Investigation of the Latent Space of Stock Market Patterns with Genetic Programming

Sungjoo Ha, Sangyeop Lee, Byung-Ro Moon

July 18th, 2018

GECCO 2018

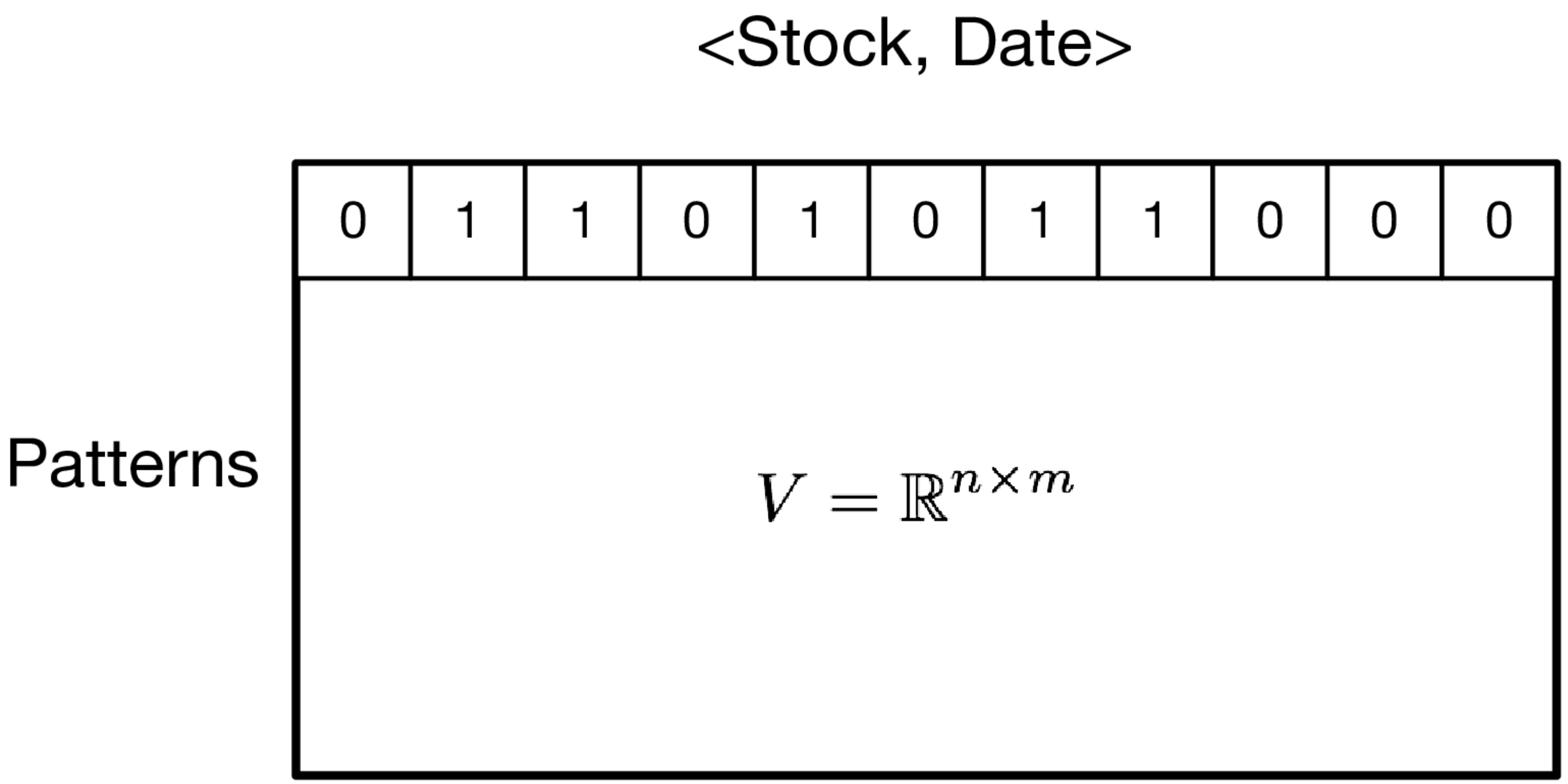
Motivation

- **Human traders** trading in a stock market
  - May have hard time **explaining** themselves
- Analyzing the behavior of different experts for new insights
  - Different approaches may share **common qualities**
  - Inspect the **latent space**

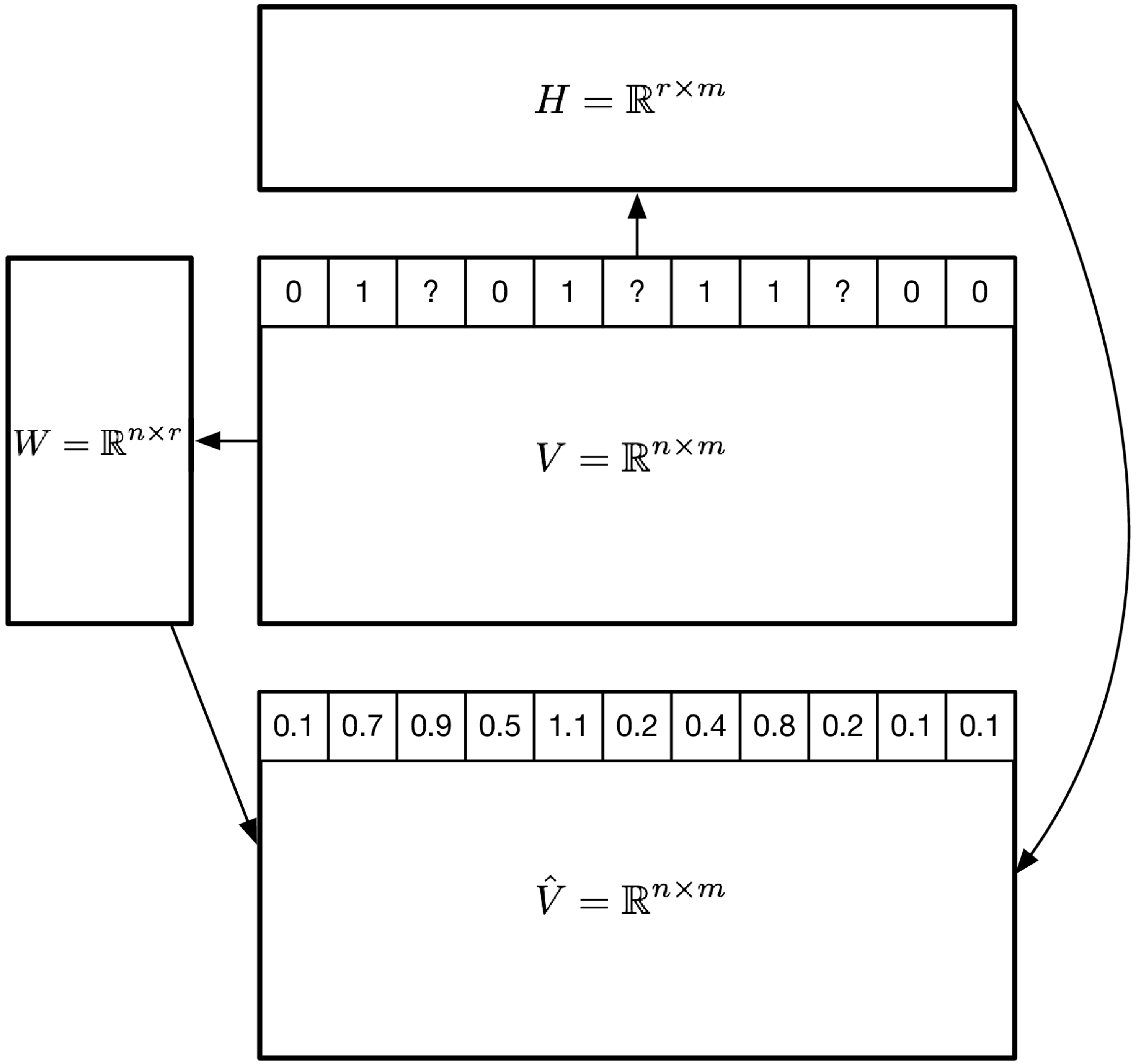
Pattern and Relation

- A **pattern** is defined to be a classifier that yields a true/false result for a given $\langle$ stock, date$\rangle$ tuple.
  - A human expert predicting the next price movement is a pattern
  - A neural network predicting the next price movement is a pattern
- A **black box pattern** is a pattern that we do not necessarily know the decision rules it applies.
- Given multiple (possibly) black box patterns and their results, we can create a **relation matrix** whose rows and columns correspond to patterns and $\langle$stock, date$\rangle$ tuples.

Relation Matrix



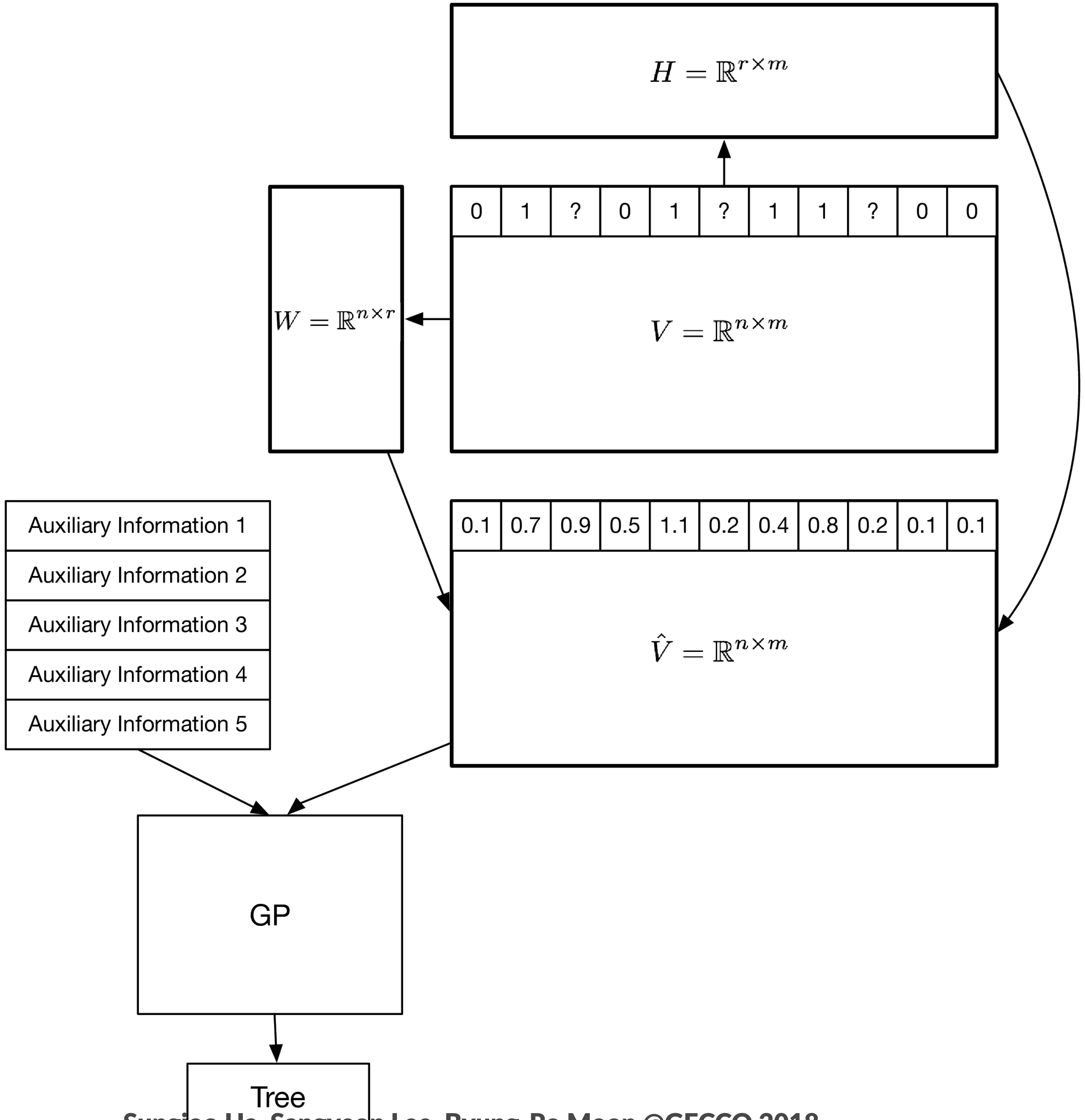
- Distinct objects can form a **relation matrix**
  - Users buying certain items
  - Pattern match results on different assets
- The collective relationship often has **underlying structure** that is difficult to observe directly



Low-Rank Matrix Factorization

$$\mathbf{V} \approx \mathbf{W} \cdot \mathbf{H}$$

- Uncover **latent structure** by low-rank matrix factorization
  - $\mathbf{V}$: Relation matrix between pattern and <stock, date> pair
  - $\mathbf{W}$: Latent pattern matrix
  - $\mathbf{H}$: Latent <stock, date> matrix
- If preference of patterns and characteristics of $\langle$stock, date$\rangle$ **share common qualities**, they will be close to one another in the latent space



Factorization and Reconstruction Approach

- $\mathbf{V}$: <pattern, (stock, date)> matrix
  - Decisions to buy a certain stock on a certain day
- Perform **matrix factorization** on V to uncover the **latent space of patterns**
- Use GP to mimic the behavior of patterns using **auxiliary information**

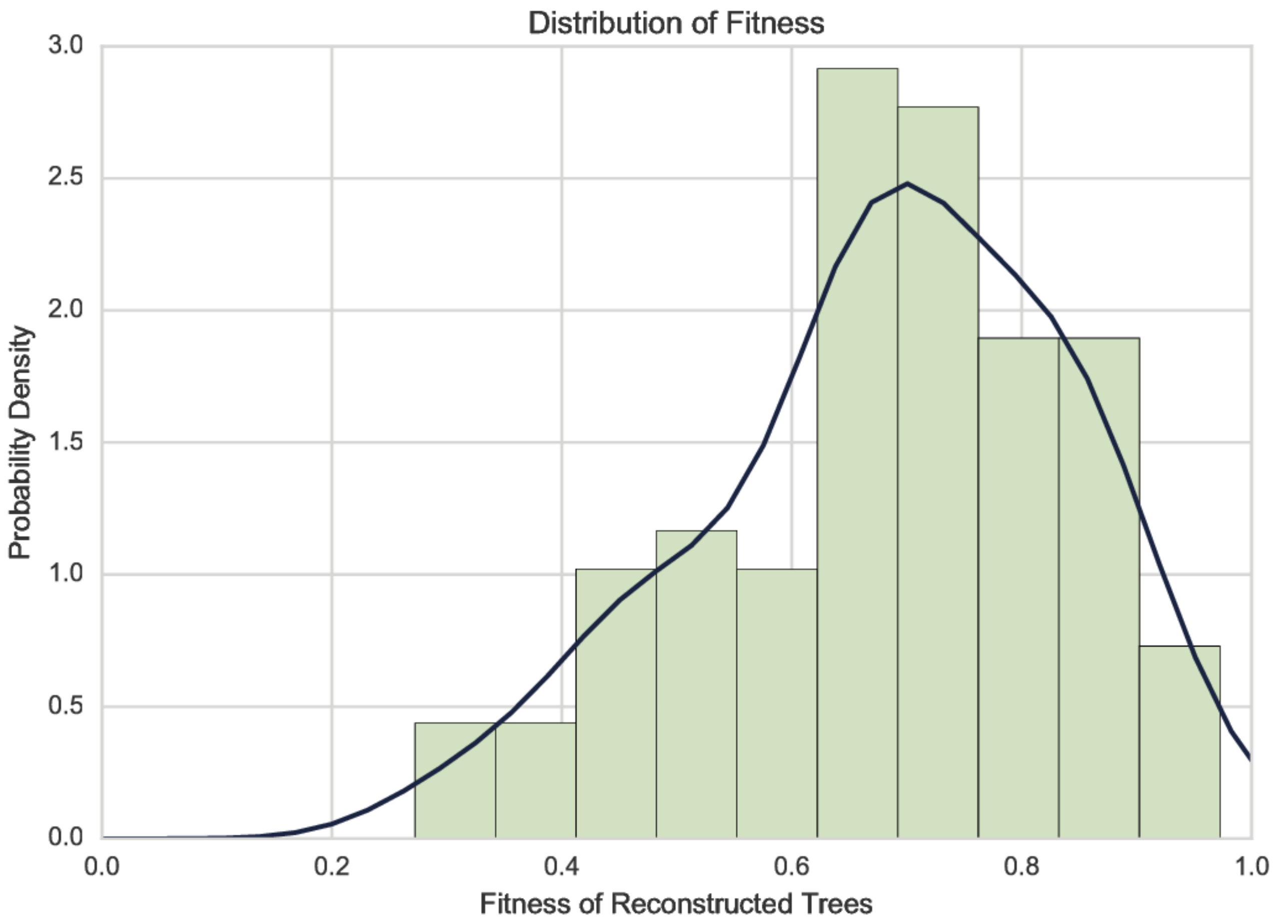
Auxiliary Information

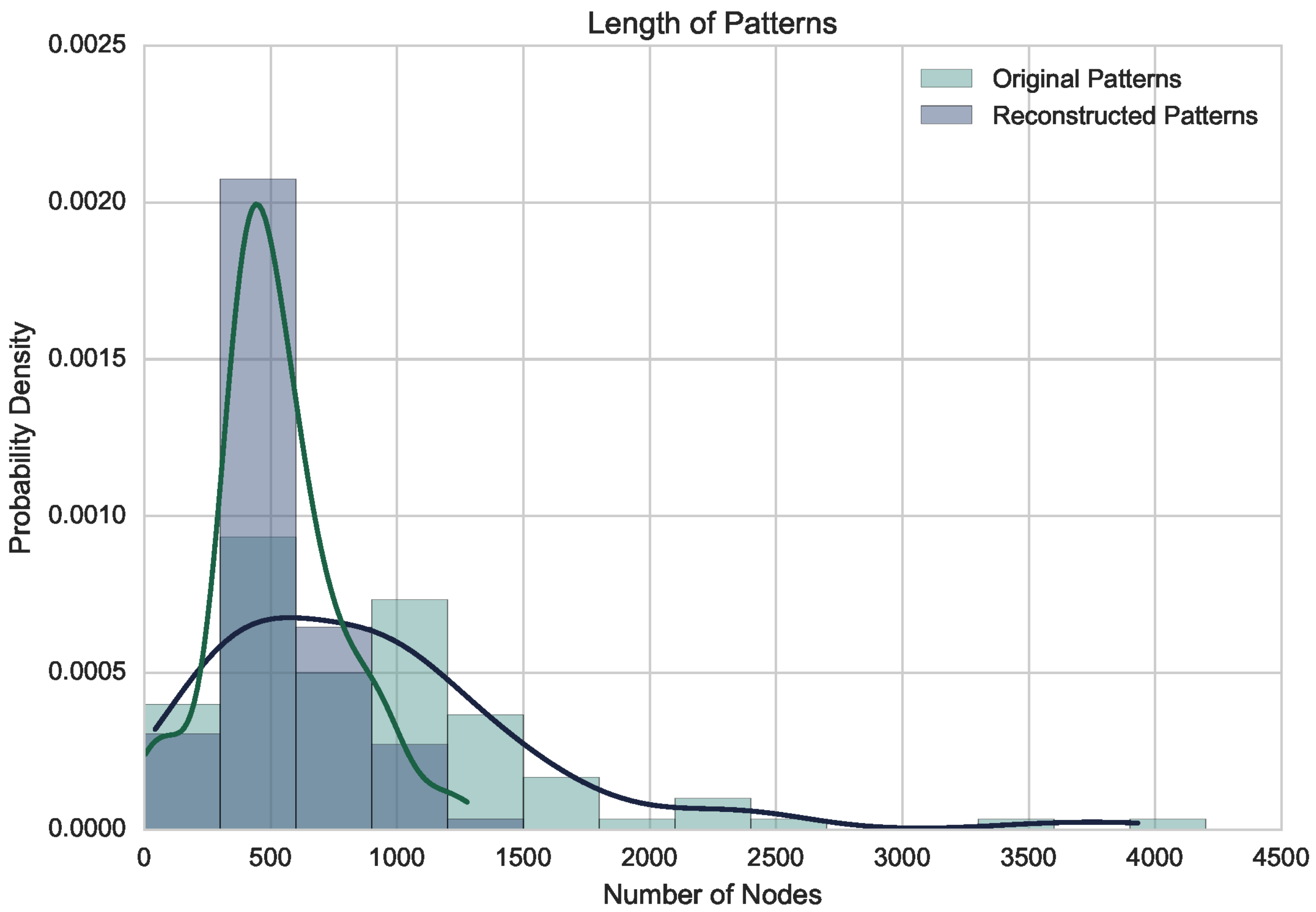
- Opening Price
- Closing Price
- Highest Price
- Lowest Price
- n-day Highest Price
- n-day Lowest Price
- n-day Moving Average
- Upper Bollinger Band
- Lower Bollinger Band

Experiments

- Use **100 patterns** mined for expected gain and frequency
  - Allows us to **study the difference** between original patterns and reconstructed patterns
- **Remove 50%** of the entries in the relation matrix
- Apply the proposed method
  - Study the behavior of **reconstructed patterns**
  - Study the **latent space** induced by matrix factorization

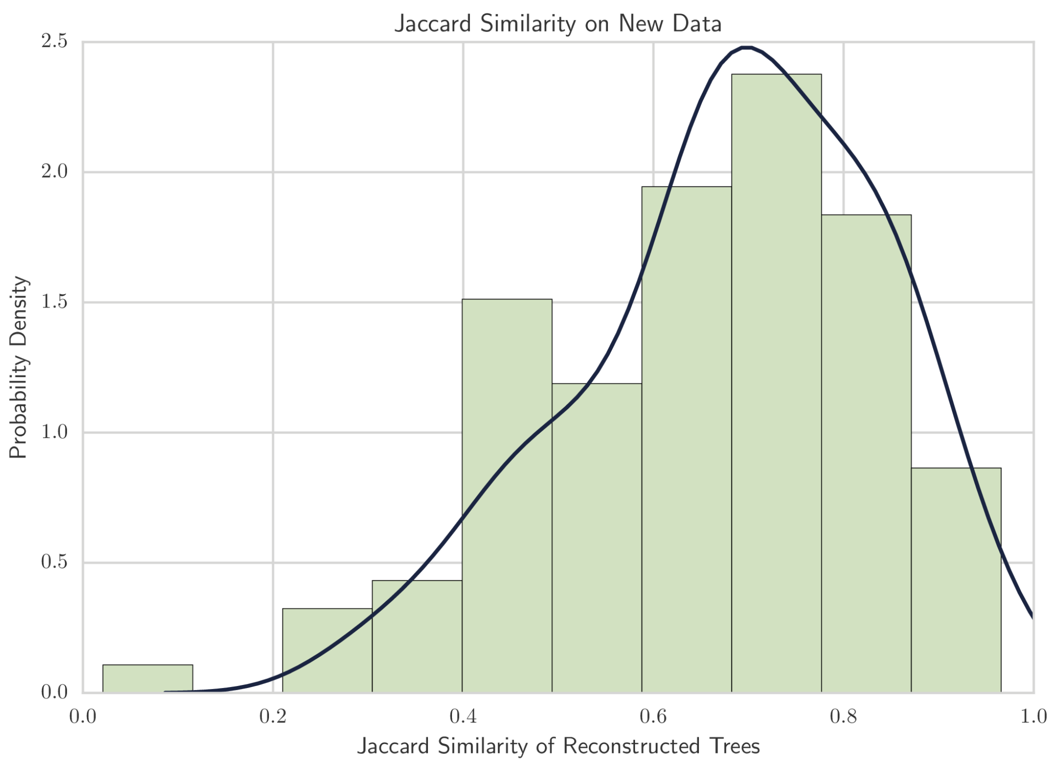
Characteristic of Reconstructed Tree

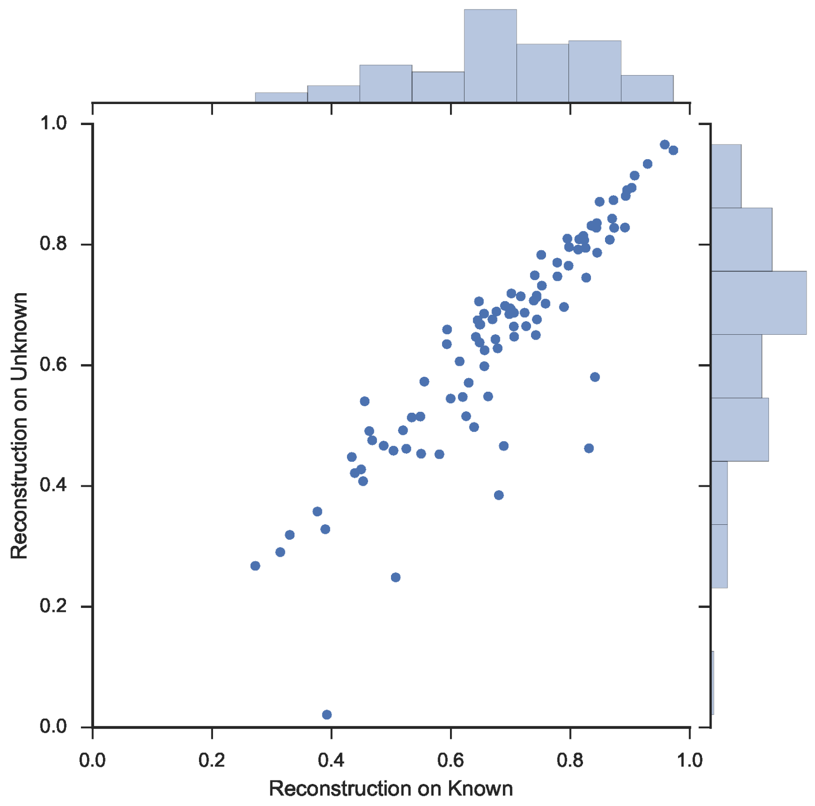




- The reconstructions are similar in semantics but are different syntactically from the originals

Recovering Original Pattern



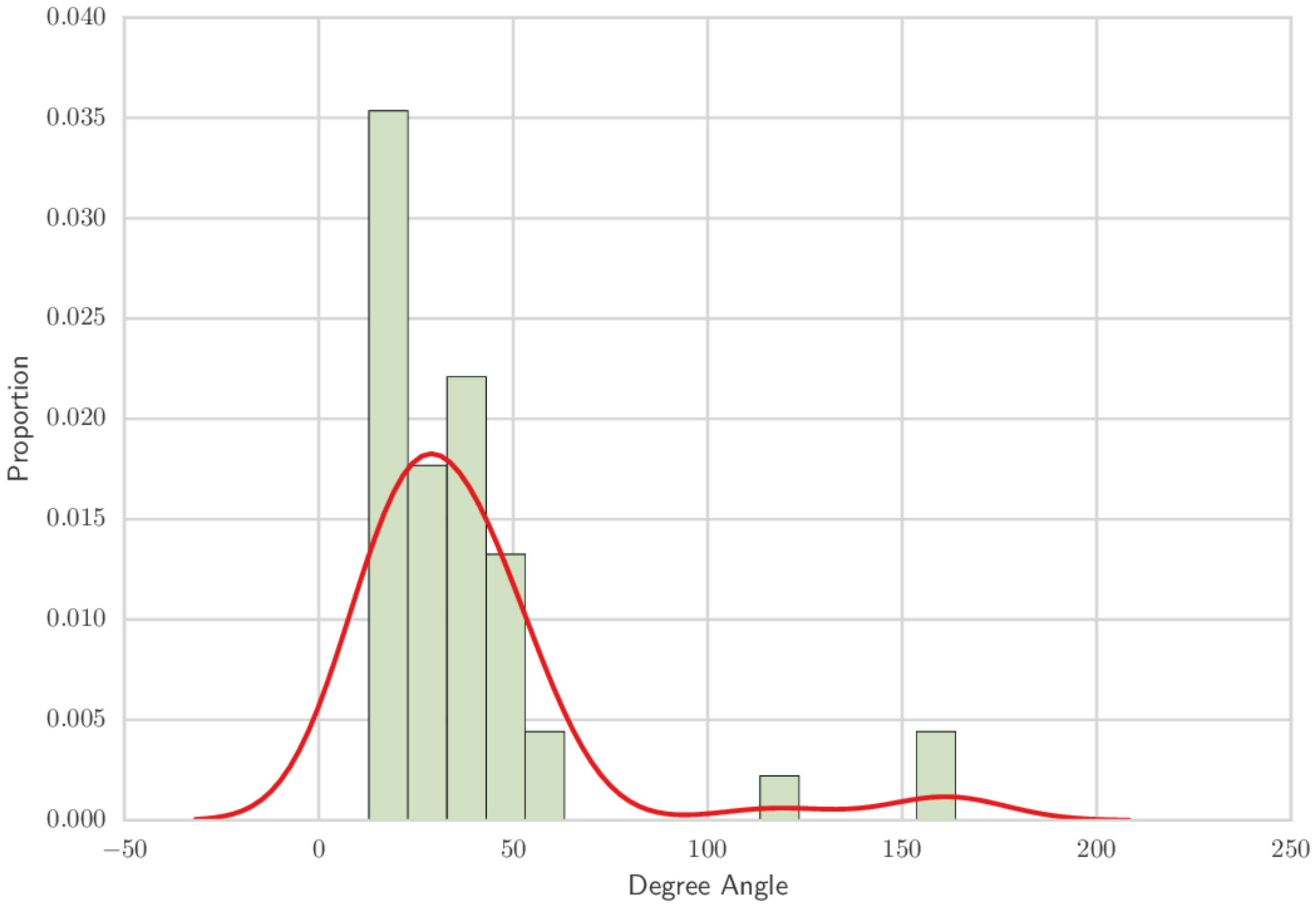


- If we can mimic the original behavior well enough, then it also **generalizes**
- Can be applied to extract **decision rules** from expert decisions
    - Exploits how others make decisions

Inspecting the Latent Space

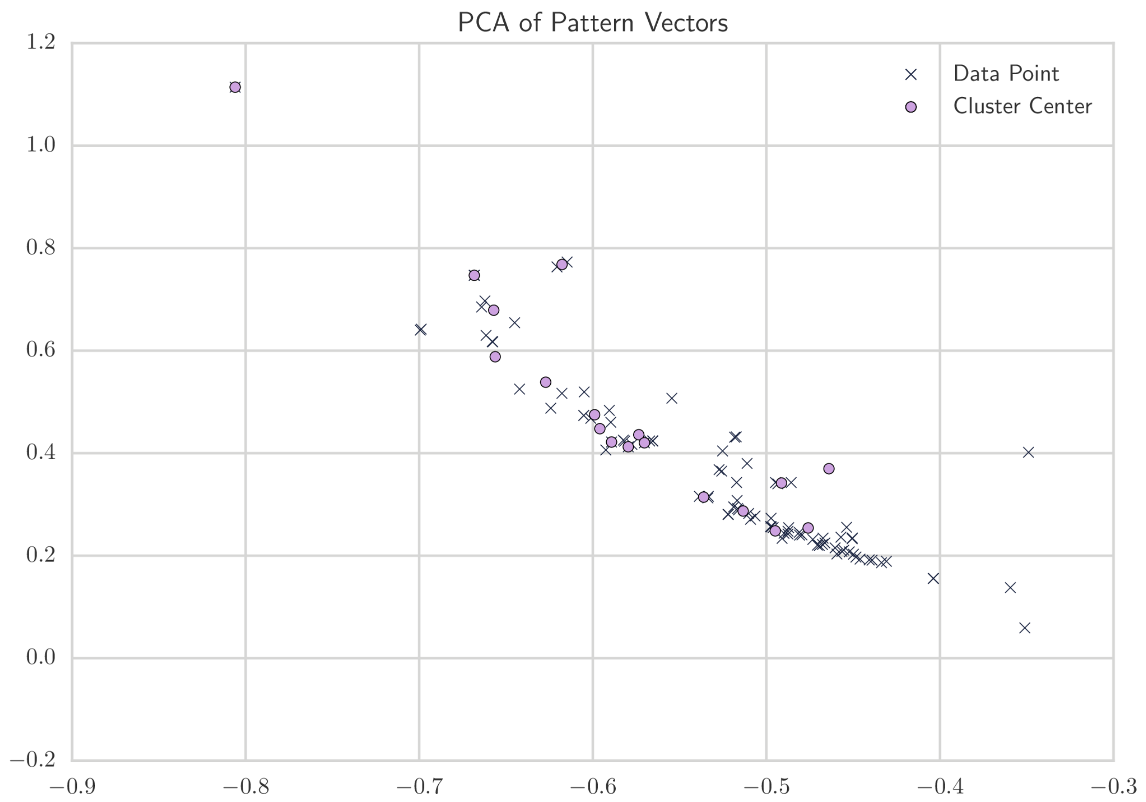
- We can **transform** a tree into a vector representation and vice versa
- We can study the **relationship** between two different representations
  - Study how **geometric crossover** is translated in latent vector space
  - **Clustering**
  - Analyze **axis vectors**

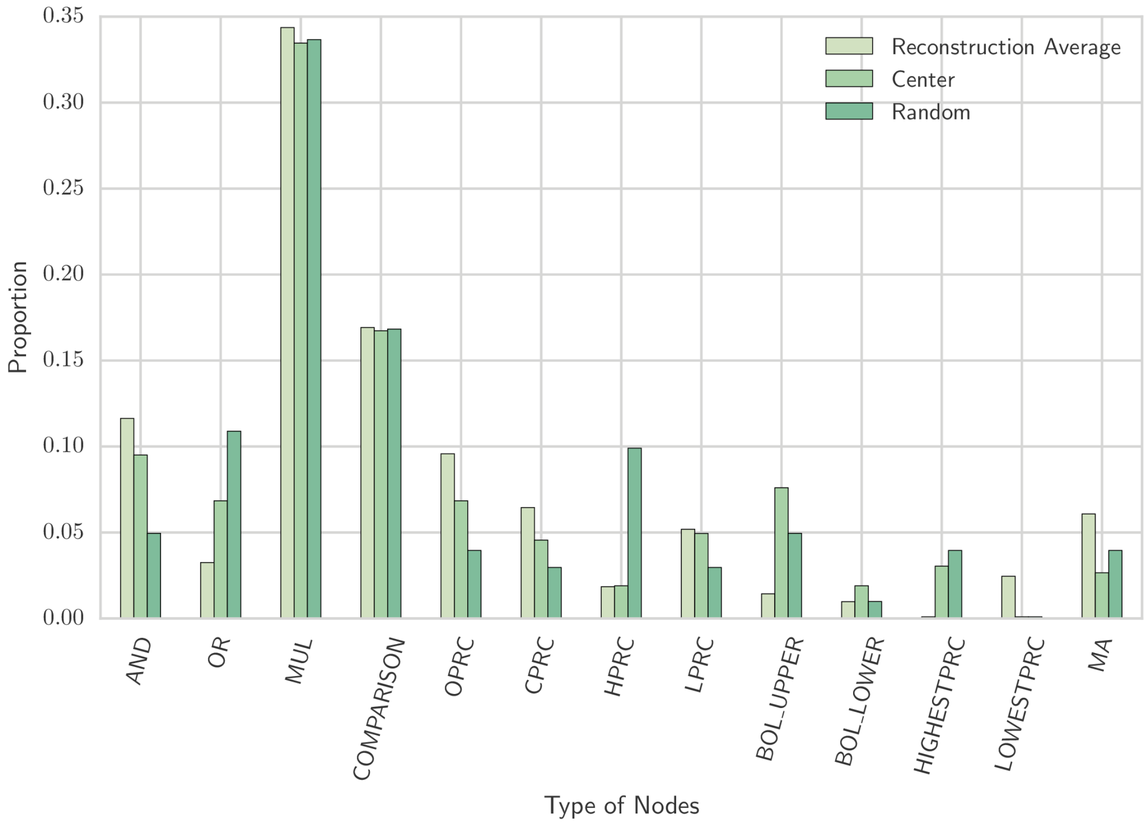
Tree Space and Latent Space



- Geometric operation in tree space
    - $T_0 = (T_1 \wedge T_r) \vee (T_2 \wedge \neg T_r)$
    - V_0 corresponding to T_0
- Corresponding latent space representation
    - $w_0 = \arg\min_w (w^T H) - V$
- 80% of the time, w_0 is inside the hypersphere created by two parents
- Usually lie within a narrow cone

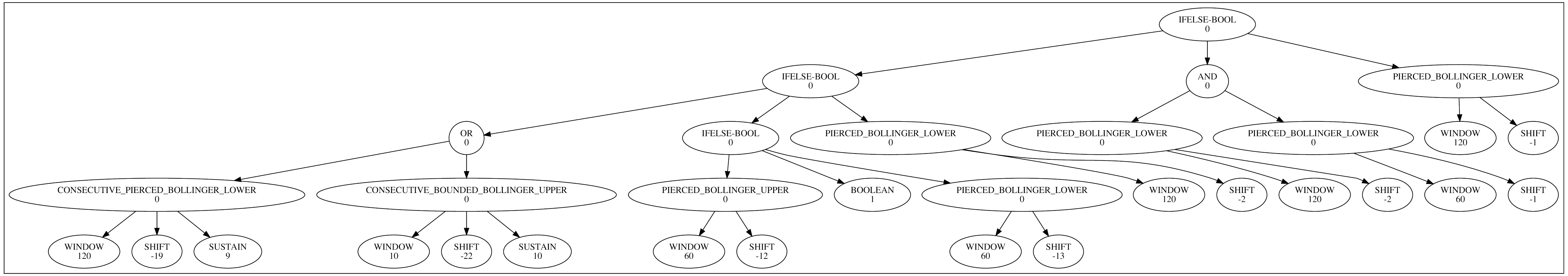
Clustering





- Clustering for **measuring diversity** of a set of patterns
- Perform clustering to obtain a **generic representation** of a pattern

Reconstructed Pattern



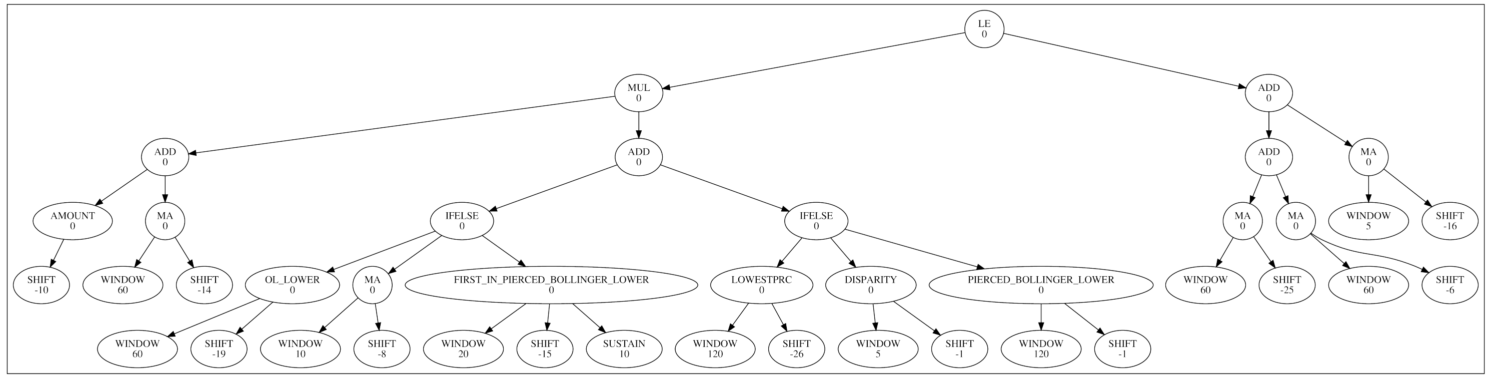
- Reconstructed pattern could be used for automation
- **Understandable**
- But is a chore to interpret every time a new pattern is included

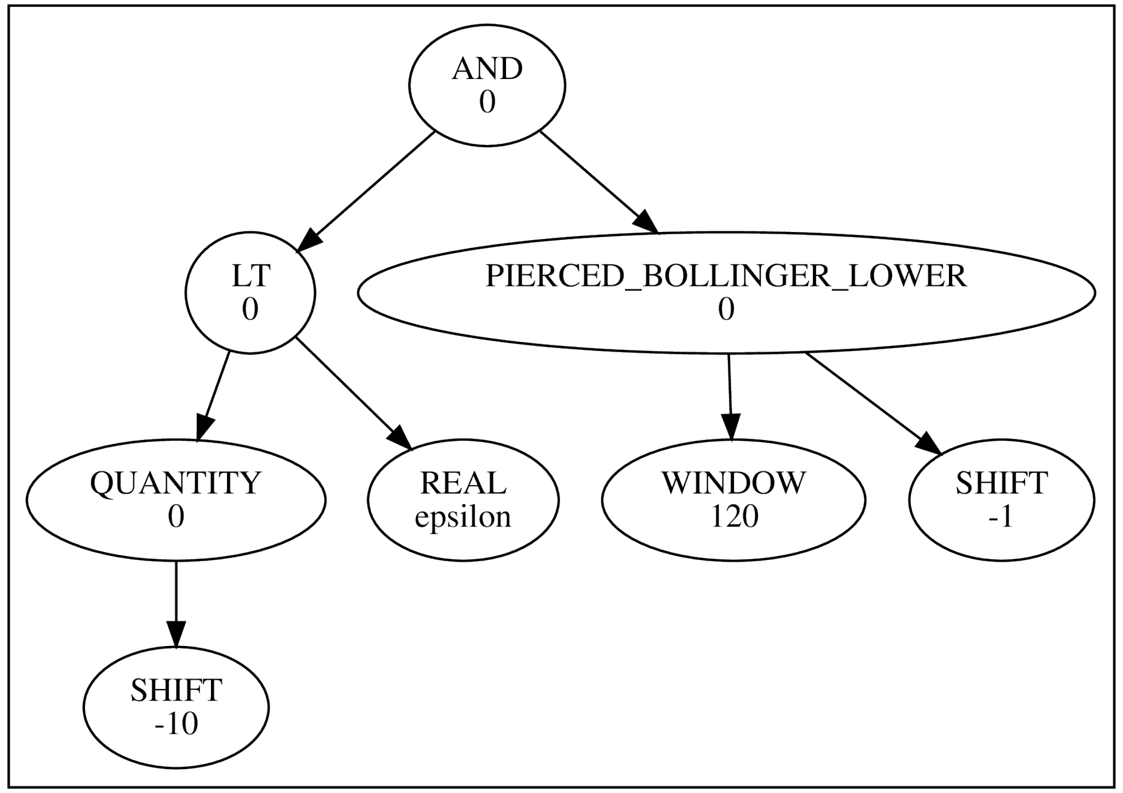
Latent Space Analysis

2.16 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1.92 1.54

- Vector representation of the previous pattern
  - **Sparse** representation due to NMF
    - 3/30 non-zero entries
    - 3 axes with > 1.0 (axis 1, 29, and 30)
  - **Additive** representation due to non-negativity

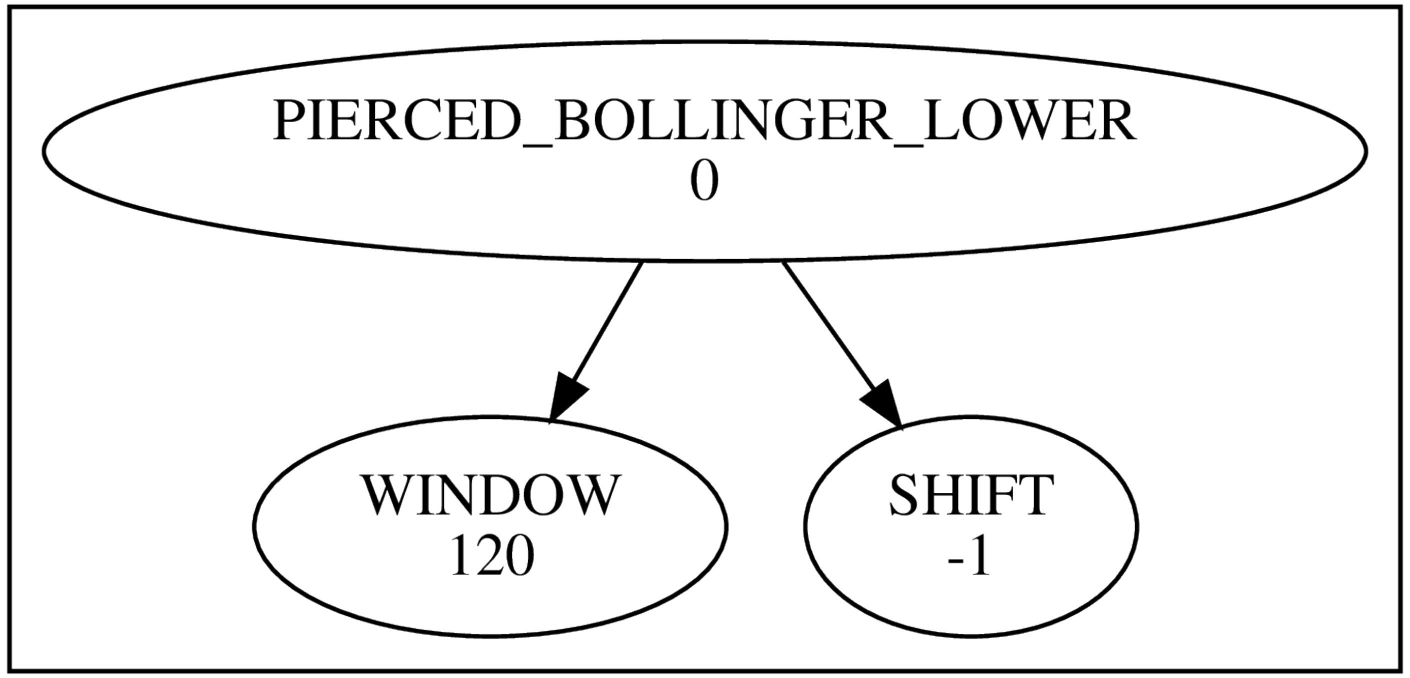
Tree Interpretation

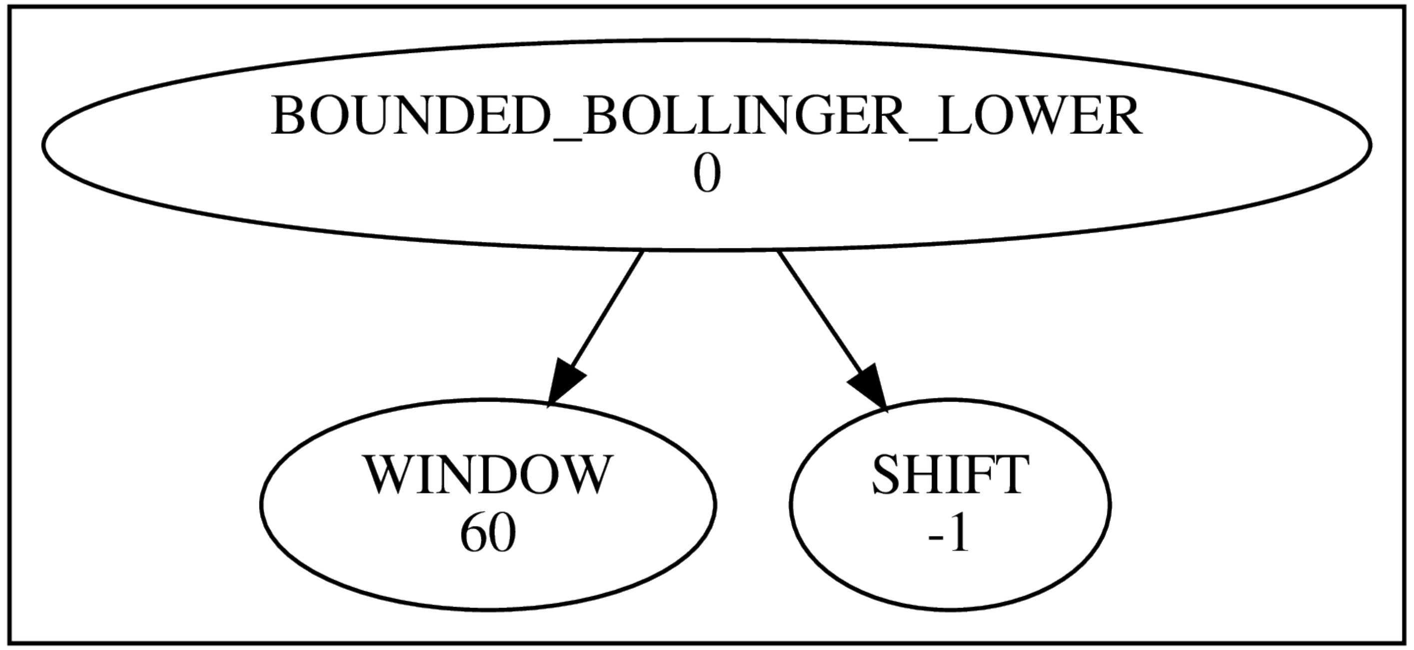




- Approximate **tree reduction** via
  - Conditional subtree plots
  - Path usage statistics

Latent Space Analysis





- Axis 1 and 30 tells us that in 120 time-step view, price has plunged
- Axis 29 tells us that in 60 time-step perspective, price hasn't plunged

Conclusion

- Using **GP** and **low-rank matrix factorization** we can
  - Create **tree representation** of black box patterns which can be used for automatic decision making
  - **Inspect the latent space** induced by low-rank matrix factorization