Weakly supervised machine learning for time series

Romain Tavenard (Université de Rennes)
@rtavenar

TSdays
Rennes, March 26th, 2019

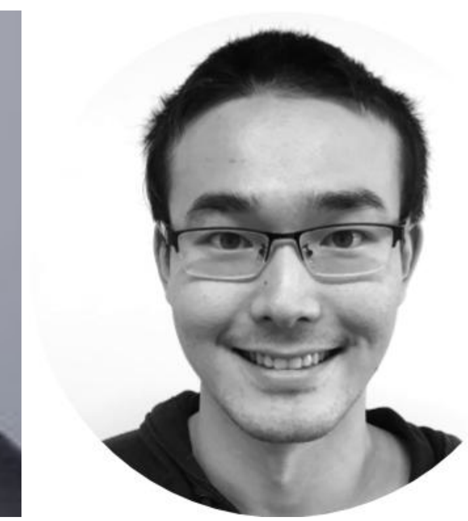
Weakly supervised machine learning for time series

R. Sperandio

A. Lods

T. Vayer

M. Rußwurm



Y. Wang

Part I

`tslearn`: A Python ML toolkit for time series

Part II

Weakly supervised learning: why & how?

Part III

Beyond time series: structured data in general

Part I

`tslearn`: A Python ML toolkit for time series

tslearn: A Python ML toolkit for time series

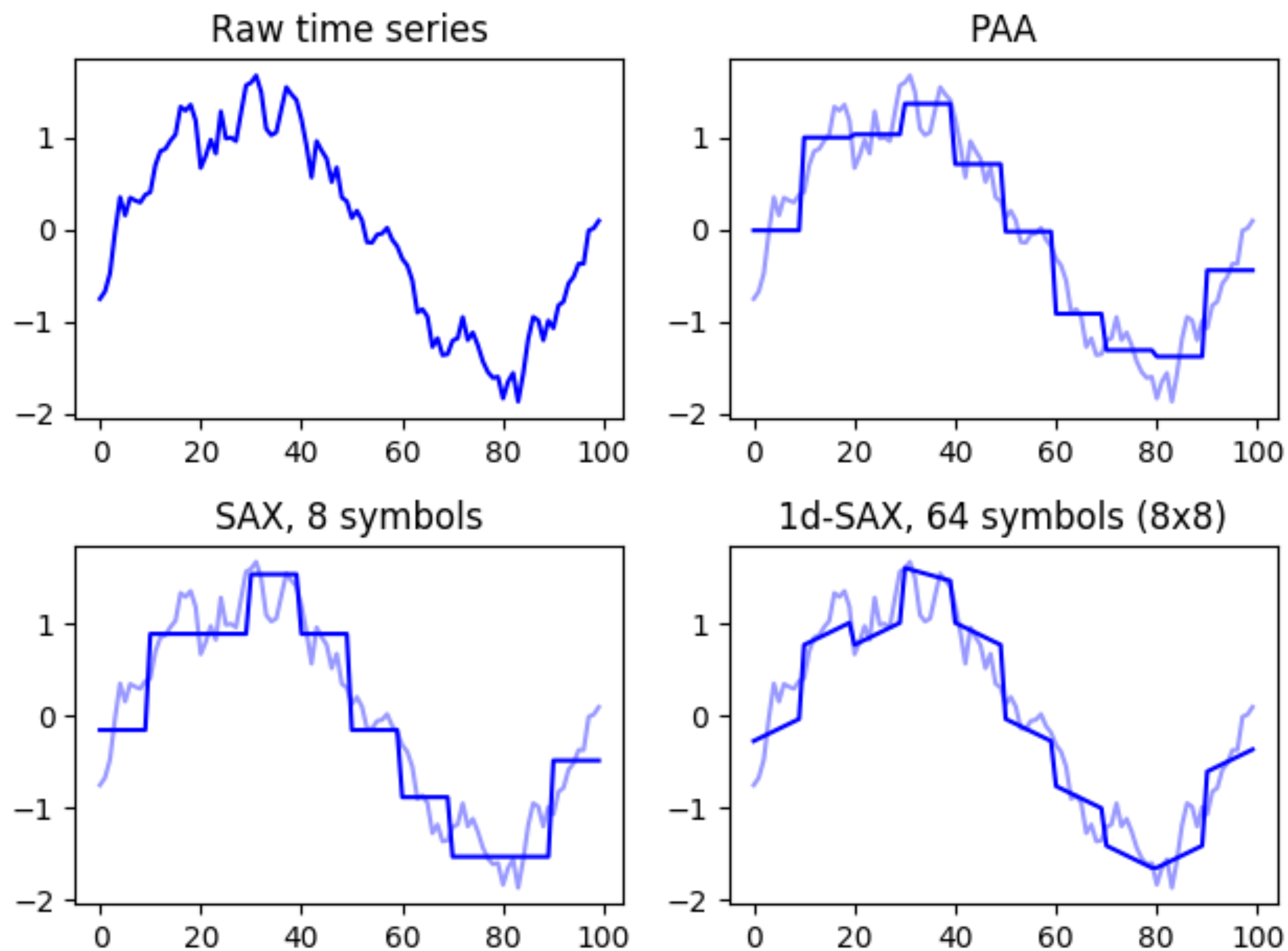
- Install it (via pip or conda)

pip install tslearn

- Play with it

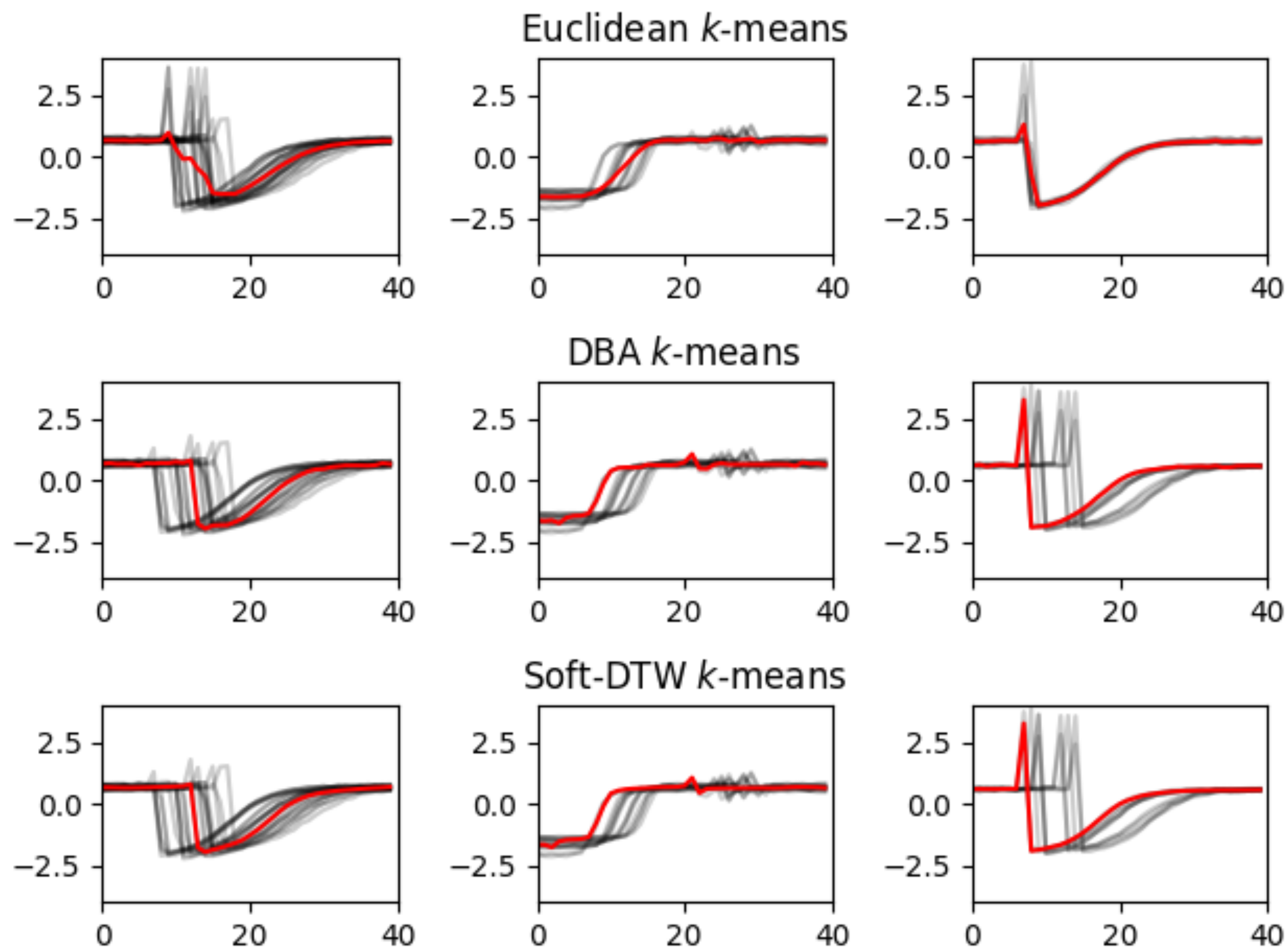
```
>>> from tslearn.datasets import UCR_UEA_datasets
>>> from tslearn.clustering import TimeSeriesKMeans
>>>
>>> X_train, y_train, X_test, y_test = UCR_UEA_datasets().load_dataset("TwoPatterns")
>>> print(X_train.shape)
(1000, 128, 1)
>>>
>>> km = TimeSeriesKMeans(n_clusters=3, metric="dtw")
>>> km.fit(X_train)
```

`tslearn`: A Python ML toolkit for time series
Samples from the gallery of examples [1/2]



Link to online notebook

`tslearn`: A Python ML toolkit for time series
Samples from the gallery of examples [2/2]



[Link to online notebook]

`tslearn`: A Python ML toolkit for time series Feel free to contribute

- All contributions are welcome (via github)
  - New feature requests
  - Bug reports
  - Bug fixes
  - Improved documentation

- Most contributors are volunteers
- 1-year dev funding scheduled in ANR MATS (applications welcome anytime)

Part II
Weakly supervised learning: why & how?

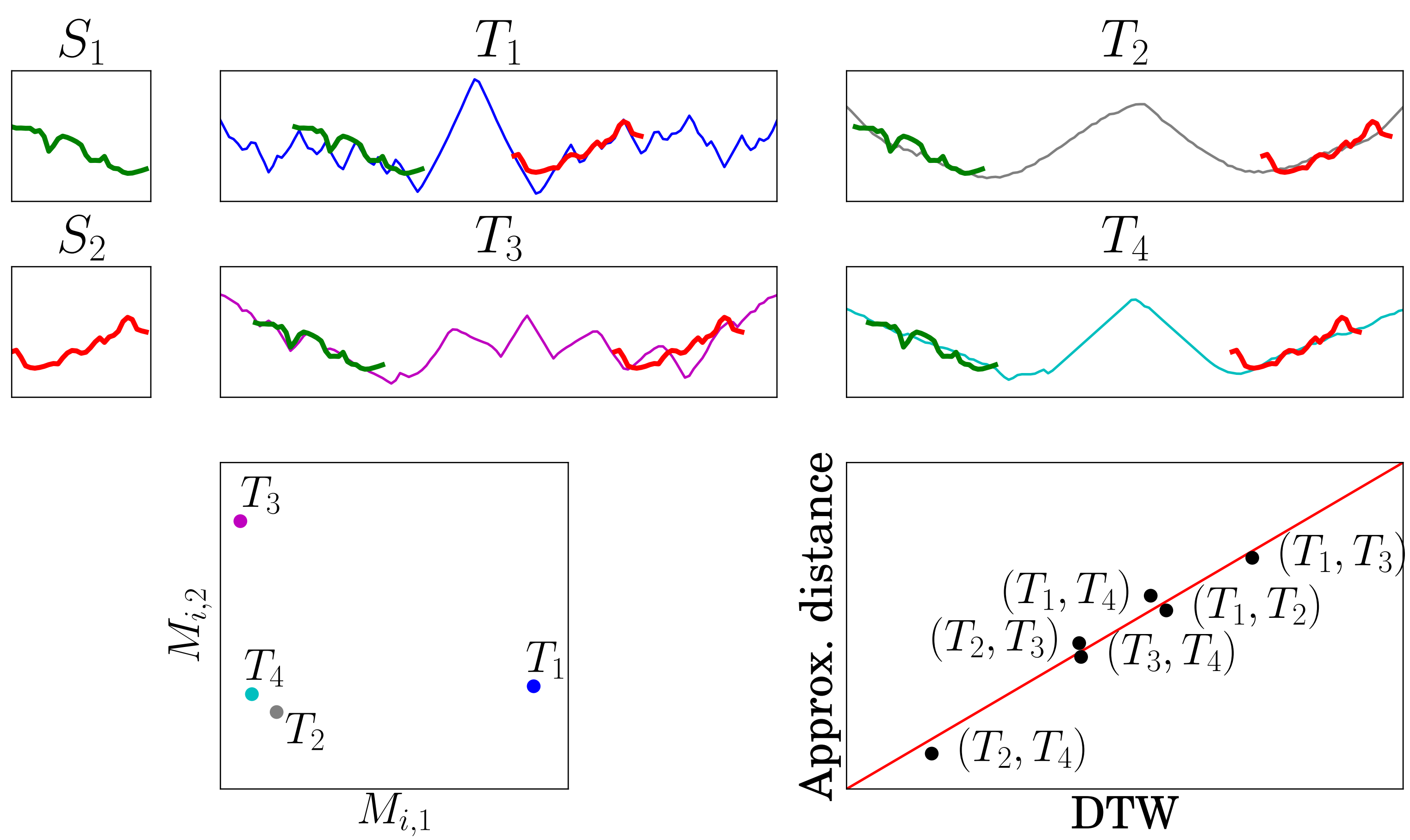
Remote sensing & ML

- Labelled training samples are…
  - Costly to acquire
  - Noisy
- Weakly supervised learning
  - Domain adaptation
    (not covered in this talk)
    - Across years
    - Across regions
    - Great stuff by Courty *et al.*
  - Representation learning
    - **Self-supervised approaches**

Image from esa.int

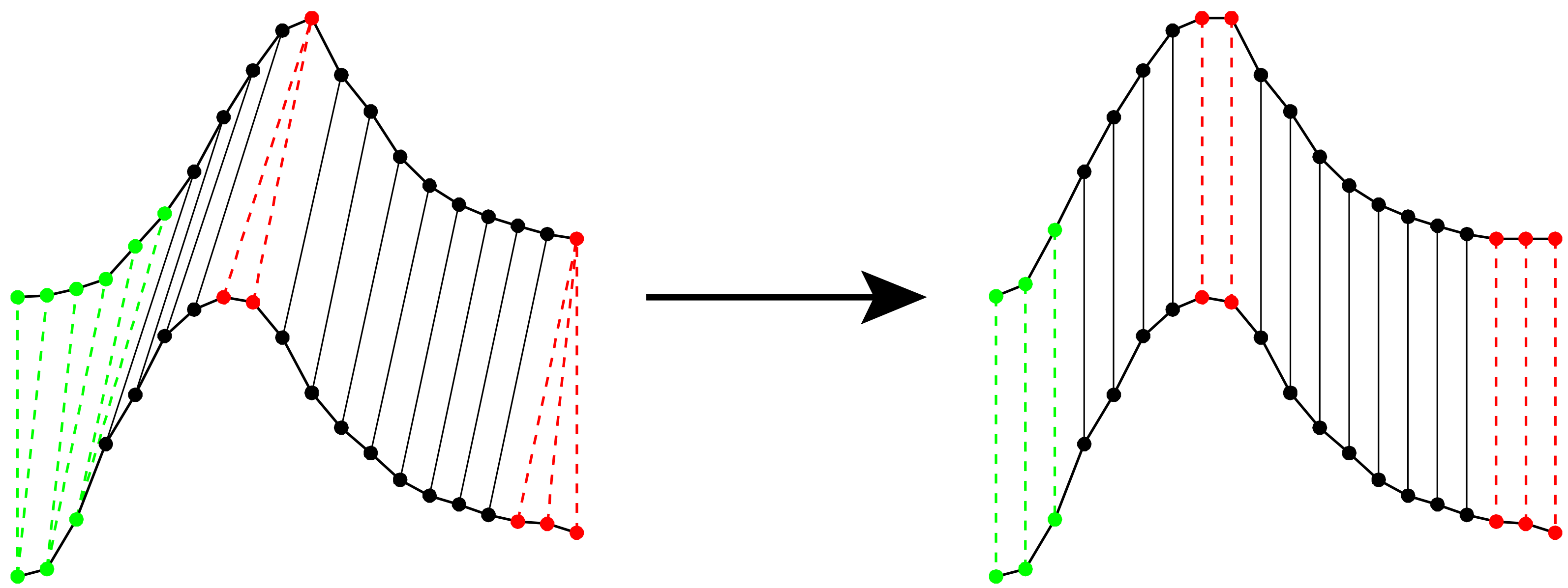
Learning DTW-Preserving Shapelets (LDPS) [Lods et al., 2017]

- **Self-supervised learning:** generate supervised information from the data itself
- **Here:** Learn a Shapelet Transform to mimic DTW



Learning DTW-Preserving Shapelets (LDPS)
Dynamic Time Warping: a tool for time series analysis

- Dynamic Time Warping (DTW)
    - *Elastic* similarity measure
    - Invariant to time shifts
    - Costly to compute



Learning DTW-Preserving Shapelets (LDPS)
k-means and Dynamic Time Warping

- **Issue:** *k*-means needs barycenters
  - Explicit formulation in Euclidean Space
  - No easy way for DTW

- Option 1: DBA k-means [Petitjean *et al.*, 2011]
- Option 2: Soft-DTW [Cuturi & Blondel, 2017]

Learning DTW-Preserving Shapelets (LDPS)
State-of-the-art: Time Series Shapelets

- Shapelets
  - **Discriminant** subseries
  - Learnt in [Grabocka *et al.*, 2014]

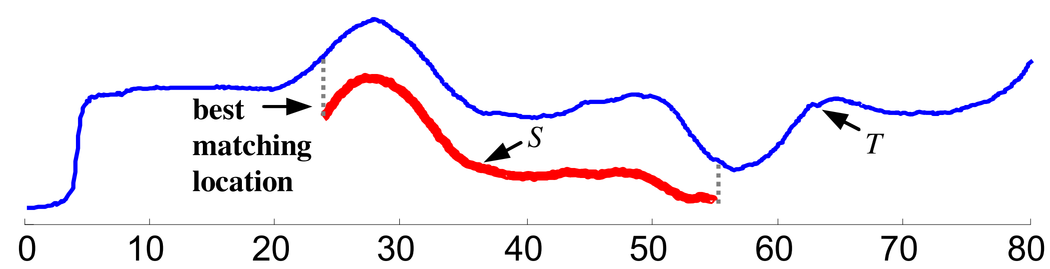


Illustration from [Ye & Keogh, 2009]

- Shapelet Transform [Hills *et al.*, 2014]
  - Efficient computation
  - Invariant to time shifts

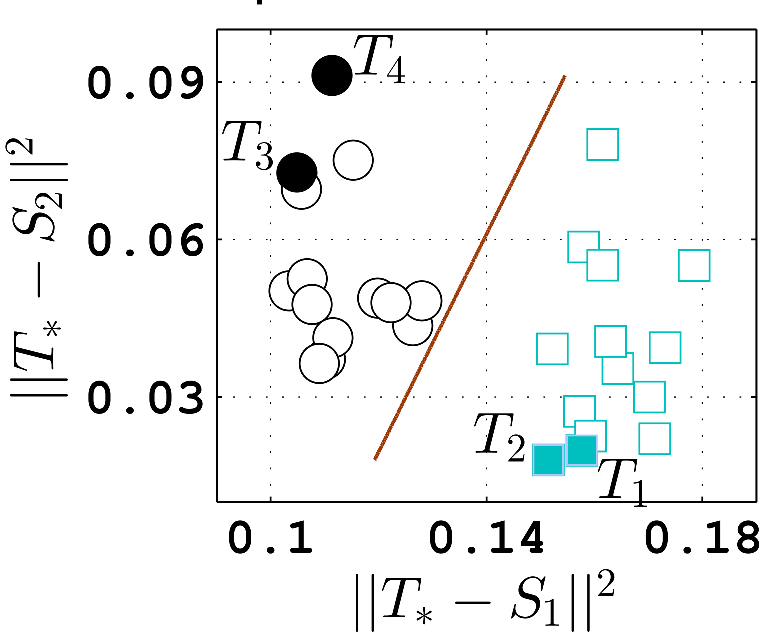
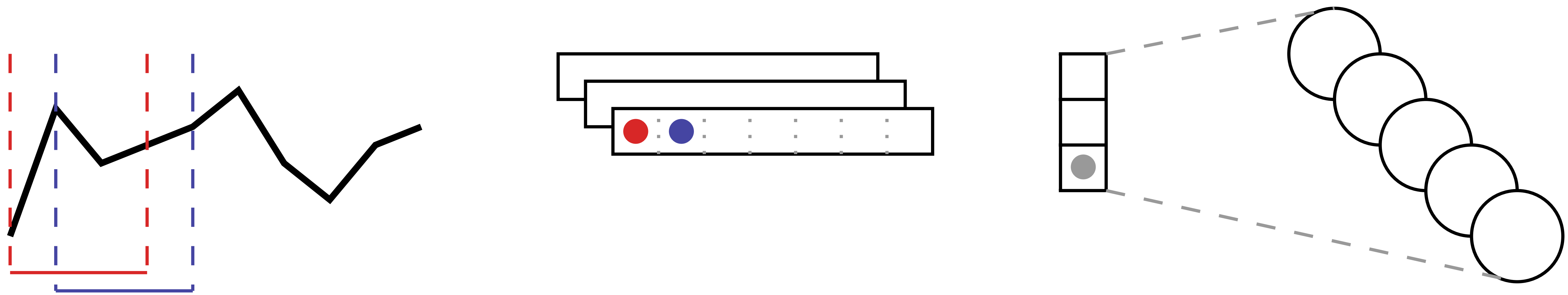


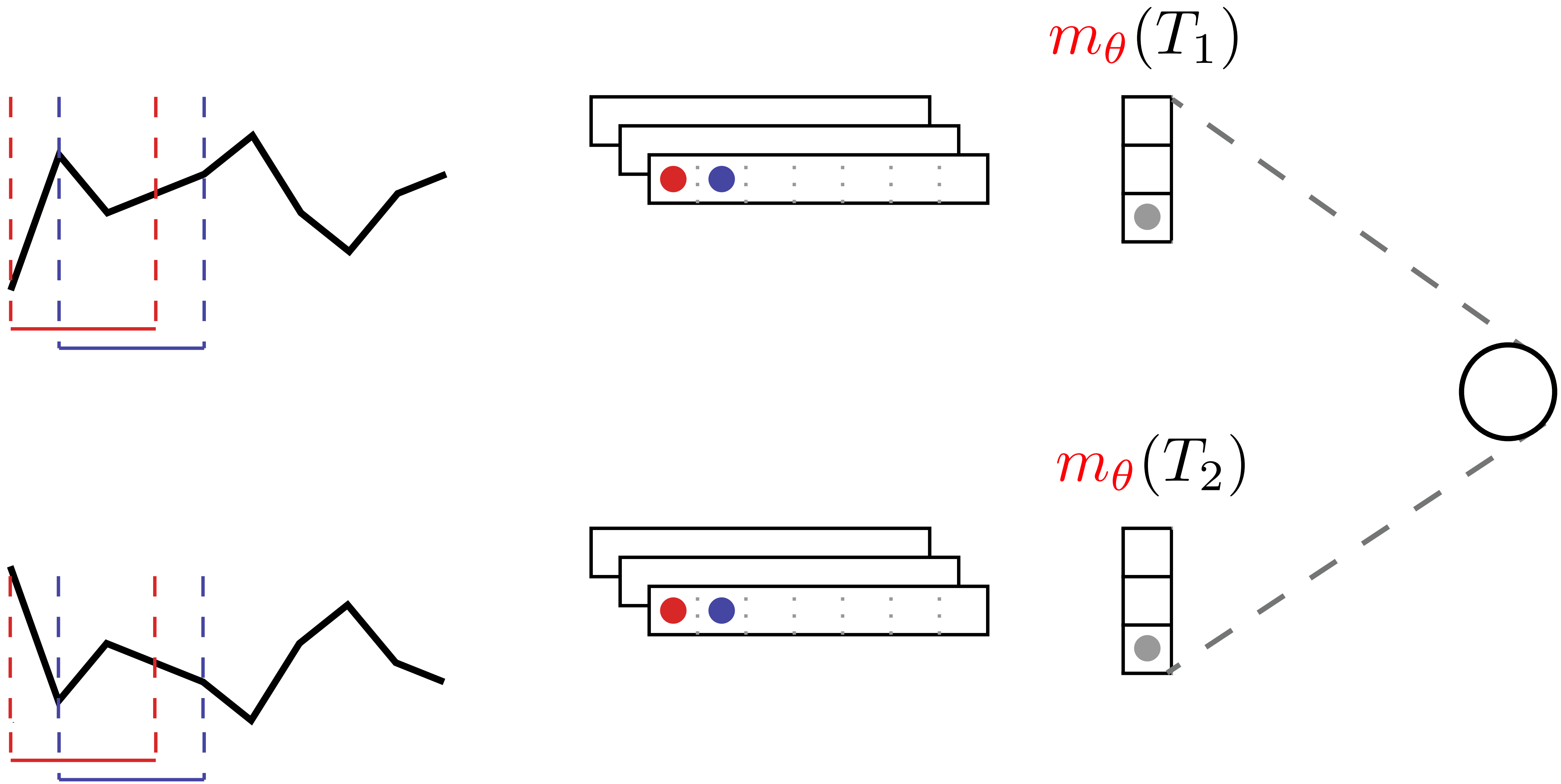
Illustration from [Grabocka *et al.*, 2014]

Learning DTW-Preserving Shapelets (LDPS)
Schematic view of our model



Learning Time Series Shapelets [Grabocka et al., 2014]

Learning DTW-Preserving Shapelets (LDPS)
Schematic view of our model



Learning DTW-Preserving Shapelets (LDPS)

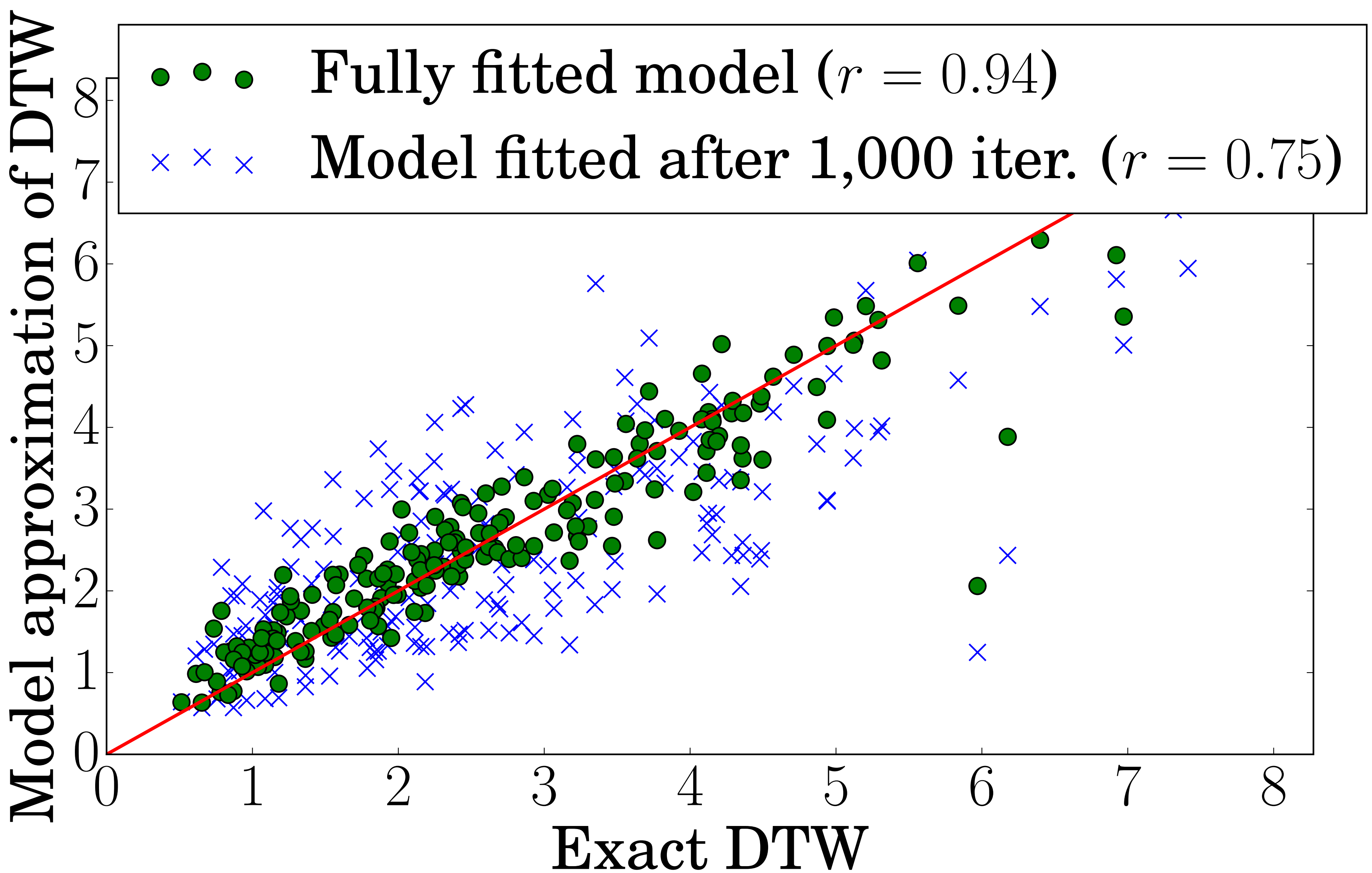
Learning DTW-Preserving Shapelets (LDPS) Problem formulation

- Loss function

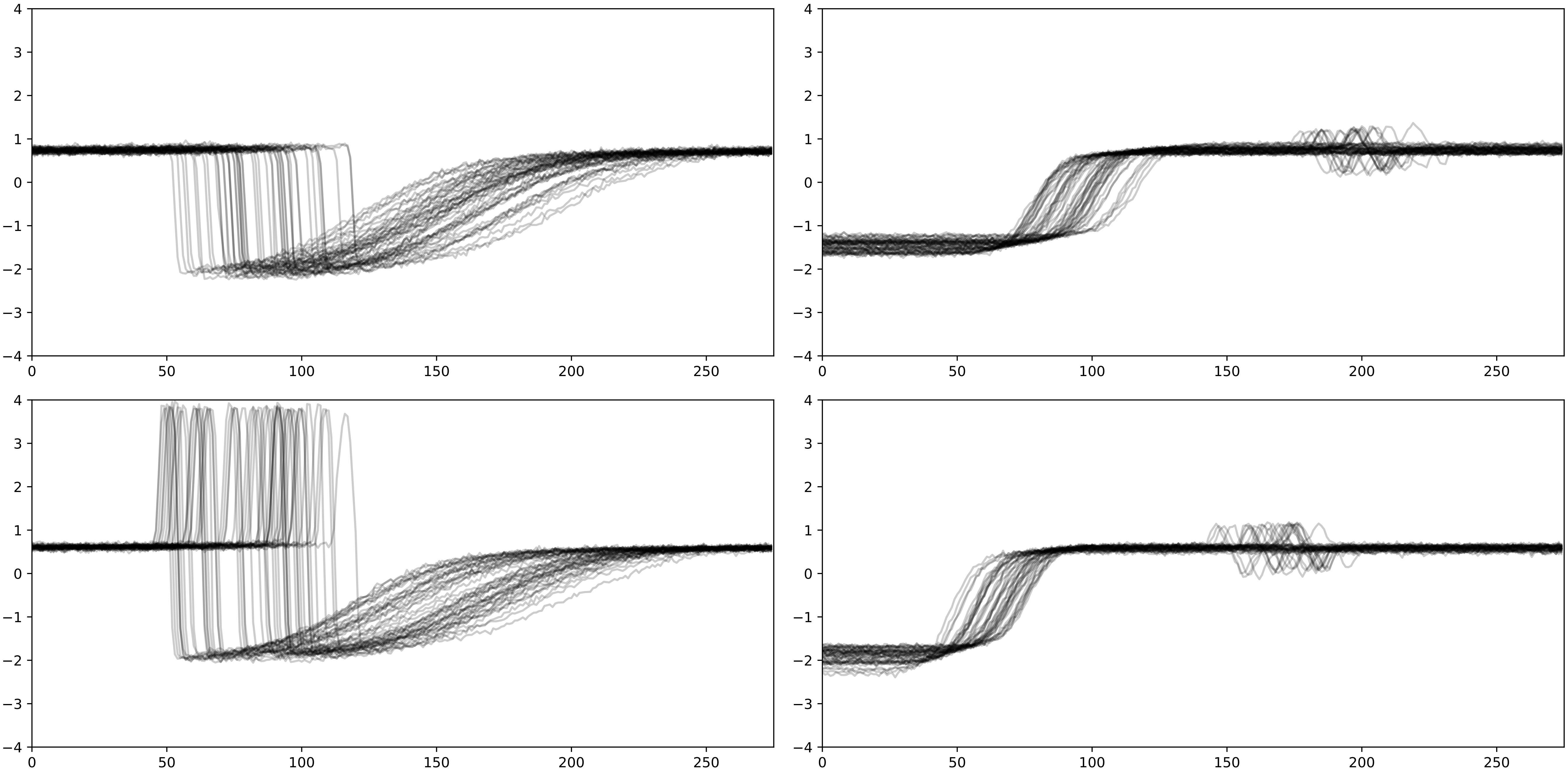
$$\mathcal{L}(\{T_i\}) \propto \sum_{i_1} \sum_{i_2} \left(DTW(T_{i_1}, T_{i_2}) - \beta \| m_\theta(T_{i_1}) - m_\theta(T_{i_2}) \|_2 \right)^2$$

- Optimize jointly on the mapping m_θ and the scaling factor β
- Self-supervised learning
  - No need for human annotation

Learning DTW-Preserving Shapelets (LDPS)
Experiment #1: Quality of fit

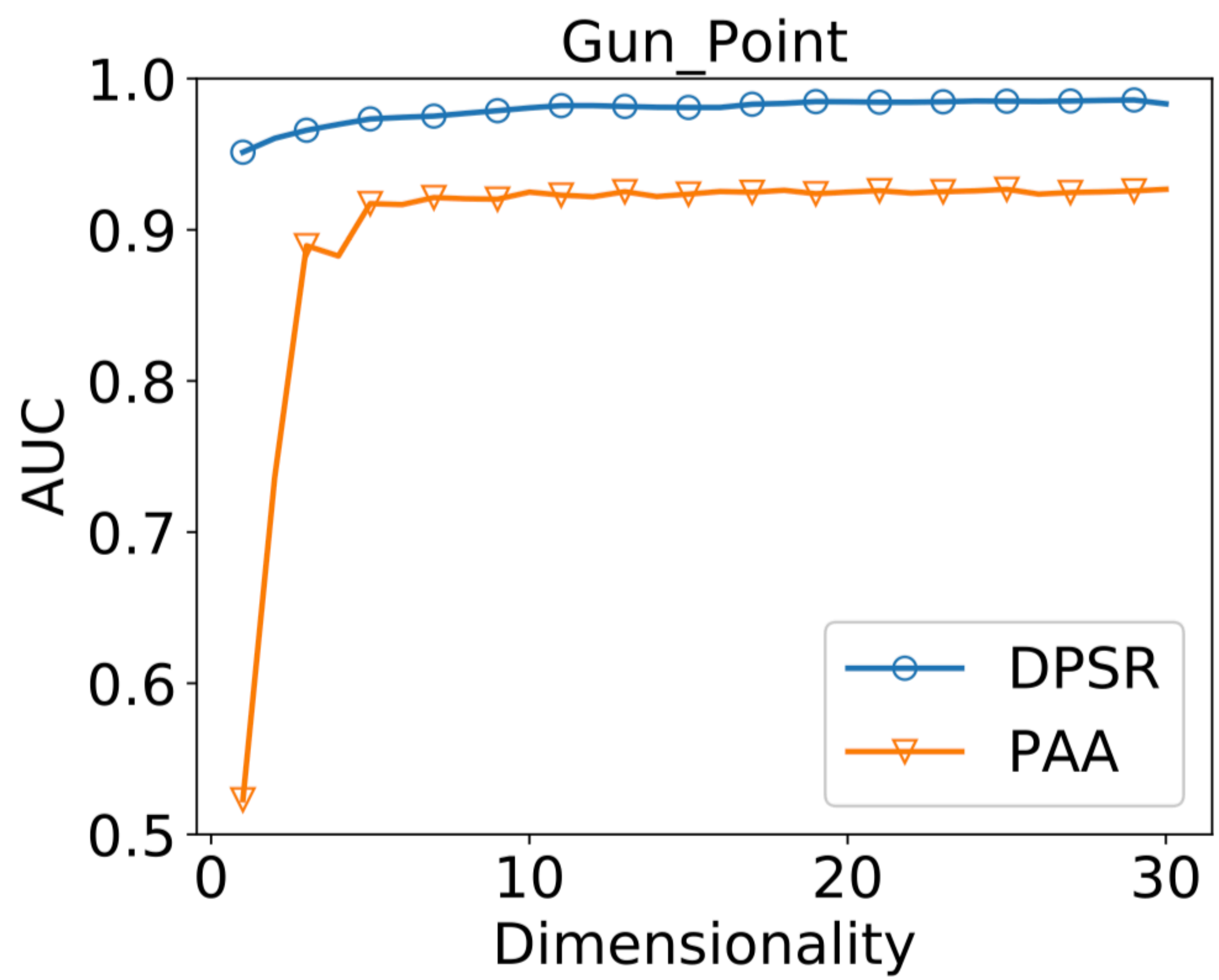


Learning DTW-Preserving Shapelets (LDPS) Experiment #2: Clustering using *k*-means



Learning DTW-Preserving Shapelets (LDPS) Experiment #3: Retrieval [Sperandio *et al.*, 2018]

- kNN search in a database
- DTW as a target metric (but too costly)



Learning DTW-Preserving Shapelets (LDPS) Take-Away slide

- Learning DTW-Preserving Shapelets (Arnaud Lods)
    - embeds time series in a metric space
        - useful for many data analysis tasks
    - uses only self-supervision
- Perspectives
    - can be extended to any similarity measure
      (pick the best for your data!)
    - LDPS + semi-supervised learning:
      Learn a mix between DTW and a good classifier
      (mixed loss function)

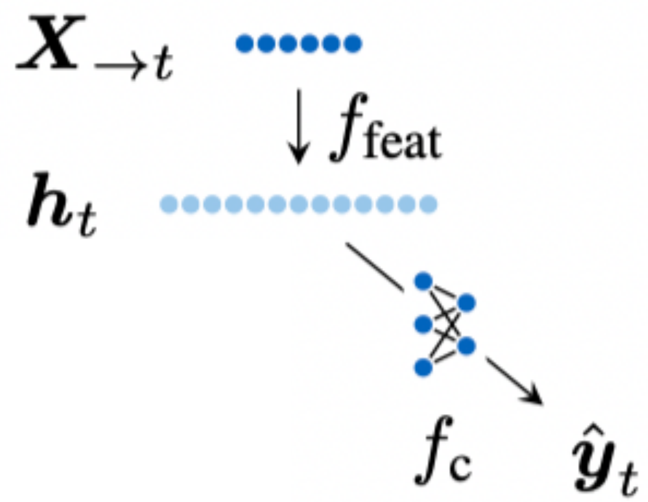
Part III
Beyond time series: structured data in general

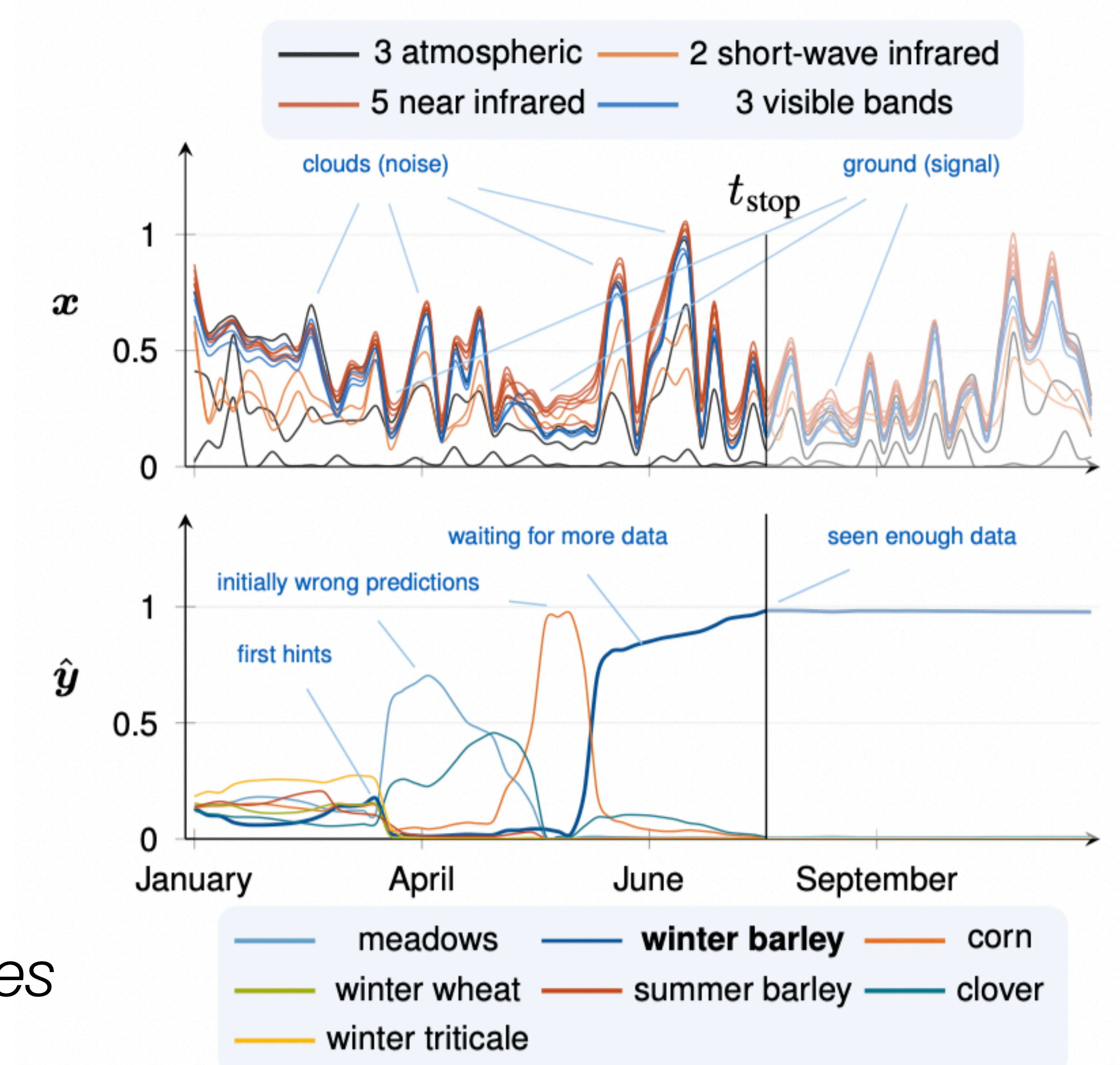
Beyond Time Series: Structured data in general
Example [1/2]: GPT-2 Language Model

We've trained a large-scale unsupervised language model which generates coherent paragraphs of text, achieves state of the art performance on many language modeling benchmarks, and performs rudimentary reading comprehension, machine translation, question answering, and summarization —

Twitter status, @openai, Feb 14th, 2019

Beyond Time Series: Structured data in general
Example [2/2]: Early classification of time series

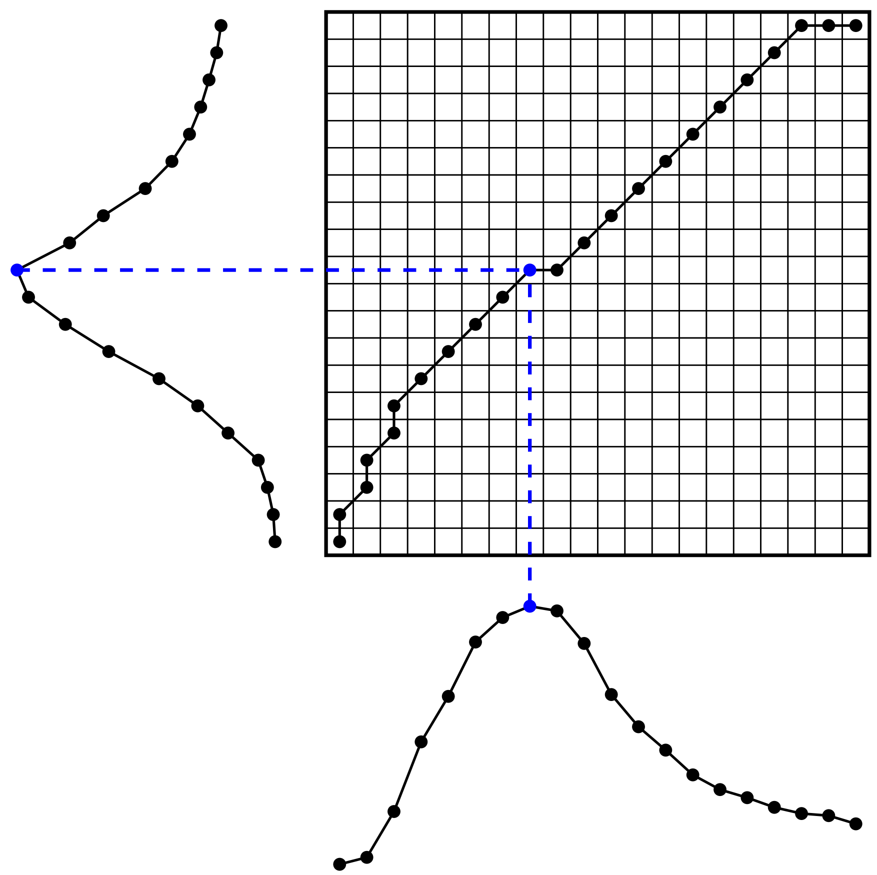




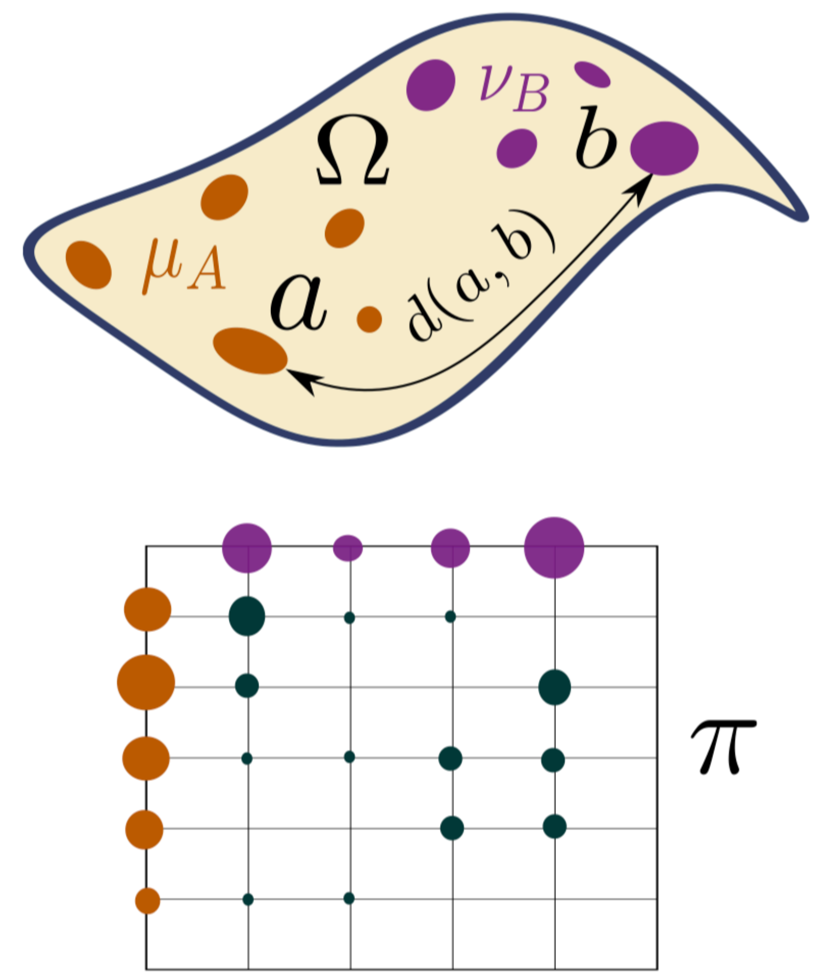
M. Rußwurm et al.
End-to-end Learning for
Early Classification of Time Series

Beyond Time Series: Structured data in general DTW & Optimal Transport (OT)

DTW problem = OT problem + strong structural constraints

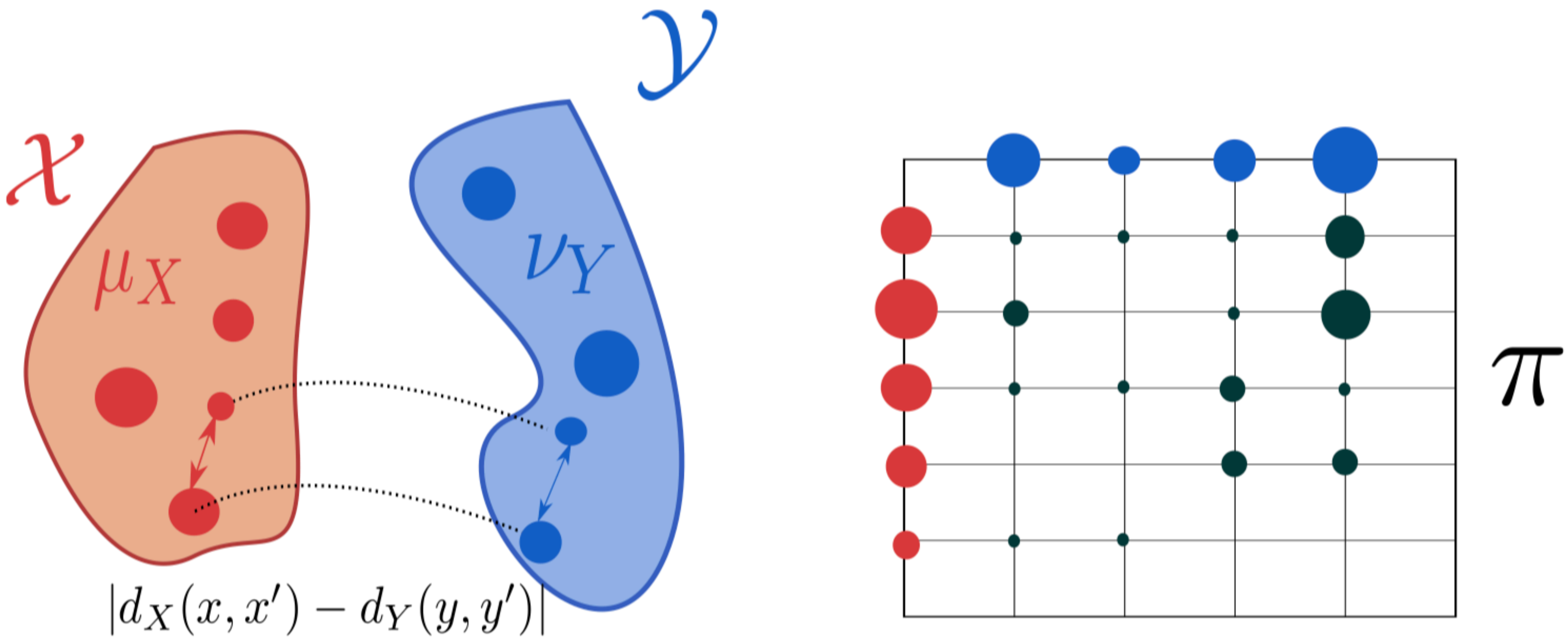


Dynamic Time Warping



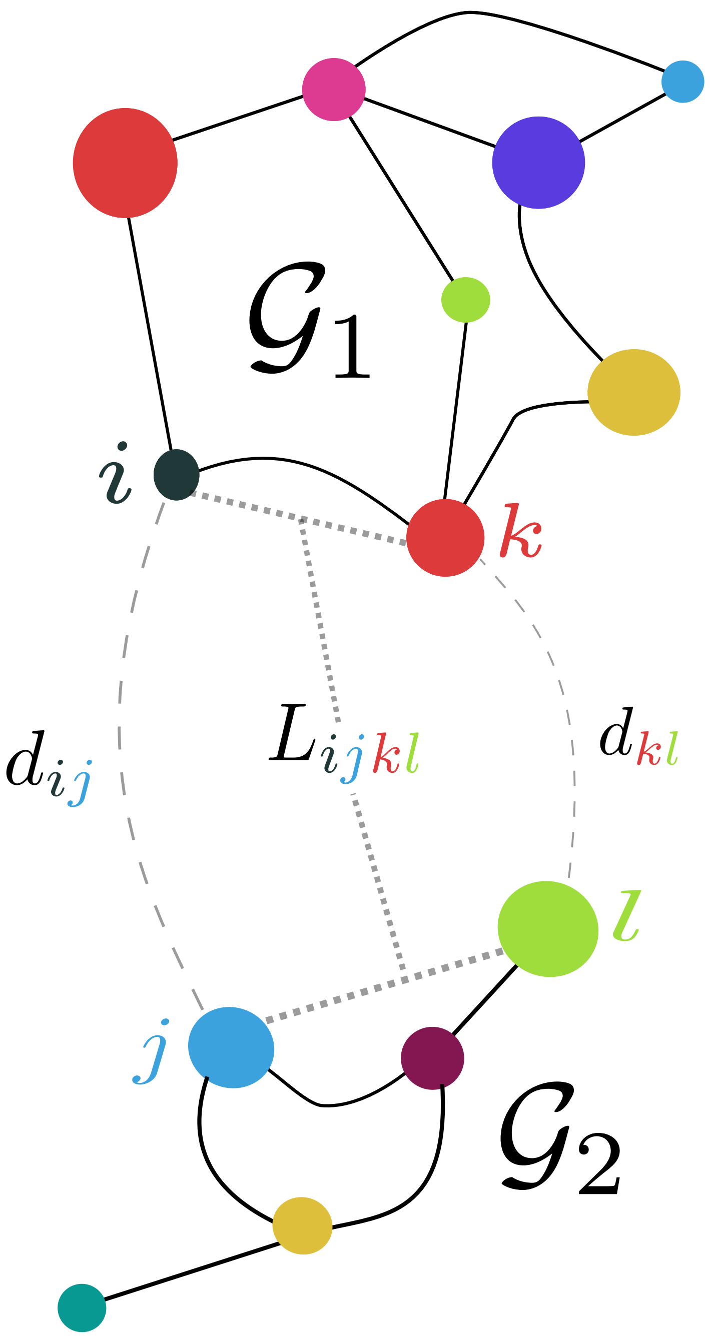
Wasserstein distance

Beyond Time Series: Structured data in general
Optimal Transport & Structure



Gromov-Wasserstein distance

Beyond Time Series: Structured data in general
Optimal Transport & Structure — cont'd



Wasserstein distance

$$\min_{\pi} \sum_{i,j} d_{ij} \pi_{i,j}$$

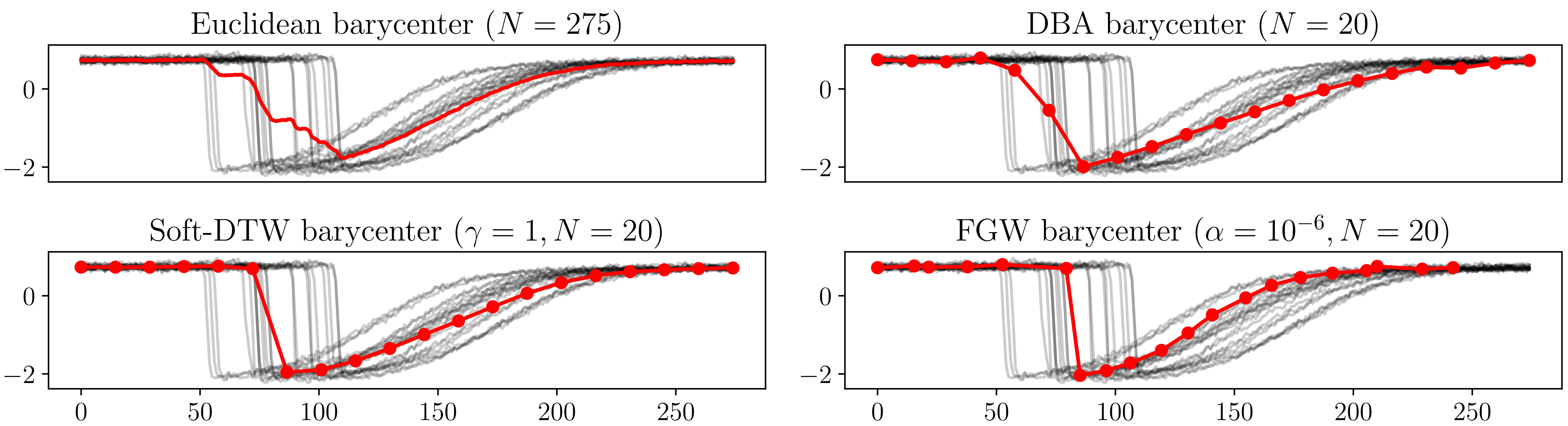
Gromov-Wasserstein distance

$$\min_{\pi} \sum_{i,j,k,l} L_{ijkl} \pi_{i,j} \pi_{k,l}$$

Fused-Gromov-Wasserstein (FGW) distance

$$\min_{\pi} \sum_{i,j,k,l} \left[(1-\alpha)\, d_{ij} + \alpha L_{ijkl}\right] \pi_{i,j} \pi_{k,l}$$

Beyond Time Series: Structured data in general
FGW & time series barycenters

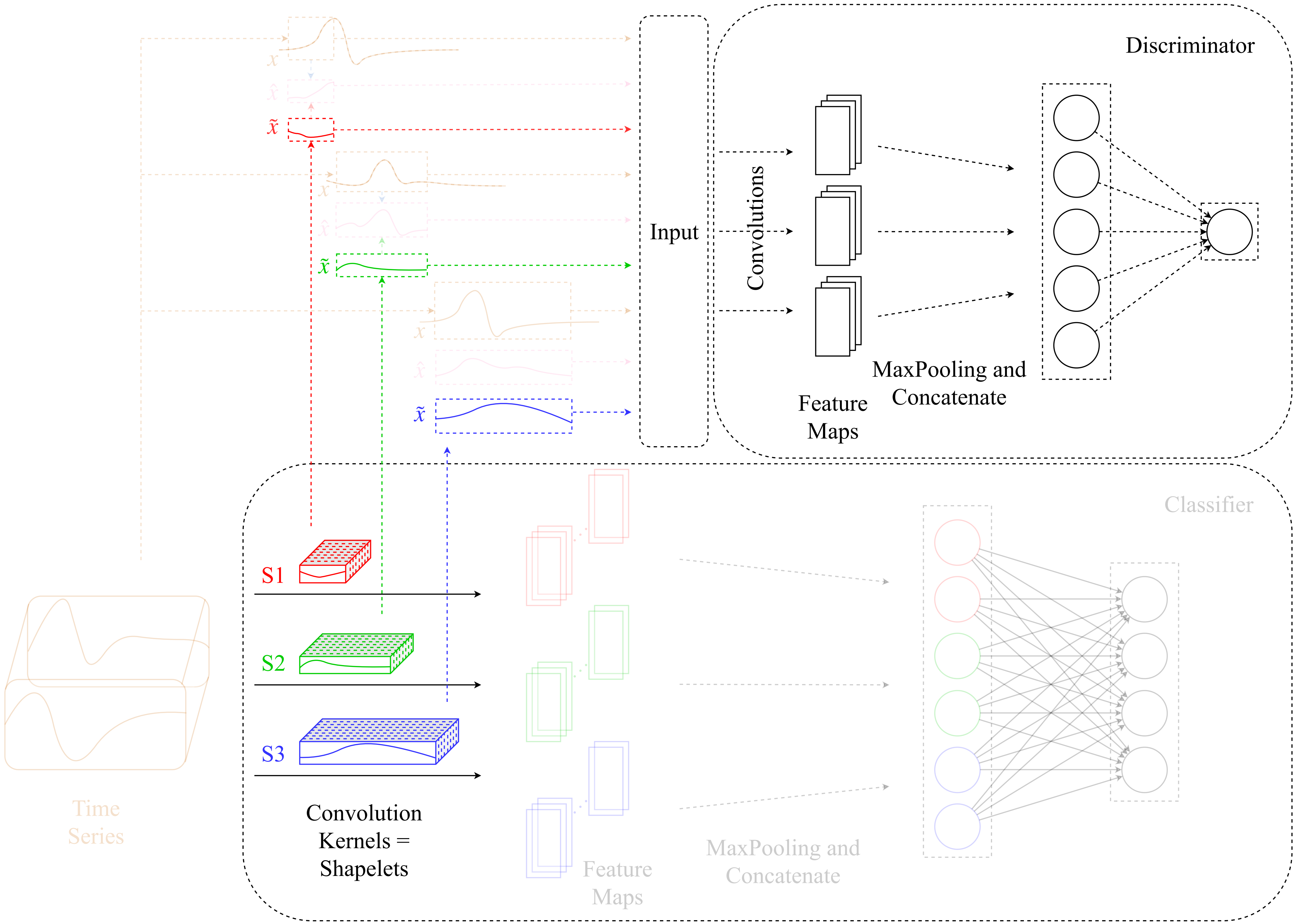


Beyond Time Series: Structured data in general
Take-Away slide

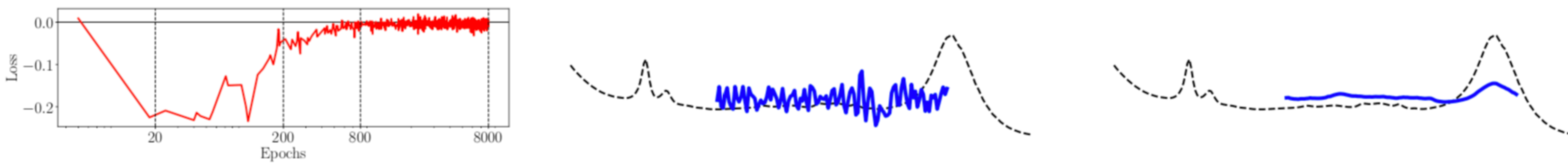
- Fused Gromov-Wasserstein (Titouan Vayer)
  - is a proper distance metric
  - is differentiable
  - can be used for
    - clustering of structured data with explicit barycenters
    - structured data classification
    - …
- Structured data
  - is awesome for weakly supervised approaches (cf. GPT-2)

Part IV
If I had more time…

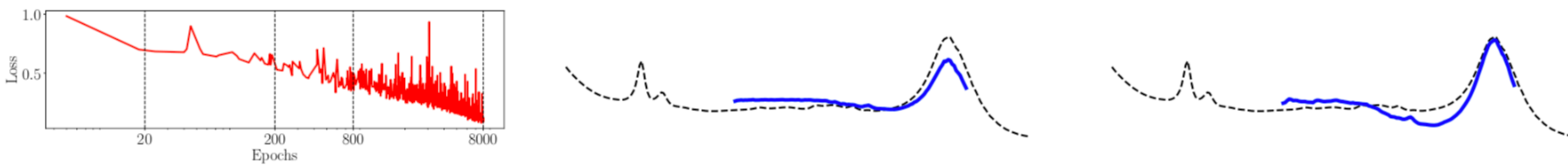
Learning Interpretable Shapelets through Adversarial Regularization [Yichang Wang — Work In Progress]



Learning Interpretable Shapelets through Adversarial Regularization [Yichang Wang — Work In Progress]



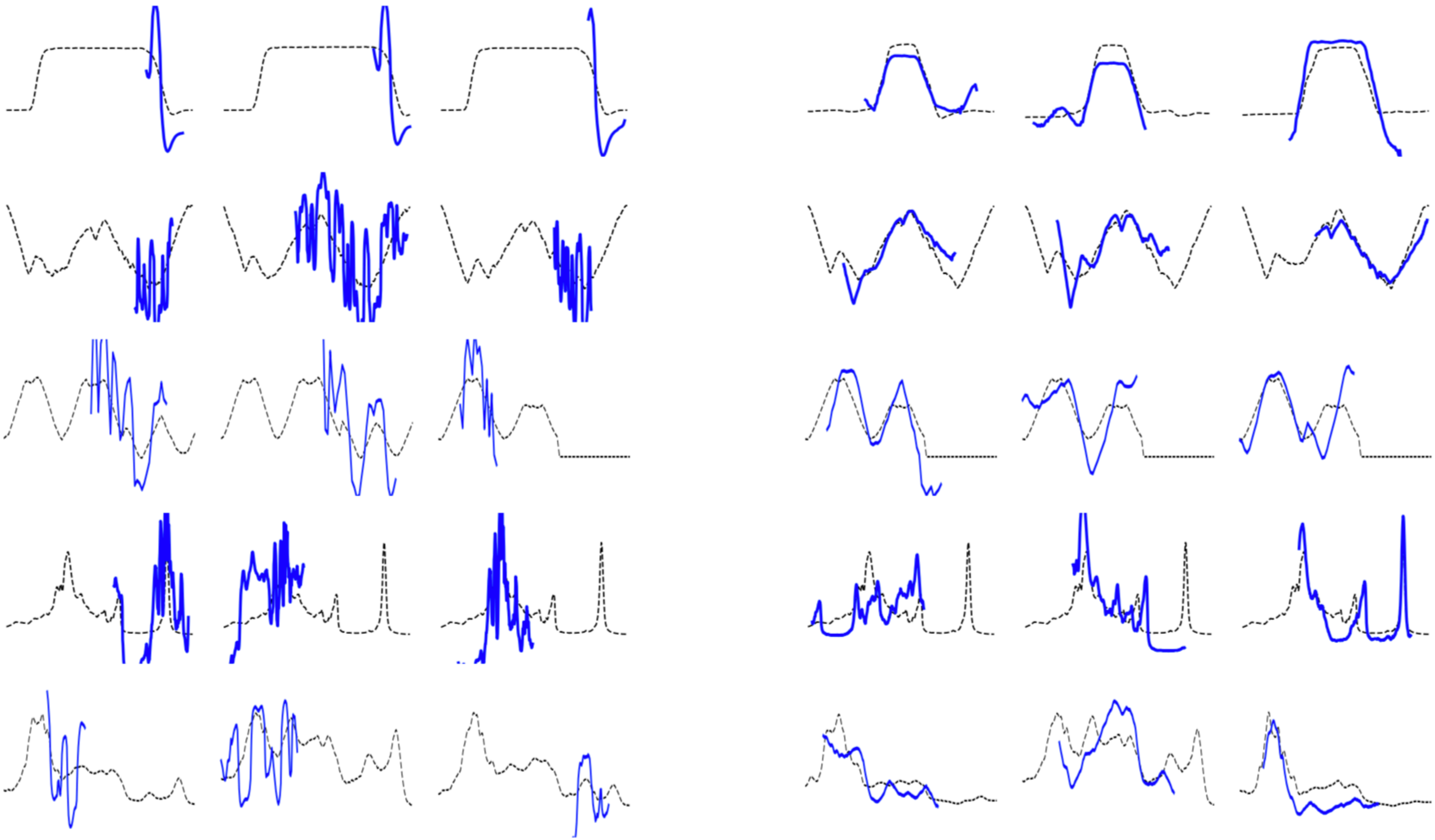
(a) **Wasserstein loss** L_d (b) **Shapelet at epoch** 20 (c) **Shapelet at epoch** 200



(d) **Cross-entropy loss** L_c (e) **Shapelet at epoch** 800 (f) **Shapelet at epoch** 8000

Figure 4: Illustration of the evolution of a shapelet during training (for the Wine dataset).

Interpretable Shapelets
[Yichang Wang — Work In Progress]



(a) **Learning Shapelets [9]** (b) **AI↔PR-CNN**

Conclusion & *Perspectives*

- `tslearn` needs you ;-)
  - Or, at least, OSS needs you
- Representation learning
  - Learn to mimic a target metric
  - *Learn a representation from structural information (generative models)*
- Interpretable models for Time Series
  - Useful for unsupervised settings

References

- R. Tavenard.
  tslearn: A machine learning toolkit dedicated to time-series data.
  https://github.com/rtavenar/tslearn

- A. Lods et al. *Learning DTW-Preserving Shapelets.* IDA 2017.

- M. Rußwurm et al. *End-to-end Learning for Early Classification of Time Series.* ArXiv 2019.

- T. Vayer et al. *Optimal Transport for structured data with application on graphs.* ArXiv 2019.