TIME-SERIES ANALYSIS USING TIME-DELAY EMBEDDINGS

Jordan Frank (jordan.frank@cs.mcgill.ca)

Reasoning and Learning Lab - McGill University

Joint work with **Doina Precup**, **Shie Mannor**, and **Joelle Pineau**

AAAI 2011 Spring Symposium on Computational Physiology - March 22, 2011

MOTIVATION

- Many applications consider data from sensors measuring complex physiological systems (wearable sensors, ECG, EKG, etc.).

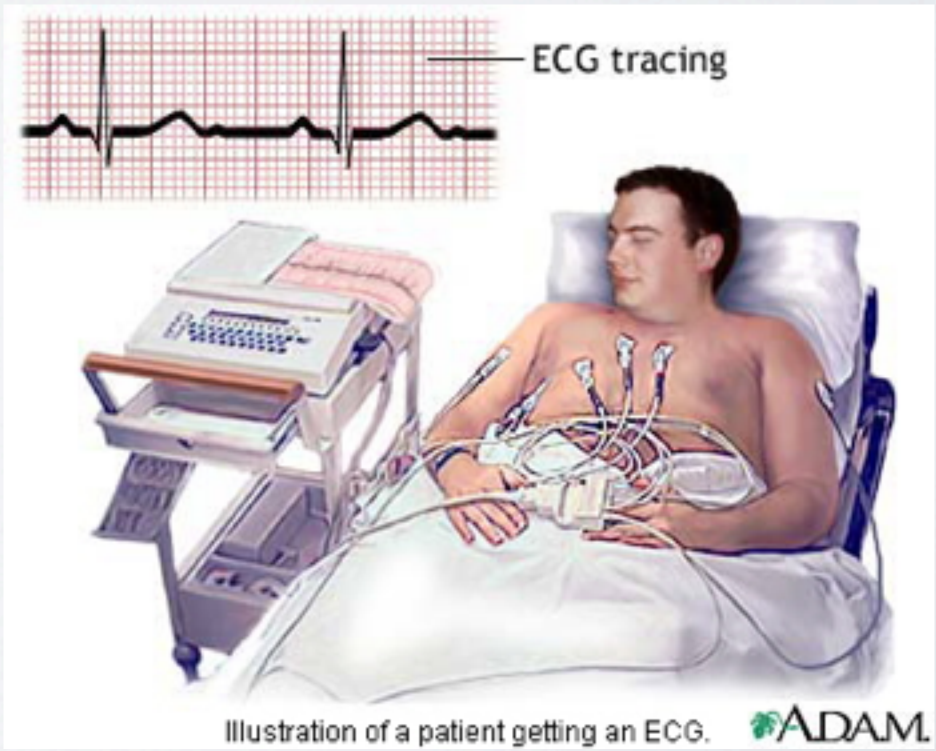
- This talk: Feature extraction that respects the inherent nonlinearities in the systems being measured.

OUTLINE

1. Overview of our approach:
   - Feature extraction from sensor data.
2. Evaluation on multiple data sets:
   - Classification of activities from wearable accelerometer and barometric pressure sensors.
   - Classification of individuals based on gait patterns collected by accelerometer sensors in mobile phones.
   - Clustering traces of accelerometer and ECG recording data.

PROBLEM FORMULATION

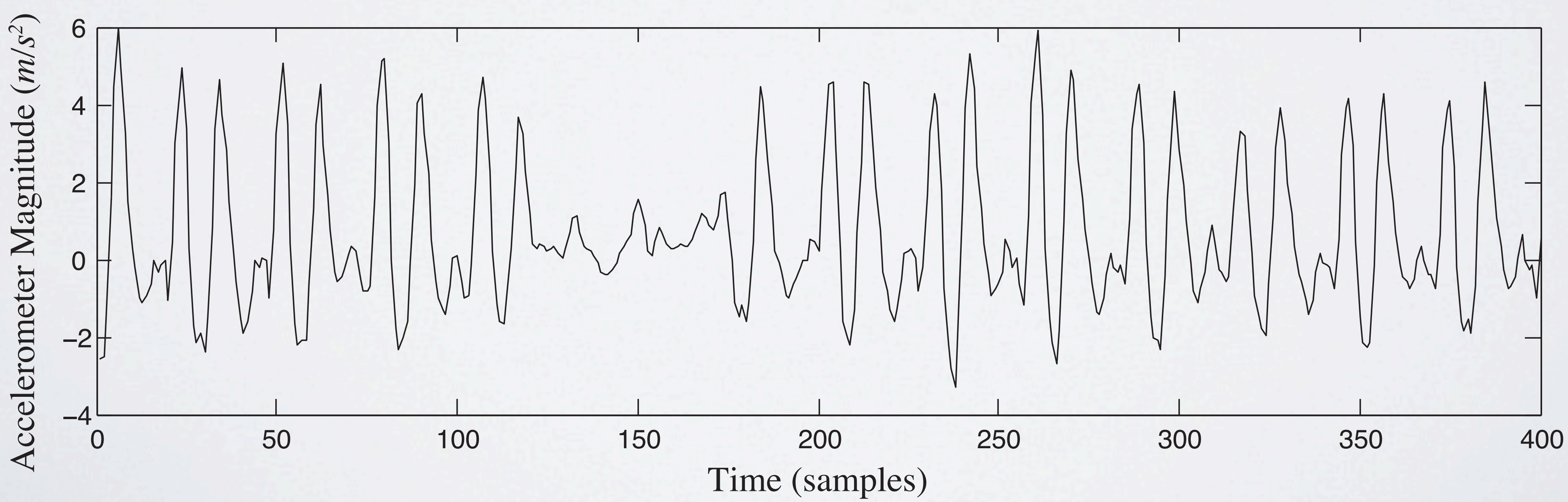
- Input: Noisy univariate observations of some high-dimensional nonlinear dynamical system.
- Output: Good features.
- Requirements: Data efficient, computationally efficient, memory efficient.



EXISTING APPROACHES

- Signal Processing: Extract lots of (linear) features from sensor data, train powerful machine learning algorithms using these features.

- Problems: Computationally expensive, requires lots of data, underlying systems certainly aren't linear or stationary. Lots of noise!



OUR APPROACH

Steps:

1. Build Models: Project segments of time series into a suitable and convenient space that preserves information about the underlying dynamical system.

2. Extract Features: Given a set of models and a segment of time series, produce informative features from the time series.

3. Play: Classification, data visualisation and exploratory data analysis.

MODELING

INTUITION

- Assumption: data represents **sequential observations** from the **steady state** of a **nonlinear** dynamical system.
- Time-delay embedding (TDE) is a technique for reconstructing state-space and dynamics models from univariate observations of a nonlinear dynamical system.
- Solid theoretical foundation (Takens, 1981) for noiseless observations.
- Seems like a good fit, provided we can handle noise.

SETUP

Setting:

- State of the dynamical system at time t: $x_t \in \mathbb{R}^k$
- Some attractor of interest $A \subset \mathbb{R}^k$ of dimension d
- Observation function: $s : \mathbb{R}^k \to \mathbb{R}$
- Time series: $\{s_t = s(x_t)\}$

Goal: Find a projection from the time series to some reconstruction space such that the dynamics of the underlying system are preserved

TIME-DELAY EMBEDDING

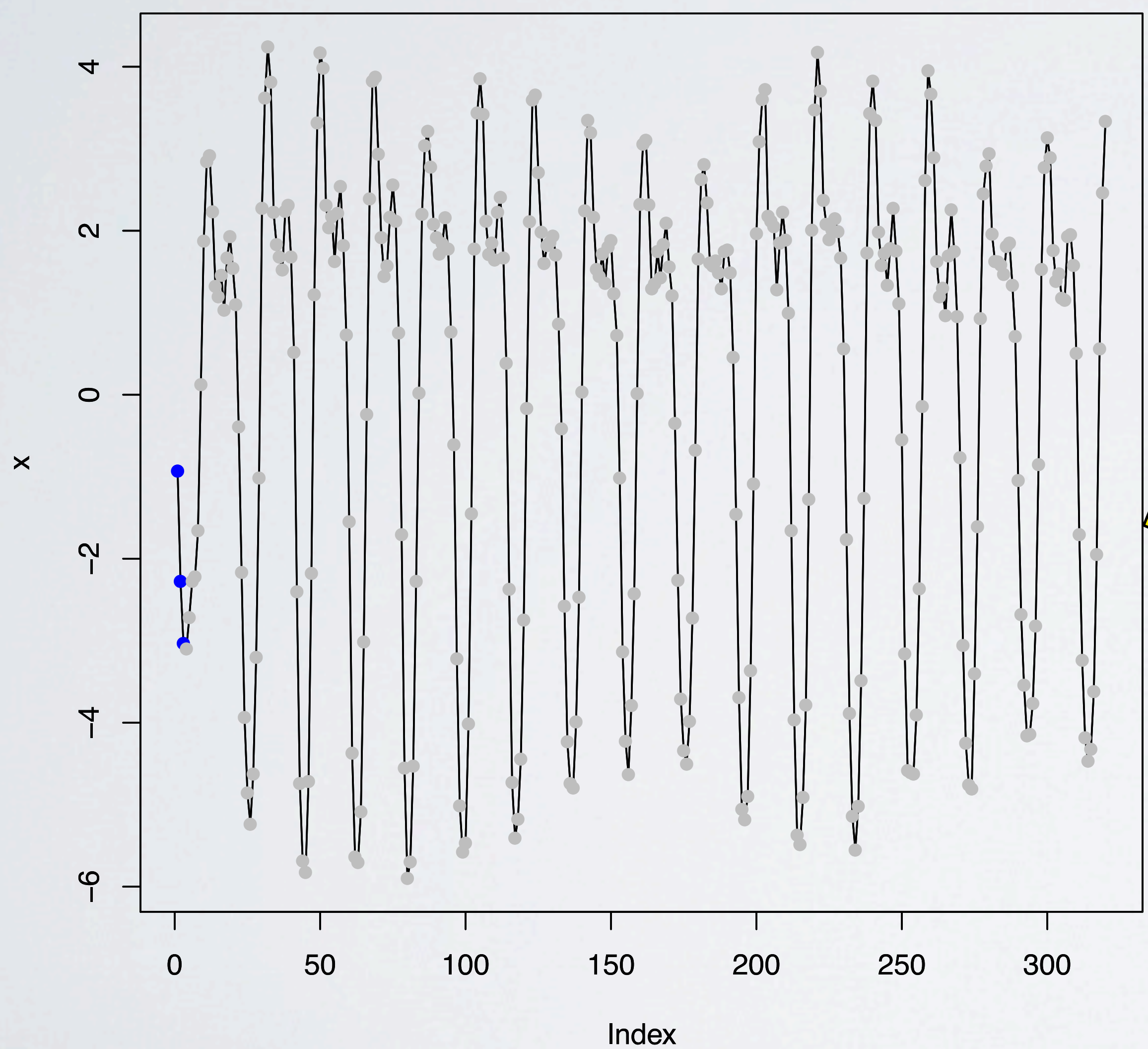
- Define the time-delay vector at time t as:

$$\mathbf{s}_t = \left(s_t, s_{t+\tau}, s_{t+2\tau}, \ldots, s_{t+(m-1)\tau}\right),$$

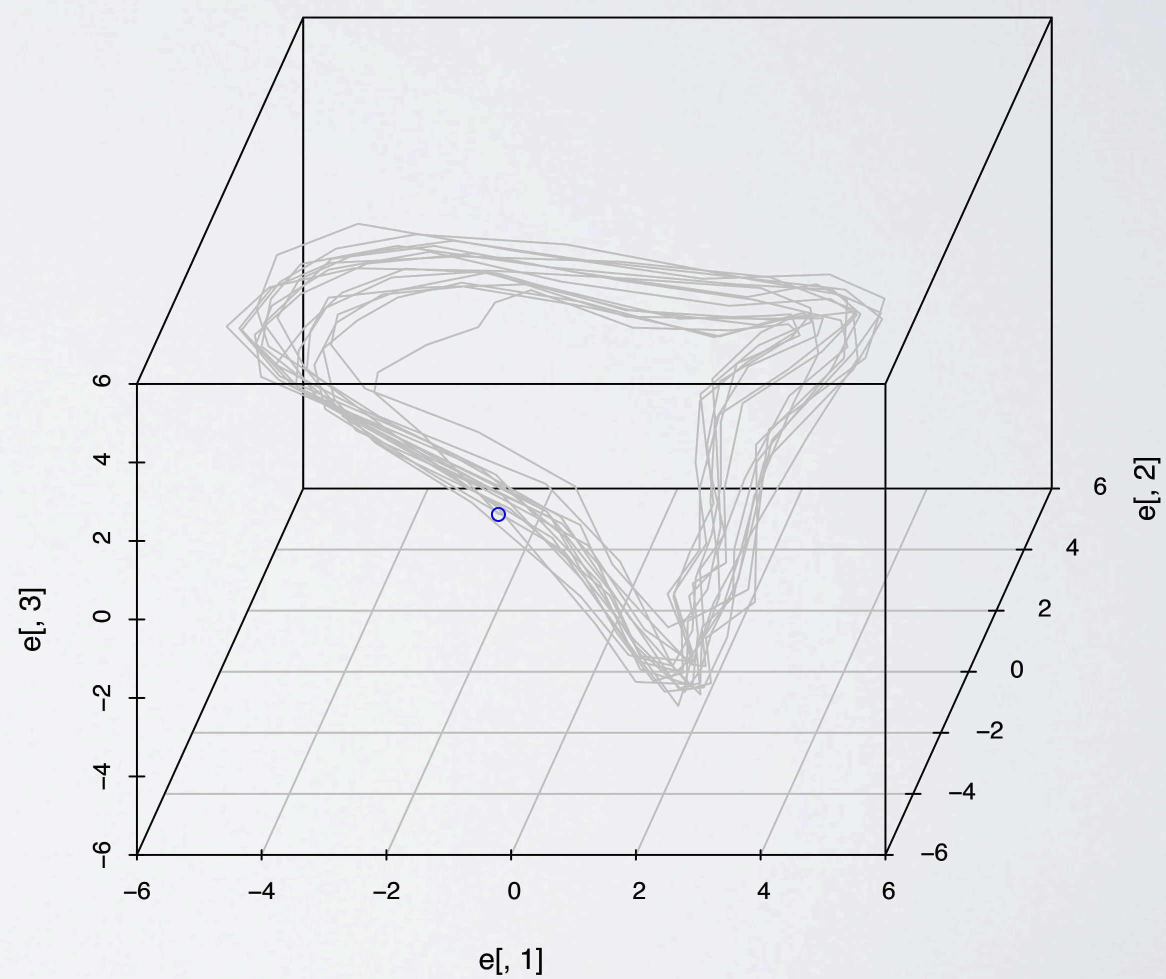
  where τ is the delay time and m is the dimension.
- This constitutes a map from $\mathbb{R}^k \to \mathbb{R}^m$.
- Call $\mathbb{R}^m$ the reconstruction space.
- An **embedding** is a map from the attractor A into reconstruction space $\mathbb{R}^m$ that is one-to-one and preserves differential information (i.e., a diffeomorphism on A).

TIME-DELAY EMBEDDING

Observations

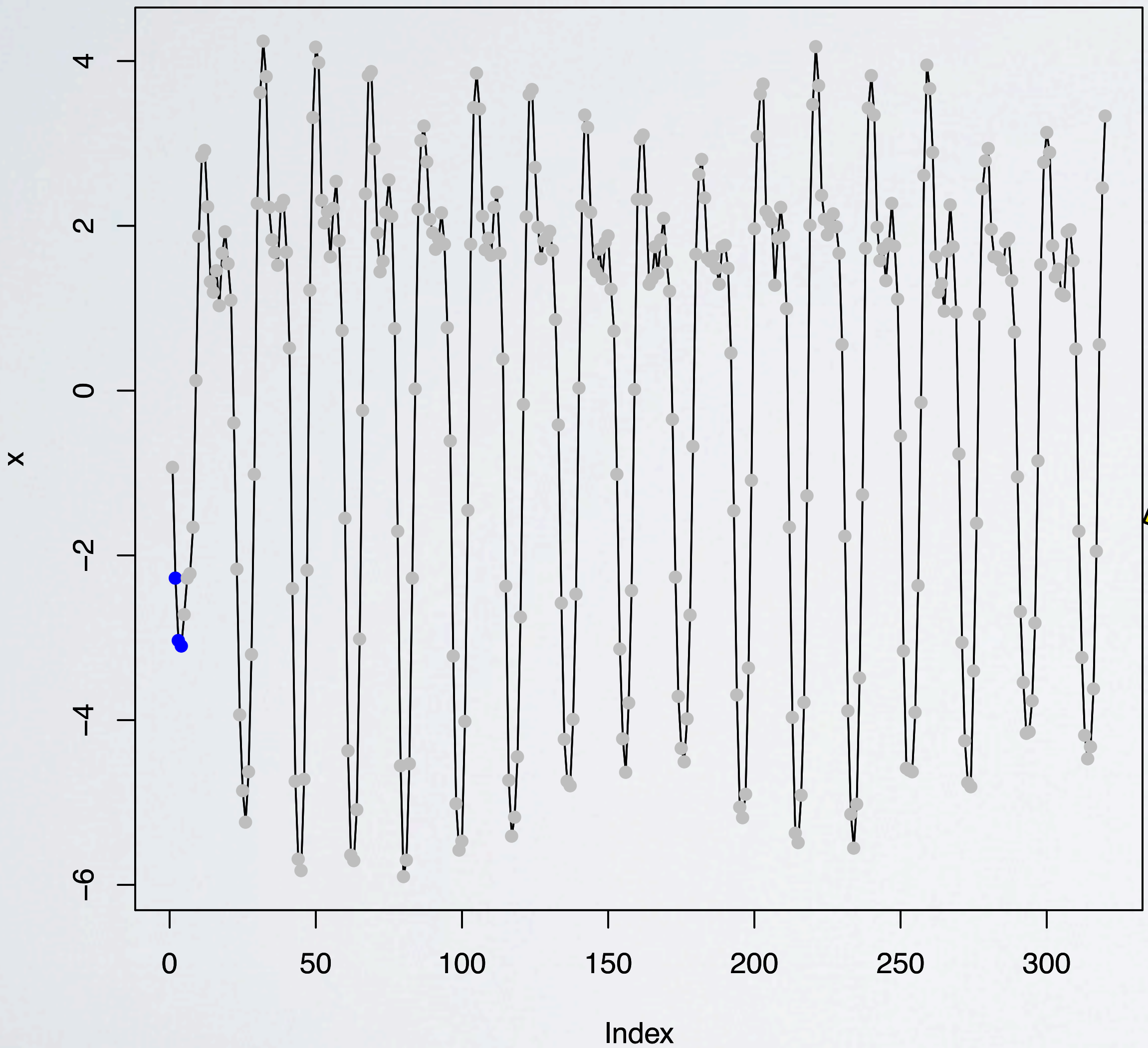


Reconstruction

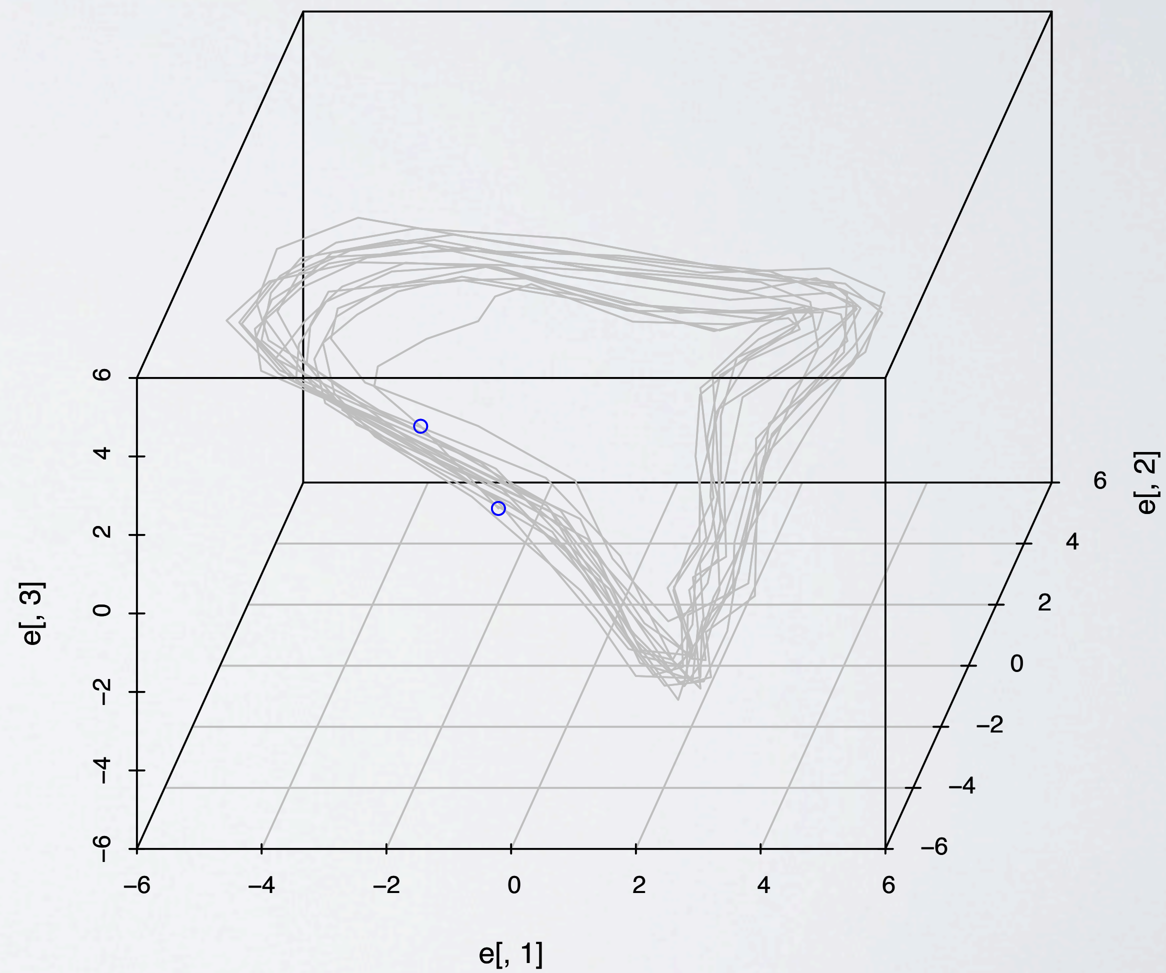


TIME-DELAY EMBEDDING

Observations

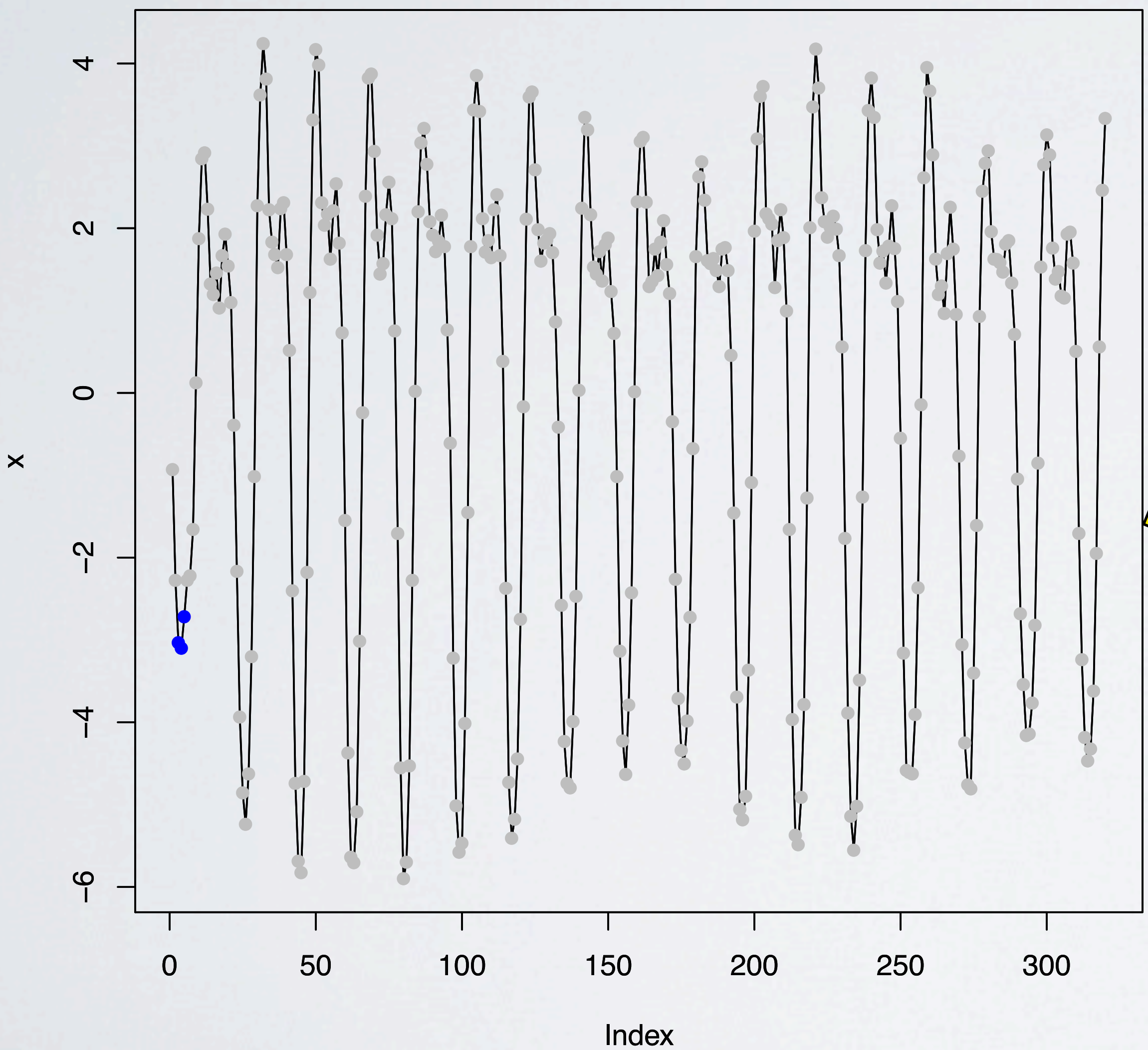


Reconstruction

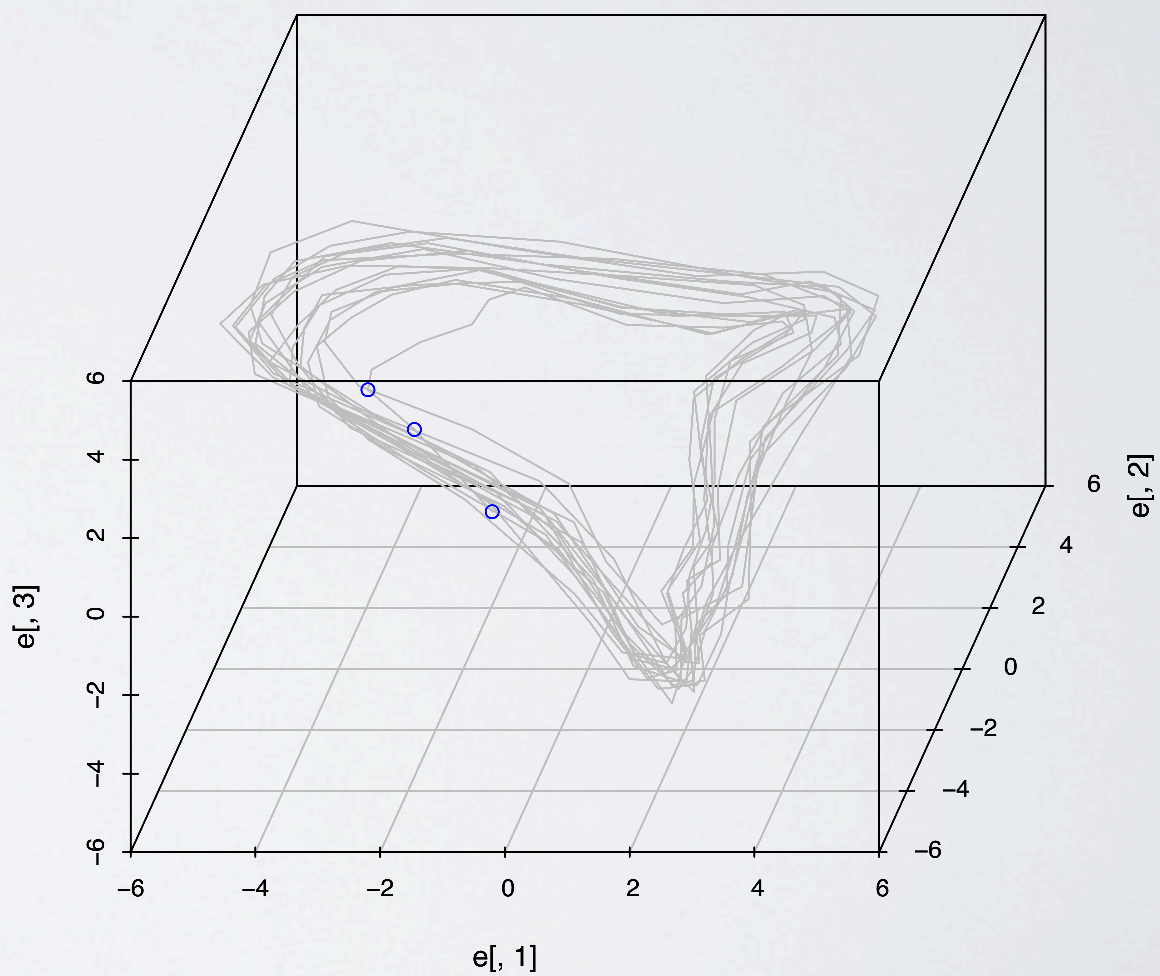


TIME-DELAY EMBEDDING

Observations

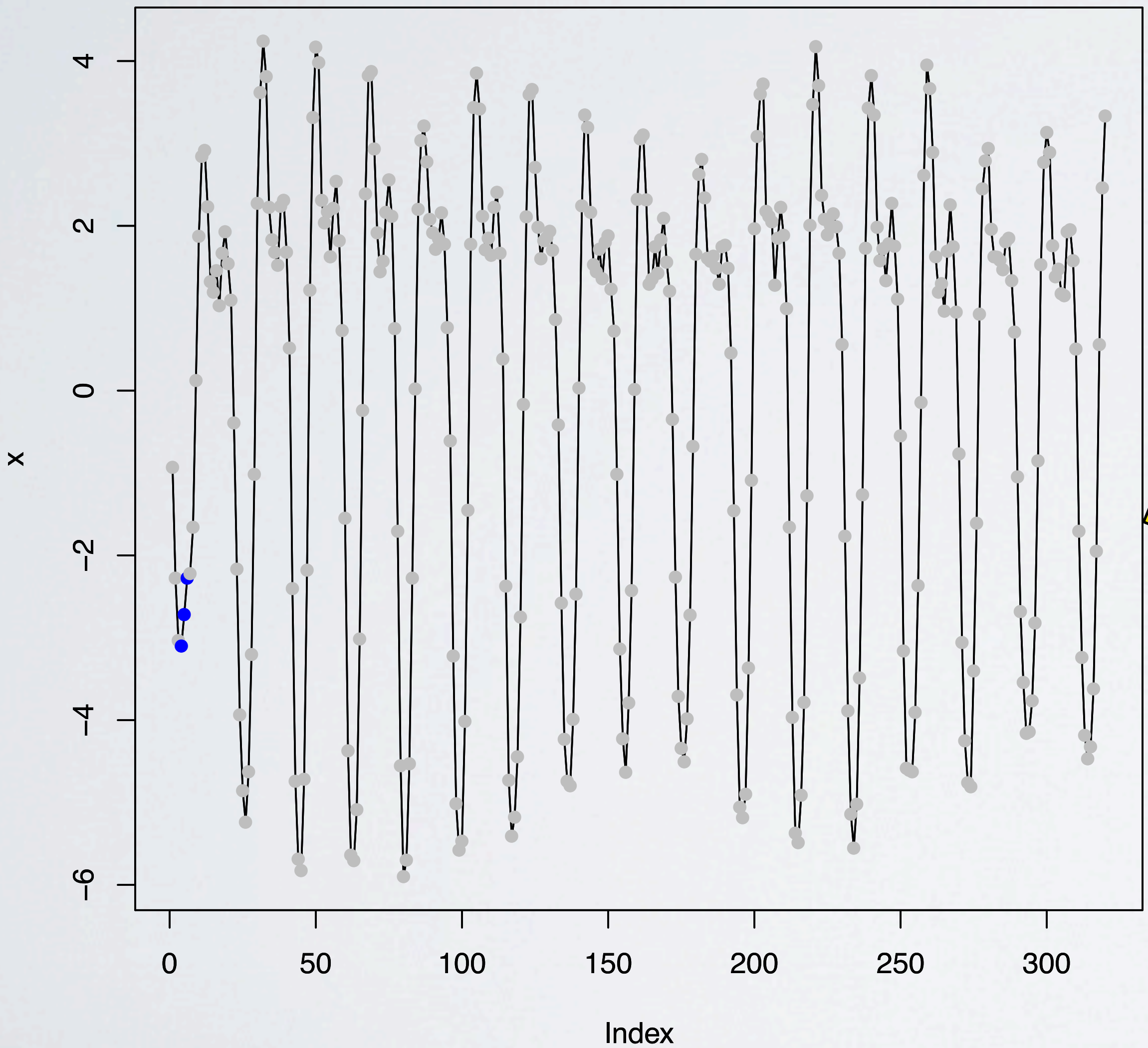


Reconstruction

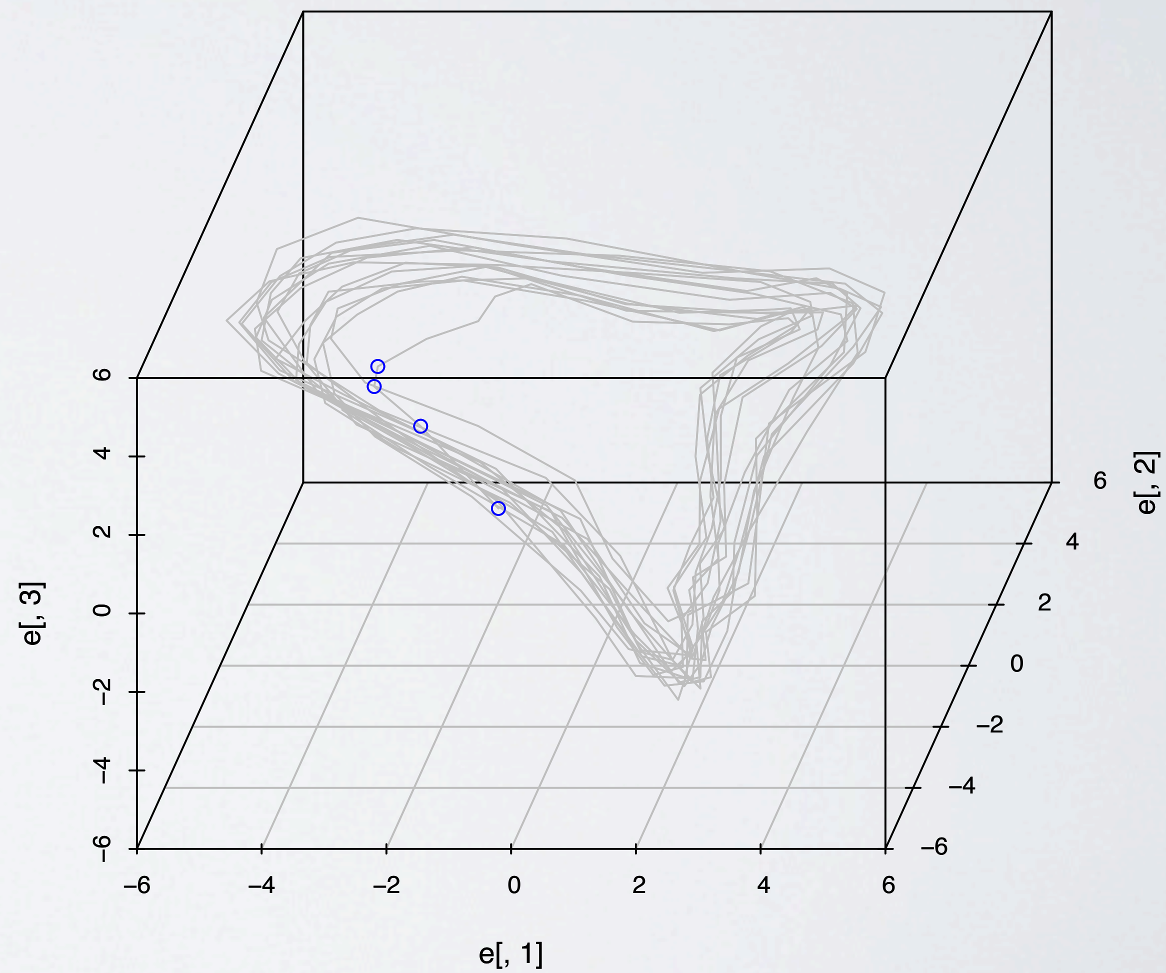


TIME-DELAY EMBEDDING

Observations

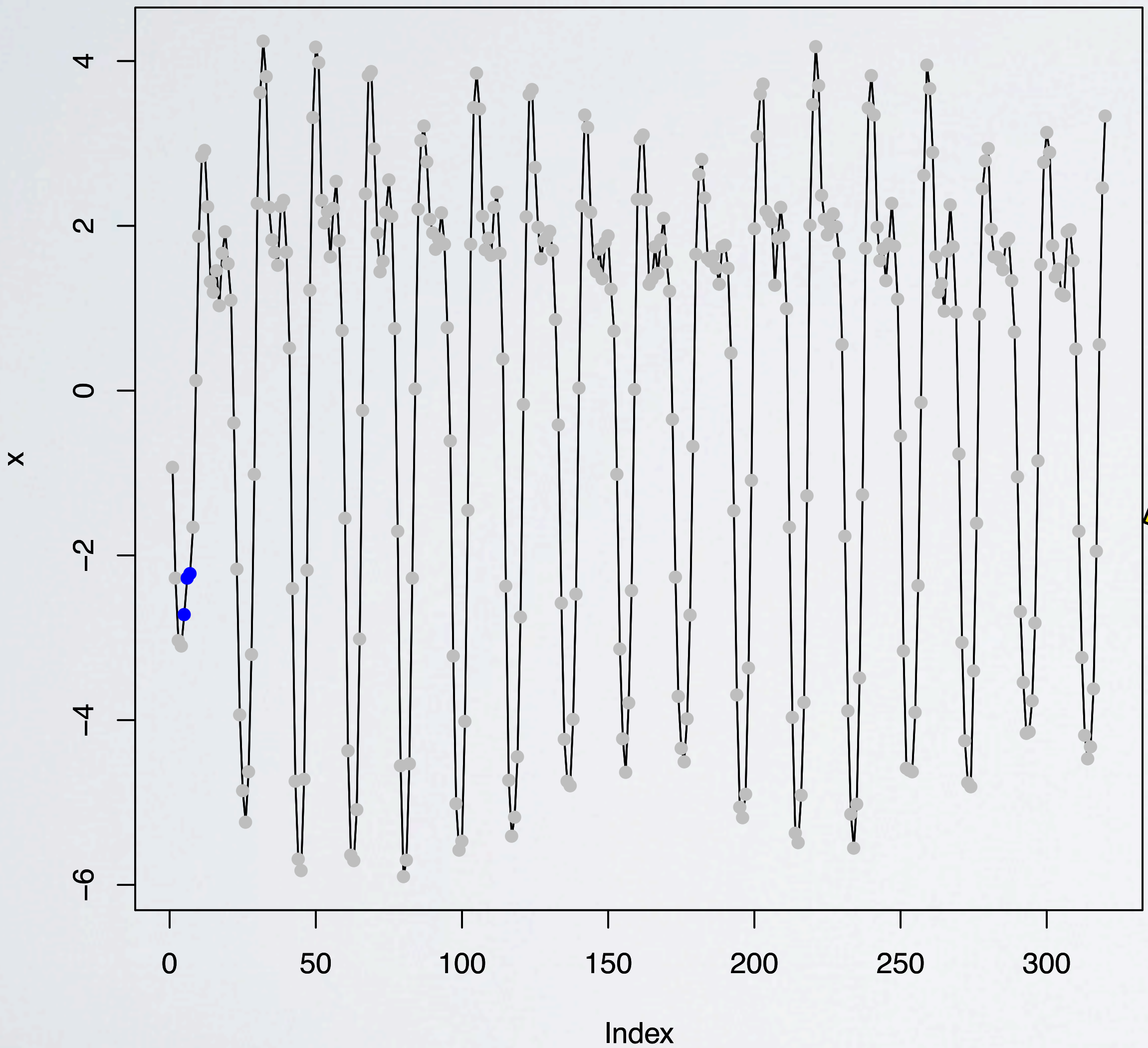


Reconstruction

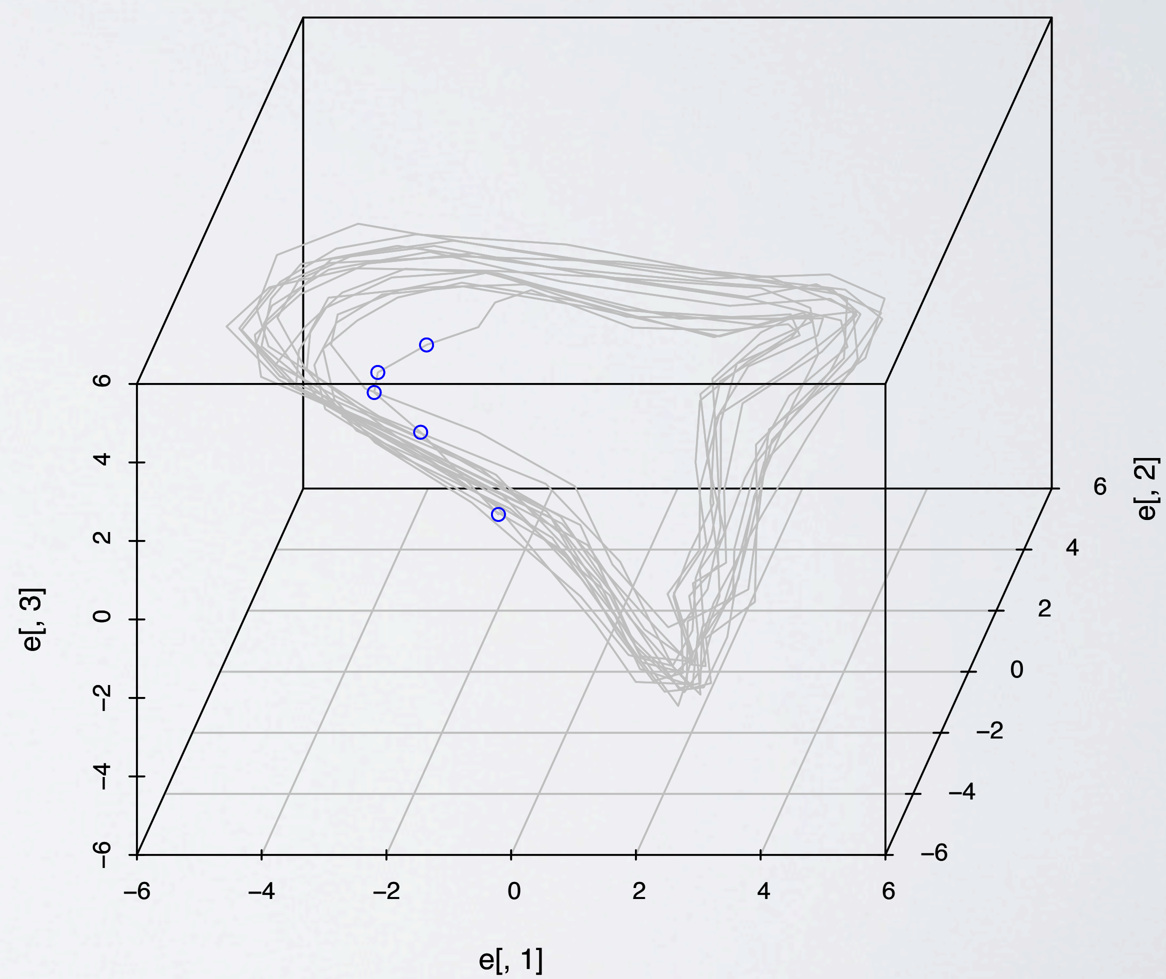


TIME-DELAY EMBEDDING

Observations

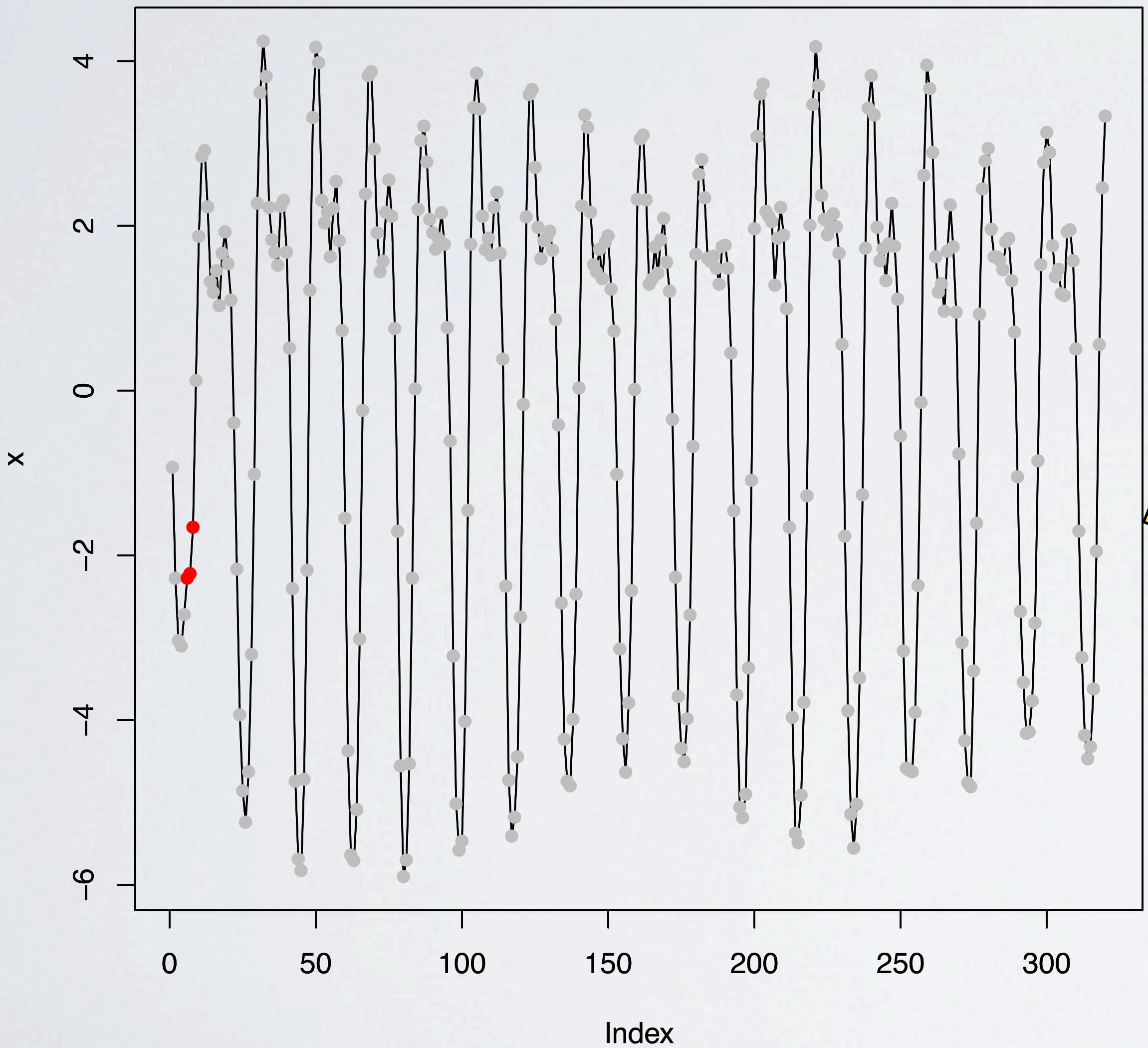


Reconstruction

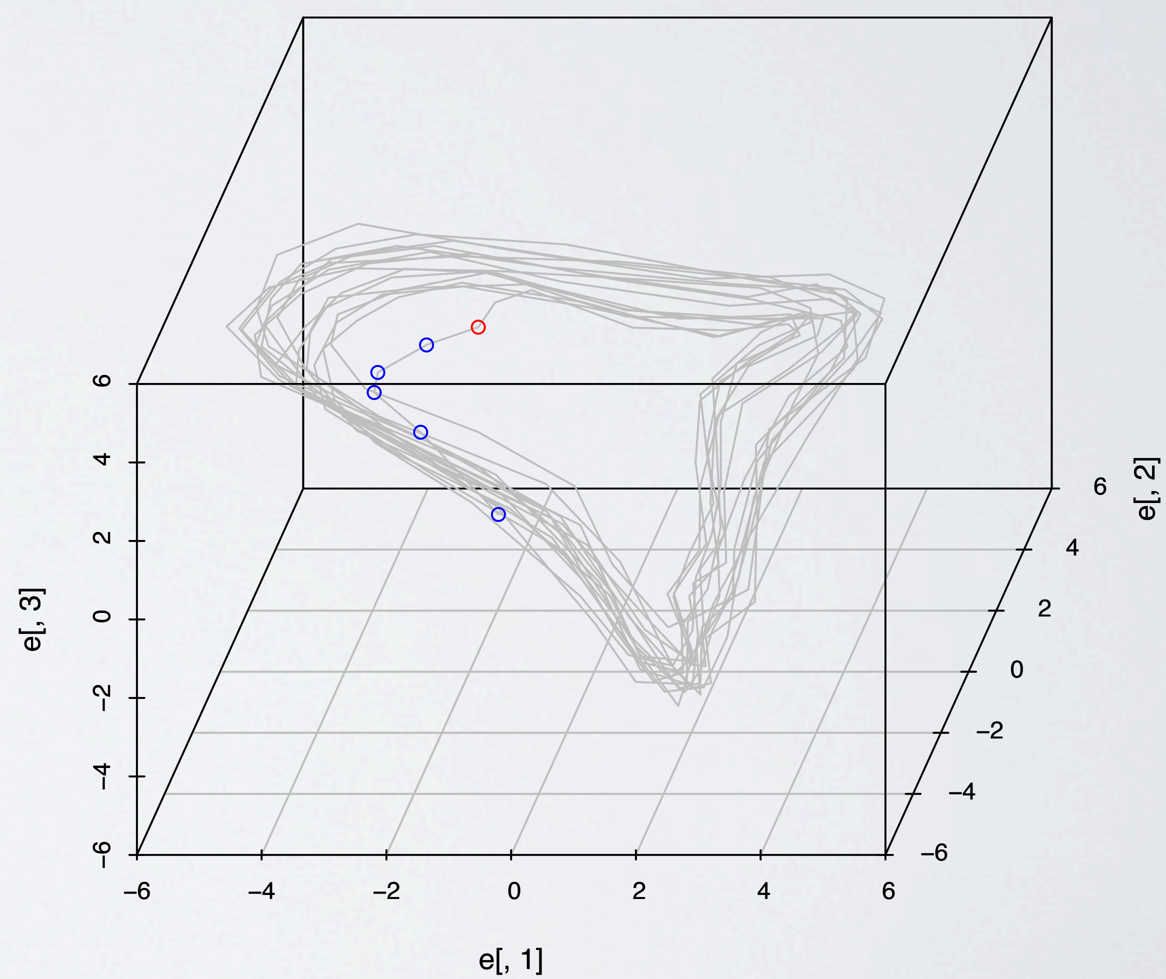


TIME-DELAY EMBEDDING

Observations

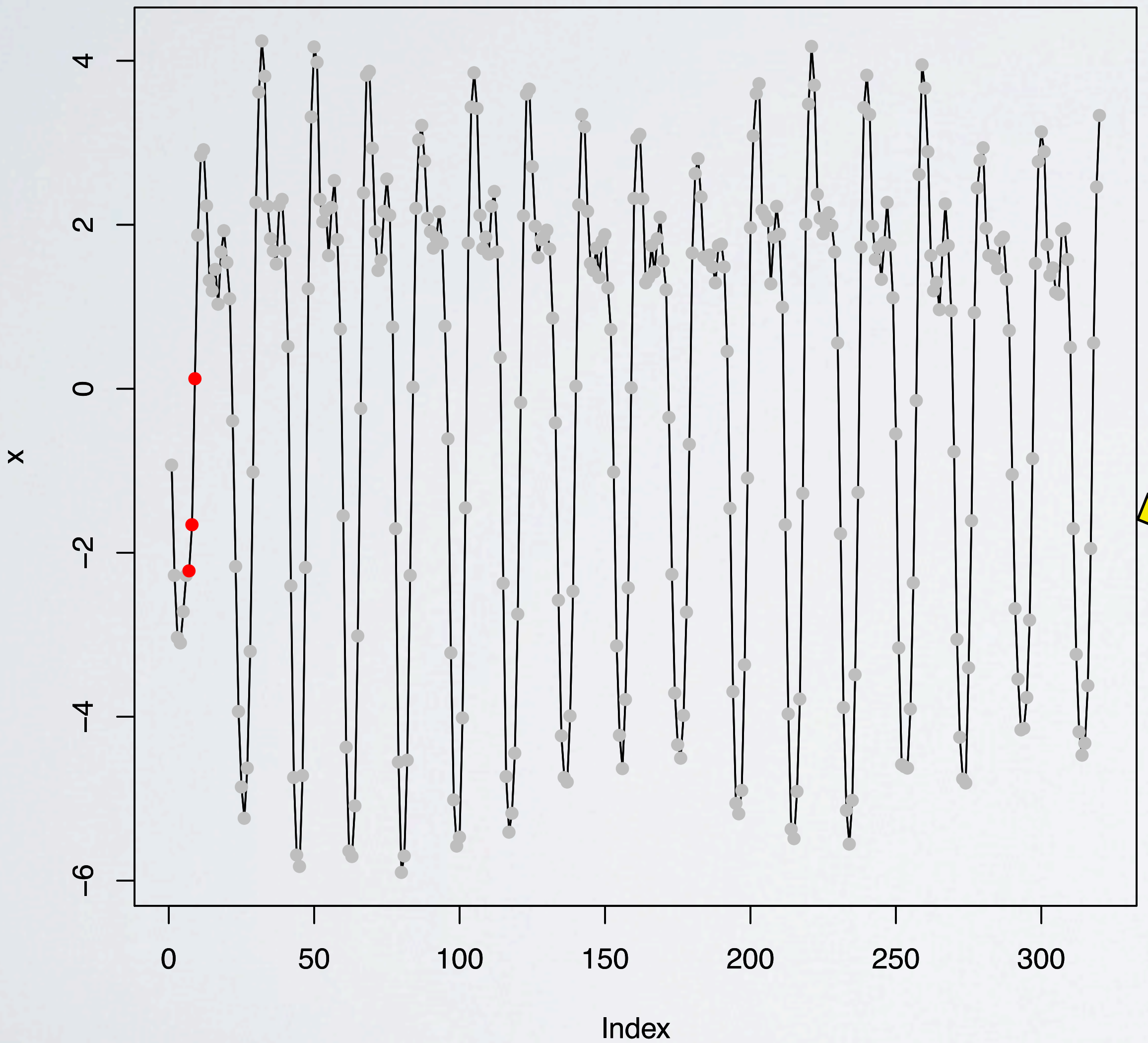


Reconstruction

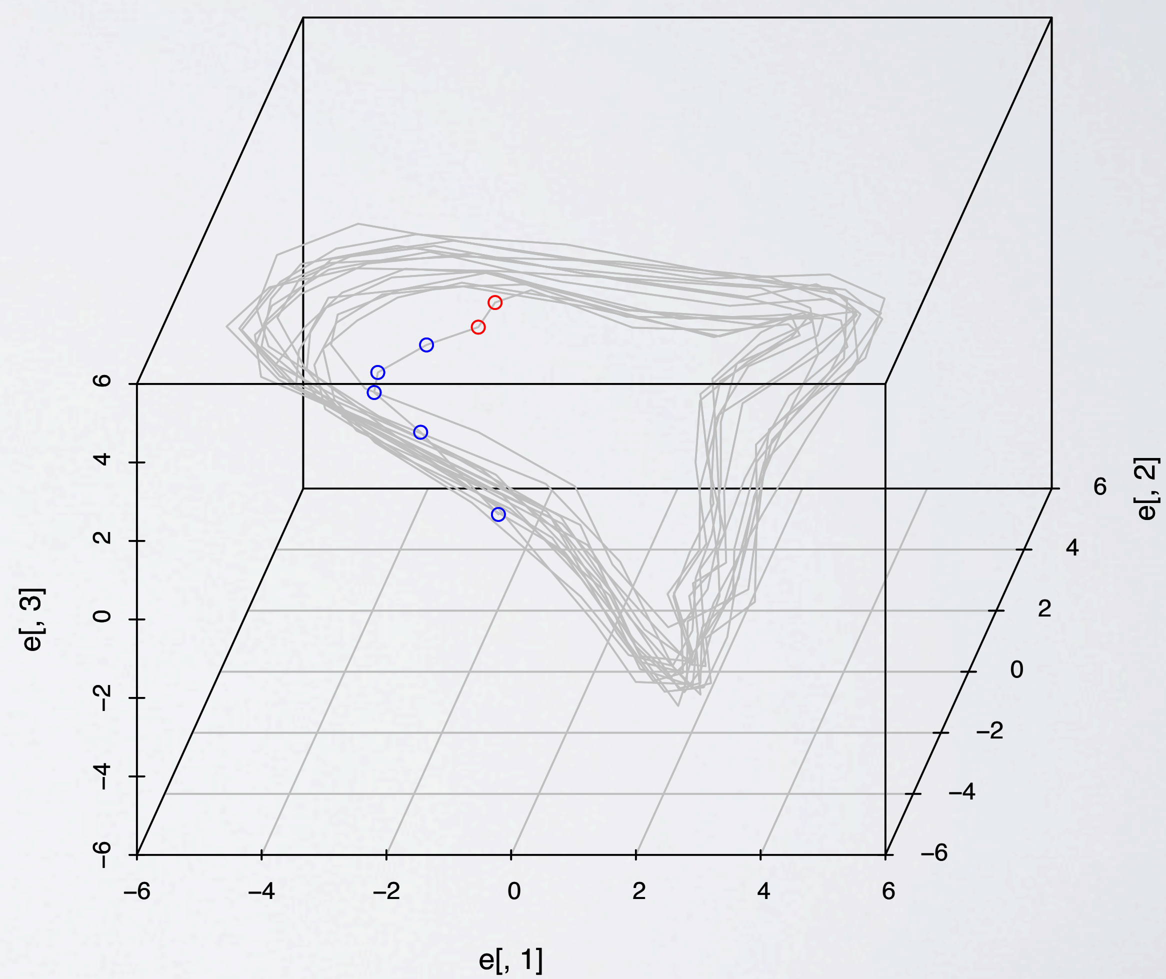


TIME-DELAY EMBEDDING

Observations

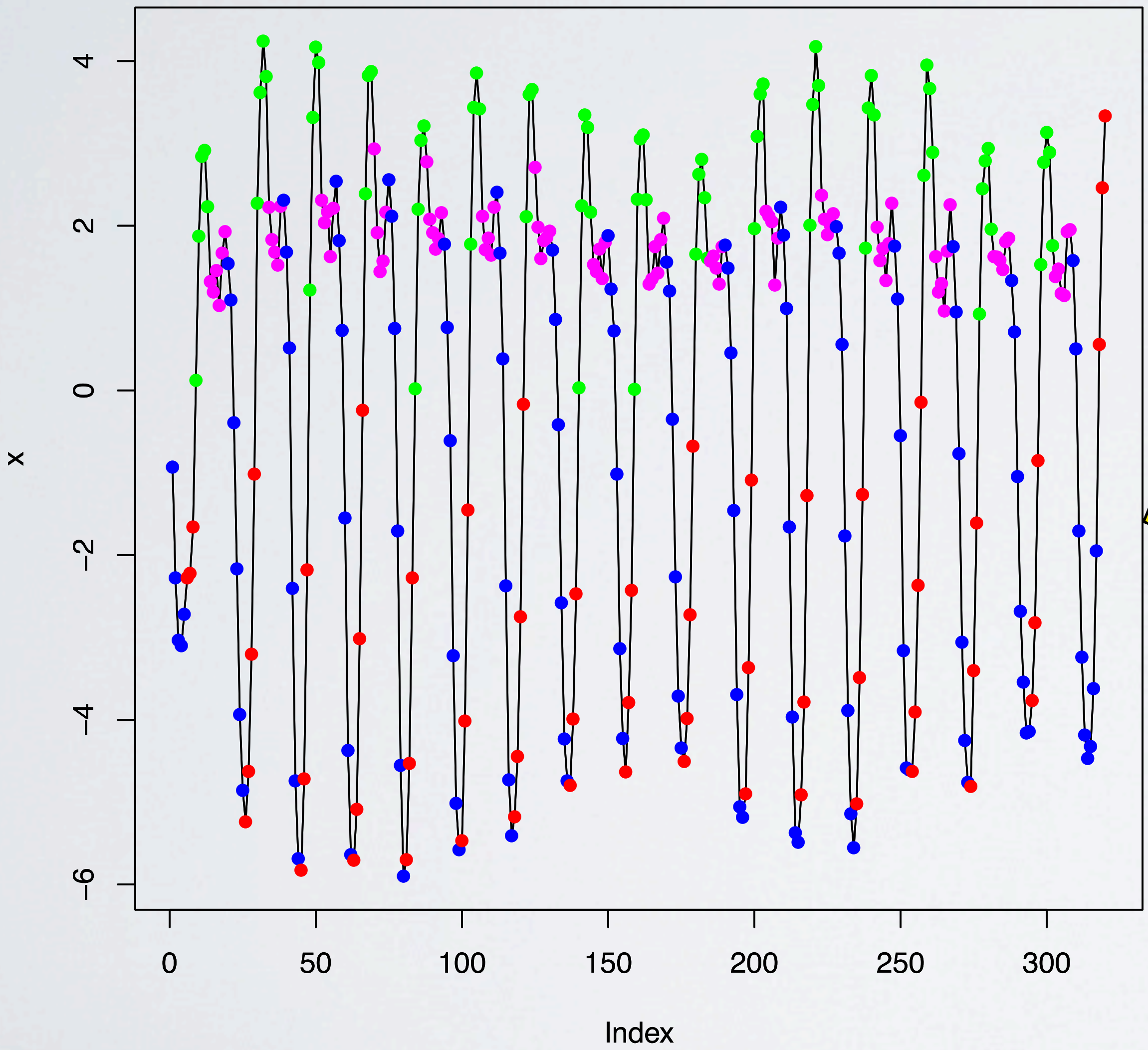


Reconstruction

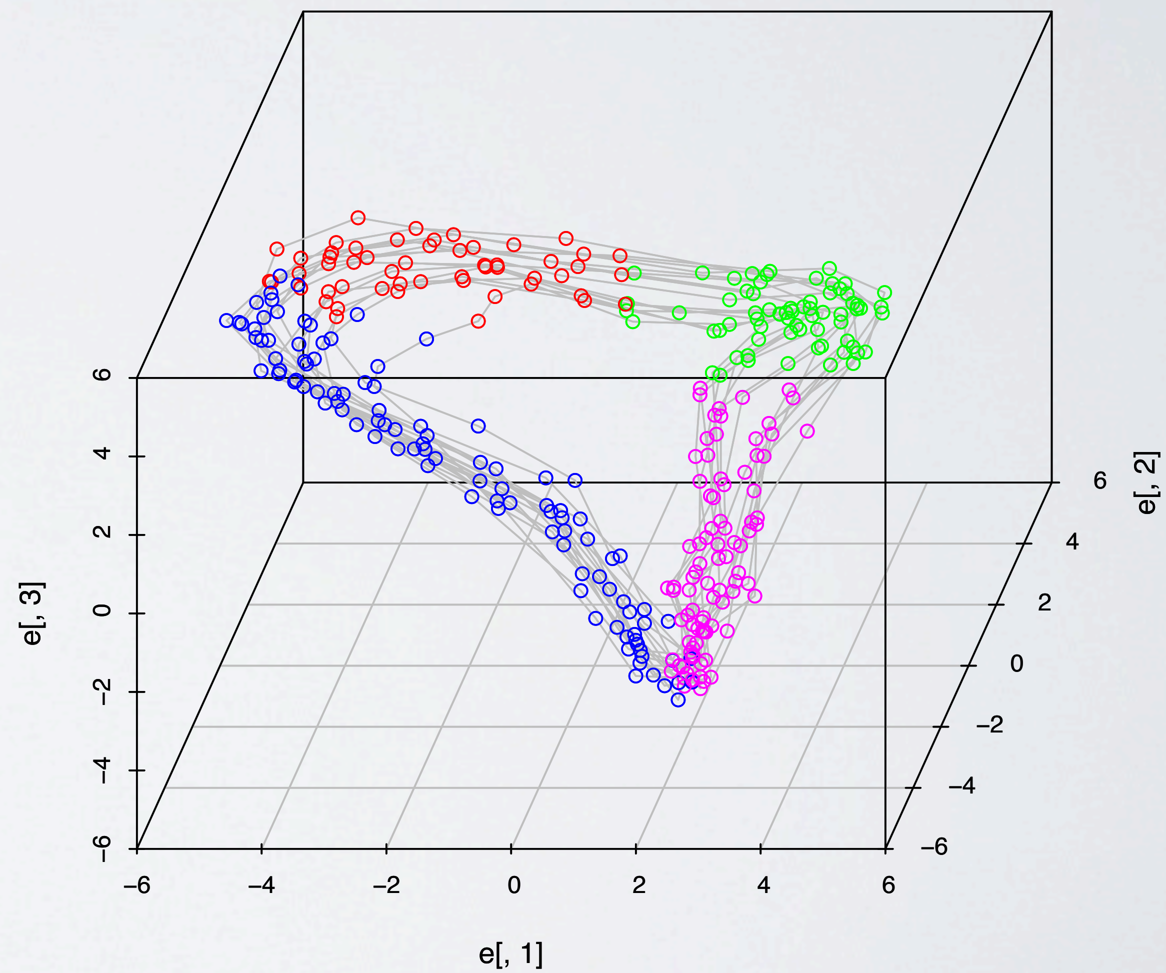


TIME-DELAY EMBEDDING

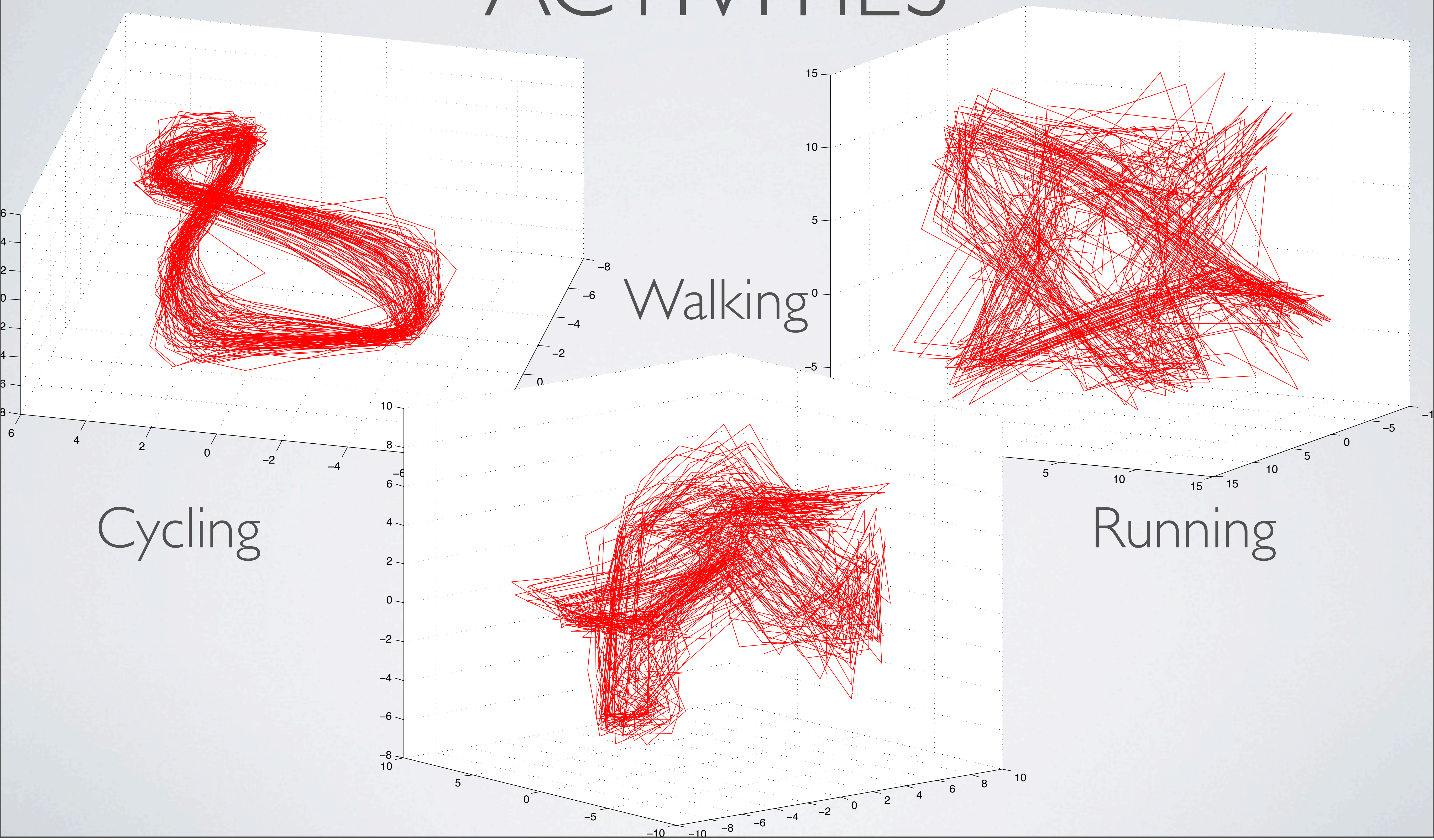
Observations



Reconstruction



MODELS FOR DIFFERENT ACTIVITIES



Cycling

Walking

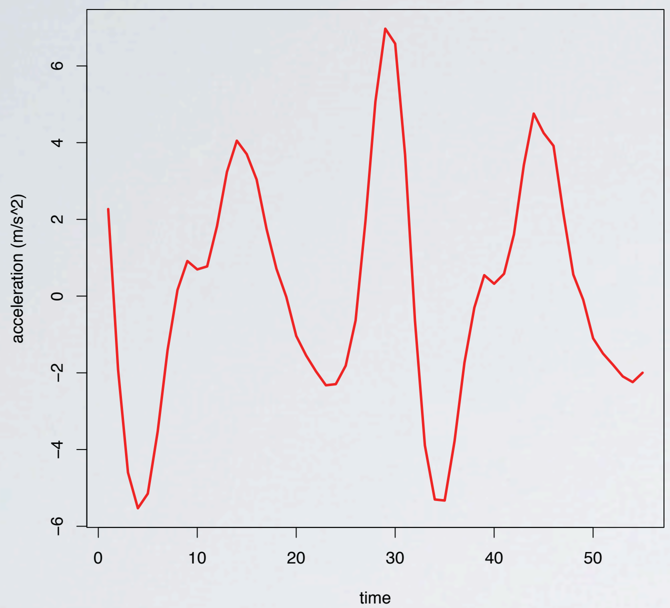
Running

FEATURE EXTRACTION

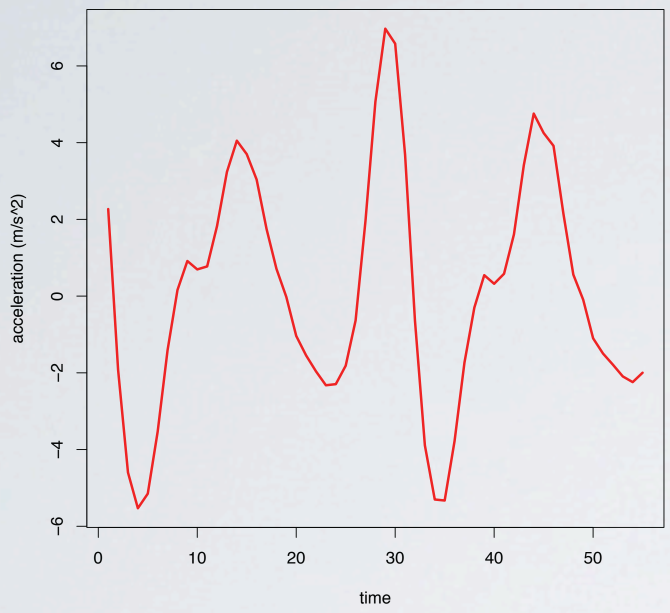
INTUITION

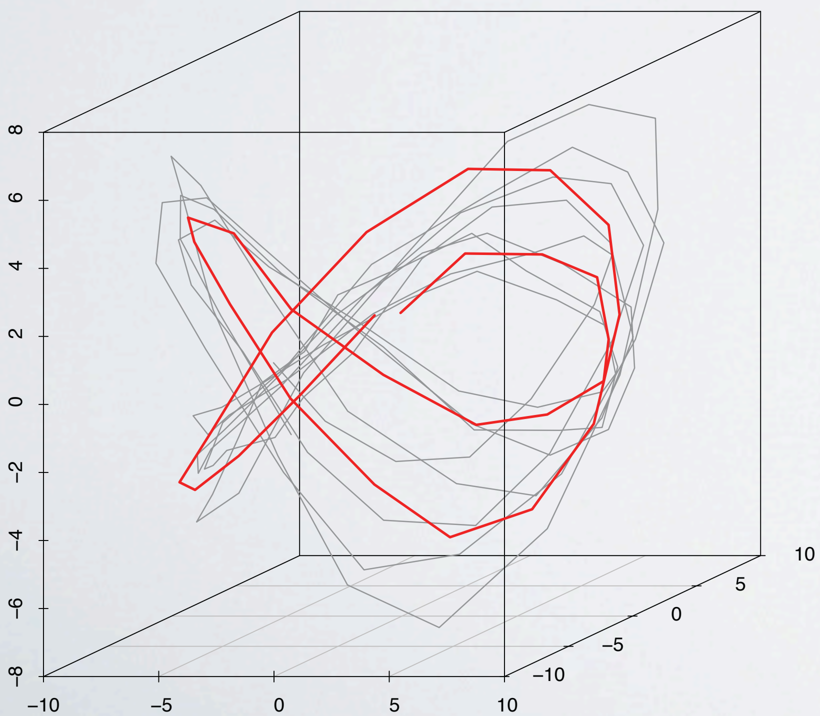
- Think of models as basis functions in their particular reconstruction spaces
- Given a set of models and a new segment, project the segment into the reconstruction space for each model and calculate a measure of similarity
- Everything is in Euclidean space, and so geometry is straightforward

TEMPLATE MATCHING

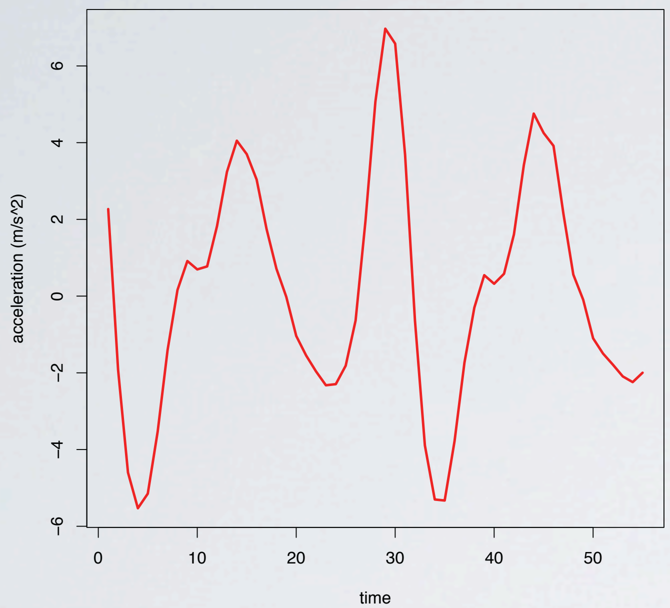


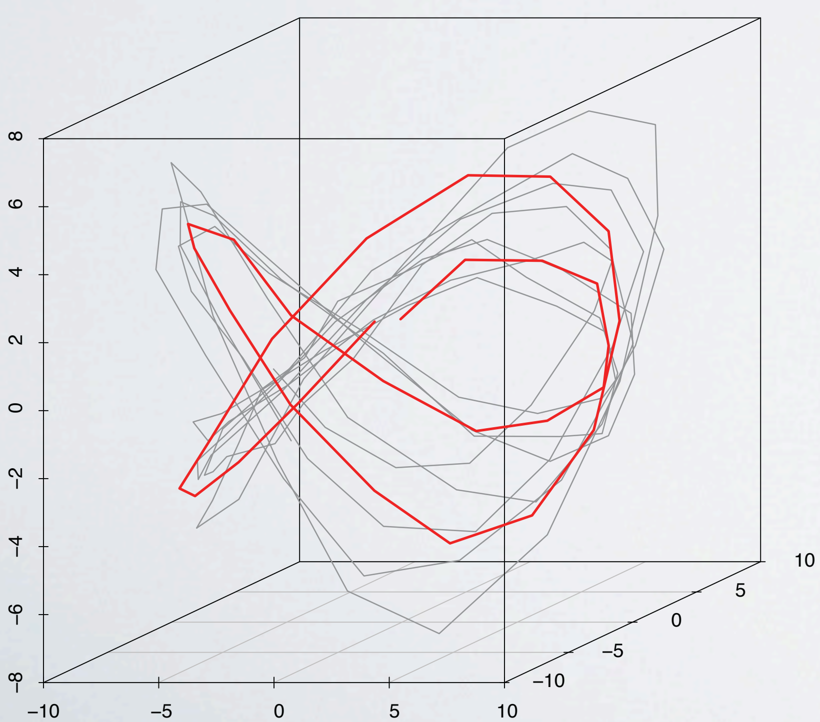
TEMPLATE MATCHING



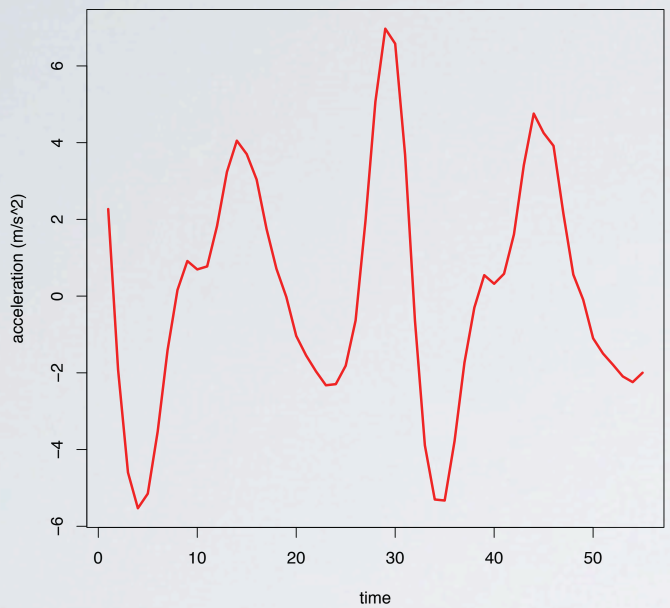


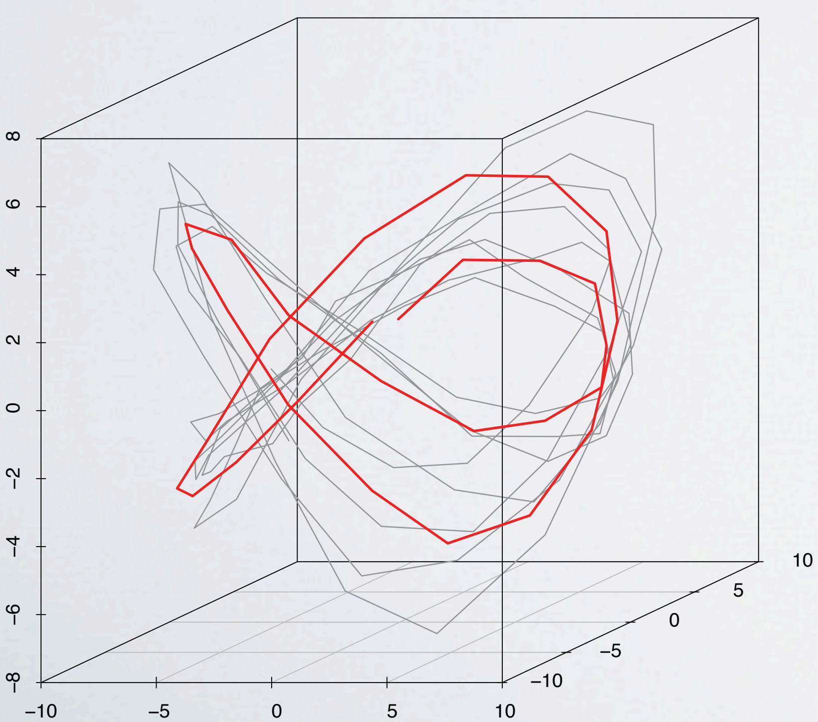
TEMPLATE MATCHING

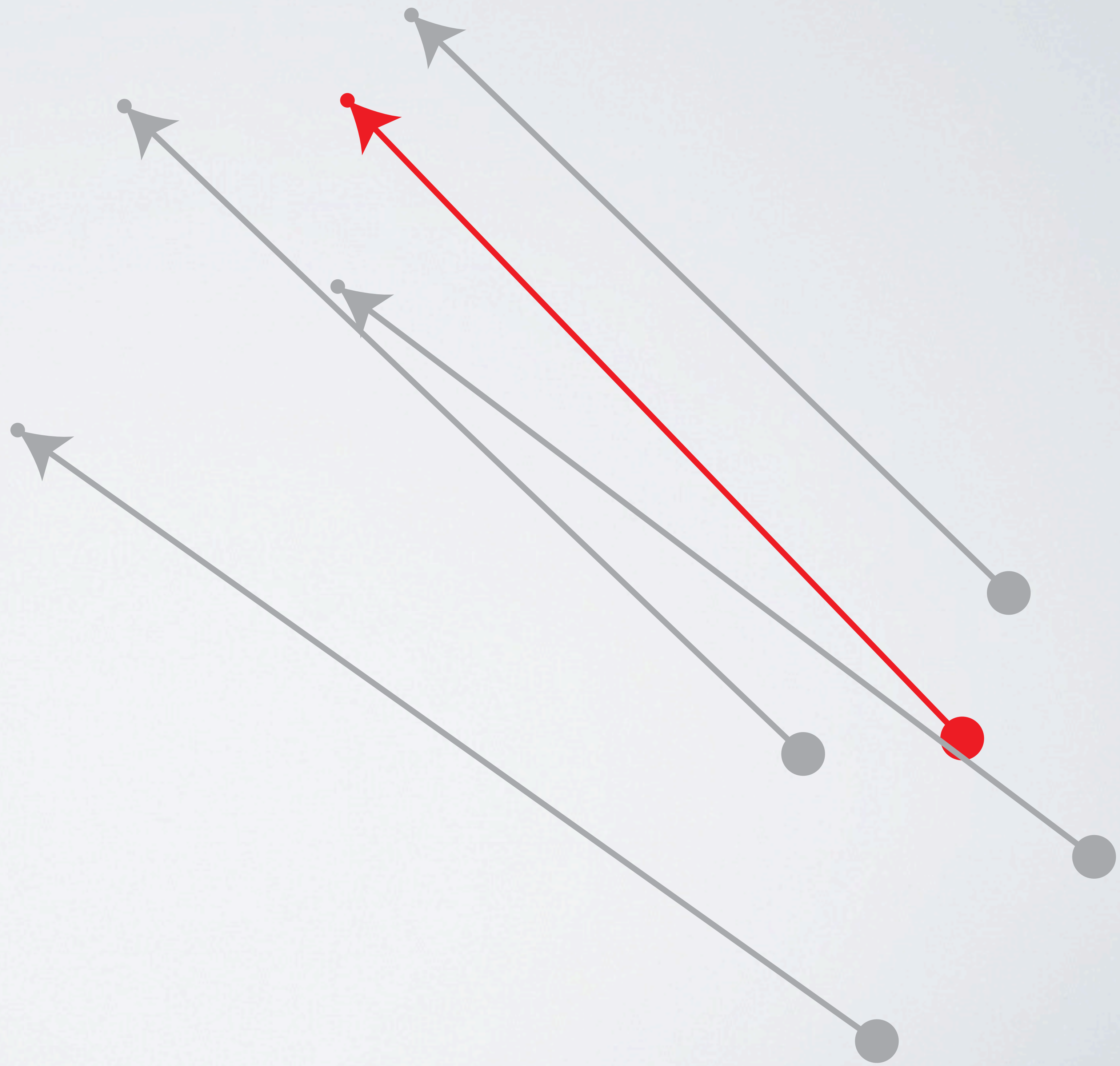




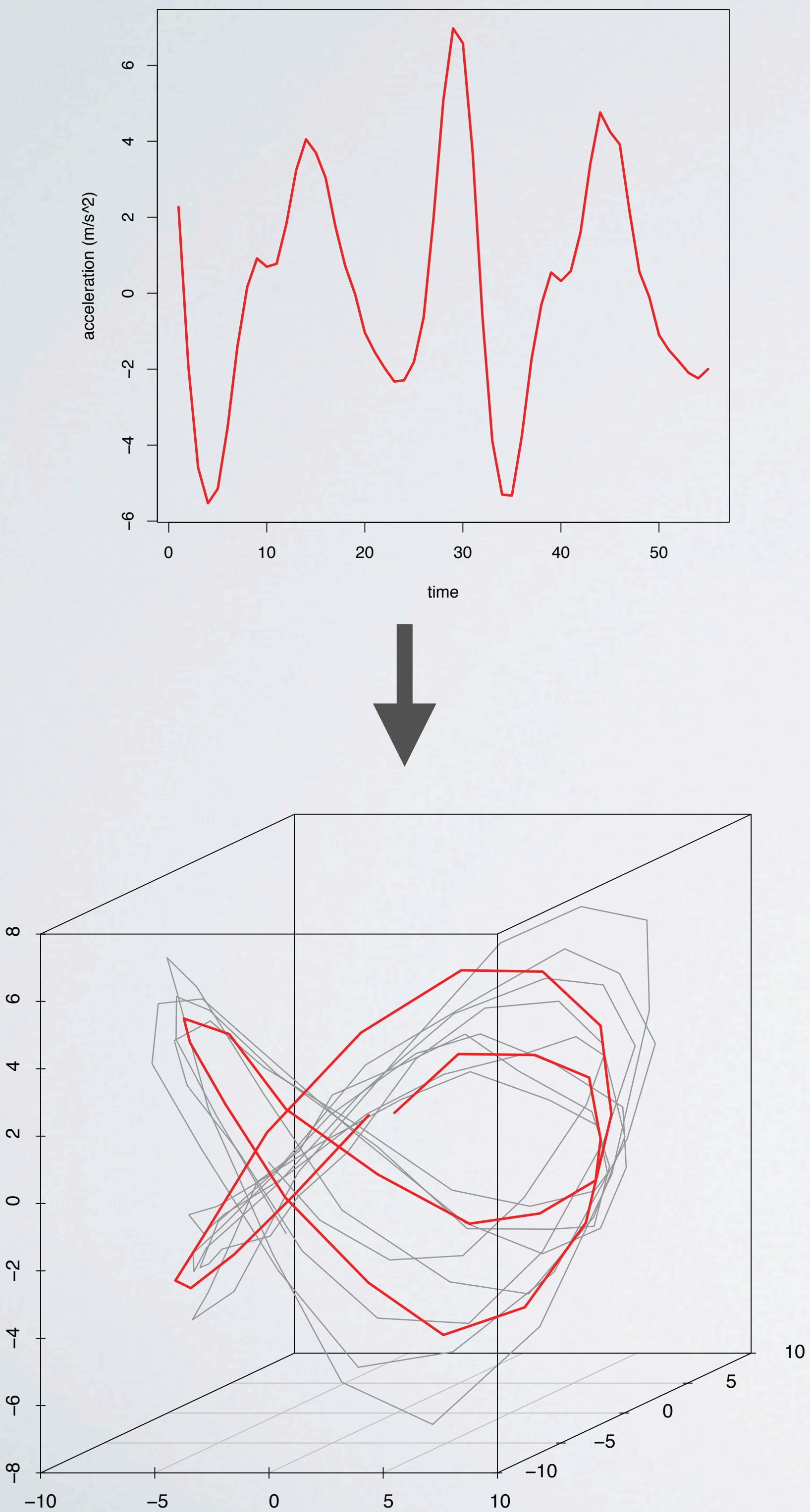
TEMPLATE MATCHING

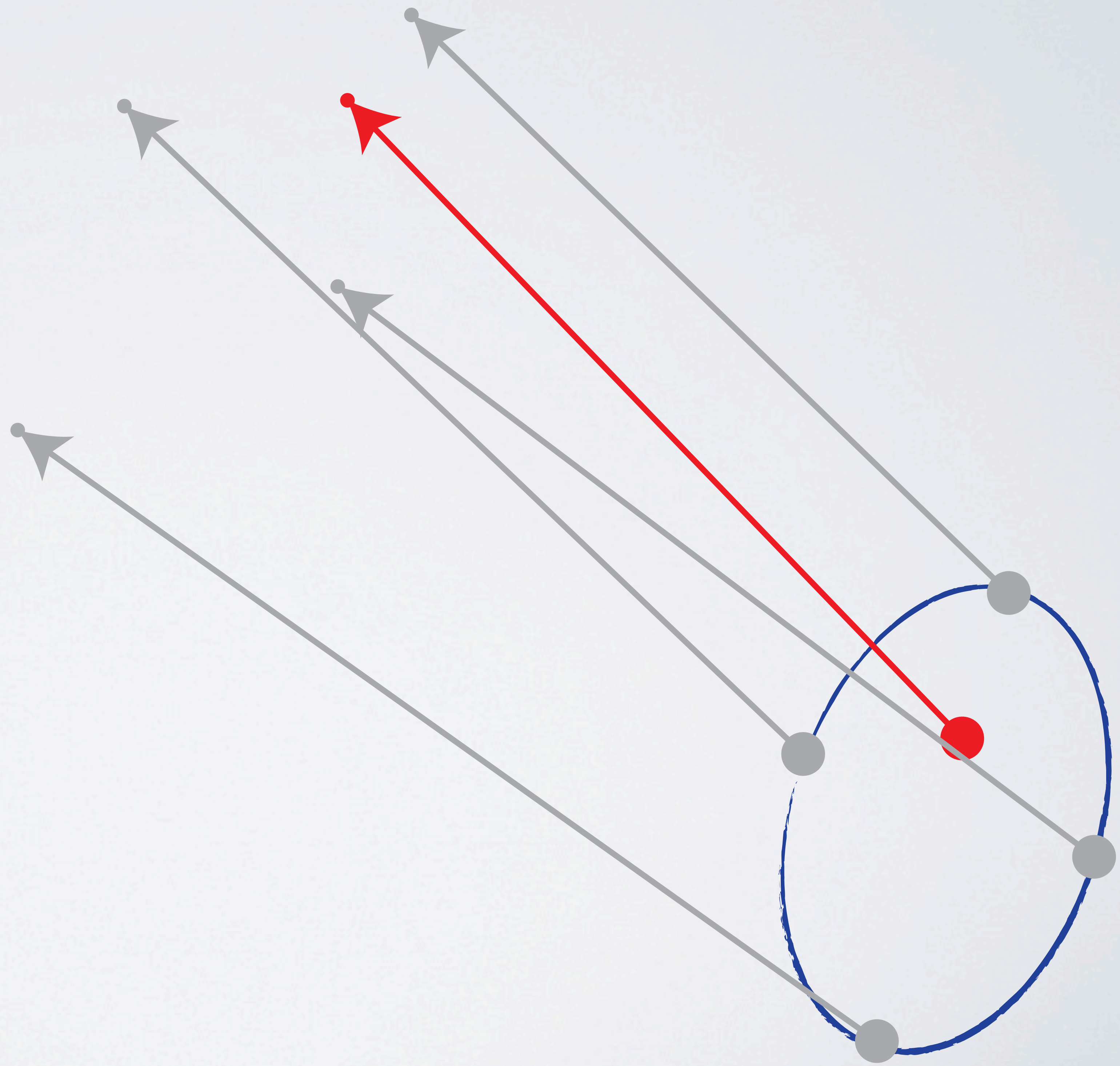




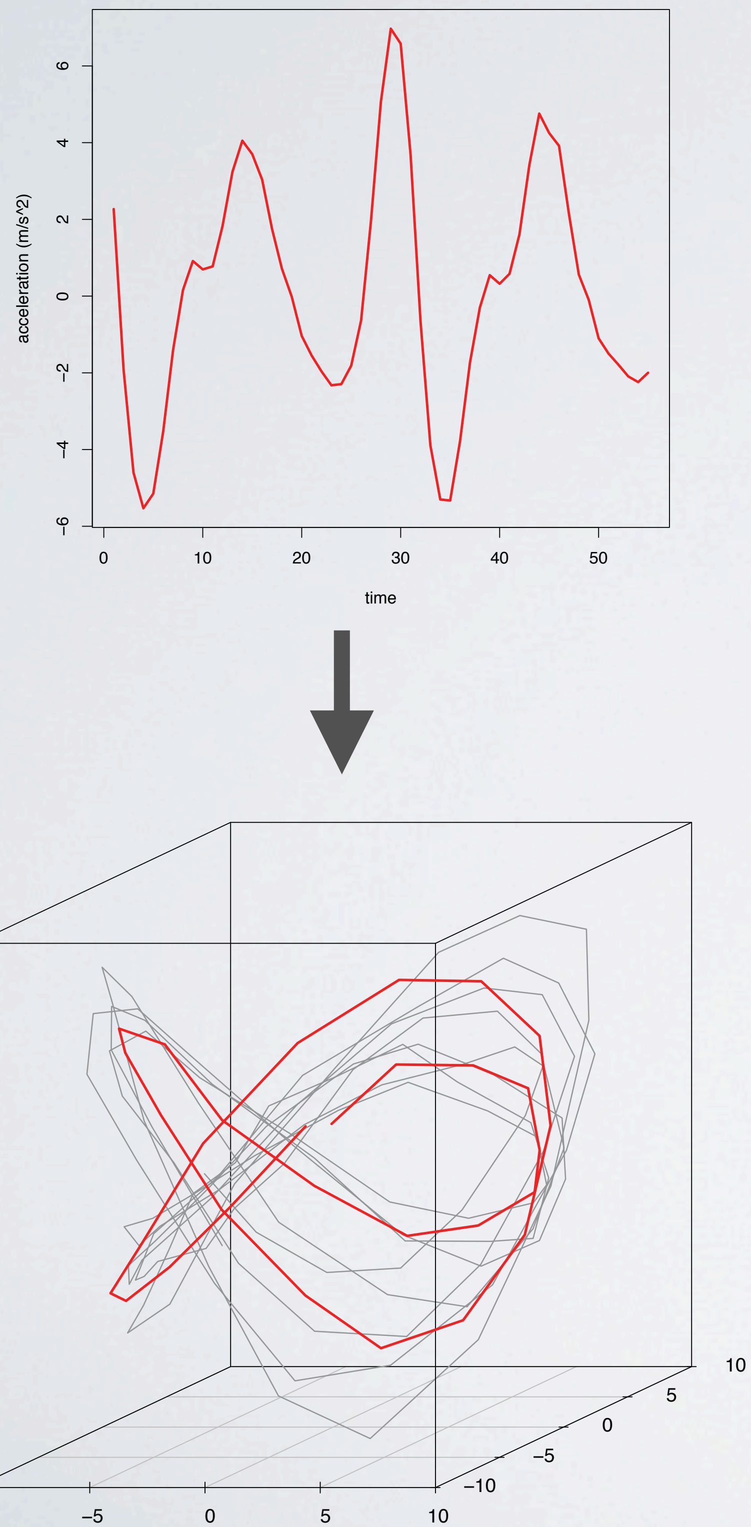


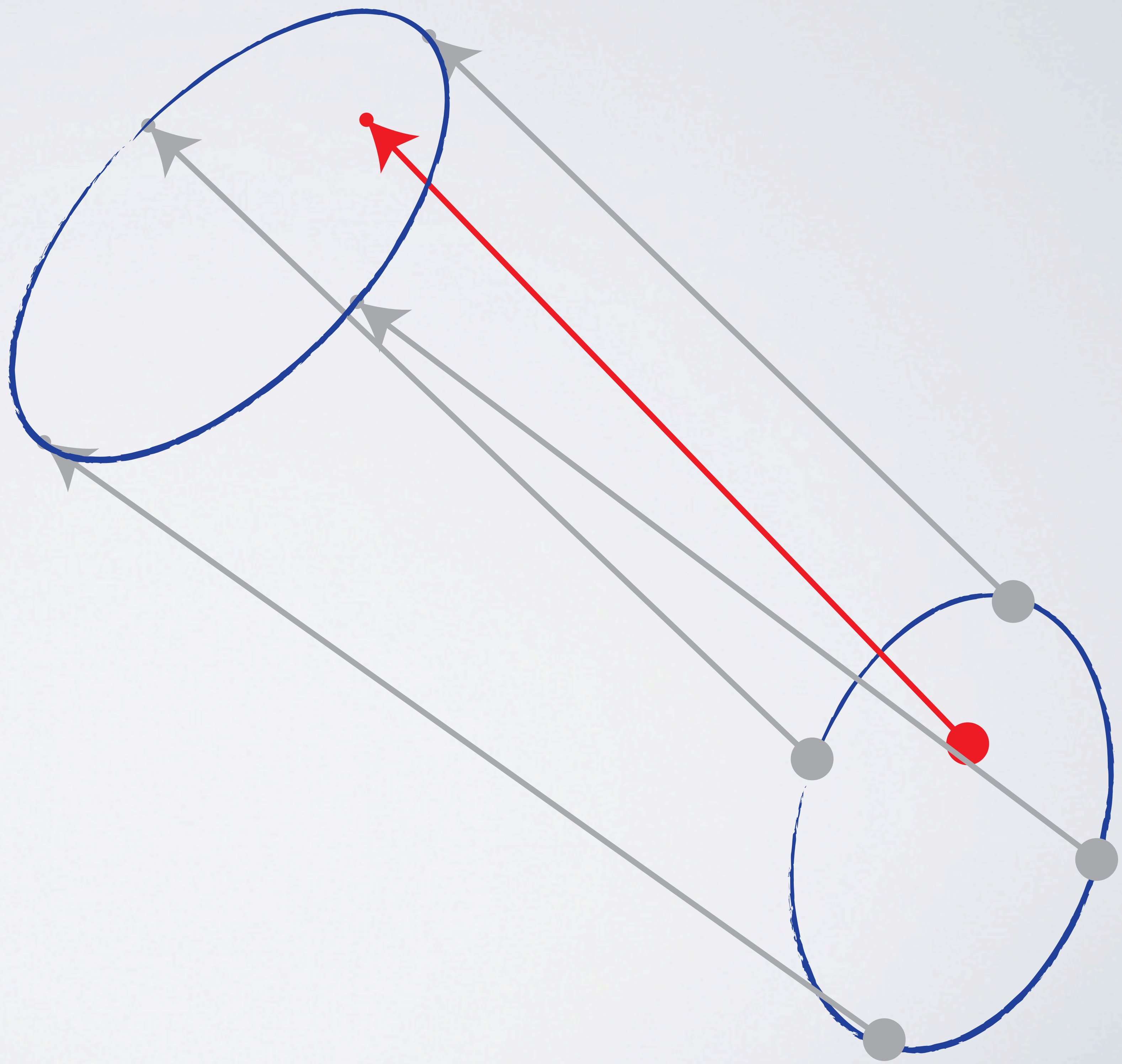
TEMPLATE MATCHING



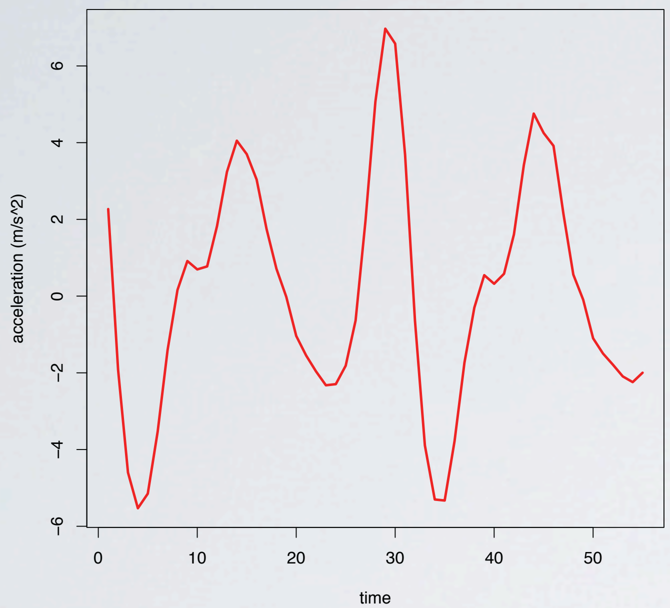


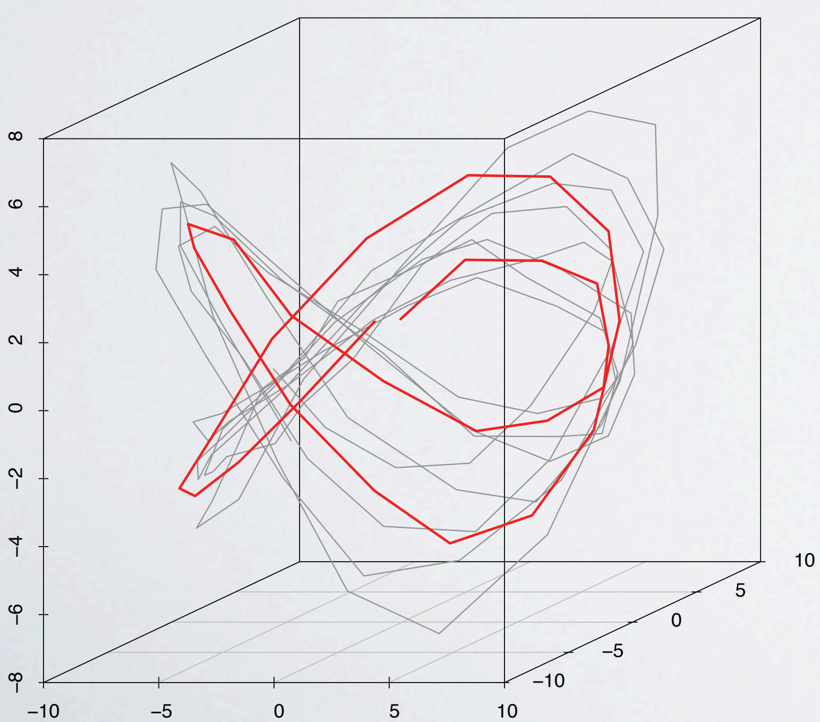
TEMPLATE MATCHING

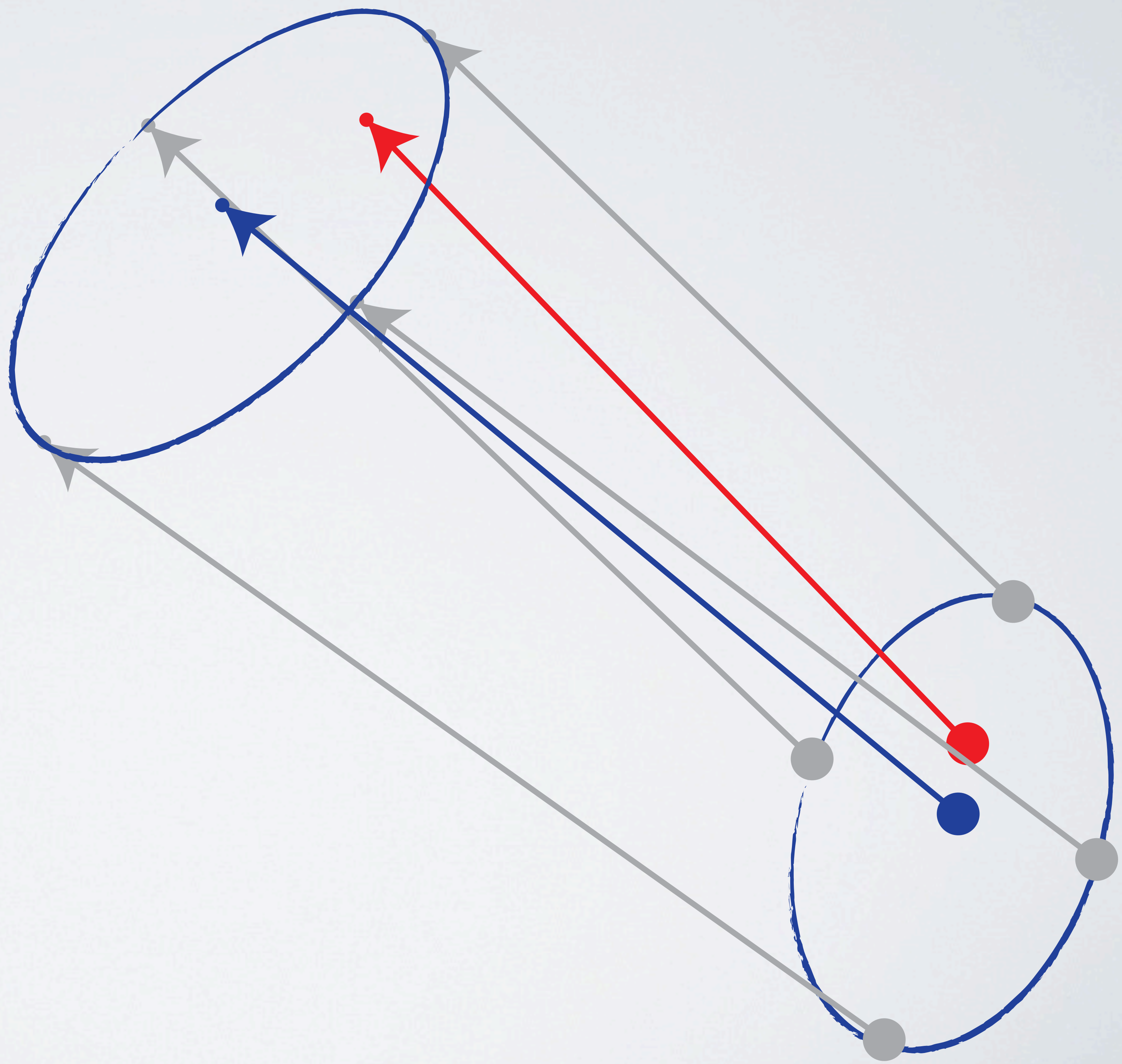




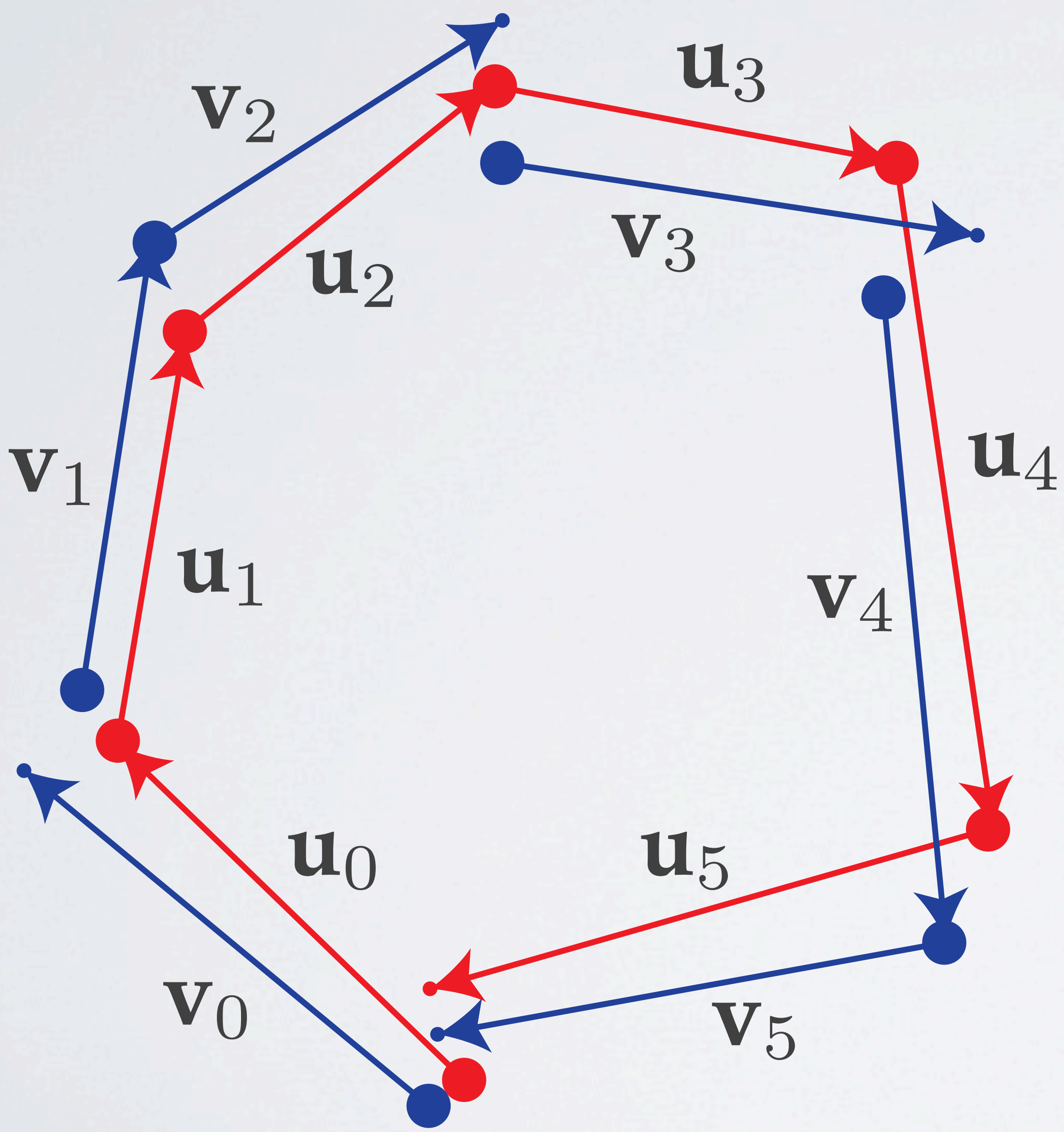
TEMPLATE MATCHING







SCORING



$$S = \sum_i \frac{\mathbf{u}_i \cdot \mathbf{v}_i}{\max(|\mathbf{u}_i|, |\mathbf{v}_i|)^2}$$

RECAP

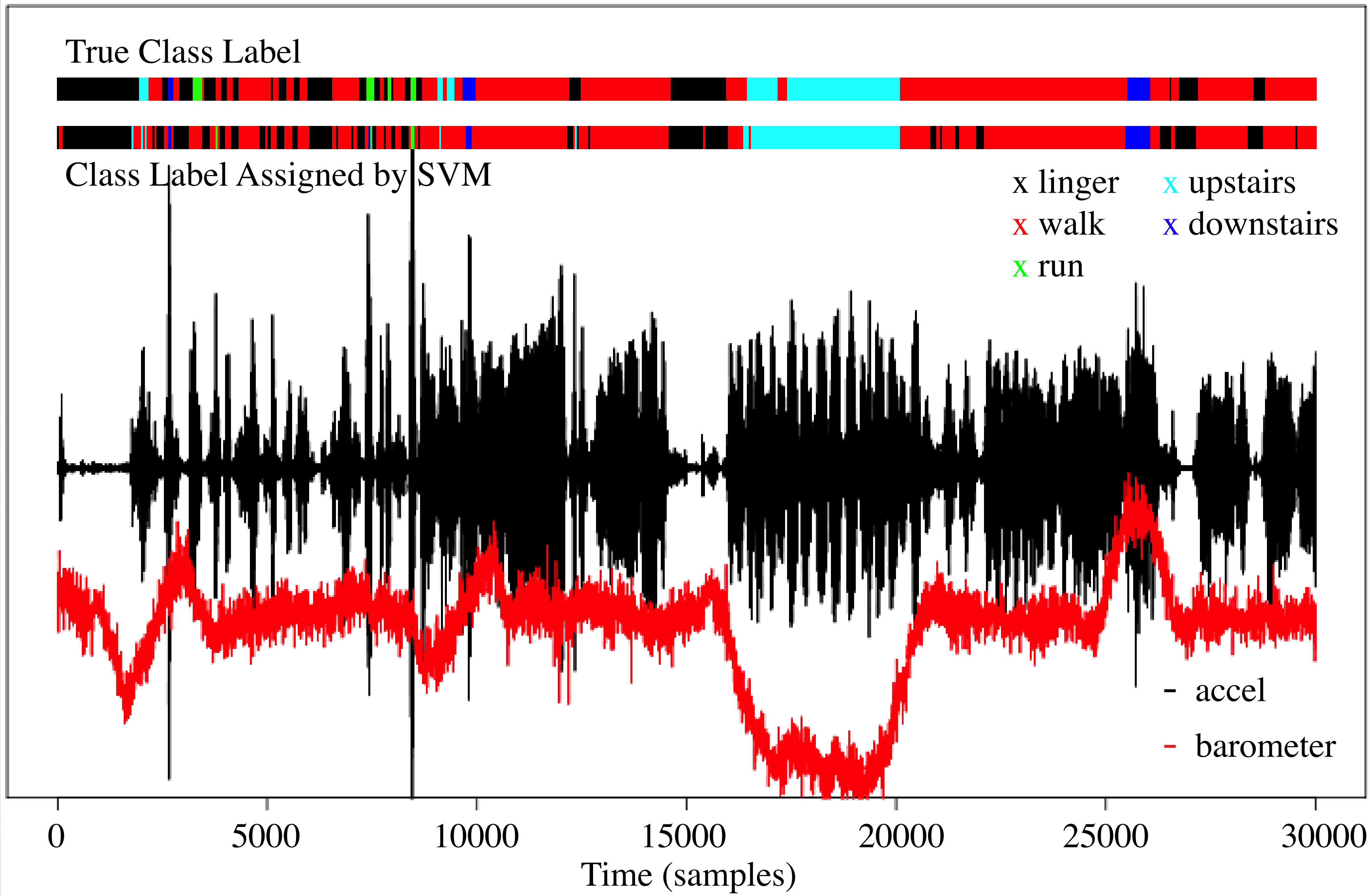
- Building models is efficient (memoization).
- Feature extraction is efficient (k-nearest neighbours).
- Features are similarity scores between a segment of data and a set of models.

PLAY

ACTIVITY RECOGNITION

- For each activity build a few (5) models from randomly selected segments of the training data
- Consider the similarity scores to be input features for training a classifier (SVM for our experiment)
- With 20 features, we achieve performance comparable to state-of-the-art systems that extract 651 features (Lester et al. 2006).
- Example: Our method (**87.89%**), baseline (87.22%).
- Fair comparisons difficult due to availability of data sets.

ACTIVITY RECOGNITION

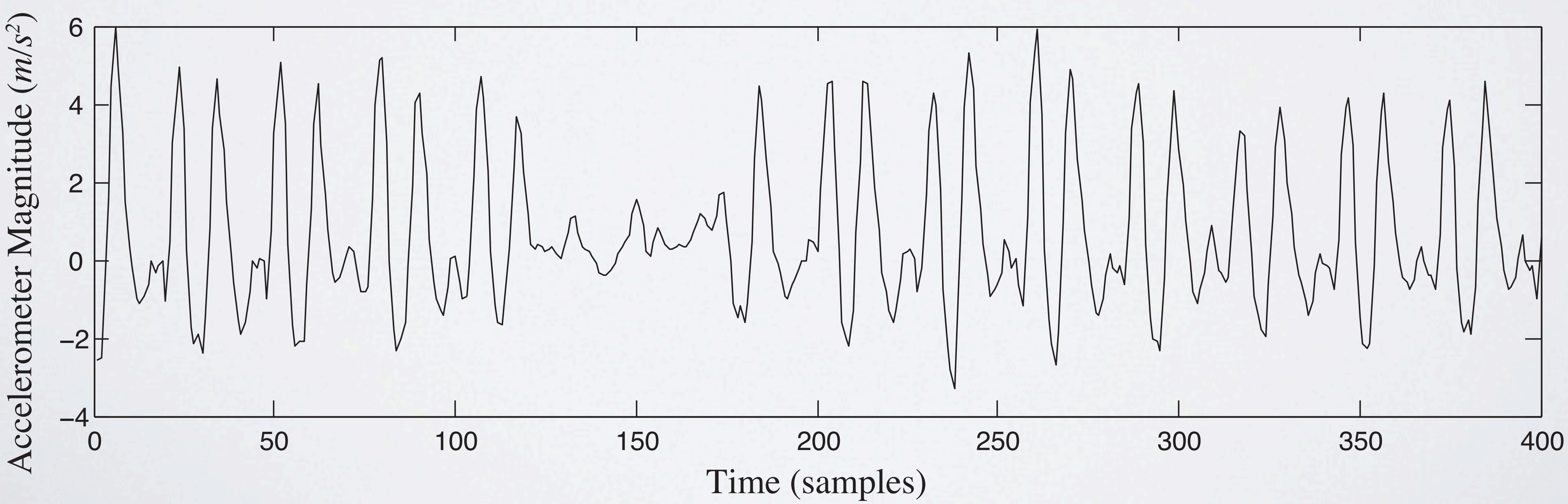


Data provided by Dieter Fox's group.

GAIT RECOGNITION (TAKE ONE)

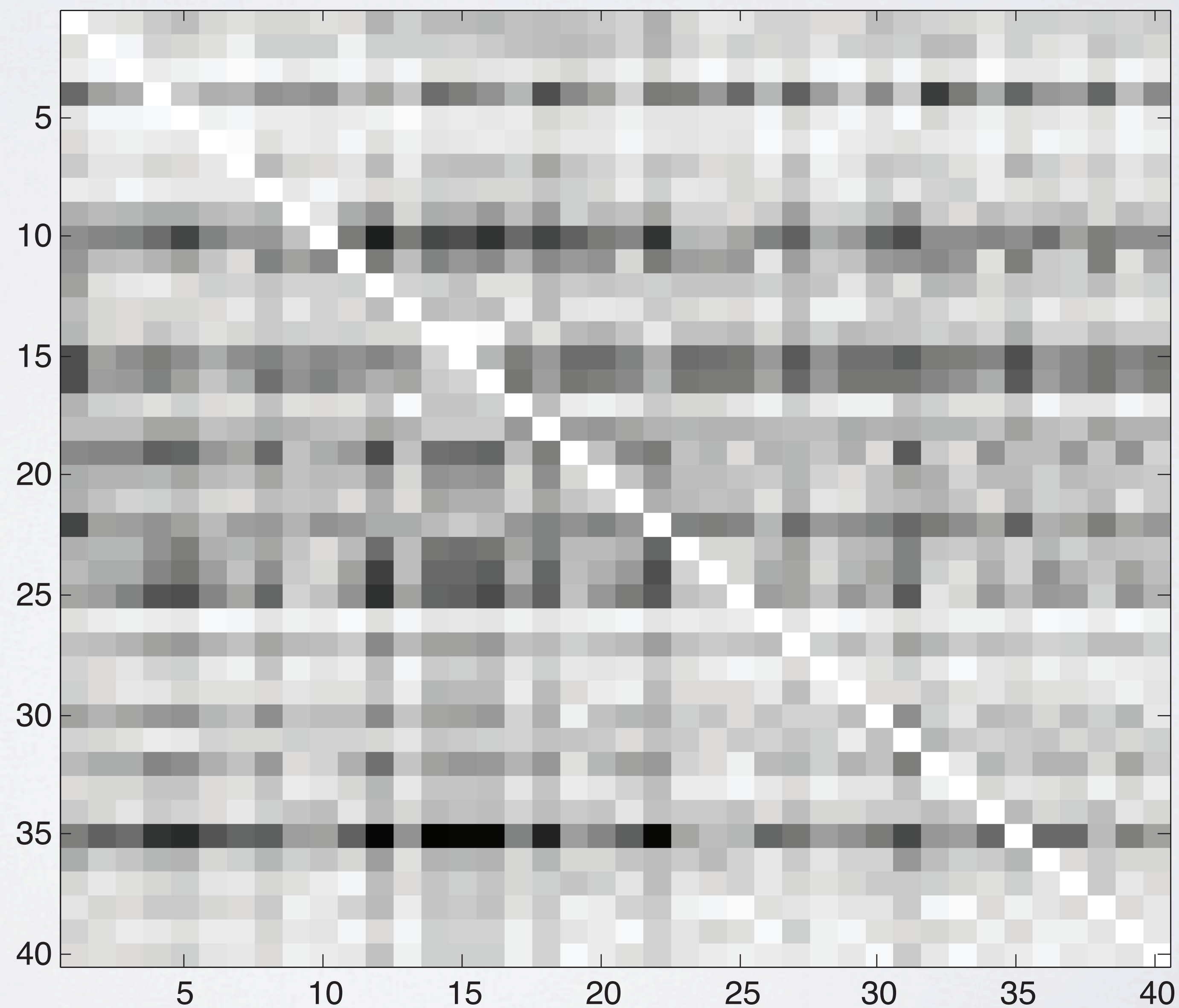
- 40 people, 12-20 seconds of data per person (walk to end of hall, walk back).

- Split each trace into training and test sets, build a model from the training set, compute score for each test set (repeat 5 times with different training sets, average scores)



GAIT RECOGNITION RESULTS

- If we predict the model with the highest score, we achieve perfect (100%) classification accuracy
- Confusion Matrix:



GAIT RECOGNITION (TAKE TWO)

- Data collected from 20 individuals (10 male, 10 female) performing two 15 minute outdoor walks on two different days. Carried a Nexus One mobile phone in their pocket.

- Subjects changed clothes between days, paused to cross the street, walked up and down hills, on grass and concrete, up and down stairs.

- Data much more representative of what a real gait recognition system would encounter.

GAIT RECOGNITION (TAKE TWO)

- Performance measured on frame-by-frame recognition.
- Train on one day, test on the other.
- Problem: How to choose segments from which to build models.
- Solution: **Boosting**. Use boosting weights to locate hard-to-classify segments and build models on these.
- One model per person (20 models). Replace one model at each round based on boosting weights. Random forest classifiers. Call our algorithm TDEBOOST.

GAIT RECOGNITION RESULTS

- Baseline used 200 features from Lester et al. (2006) and random forest classifiers. Used more trees per forest as there were 10 times as many features.

- TDEBOOST Accuracy: **42%**
  Baseline Accuracy: 20%

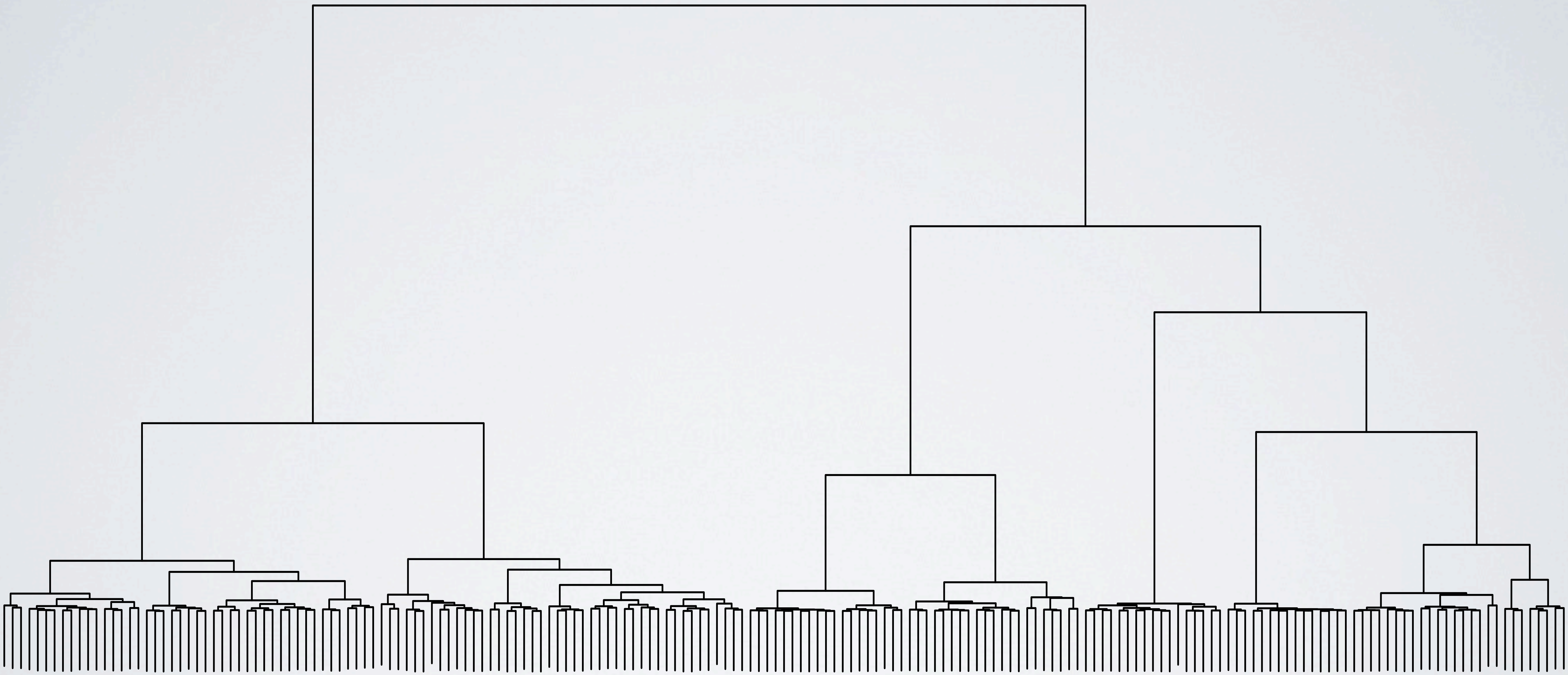
- For 16 of the 20 individuals, TDEBOOST has higher precision and recall than the baseline.

- This data is freely available on my website.

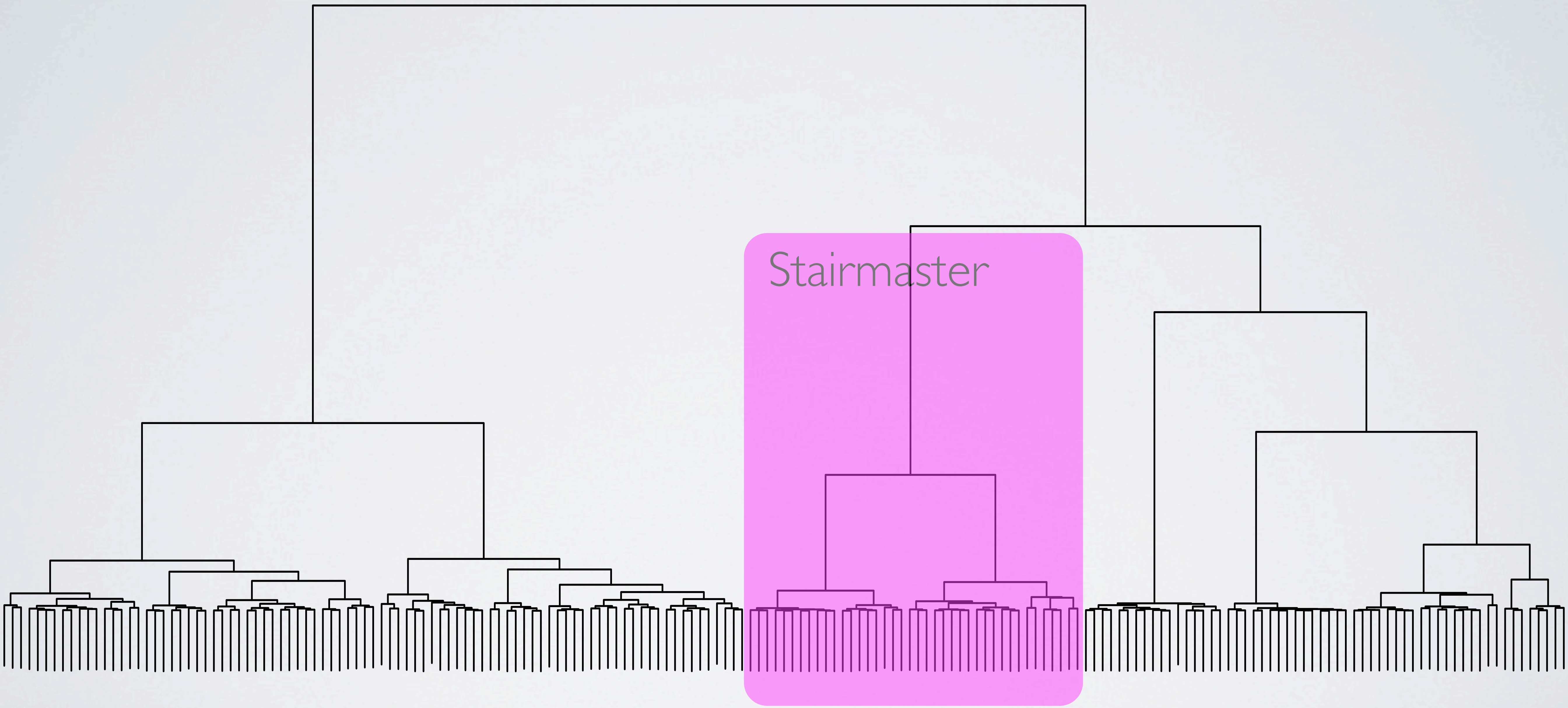
UNSUPERVISED LEARNING

- We have a method for computing the similarity (difference) between two segments of data.
- Treat this as a distance function, and use clustering techniques for data exploration.

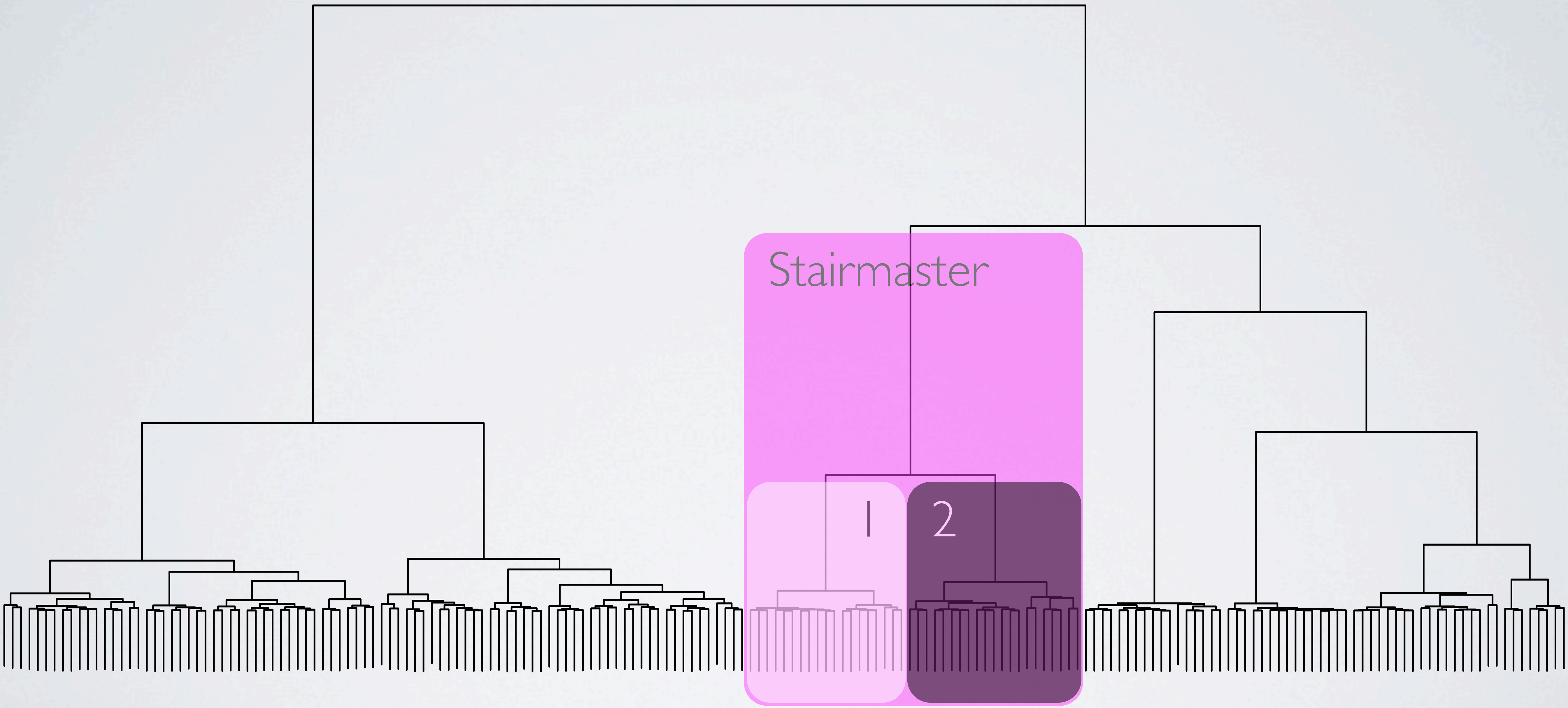
HIERARCHICAL CLUSTERING (EXERCISE ROUTINE DATA)



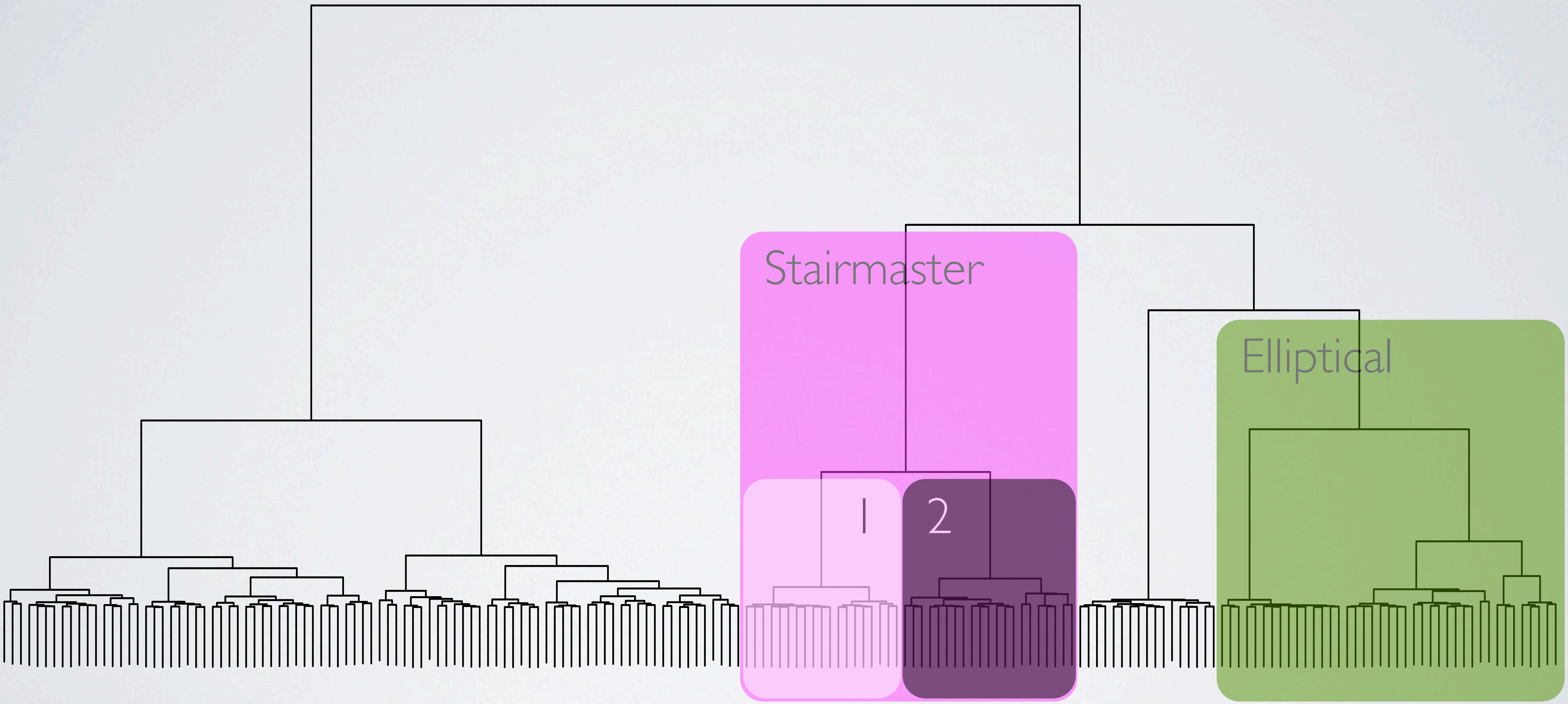
HIERARCHICAL CLUSTERING (EXERCISE ROUTINE DATA)



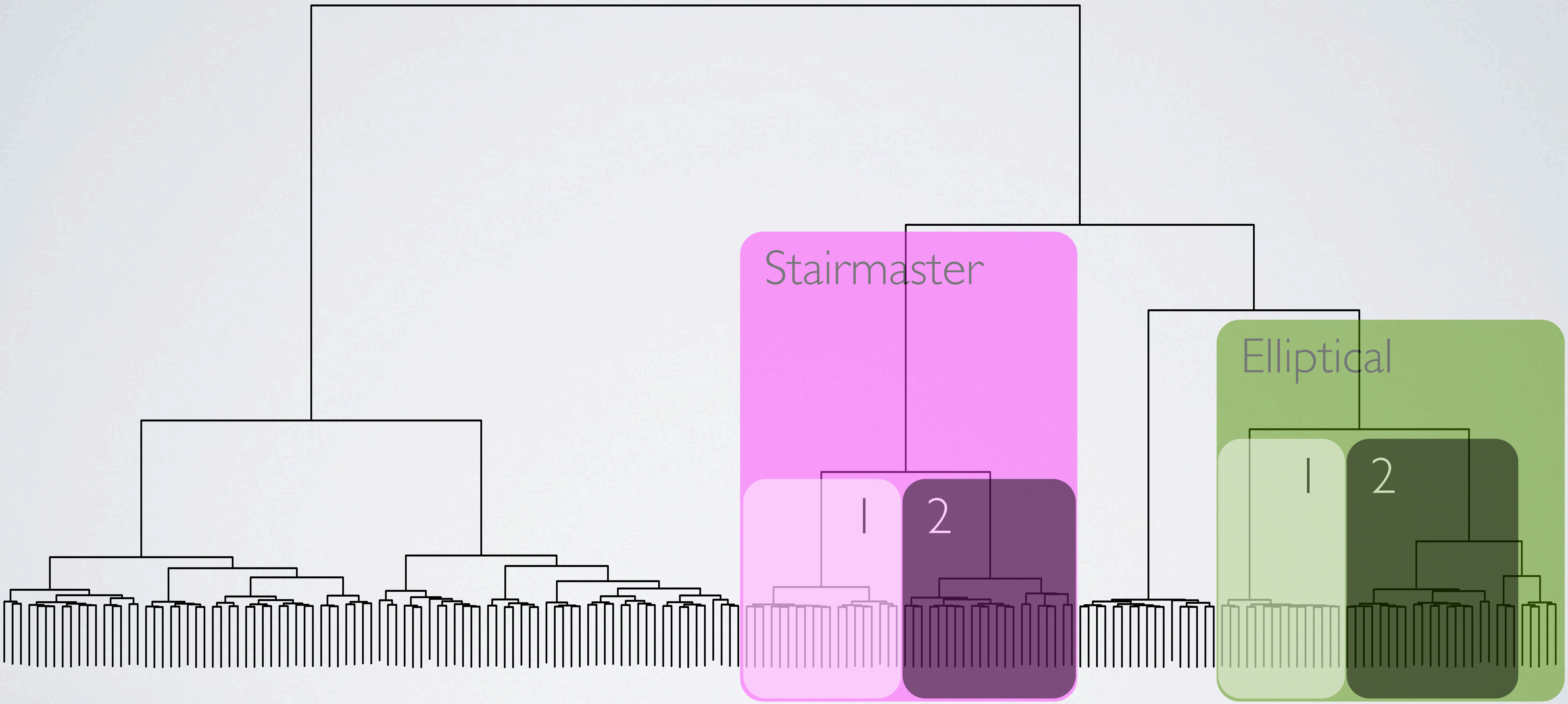
HIERARCHICAL CLUSTERING (EXERCISE ROUTINE DATA)



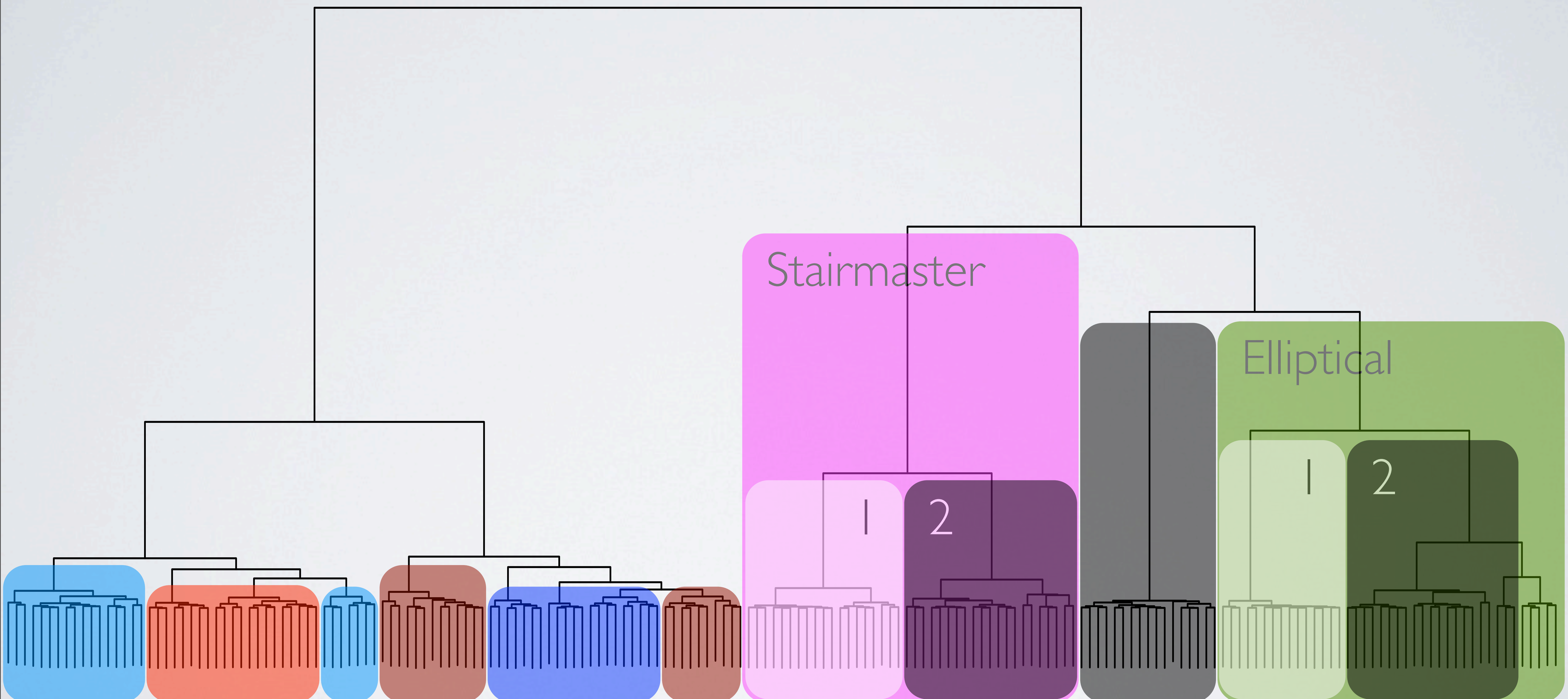
HIERARCHICAL CLUSTERING (EXERCISE ROUTINE DATA)



HIERARCHICAL CLUSTERING (EXERCISE ROUTINE DATA)



HIERARCHICAL CLUSTERING (EXERCISE ROUTINE DATA)



SEMI-SUPERVISED LEARNING

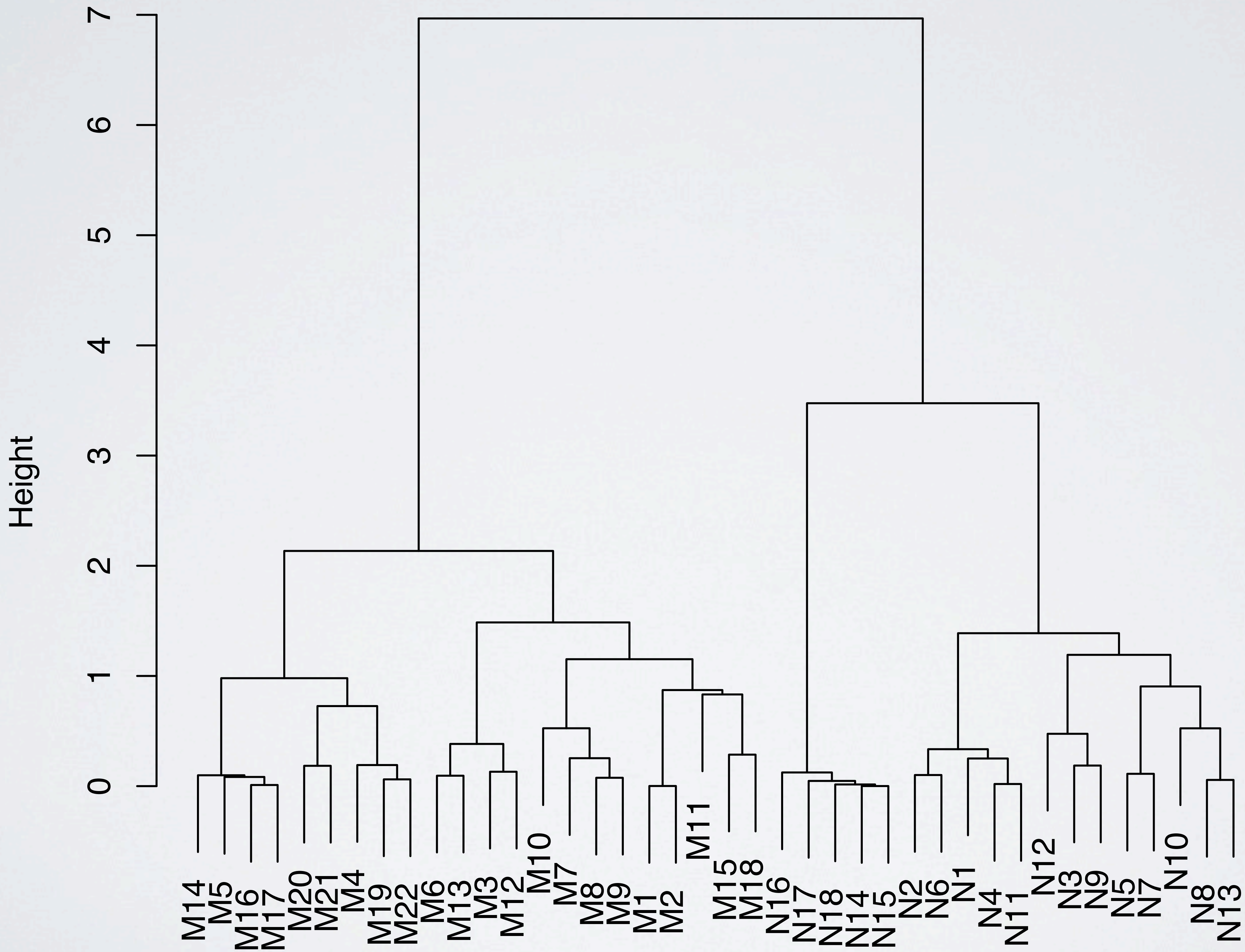
- With only 1 label per subject-activity pair, we can correctly label 87% of the unlabeled data. Jumps to 91% with 3 labeled examples per subject-activity pair.

- With only 2 labels per activity for only one subject, we can correctly label 86% of the unlabeled data for all subjects.

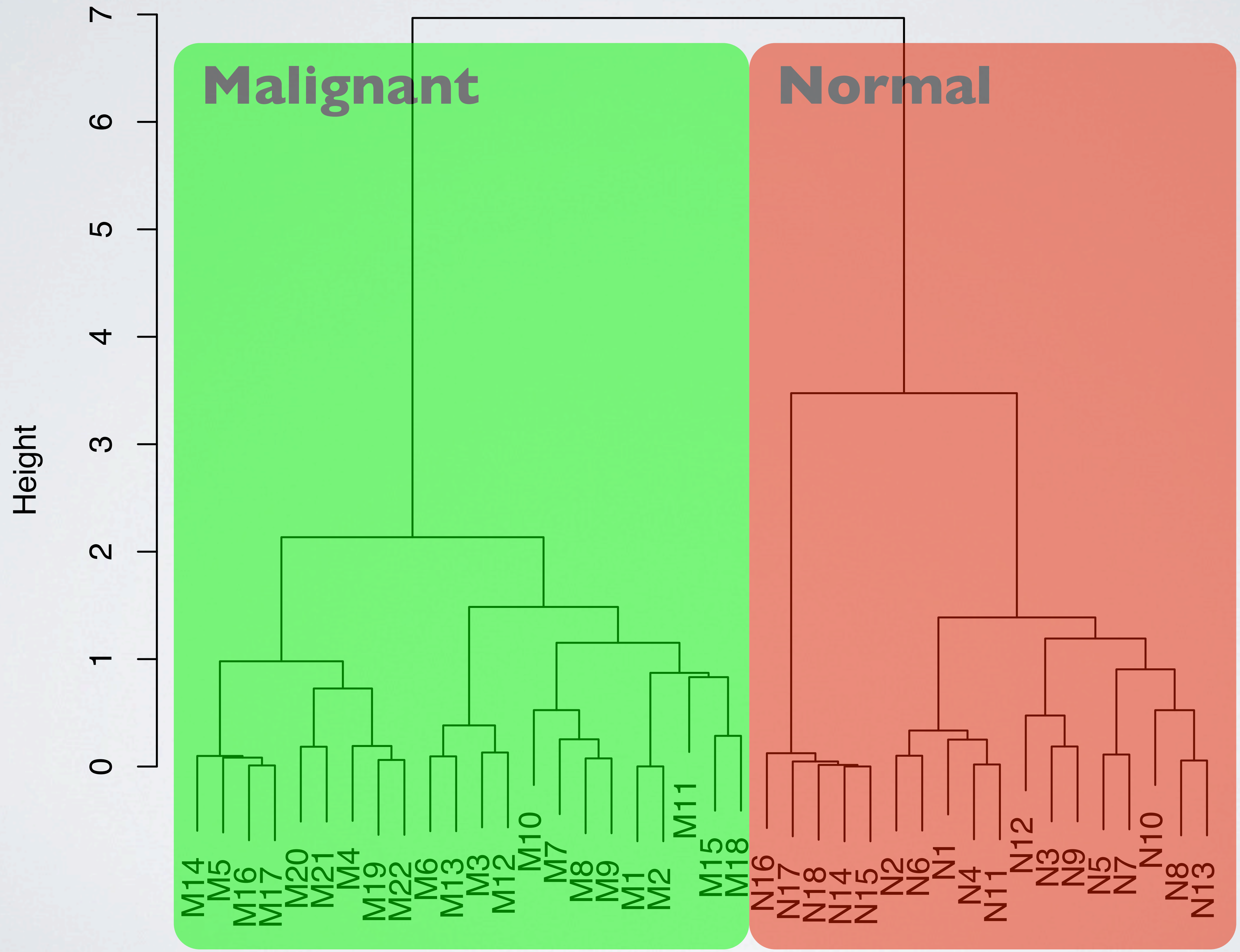
CLUSTERING ECG DATA

- ECG Data from PhysioNet ECG Database.
- 40 two-minute time series, 18 from people with normal sinus rhythm, 22 having malignant ventricular arrhythmia.
- Kalpakis et al. (2001) tried clustering with a number of different distance measures.
- Best results reported: 3 malignant mislabeled, 1 normal mislabeled.
- Authors: Mislabeled malignant traces "look more similar to the normal time-series than to the malignant arrhythmia time-series."

CLUSTERING ECG DATA



CLUSTERING ECG DATA



CONCLUSION

- Efficient way to build light-weight models of time series data.
- Efficient feature extraction algorithm.
- With no noise, we have nice theoretical properties.
- Seems to hold up well on real (noisy) data.
- Currently looking into novel activity detection, more applications (possibly in the medical domain), more sensors (BodyMedia Armbands...David?).

THANKS! QUESTIONS?

References:

J. Frank, S. Mannor, and D. Precup. Activity and Gait Recognition with Time-Delay Embeddings. *AAAI*, 2010.

J. Frank, S. Mannor, and D. Precup. A Novel Similarity Measure for Time Series Data with Applications to Gait and Activity Recognition. *UBICOMP adjunct proceedings*, 2010.

K. Kalpakis, D. Gada, and V. Puttagunta. Distance measures for effective clustering of ARIMA time-series. *ICDM*, 2001.

J. Lester, T. Choudhury, and G. Borriello. A practical approach to recognizing physical activities. *LNCS*, 2006.

F. Takens. Detecting strange attractors in turbulence. *Dynamical Systems and Turbulence.* 1981.

H. Kantz and T. Schreiber. *Nonlinear Time Series Analysis.* 2004.

TDE Code and Gait Data available on my website:

http://www.cs.mcgill.ca/~jfrank8/

This work is supported by NSERC and the Israel Science Foundation.

WHY IS THIS HARD?

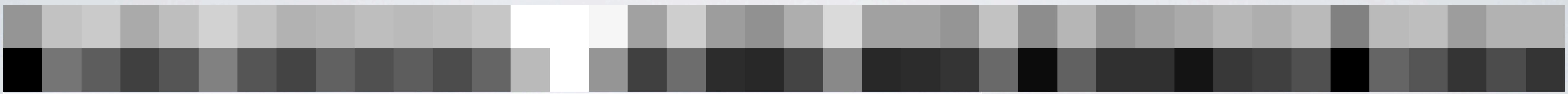
- Time series data is difficult to deal with in general.
- Data is non-stationary, so looking at spectrum doesn't really work.
- Periods aren't very different between activities (e.g., running, walking, both approximately 1Hz).
- Real-world data is noisy.
- Resources are limited, or at least we should consider them as such.

THE FINE PRINT

Theorem (Takens, 1981): If A is a d dimensional smooth compact manifold, then if $m > 2d$ and τ is chosen as to not coincide with any periodic orbits, then for almost every smooth observation function s, the map from $\mathbb{R}^k$ to the time-delay reconstruction in $\mathbb{R}^m$ is an embedding.

PERFECT ACCURACY, BUT...

Sometimes the top scoring models did not stand out.



Sometimes the winner was clear.

- In terms of the empirical standard deviation over 5 runs:
  - Average difference between top two scores: 0.81
  - Average difference between top score and fifth score: 1.37
- Neither is statistically significant.
- Data collected in a controlled environment.

GAIT RECOGNITION RESULTS

- Baseline used 200 features from Lester et al. (2006) and Random Forest classifiers.

CLASS	0	1	2	3	4	5	6	7	8	9	10
TDEBOOST PRECISION	0.81	**0.85**	0.15	**0.50**	**0.10**	**0.29**	**0.26**	**0.26**	**0.30**	**0.88**	**0.68**
BASELINE PRECISION	**0.86**	0.68	**0.30**	0.33	0.02	0.04	0.02	0.01	0.16	0.06	0.03
TDEBOOST RECALL	**0.96**	0.46	0.21	**0.84**	**0.01**	**0.36**	**0.14**	**0.09**	**0.48**	**0.71**	**0.53**
BASELINE RECALL	0.94	**0.77**	**0.84**	0.56	0.00	0.00	0.02	0.00	0.34	0.13	0.07

CLASS	11	12	13	14	15	16	17	18	19	20
TDEBOOST PRECISION	0.58	0.04	**0.93**	**0.05**	**0.26**	**0.96**	0.22	**0.25**	**0.71**	**0.88**
BASELINE PRECISION	**0.95**	**0.35**	0.00	0.00	0.05	0.25	**0.37**	0.00	0.00	0.49
TDEBOOST RECALL	**0.60**	0.01	**0.52**	**0.01**	**0.61**	**0.78**	0.64	**0.61**	**0.59**	**0.57**
BASELINE RECALL	0.03	**0.14**	0.00	0.00	0.00	0.35	**0.78**	0.00	0.00	0.33

- This data is freely available on my website.

WORKOUT DATA