A generalized signature method for multivariate time series classification

Second Symposium on Machine Learning and Dynamical Systems, Fields Institute, Toronto

Adeline Fermanian

September 25th 2020

Joint work with

Benoît Cadre

University Rennes 2

Gérard Biau

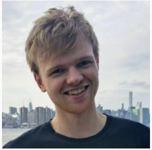
Sorbonne University

Terry Lyons

University of Oxford

Patrick Kidger

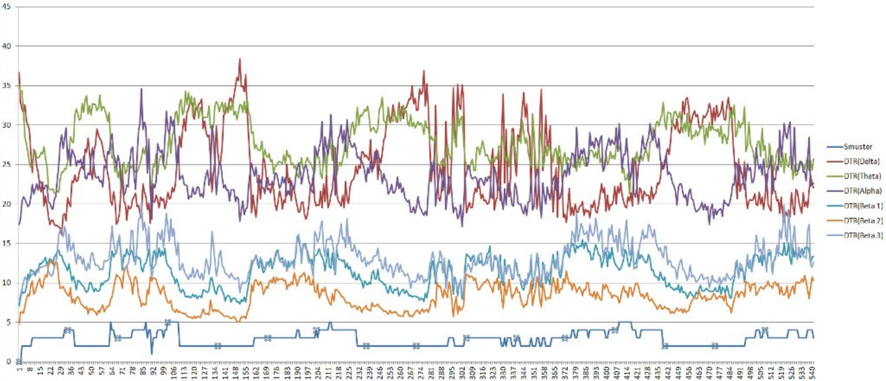
University of Oxford



James Morrill

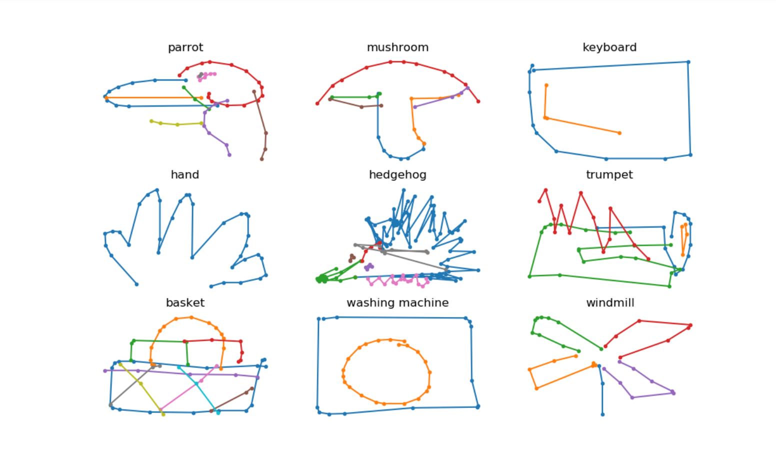
University of Oxford

Time series classification



Automated medical diagnosis from sensor data

Time series classification

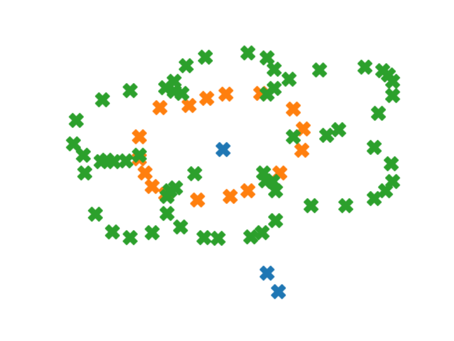


Characters recognition

Common feature

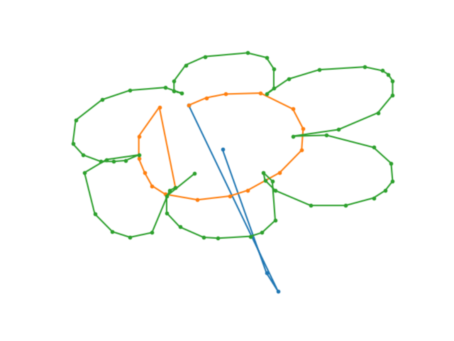
The predictor is a path $X : [a, b] \to \mathbb{R}^d$.

Data representation



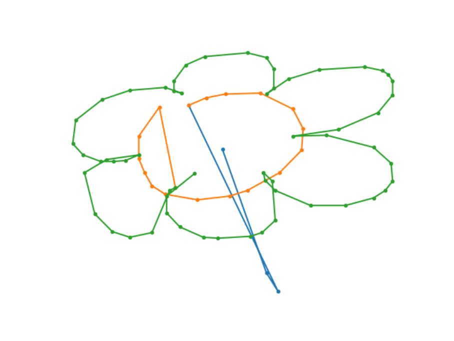
A sample from the class flower

Data representation

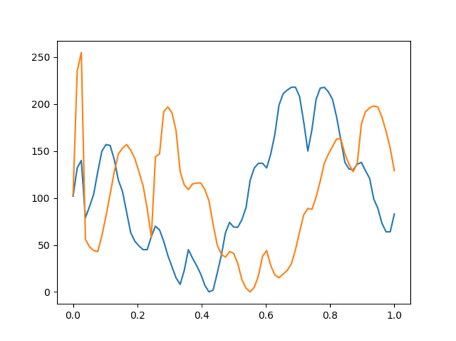


A sample from the class flower

Data representation

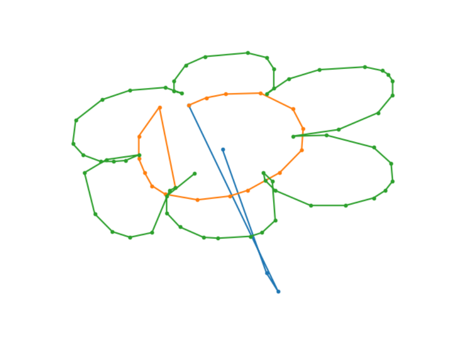


A sample from the class flower

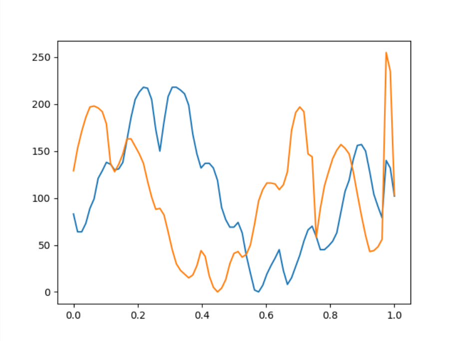


x and y coordinates

Data representation

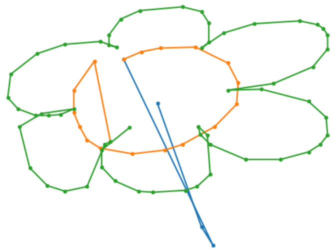


A sample from the class flower

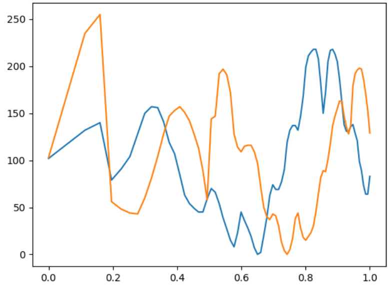


Time reversed

Data representation



A sample from the class flower



x and y at a different speed

The signature will overcome some of these problems.

The signature will overcome some of these problems.

- It is a transformation from a path to a sequence of coefficients.

The signature will overcome some of these problems.

- It is a transformation from a path to a sequence of coefficients.
- Independent of time parameterization.

The signature will overcome some of these problems.

- It is a transformation from a path to a sequence of coefficients.
- Independent of time parameterization.
- Encodes geometric properties of the path.

The signature will overcome some of these problems.

- ▷ It is a transformation from a path to a sequence of coefficients.
- ▷ Independent of time parameterization.
- ▷ Encodes geometric properties of the path.
- ▷ No loss of information.

Definition and basic properties

Mathematical setting

- A path $X : [0,1] \to \mathbb{R}^d$. Notation: X_t.

Mathematical setting

- A path $X : [0,1] \to \mathbb{R}^d$. Notation: X_t.
- Assumption: $\|X\|_{\text{1-var}} < \infty$.
- Path integral:

Mathematical setting

- A path $X : [0,1] \to \mathbb{R}^d$. Notation: X_t.
- Assumption: $\|X\|_{\text{1-var}} < \infty$.
- Path integral:
  - X is differentiable, $Y : [0,1] \to \mathbb{R}$ a continuous path.

$$\int_0^1 Y_t dX_t^i = \int_0^1 Y_t \dot{X}_t^i dt.$$

Mathematical setting

- A path $X : [0,1] \to \mathbb{R}^d$. Notation: X_t.
- Assumption: $\|X\|_{\text{1-var}} < \infty$.
- Path integral:
  - X is differentiable, $Y : [0,1] \to \mathbb{R}$ a continuous path.

$$\int_0^1 Y_t dX_t^i = \int_0^1 Y_t \dot{X}_t^i dt.$$

  - Generalization to $\|X\|_{\text{1-var}} < \infty$ via the Riemann-Stieltjes integral.

Iterated integrals

- $X : [0, 1] \to \mathbb{R}^d$, $X = (X^1, \ldots, X^d)$.

Iterated integrals

- $X : [0,1] \to \mathbb{R}^d$, $X = (X^1, \dots, X^d)$.
- For $i \in \{1, \dots, d\}$,

$$S^i(X)_{[0,t]} = \int_{0<s<t} dX_s^i = X_t^i - X_0^i$$

Iterated integrals

- $X : [0,1] \to \mathbb{R}^d$, $X = (X^1, \dots, X^d)$.
- For $i \in \{1, \dots, d\}$,

$$S^i(X)_{[0,t]} = \int_{0<s<t} dX_s^i = X_t^i - X_0^i \quad \to \text{a path!}$$

Iterated integrals

- $X : [0,1] \to \mathbb{R}^d$, $X = (X^1, \dots, X^d)$.
- For $i \in \{1, \dots, d\}$,

$$S^i(X)_{[0,t]} = \int_{0<s<t} dX_s^i = X_t^i - X_0^i \quad \to \text{a path!}$$

- For $(i,j) \in \{1, \dots, d\}^2$,

$$S^{i,j}(X)_{[0,t]} = \int_{0<s<t} S^i(X)_{[0,s]} dX_s^j = \int_{0<r<s<t} dX_r^i dX_s^j$$

Iterated integrals

- $X : [0,1] \to \mathbb{R}^d$, $X = (X^1, \dots, X^d)$.
- For $i \in \{1, \dots, d\}$,

$$S^i(X)_{[0,t]} = \int_{0<s<t} dX_s^i = X_t^i - X_0^i \quad \to \text{a path!}$$

- For $(i,j) \in \{1, \dots, d\}^2$,

$$S^{i,j}(X)_{[0,t]} = \int_{0<s<t} S^i(X)_{[0,s]} dX_s^j = \int_{0<r<s<t} dX_r^i dX_s^j \quad \to \text{a path!}$$

Iterated integrals

- $X : [0,1] \to \mathbb{R}^d$, $X = (X^1, \ldots, X^d)$.
- For $i \in \{1, \ldots, d\}$,

$$S^i(X)_{[0,t]} = \int_{0<s<t} dX_s^i = X_t^i - X_0^i \quad \to \text{a path!}$$

- For $(i,j) \in \{1, \ldots, d\}^2$,

$$S^{i,j}(X)_{[0,t]} = \int_{0<s<t} S^i(X)_{[0,s]} dX_s^j = \int_{0<r<s<t} dX_r^i dX_s^j \quad \to \text{a path!}$$

- Recursively, for $(i_1, \ldots, i_k) \in \{1, \ldots, d\}^k$,

$$S^{(i_1,\ldots,i_k)}(X)_{[0,t]} = \int_{0<t_1<t_2<\cdots<t_k<t} dX_{t_1}^{i_1} \ldots dX_{t_k}^{i_k}.$$

Iterated integrals

- $X : [0,1] \to \mathbb{R}^d$, $X = (X^1, \ldots, X^d)$.
- For $i \in \{1, \ldots, d\}$,

$$S^i(X)_{[0,t]} = \int_{0<s<t} dX_s^i = X_t^i - X_0^i \quad \to \text{a path!}$$

- For $(i,j) \in \{1, \ldots, d\}^2$,

$$S^{i,j}(X)_{[0,t]} = \int_{0<s<t} S^i(X)_{[0,s]} dX_s^j = \int_{0<r<s<t} dX_r^i dX_s^j \quad \to \text{a path!}$$

- Recursively, for $(i_1, \ldots, i_k) \in \{1, \ldots, d\}^k$,

$$S^{(i_1,\ldots,i_k)}(X)_{[0,t]} = \int_{0<t_1<t_2<\cdots<t_k<t} dX_{t_1}^{i_1} \ldots dX_{t_k}^{i_k}.$$

- $S^{(i_1,\ldots,i_k)}(X)_{[0,1]}$ is the k-fold iterated integral of X along $i_1, \ldots, i_k$.

Signature

Definition
The signature of X is the sequence of real numbers

$$S(X) = (1, S^{1}(X), \ldots, S^{d}(X), S^{(1,1)}(X), S^{(1,2)}(X), \ldots).$$

Signature

Definition
The signature of X is the sequence of real numbers

$$S(X) = (1, S^1(X), \ldots, S^d(X), S^{(1,1)}(X), S^{(1,2)}(X), \ldots).$$

- $d = 3 \rightarrow (1, 2, 3, 11, 12, 13, 21, 22, 23, 31, 32, 33, 111, 112, 113, \ldots)$

Signature

Definition
The signature of X is the sequence of real numbers

$$S(X) = (1, S^{1}(X), \ldots, S^{d}(X), S^{(1,1)}(X), S^{(1,2)}(X), \ldots).$$

- $d = 3 \rightarrow (1, 2, 3, 11, 12, 13, 21, 22, 23, 31, 32, 33, 111, 112, 113, \ldots)$
- Tensor notation:

$$\mathrm{X}^{\mathrm{k}} = \sum_{(i_1,\ldots,i_k)\subset\{1,\ldots,d\}^k} S^{(i_1,\ldots,i_k)}(X) e_{i_1} \otimes \cdots \otimes e_{i_k}.$$

Signature

Definition
The signature of X is the sequence of real numbers

$$S(X) = (1, S^1(X), \ldots, S^d(X), S^{(1,1)}(X), S^{(1,2)}(X), \ldots).$$

- $d = 3 \to (1, 2, 3, 11, 12, 13, 21, 22, 23, 31, 32, 33, 111, 112, 113, \ldots)$
- Tensor notation:

$$\mathrm{X}^k = \sum_{(i_1,\ldots,i_k)\subset\{1,\ldots,d\}^k} S^{(i_1,\ldots,i_k)}(X) e_{i_1} \otimes \cdots \otimes e_{i_k}.$$

- Signature:

$$S(X) = (1, \mathrm{X}^1, \mathrm{X}^2, \ldots, \mathrm{X}^k, \ldots) \in T(\mathbb{R}^d),$$

Signature

Definition
The signature of X is the sequence of real numbers

$$S(X) = (1, S^1(X), \ldots, S^d(X), S^{(1,1)}(X), S^{(1,2)}(X), \ldots).$$

- $d = 3 \to (1, 2, 3, 11, 12, 13, 21, 22, 23, 31, 32, 33, 111, 112, 113, \ldots)$
- Tensor notation:

$$\mathrm{X}^{\mathrm{k}} = \sum_{(i_1,\ldots,i_k)\subset\{1,\ldots,d\}^k} S^{(i_1,\ldots,i_k)}(X) e_{i_1} \otimes \cdots \otimes e_{i_k}.$$

- Signature:

$$S(X) = (1, \mathrm{X}^1, \mathrm{X}^2, \ldots, \mathrm{X}^{\mathrm{k}}, \ldots) \in T(\mathbb{R}^d),$$

where

$$T(\mathbb{R}^d) = 1 \oplus \mathbb{R}^d \oplus (\mathbb{R}^d)^{\otimes 2} \oplus \cdots \oplus (\mathbb{R}^d)^{\otimes k} \oplus \cdots$$

Example

For $X_t = (X_t^1, X_t^2)$,

$$\mathrm{X}^1 = \begin{pmatrix} \int_0^1 dX_t^1 & \int_0^1 dX_t^2 \end{pmatrix} = \begin{pmatrix} X_1^1 - X_0^1 & X_1^2 - X_0^2 \end{pmatrix}$$

Example

For $X_t = (X_t^1, X_t^2)$,

$$\mathbb{X}^1 = \begin{pmatrix} \int_0^1 dX_t^1 & \int_0^1 dX_t^2 \end{pmatrix} = \begin{pmatrix} X_1^1 - X_0^1 & X_1^2 - X_0^2 \end{pmatrix}$$

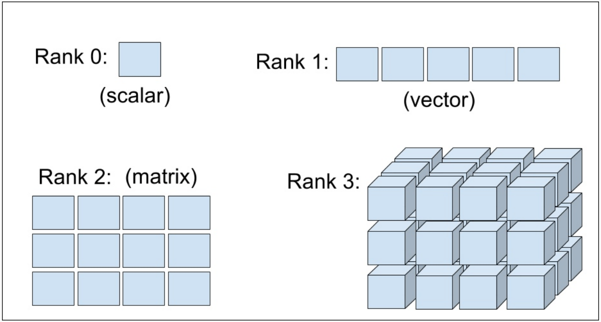
$$\mathbb{X}^2 = \begin{pmatrix} \int_0^1 \int_0^t dX_s^1 dX_t^1 & \int_0^1 \int_0^t dX_s^1 dX_t^2 \\ \int_0^1 \int_0^t dX_s^2 dX_t^1 & \int_0^1 \int_0^t dX_s^2 dX_t^2 \end{pmatrix}$$

Example

For $X_t = (X_t^1, X_t^2)$,

$$\mathbb{X}^1 = \begin{pmatrix} \int_0^1 dX_t^1 & \int_0^1 dX_t^2 \end{pmatrix} = \begin{pmatrix} X_1^1 - X_0^1 & X_1^2 - X_0^2 \end{pmatrix}$$

$$\mathbb{X}^2 = \begin{pmatrix} \int_0^1 \int_0^t dX_s^1 dX_t^1 & \int_0^1 \int_0^t dX_s^1 dX_t^2 \\ \int_0^1 \int_0^t dX_s^2 dX_t^1 & \int_0^1 \int_0^t dX_s^2 dX_t^2 \end{pmatrix}$$



Truncated signature

- Truncated signature at order m:

$$S^m(X) = (1, \mathrm{X}^1, \mathrm{X}^2, \ldots, \mathrm{X}^\mathrm{m}).$$

Truncated signature

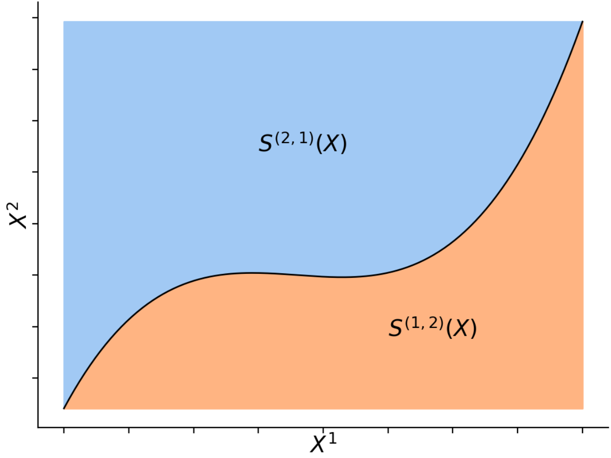
- Truncated signature at order m:

$$S^m(X) = (1, \mathrm{X}^1, \mathrm{X}^2, \dots, \mathrm{X}^\mathrm{m}).$$

- Dimension:

$$s_d(m) = \sum_{k=0}^{m} d^k = \frac{d^{m+1} - 1}{d - 1}.$$

Geometric interpretation



Important example

Linear path

- $X : [0,1] \to \mathbb{R}^d$ a linear path.

Important example

Linear path

- $X : [0, 1] \to \mathbb{R}^d$ a linear path.
- $X_t = X_0 + (X_1 - X_0)t$.

Important example

Linear path

- $X : [0,1] \to \mathbb{R}^d$ a linear path.
- $X_t = X_0 + (X_1 - X_0)t$.
- For any $I = (i_1, \ldots, i_k)$,

$$S^I(X) = \frac{1}{k!} \prod_{j=1}^{k} (X_1 - X_0)^{i_j}.$$

Important example

Linear path

- $X : [0,1] \to \mathbb{R}^d$ a linear path.
- $X_t = X_0 + (X_1 - X_0)t$.
- For any $I = (i_1, \ldots, i_k)$,

$$S^I(X) = \frac{1}{k!} \prod_{j=1}^{k} (X_1 - X_0)^{i_j}.$$

▷ Very useful: in practice, we always deal with piecewise linear paths.

▷ Needed: concatenation operations.

Properties 1

Chen's identity

- $X : [a, b] \to \mathbb{R}^d$ and $Y : [b, c] \to \mathbb{R}^d$ paths.

Properties 1

Chen's identity

- $X : [a, b] \to \mathbb{R}^d$ and $Y : [b, c] \to \mathbb{R}^d$ paths.
- $X * Y : [a, c] \to \mathbb{R}^d$ the concatenation.

Properties 1

Chen's identity

- $X : [a, b] \to \mathbb{R}^d$ and $Y : [b, c] \to \mathbb{R}^d$ paths.
- $X * Y : [a, c] \to \mathbb{R}^d$ the concatenation.
- Then
$$S(X * Y) = S(X) \otimes S(Y).$$

Properties 1

Chen's identity

- $X : [a, b] \to \mathbb{R}^d$ and $Y : [b, c] \to \mathbb{R}^d$ paths.
- $X * Y : [a, c] \to \mathbb{R}^d$ the concatenation.
- Then

$$S(X * Y) = S(X) \otimes S(Y).$$

▷ We can compute the signature of piecewise linear paths!

▷ Data stream of p points and truncation at m: $O(pd^m)$ operations.

▷ Fast packages and libraries available in `C++` and `Python`.

Properties 2

Uniqueness
If X has at least one monotone coordinate, then $S(X)$ determines X uniquely up to translations and reparametrizations.

Properties 2

Uniqueness
If X has at least one monotone coordinate, then $S(X)$ determines X uniquely up to translations and reparametrizations.

- The signature characterizes paths.
- Trick: add a dummy monotone component to X.
- Important concept of augmentation.

Properties 3

Signature approximation

- D compact subset of paths from $[0,1]$ to $\mathbb{R}^d$ that are not tree-like equivalent.

Properties 3

Signature approximation

- D compact subset of paths from $[0, 1]$ to $\mathbb{R}^d$ that are not tree-like equivalent.
- $f : D \to \mathbb{R}$ continuous.

Properties 3

Signature approximation

- D compact subset of paths from $[0,1]$ to $\mathbb{R}^d$ that are not tree-like equivalent.
- $f : D \to \mathbb{R}$ continuous.
- Then, for every $\varepsilon > 0$, there exists $w \in T(\mathbb{R}^d)$ such that, for any $X \in D$,

$$\big| f(X) - \langle w, S(X) \rangle \big| \leq \varepsilon.$$

Properties 3

Signature approximation

- D compact subset of paths from $[0, 1]$ to $\mathbb{R}^d$ that are not tree-like equivalent.
- $f : D \to \mathbb{R}$ continuous.
- Then, for every $\varepsilon > 0$, there exists $w \in T(\mathbb{R}^d)$ such that, for any $X \in D$,

$$\big|f(X) - \langle w, S(X)\rangle\big| \leq \varepsilon.$$

▷ Signature and linear model are happy together!

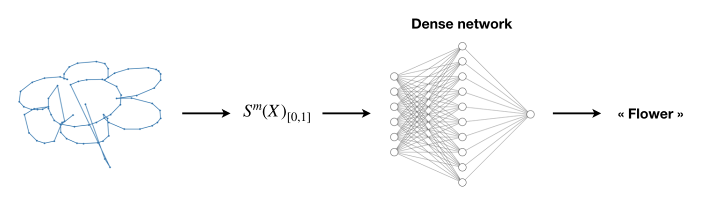
▷ This raises many interesting statistical issues.

Conclusion

- The signature is a good descriptor of a path: combine it with any machine learning “black box” algorithm.

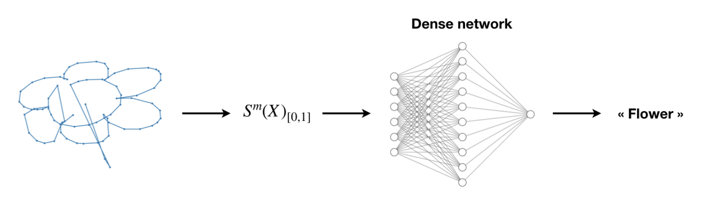
Conclusion

- The signature is a good descriptor of a path: combine it with any machine learning "black box" algorithm.



Conclusion

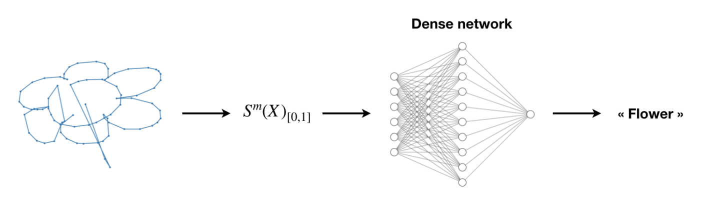
- The signature is a good descriptor of a path: combine it with any machine learning “black box” algorithm.



- Flexible tool and many choices: transformations to the path, domain of integration...

Conclusion

- The signature is a good descriptor of a path: combine it with any machine learning "black box" algorithm.



- Flexible tool and many choices: transformations to the path, domain of integration...
- Could we find a canonical signature pipeline that would be a domain-agnostic starting point for practitioners?

A generalized signature method

Overview

- Goal: systematic comparison of the different variations of the signature method.

Overview

- Goal: systematic comparison of the different variations of the signature method.
- Empirical study over 26 datasets of time series classification.

Overview

- Goal: systematic comparison of the different variations of the signature method.
- Empirical study over 26 datasets of time series classification.
- Define a generalised signature method as a framework to capture all these variations.

Overview

- Goal: systematic comparison of the different variations of the signature method.
- Empirical study over 26 datasets of time series classification.
- Define a generalised signature method as a framework to capture all these variations.
- Give practitioners some simple, domain-agnostic guidelines for a first signature algorithm.

Framework

- Input: a sequence $\mathrm{x} \in \mathcal{S}(\mathbb{R}^d)$, where

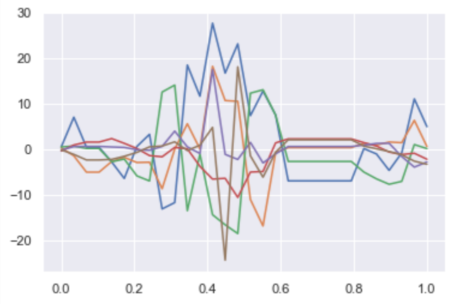
$$\mathcal{S}(\mathbb{R}^d) = \{(x_1, \ldots, x_n) \,|\, x_i \in \mathbb{R}^d, n \in \mathbb{N}\}.$$

Framework

- Input: a sequence $x \in \mathcal{S}(\mathbb{R}^d)$, where

$$\mathcal{S}(\mathbb{R}^d) = \{(x_1, \ldots, x_n) \,|\, x_i \in \mathbb{R}^d, n \in \mathbb{N}\}.$$

Racketsports dataset



A sample x with $d = 6,\ n = 30$

Framework

- Input: a sequence $\mathrm{x} \in \mathcal{S}(\mathbb{R}^d)$, where

$$\mathcal{S}(\mathbb{R}^d) = \{(x_1, \ldots, x_n) \,|\, x_i \in \mathbb{R}^d, n \in \mathbb{N}\}.$$

- Output: a label $y \in \{1, \ldots, q\}$.

Framework

- Input: a sequence $x \in \mathcal{S}(\mathbb{R}^d)$, where

$$\mathcal{S}(\mathbb{R}^d) = \{(x_1, \ldots, x_n) \,|\, x_i \in \mathbb{R}^d, n \in \mathbb{N}\}.$$

- Output: a label $y \in \{1, \ldots, q\}$.
  - ▷ Racketsports dataset: $q = 4$
    $\rightarrow$ forehand squash, backhand squash, clear badminton, smash badminton.

Framework

- For some $e, p \in \mathbb{N}$, an augmentation is a map

$$\phi = (\phi^1, \ldots, \phi^p)\colon \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}(\mathbb{R}^e)^p.$$

Framework

- For some $e, p \in \mathbb{N}$, an augmentation is a map

$$\phi = (\phi^1, \dots, \phi^p)\colon \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}(\mathbb{R}^e)^p.$$

- For some $q \in \mathbb{N}$, a window is a map

$$W = (W^1, \dots, W^q)\colon \mathcal{S}(\mathbb{R}^e) \to \mathcal{S}(\mathcal{S}(\mathbb{R}^e))^q.$$

Framework

- For some $e, p \in \mathbb{N}$, an augmentation is a map

$$\phi = (\phi^1, \dots, \phi^p) \colon \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}(\mathbb{R}^e)^p.$$

- For some $q \in \mathbb{N}$, a window is a map

$$W = (W^1, \dots, W^q) \colon \mathcal{S}(\mathbb{R}^e) \to \mathcal{S}(\mathcal{S}(\mathbb{R}^e))^q.$$

- Signature or logsignature transform: S^m.

Framework

- For some $e, p \in \mathbb{N}$, an augmentation is a map

$$\phi = (\phi^1, \dots, \phi^p) \colon \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}(\mathbb{R}^e)^p.$$

- For some $q \in \mathbb{N}$, a window is a map

$$W = (W^1, \dots, W^q) \colon \mathcal{S}(\mathbb{R}^e) \to \mathcal{S}(\mathcal{S}(\mathbb{R}^e))^q.$$

- Signature or logsignature transform: S^m.
- Rescaling operation ρ_{post} or ρ_{pre}.

Framework

- For some $e, p \in \mathbb{N}$, an augmentation is a map

$$\phi = (\phi^1, \dots, \phi^p) \colon \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}(\mathbb{R}^e)^p.$$

- For some $q \in \mathbb{N}$, a window is a map

$$W = (W^1, \dots, W^q) \colon \mathcal{S}(\mathbb{R}^e) \to \mathcal{S}(\mathcal{S}(\mathbb{R}^e))^q.$$

- Signature or logsignature transform: S^m.
- Rescaling operation ρ_{post} or ρ_{pre}.

Feature set

$$\mathrm{y}_{i,j,k} = (\rho_{\mathrm{post}} \circ S^m \circ \rho_{\mathrm{pre}} \circ W^{i,j} \circ \phi^k)(\mathrm{x}).$$

Framework

- For some $e, p \in \mathbb{N}$, an augmentation is a map

$$\phi = (\phi^1, \ldots, \phi^p) \colon \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}(\mathbb{R}^e)^p.$$

- For some $q \in \mathbb{N}$, a window is a map

$$W = (W^1, \ldots, W^q) \colon \mathcal{S}(\mathbb{R}^e) \to \mathcal{S}(\mathcal{S}(\mathbb{R}^e))^q.$$

- Signature or logsignature transform: S^m.
- Rescaling operation ρ_{post} or ρ_{pre}.

Feature set

$$y_{i,j,k} = (\rho_{\mathrm{post}} \circ S^m \circ \rho_{\mathrm{pre}} \circ W^{i,j} \circ \phi^k)(x).$$

Augmentations

Different goals:

- ▷ Sensitivity to properties of the path.

Augmentations

Different goals:

- Sensitivity to properties of the path.
- Dimension reduction.

Augmentations

Different goals:

- Sensitivity to properties of the path.
- Dimension reduction.
- Data-dependent transformation $\rightarrow$ make the path adapted to signatures.

Augmentations

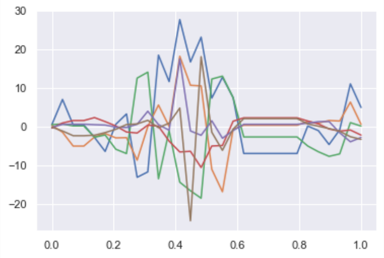
- Time augmentation

$$\phi(\mathrm{x}) = \big((1, x_1), \ldots, (n, x_n)\big) \in \mathcal{S}(\mathbb{R}^{d+1}).$$

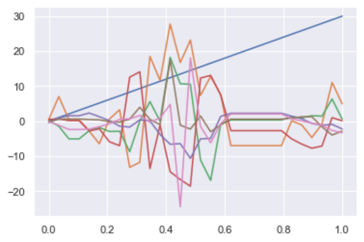
Augmentations

- Time augmentation

$$\phi(x) = \big((1, x_1), \ldots, (n, x_n)\big) \in \mathcal{S}(\mathbb{R}^{d+1}).$$



Sample $x \in \mathcal{S}(\mathbb{R}^6)$

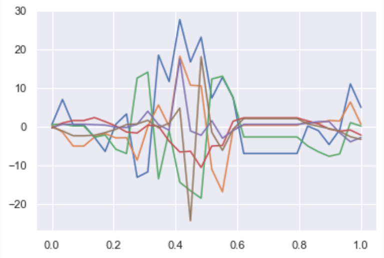


Augmented path $\phi(x) \in \mathcal{S}(\mathbb{R}^7)$

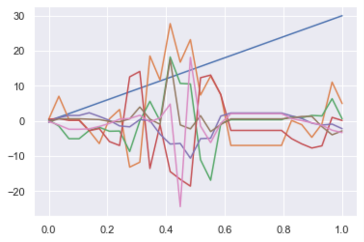
Augmentations

- Time augmentation

$$\phi(x) = \big((1, x_1), \ldots, (n, x_n)\big) \in \mathcal{S}(\mathbb{R}^{d+1}).$$



Sample $x \in \mathcal{S}(\mathbb{R}^6)$



Augmented path $\phi(x) \in \mathcal{S}(\mathbb{R}^7)$

▷ Sensitivity to parametrization and ensures signature uniqueness.

Augmentations

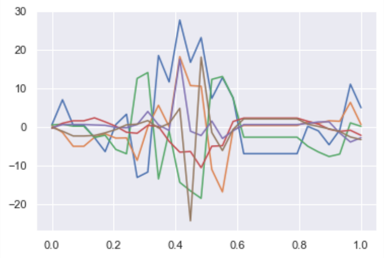
- Lead-lag augmentation

$$\phi(\mathsf{x}) = ((x_1, x_1), (x_2, x_1), (x_2, x_2), \ldots, (x_n, x_n)) \in \mathcal{S}(\mathbb{R}^{2d}).$$

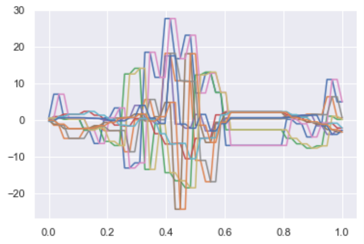
Augmentations

- Lead-lag augmentation

$$\phi(x) = ((x_1, x_1), (x_2, x_1), (x_2, x_2), \ldots, (x_n, x_n)) \in \mathcal{S}(\mathbb{R}^{2d}).$$



Sample $x \in \mathcal{S}(\mathbb{R}^6)$

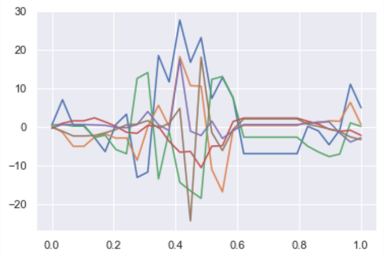


Augmented path $\phi(x) \in \mathcal{S}(\mathbb{R}^{12})$

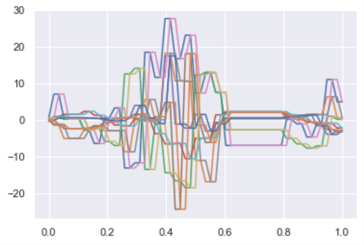
Augmentations

- Lead-lag augmentation

$$\phi(x) = ((x_1, x_1), (x_2, x_1), (x_2, x_2), \ldots, (x_n, x_n)) \in \mathcal{S}(\mathbb{R}^{2d}).$$



Sample $x \in \mathcal{S}(\mathbb{R}^6)$



Augmented path $\phi(x) \in \mathcal{S}(\mathbb{R}^{12})$

▷ Captures the quadratic variation of a process.

Augmentations

- Basepoint augmentation

$$\phi(\mathrm{x}) = (0, x_1, \ldots, x_n) \in \mathcal{S}(\mathbb{R}^d).$$

Augmentations

- Basepoint augmentation

$$\phi(\mathrm{x}) = (0, x_1, \ldots, x_n) \in \mathcal{S}(\mathbb{R}^d).$$

- Invisibility-reset augmentation

$$\phi(\mathrm{x}) = \big((1, x_1), \ldots, (1, x_{n-1}), (1, x_n), (0, x_n), (0, 0)\big) \in \mathcal{S}(\mathbb{R}^{d+1}).$$

Augmentations

- Basepoint augmentation

$$\phi(\mathsf{x}) = (0, x_1, \ldots, x_n) \in \mathcal{S}(\mathbb{R}^d).$$

- Invisibility-reset augmentation

$$\phi(\mathsf{x}) = \big((1, x_1), \ldots, (1, x_{n-1}), (1, x_n), (0, x_n), (0, 0)\big) \in \mathcal{S}(\mathbb{R}^{d+1}).$$

▷ Sensitivity to translations.

Framework

- For some $e, p \in \mathbb{N}$, an augmentation is a map

$$\phi = (\phi^1, \ldots, \phi^p) \colon \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}(\mathbb{R}^e)^p.$$

- For some $q \in \mathbb{N}$, a window is a map

$$W = (W^1, \ldots, W^q) \colon \mathcal{S}(\mathbb{R}^e) \to \mathcal{S}(\mathcal{S}(\mathbb{R}^e))^q.$$

- Signature or logsignature transform: S^m.
- Rescaling operation ρ_{post} or ρ_{pre}.

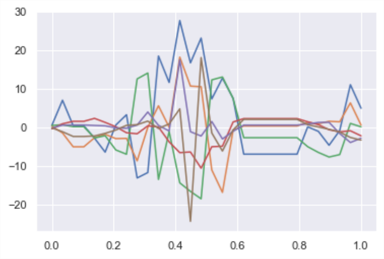
Feature set

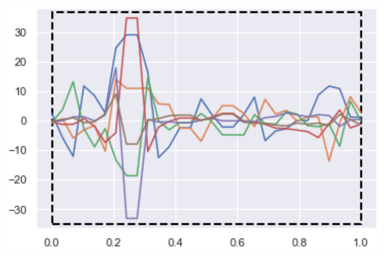
$$\mathrm{y}_{i,j,k} = (\rho_{\mathrm{post}} \circ S^m \circ \rho_{\mathrm{pre}} \circ W^{i,j} \circ \phi^k)(\mathrm{x}).$$

Windows

- Global window

$$W(\mathrm{x}) = (\mathrm{x}) \in \mathcal{S}(\mathbb{R}^e),$$

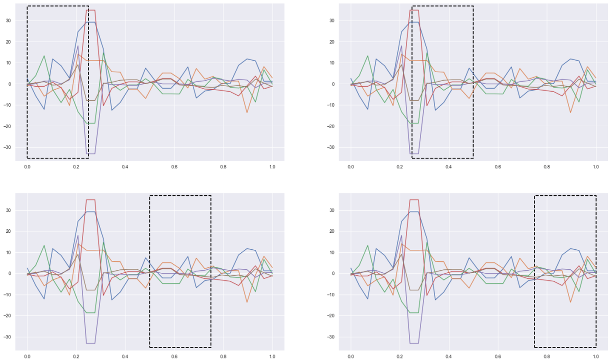




Windows

- Sliding window

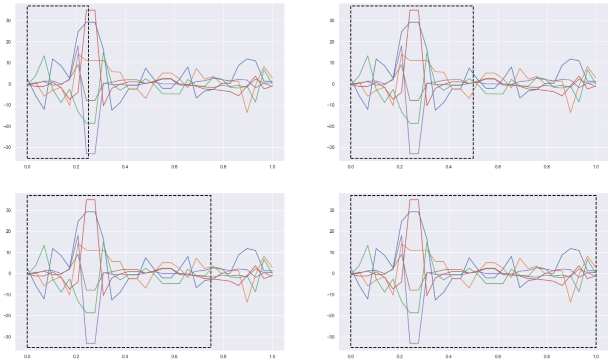
$$W(x) = (x_{1,\ell}, x_{l+1,l+\ell}, x_{2l+1,2l+\ell}, \ldots) \in \mathcal{S}(\mathcal{S}(\mathbb{R}^e)),$$



Windows

- Expanding window

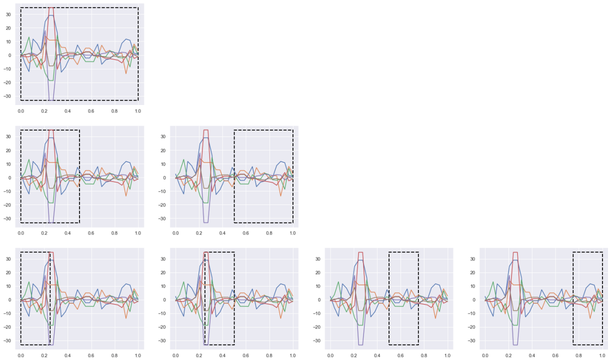
$$W(x) = (x_{1,\ell}, x_{1,l+\ell}, x_{1,2l+\ell}, \dots) \in \mathcal{S}(\mathcal{S}(\mathbb{R}^e)).$$



Windows

- Dyadic window

$$W(x) = (W^1(x), \ldots, W^q(x)) \in \mathcal{S}(\mathcal{S}(\mathbb{R}^e))^q.$$



Framework

- For some $e, p \in \mathbb{N}$, an augmentation is a map

$$\phi = (\phi^1, \dots, \phi^p) \colon \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}(\mathbb{R}^e)^p.$$

- For some $q \in \mathbb{N}$, a window is a map

$$W = (W^1, \dots, W^q) \colon \mathcal{S}(\mathbb{R}^e) \to \mathcal{S}(\mathcal{S}(\mathbb{R}^e))^q.$$

- Signature or logsignature transform: S^m.
- Rescaling operation ρ_{post} or ρ_{pre}.

Feature set

$$\mathrm{y}_{i,j,k} = (\rho_{\mathrm{post}} \circ S^m \circ \rho_{\mathrm{pre}} \circ W^{i,j} \circ \phi^k)(\mathrm{x}).$$

Framework

- Signature transform

$$S^m(\mathrm{x}) = (1, \mathrm{X}^1, \mathrm{X}^2, \dots, \mathrm{X}^\mathrm{m}).$$

Framework

- Signature transform

$$S^m(\mathrm{x}) = (1, \mathrm{X}^1, \mathrm{X}^2, \ldots, \mathrm{X}^{\mathrm{m}}).$$

- Logsignature transform $\log(S^m(\mathrm{x}))$, where for any $a \in T((\mathbb{R}^d))$,

$$\log(a) = \sum_{k \geq 0} \frac{(-1)^k}{k} (1 - a)^{\otimes k}.$$

Framework

- Signature transform

$$S^m(\mathrm{x}) = (1, \mathrm{X}^1, \mathrm{X}^2, \dots, \mathrm{X}^\mathrm{m}).$$

- Logsignature transform $\log(S^m(\mathrm{x}))$, where for any $a \in T((\mathbb{R}^d))$,

$$\log(a) = \sum_{k \geq 0} \frac{(-1)^k}{k} (1 - a)^{\otimes k}.$$

▷ Same information and logsignature less dimensional but no linear approximation property.

Signature versus logsignature

Table 1: Typical dimensions of $S^m(\mathrm{x})$ and $\log(S^m(\mathrm{x}))$.

	$d = 2$	$d = 3$	$d = 6$
$m = 1$	2 / 2	3 / 3	6 / 6
$m = 2$	6 / 3	12 / 6	42 / 21
$m = 5$	62 / 14	363 / 80	9330 / 1960
$m = 7$	254 / 41	3279 / 508	335922 / 49685

Framework

- For some $e, p \in \mathbb{N}$, an augmentation is a map

$$\phi = (\phi^1, \ldots, \phi^p) \colon \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}(\mathbb{R}^e)^p.$$

- For some $q \in \mathbb{N}$, a window is a map

$$W = (W^1, \ldots, W^q) \colon \mathcal{S}(\mathbb{R}^e) \to \mathcal{S}(\mathcal{S}(\mathbb{R}^e))^q.$$

- Signature or logsignature transform: S^m.
- Rescaling operation ρ_{post} or ρ_{pre}.

Feature set

$$\mathrm{y}_{i,j,k} = (\rho_{\text{post}} \circ S^m \circ \rho_{\text{pre}} \circ W^{i,j} \circ \phi^k)(\mathrm{x}).$$

Framework

- For some $e, p \in \mathbb{N}$, an augmentation is a map

$$\phi = (\phi^1, \dots, \phi^p) \colon \mathcal{S}(\mathbb{R}^d) \to \mathcal{S}(\mathbb{R}^e)^p.$$

- For some $q \in \mathbb{N}$, a window is a map

$$W = (W^1, \dots, W^q) \colon \mathcal{S}(\mathbb{R}^e) \to \mathcal{S}(\mathcal{S}(\mathbb{R}^e))^q.$$

- Signature or logsignature transform: S^m.
- Rescaling operation ρ_{post} or ρ_{pre}.

Feature set

$$\mathrm{y}_{i,j,k} = (\rho_{\mathrm{post}} \circ S^m \circ \rho_{\mathrm{pre}} \circ W^{i,j} \circ \phi^k)(\mathrm{x}).$$

Empirical study methodology

- 26 datasets: Human Activities and Postural Transitions, Speech Commands and 24 datasets from the UEA archive.

Empirical study methodology

- 26 datasets: Human Activities and Postural Transitions, Speech Commands and 24 datasets from the UEA archive.
- Definition of a baseline:

Empirical study methodology

- 26 datasets: Human Activities and Postural Transitions, Speech Commands and 24 datasets from the UEA archive.
- Definition of a baseline: time augmentation + global window + signature of depth 3 + pre-signature scaling

Empirical study methodology

- 26 datasets: Human Activities and Postural Transitions, Speech Commands and 24 datasets from the UEA archive.
- Definition of a baseline: time augmentation + global window + signature of depth 3 + pre-signature scaling

$$(S^3 \circ \rho_{\mathrm{pre}} \circ \phi)(x).$$

Empirical study methodology

- 26 datasets: Human Activities and Postural Transitions, Speech Commands and 24 datasets from the UEA archive.
- Definition of a baseline: time augmentation + global window + signature of depth 3 + pre-signature scaling

$$(S^3 \circ \rho_{\mathrm{pre}} \circ \phi)(x).$$

- Vary each group of options with regards to this baseline.

Empirical study methodology

- 26 datasets: Human Activities and Postural Transitions, Speech Commands and 24 datasets from the UEA archive.
- Definition of a baseline: time augmentation + global window + signature of depth 3 + pre-signature scaling

$$(S^3 \circ \rho_{\mathrm{pre}} \circ \phi)(x).$$

- Vary each group of options with regards to this baseline.
- 4 classifiers: logistic regression, random forest, GRU, CNN.

Empirical study methodology

- 26 datasets: Human Activities and Postural Transitions, Speech Commands and 24 datasets from the UEA archive.
- Definition of a baseline: time augmentation + global window + signature of depth 3 + pre-signature scaling

$$(S^3 \circ \rho_{\mathrm{pre}} \circ \phi)(x).$$

- Vary each group of options with regards to this baseline.
- 4 classifiers: logistic regression, random forest, GRU, CNN.

$\rightarrow$ 9984 combinations.

Empirical study methodology

- Evaluation of one combination: compute the best accuracy accross the 4 classifiers.

Empirical study methodology

- Evaluation of one combination: compute the best accuracy accross the 4 classifiers.

	Window			
Dataset	Global	Sliding	Expanding	Dyadic
ArticularyWordRecognition	96.3%	89.3%	99.0%	99.0%
AtrialFibrillation	46.7%	46.7%	46.7%	60.0%
BasicMotions	100.0%	100.0%	100.0%	100.0%
CharacterTrajectories	93.2%	94.6%	96.9%	97.1%
Cricket	97.2%	93.1%	97.2%	95.8%
⋮	⋮	⋮	⋮	⋮

Empirical study methodology

- Evaluation of one combination: compute the best accuracy accross the 4 classifiers.
- Compute the average rank over all datasets.

	Window			
Dataset	Global	Sliding	Expanding	Dyadic
ArticularyWordRecognition	96.3%	89.3%	99.0%	99.0%
AtrialFibrillation	46.7%	46.7%	46.7%	60.0%
BasicMotions	100.0%	100.0%	100.0%	100.0%
CharacterTrajectories	93.2%	94.6%	96.9%	97.1%
Cricket	97.2%	93.1%	97.2%	95.8%
⋮	⋮	⋮	⋮	⋮

Empirical study methodology

- Evaluation of one combination: compute the best accuracy accross the 4 classifiers.
- Compute the average rank over all datasets.

	Window			
Dataset	Global	Sliding	Expanding	Dyadic
ArticularyWordRecognition	3	4	1.5	1.5
AtrialFibrillation	3	3	3	1
BasicMotions	2	2	2	2
CharacterTrajectories	4	3	2	1
Cricket	1.5	4	1.5	3
⋮	⋮	⋮	⋮	⋮

Empirical study methodology

- Evaluation of one combination: compute the best accuracy accross the 4 classifiers.
- Compute the average rank over all datasets.

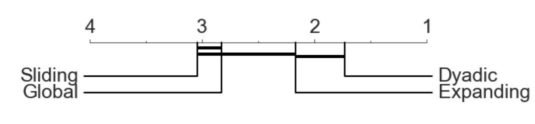
	Window			
Dataset	Global	Sliding	Expanding	Dyadic
ArticularyWordRecognition	3	4	1.5	1.5
AtrialFibrillation	3	3	3	1
BasicMotions	2	2	2	2
CharacterTrajectories	4	3	2	1
Cricket	1.5	4	1.5	3
⋮	⋮	⋮	⋮	⋮
Average ranks	2.83	3.04	2.17	**1.73**

Empirical study methodology

- Evaluation of one combination: compute the best accuracy accross the 4 classifiers.
- Compute the average rank over all datasets.
- Use a critical differences plot, a global Friedman test, and pairwise Wilcoxon signed-rank tests at 5% with Holm's alpha correction.

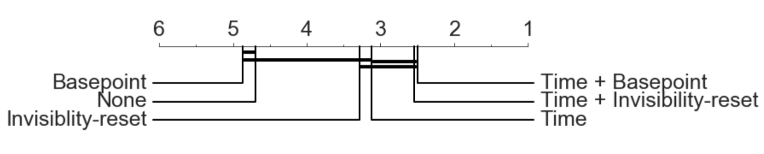
Results

- ▷ Windows:



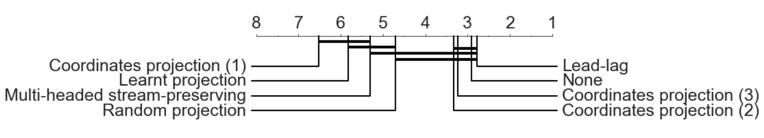
Results

- ▷ Invariance-removing augmentations:



Results

- ▷ Other augmentations:



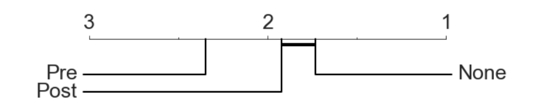
Results

- Signature versus logsignature transform:

	Signature	Logsignature
Average ranks	**1.25**	1.75
p-value		0.01

Results

- ▷ Rescalings:



Canonical signature pipeline

- ▷ Unless the problem is known to be translation or reparameterisation invariant, then use the time and basepoint augmentations.

Canonical signature pipeline

- Unless the problem is known to be translation or reparameterisation invariant, then use the time and basepoint augmentations.
- The lead-lag augmentation should be considered, but we do not recommend it in general, due to its computational cost.

Canonical signature pipeline

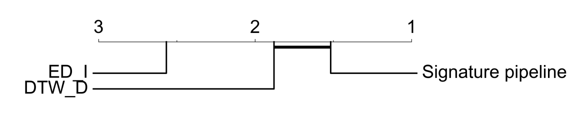
- Unless the problem is known to be translation or reparameterisation invariant, then use the time and basepoint augmentations.
- The lead-lag augmentation should be considered, but we do not recommend it in general, due to its computational cost.
- Use hierarchical dyadic windows, and the signature transform; both have a depth hyperparameter that must be optimised.

Canonical signature pipeline

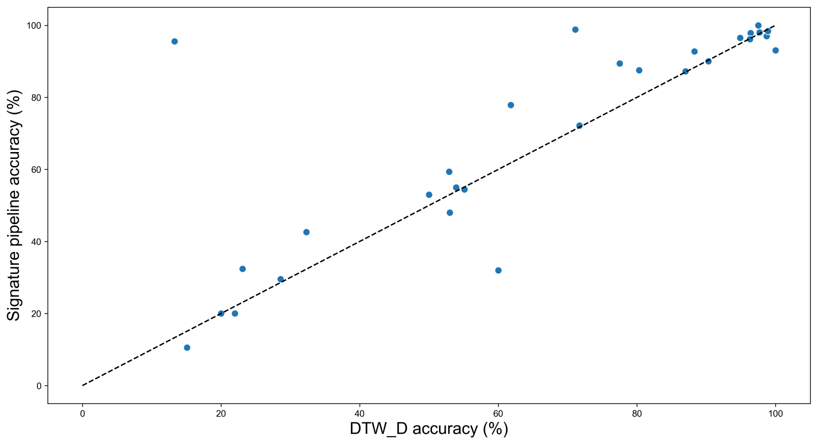
$\rightarrow$ Implement this pipeline on the 30 datasets from the UEA archive and compare it to benchmark classifiers.

Canonical signature pipeline

→ Implement this pipeline on the 30 datasets from the UEA archive and compare it to benchmark classifiers.



Canonical signature pipeline



Conclusion

- Signatures are a flexible tool.

Conclusion

- Signatures are a flexible tool.
- We are able to find a canonical signature method that represents a domain-agnostic starting point.

Conclusion

- Signatures are a flexible tool.
- We are able to find a canonical signature method that represents a domain-agnostic starting point.
- The combination "signature + generic algorithm" $\approx$ state-of-the-art.

Conclusion

- Signatures are a flexible tool.
- We are able to find a canonical signature method that represents a domain-agnostic starting point.
- The combination "signature + generic algorithm" $\approx$ state-of-the-art.
- Few computing resources and no domain-specific knowledge.

Conclusion

- Signatures are a flexible tool.
- We are able to find a canonical signature method that represents a domain-agnostic starting point.
- The combination "signature + generic algorithm" $\approx$ state-of-the-art.
- Few computing resources and no domain-specific knowledge.
- A lot of open questions and potential applications.

Thank you!