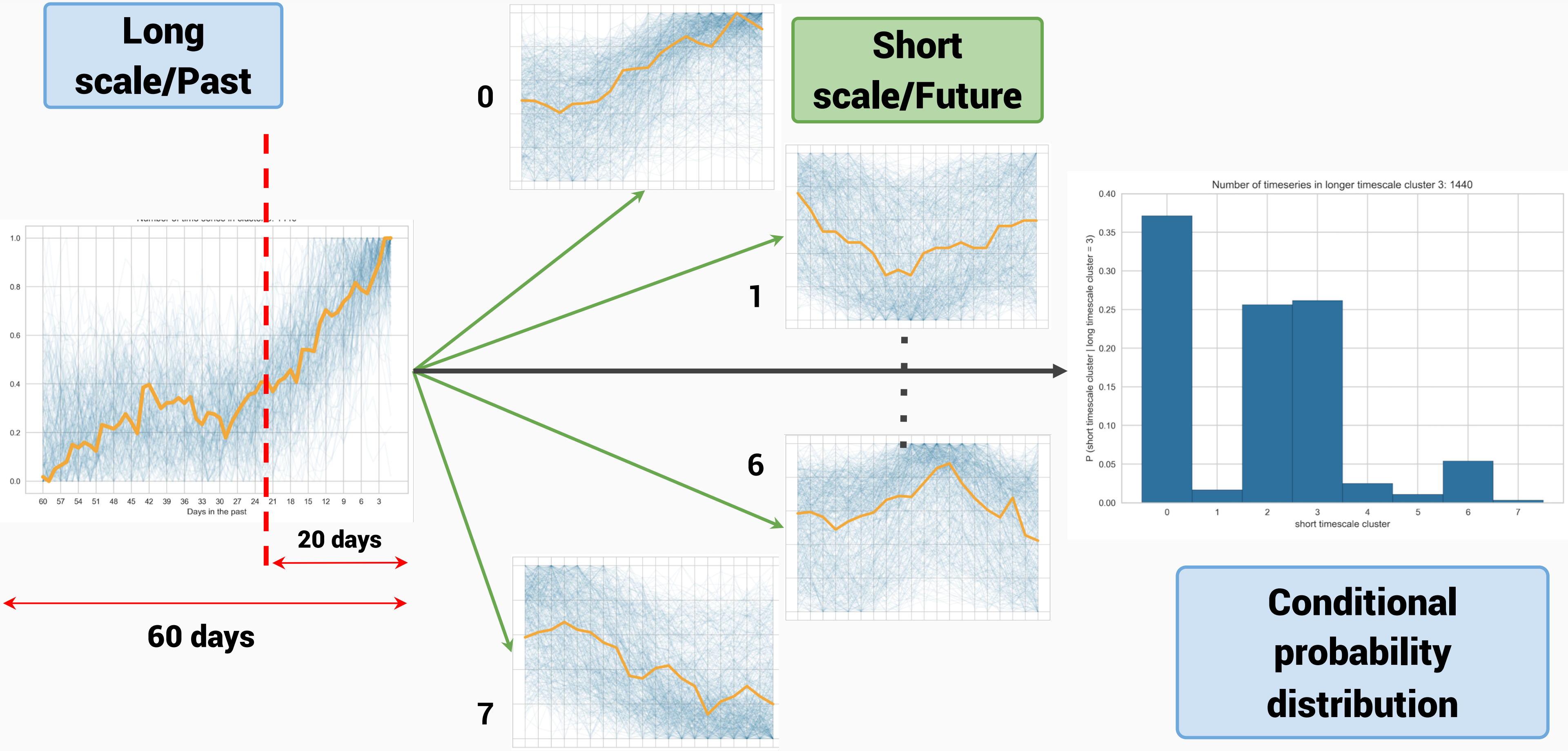
- insights on predictive power of clusters
- how short term (or future) trends relate to long term (or past) trends

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( cluster_2nd_set | cluster_1st_set = c )`

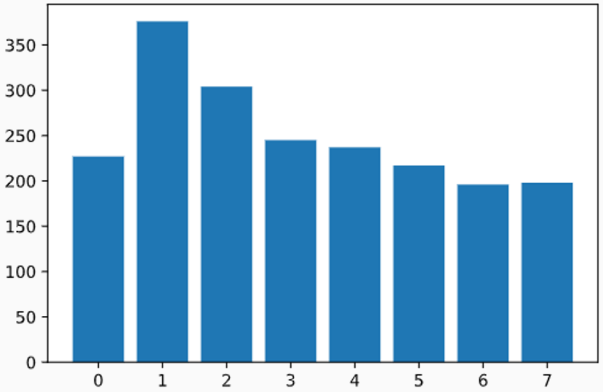
Conditional Probability Analyses

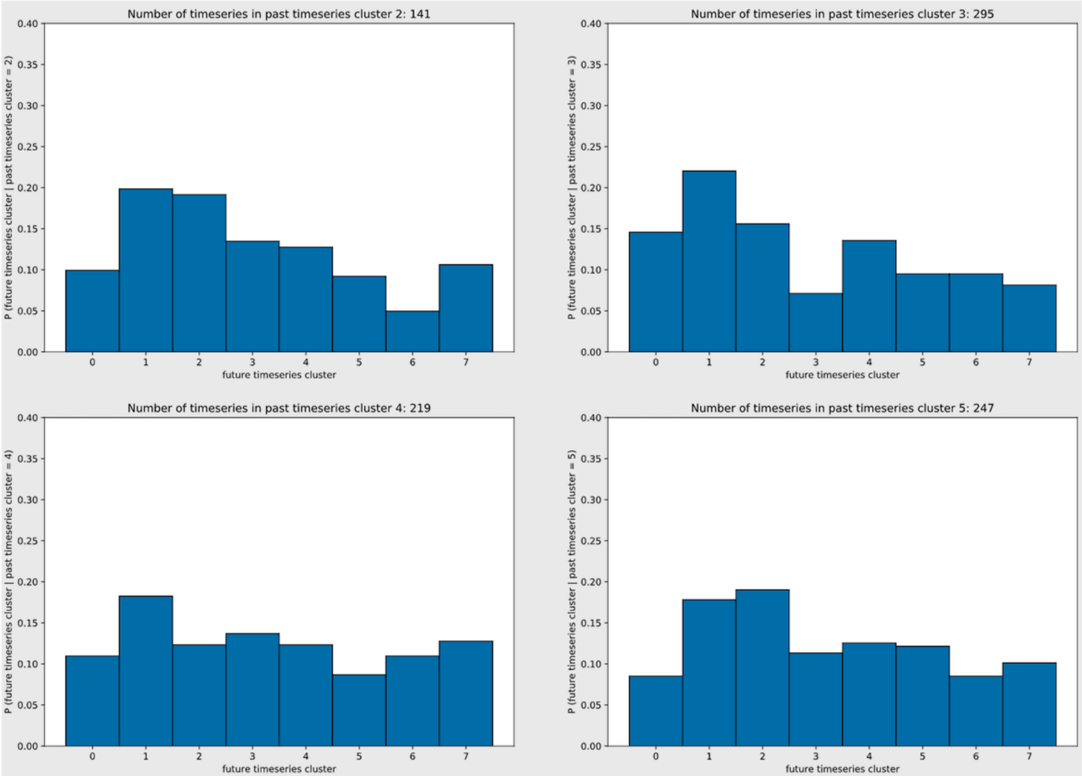


Conditional Probability Analyses: k-means

Past time-series clusters (4 shown here)

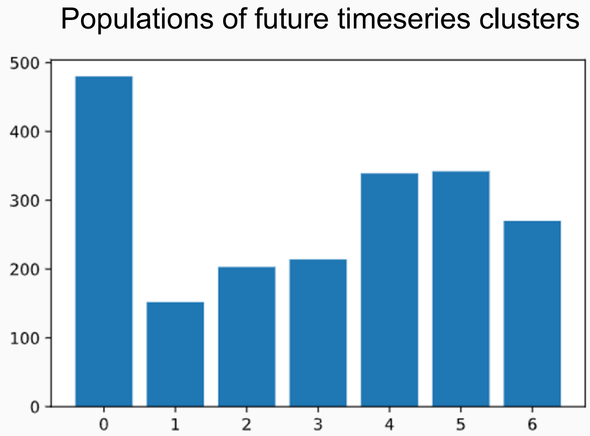
Populations of future timeseries clusters

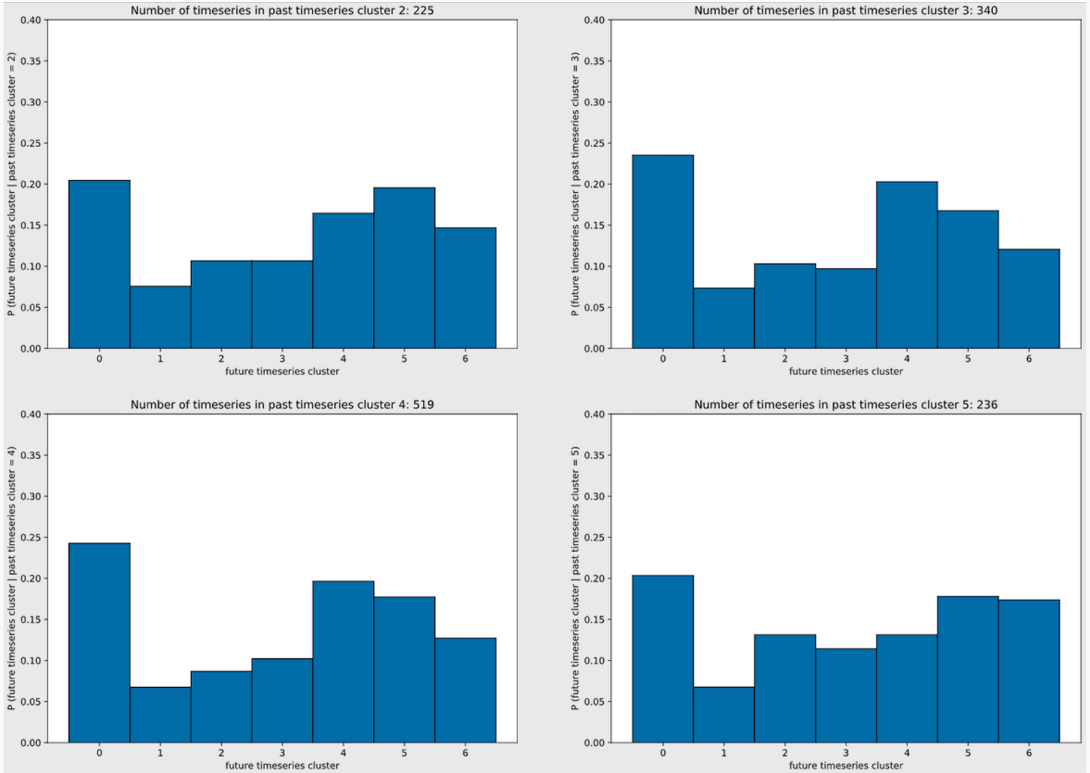




Conditional Probability Analyses: Autoencoders

Past time-series clusters (4 shown here)





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 = number of clusters

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

Pipeline	40/40	50/30	60/20
k-means	0.978	0.975	0.986
Autoencoders	0.962	0.949	0.978

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!