Stats and ML for Persistence

Lecture 14 - CMSE 890

Prof. Elizabeth Munch

Michigan State University
::
Dept of Computational Mathematics, Science & Engineering

Thurs, Oct 16, 2025

Goals for today

Goals for today:

- Today: ML for Persistence
- Sorta follows Ch 13.1 but not entirely

Questions and tasks

Tasks

- Classification
- Regression

The Issue

- Most ML methods want vectors in $\mathbb{R}^n$.
- That implies some sorting or ordering, which we don't want for the persistence diagrams.
- Implies the same n for all data points, but we have a variable number of points in the diagrams.
- Usually, we need the data points to come from a vector space, but persistence diagrams don't.

It's even worse

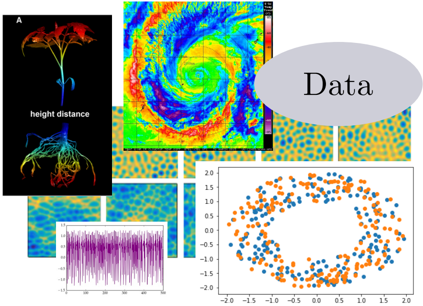
- Usually want ML inputs to come from a Banach space (complete normed vector space) or a Hilbert space (vector space with inner product and limits).
- Note all Hilbert spaces are Banach spaces, but not all Banach spaces are Hilbert spaces.

Theorem (Bubenik Wagner 2020)

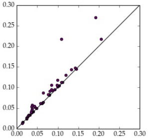
$(\mathcal{D}, W_p)$ does not admit an isometric embedding into a Hilbert space for any $1 \leq p \leq \infty$.

Featurization: the idea

Here be dragons.....

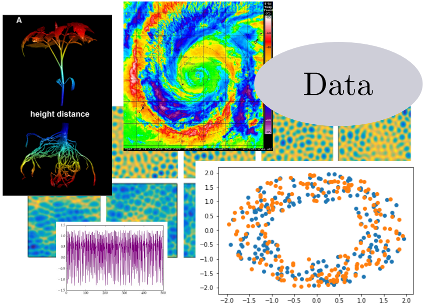


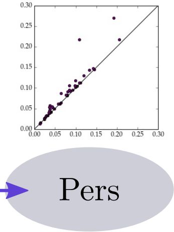
Pers



Featurization: the idea

Here be dragons.....

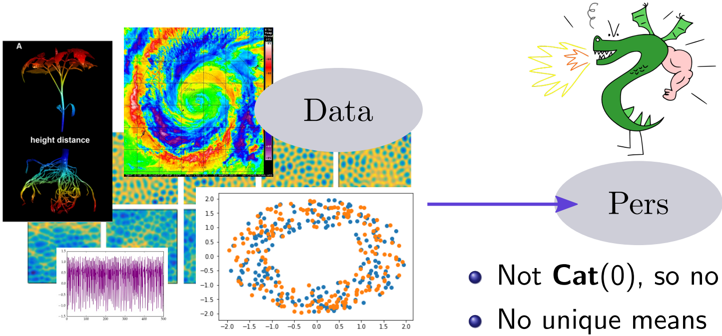




- Not **Cat**(0), so no unique geodesics
- No unique means
- No inner product

Featurization: the idea

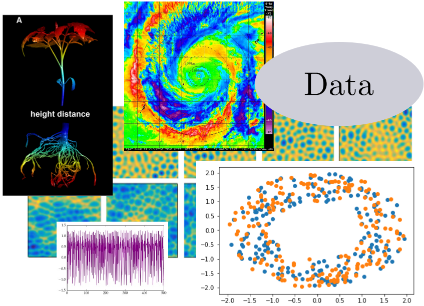
Here be dragons.....

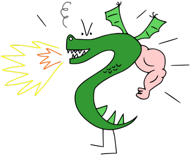


- Not **Cat**(0), so no unique geodesics
- No unique means
- No inner product

Featurization: the idea

Here be dragons.....





$\longrightarrow$ Pers $\longrightarrow \mathbb{R}^d$

- Not **Cat**(0), so no unique geodesics
- No unique means
- No inner product

Turning persistence diagrams into something else

Section 1

Algebraic structures

What if we just forget about the diagram?

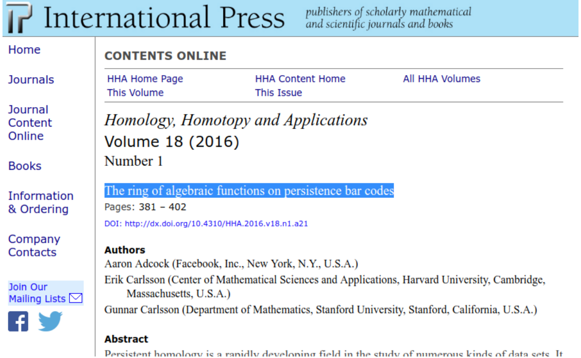
- Just treat a persistence diagram as a set of points in $\mathbb{R}^2$

$$\{(x_1, y_1), \cdots, (x_n, y_n)\} \rightsquigarrow (x_1, y_1, x_2, y_2, \cdots, x_n, y_n) \in \mathbb{R}^{2n}$$

- Problems:
  - Order shouldn't matter
  - n isn't fixed

Algebraic geometry to the rescue!

- Associate persistence diagrams to $k[x_1, y_1, x_2, y_2, \cdots]/\sim$
- Come up with functions on the points that don't care about order



International Press — publishers of scholarly mathematical and scientific journals and books

Home
Journals
Journal Content Online
Books
Information & Ordering
Company Contacts
Join Our Mailing Lists

CONTENTS ONLINE

HHA Home Page | HHA Content Home | All HHA Volumes
This Volume | This Issue

Homology, Homotopy and Applications
Volume 18 (2016)
Number 1

The ring of algebraic functions on persistence bar codes

Pages: 381 – 402

DOI: http://dx.doi.org/10.4310/HHA.2016.v18.n1.a21

Authors

Aaron Adcock (Facebook, Inc., New York, N.Y., U.S.A.)

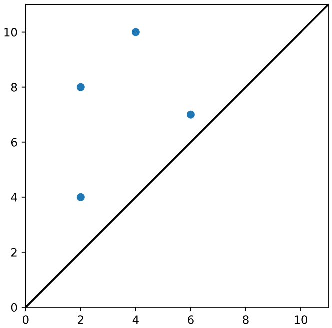
Erik Carlsson (Center of Mathematical Sciences and Applications, Harvard University, Cambridge, Massachusetts, U.S.A.)

Gunnar Carlsson (Department of Mathematics, Stanford University, Stanford, California, U.S.A.)

Abstract

Persistent homology is a rapidly developing field in the study of numerous kinds of data sets. It

Example



$$\{(2,4),(2,8),(4,10),(6,7)\}$$

- Ex. $\displaystyle\sum_i (x_i)(y_i - x_i)$

- Ex. $\displaystyle\sum_i (y_{\text{max}} - y_i)(y_i - x_i)$

Mnist Data

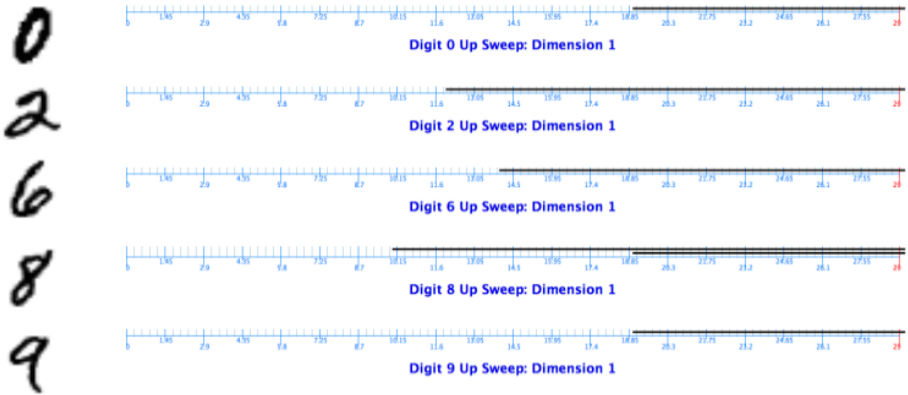
Build complex:

- Put vertex at each white pixel
- Connect vertices if their pixels touch.
- My assumption: clique complex after that

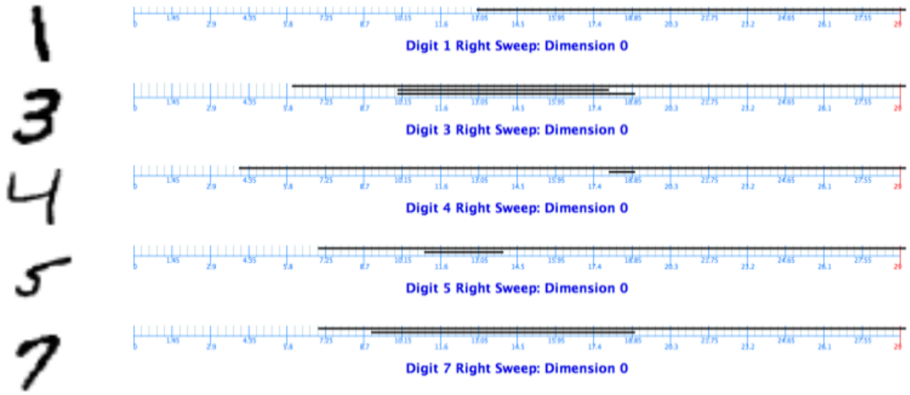
Filtration:

- Pick cardinal direction
- Add vertices in order in that direction
- Gives 4 each 0- and 1-dimensional diagrams

Results - Loops



Results - Non-loops



Features - Digits example

For each digit

Diagrams:

- 4 directions
- 0- and 1-dimensional diagrams for each direction
- $=8$ diagrams each

Features:

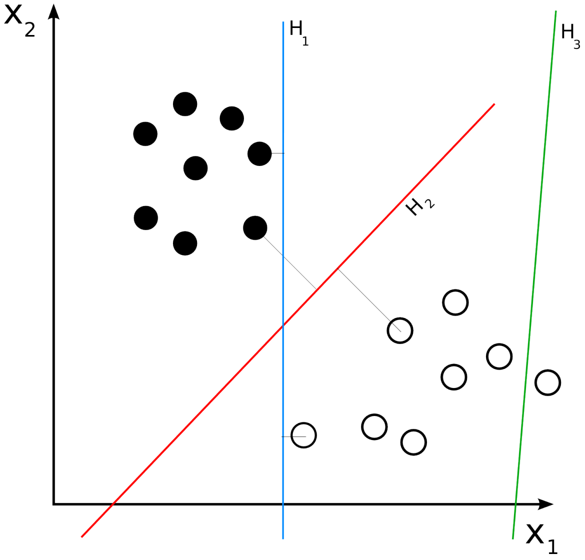
- Each Diagram has 4 features
- $=32$ features total

Features

$y_{\max} =$ max death for all diagrams
Diagrams $X = \{(x_i, y_i)\}$.

- $\sum x_i(y_i - x_i)$
- $\sum (y_{max} - y_i)(y_i - x_i)$
- $\sum x_i^2(y_i - x_i)^4$
- $\sum (y_{max} - y_i)^2(y_i - x_i)^4$

Support Vector Machine (SVM)



Results

Table 1: Classification Accuracy of two SVM Kernels

SVM	1000 Digits	5000 Digits	10000 Digits
Gaussian	87.70%	91.54%	92.04%
Polynomial	88.00%	91.62%	92.10%

(a) Stylistic Problems

(b) Spurious Topological Changes

Figure 5: Common Misclassifications

Pros & Cons

Pros

- Incredibly simple to explain
- Works well in lots of simple cases
- Fast to compute

Cons

- Not stable with respect to distances

Section 2

Landscapes

Paper

Journal of Machine Learning Research 16 (2015) 77-102 Submitted 7/14; Published 1/15

Statistical Topological Data Analysis using Persistence Landscapes

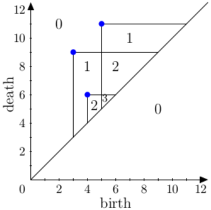
Peter Bubenik PETER.BUBENIK@GMAIL.COM
Department of Mathematics
Cleveland State University
Cleveland, OH 44115-2214, USA

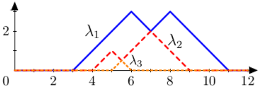
Editor: David Dunson

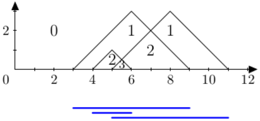
Abstract

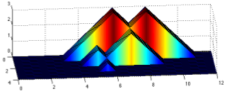
We define a new topological summary for data that we call the persistence landscape. Since this summary lies in a vector space, it is easy to combine with tools from statistics and machine learning, in contrast to the standard topological summaries. Viewed as a random variable with values in a Banach space, this summary obeys a strong law of large numbers and a central limit theorem. We show how a number of standard statistical tests can be used for statistical inference using this summary. We also prove that this summary is

Definition by picture



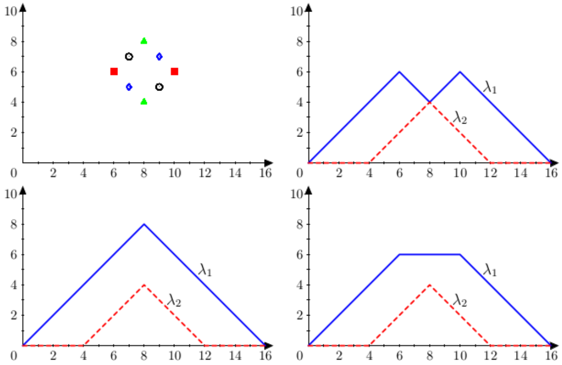




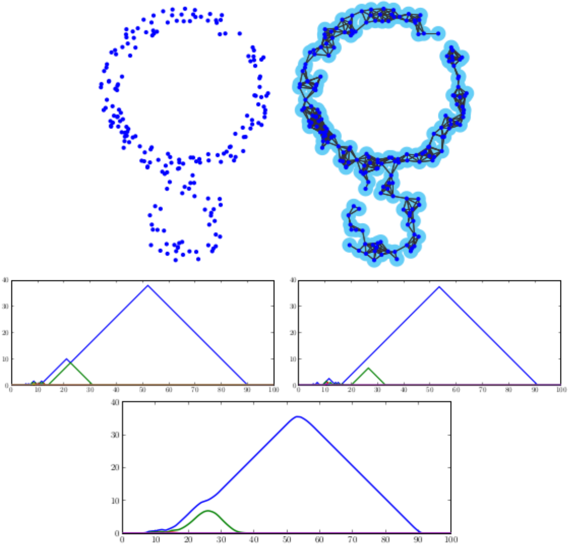


- Take diagram
- Rank function:
  $\beta^{a,b} = \dim(\mathrm{Im}\ (H_k(X_a) \to H_k(X_b)))$
- Rotate $(x, y) \mapsto \left(\frac{x+y}{2}, \frac{y-x}{2}\right)$
- $\lambda : \mathbb{N} \times \mathbb{R} \to \mathbb{R}$, $(k, t) \to \lambda_k(t)$
- $\lambda_k(t) = \sup(m \geq 0 \mid \beta^{t-m,t+m} \geq k)$

Everyone's favorite example

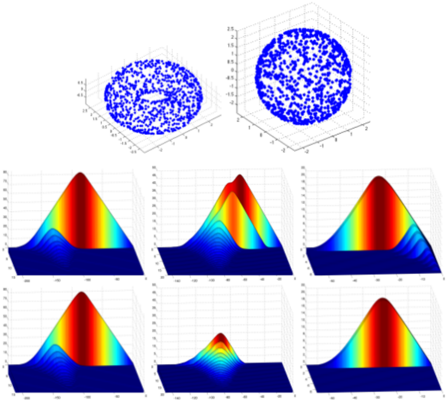


200 points, repeated 100 times



- Top two are example landscapes
- Bottom is average over all landscapes

Torus and Sphere



Average landscapes:

- Row 1: torus; Row 2: Sphere
- Col: 0-, 1-, and 2-dimensional diagram

Stability

Define

$$\|\lambda^D\|_p = \left(\sum_{k=1}^{\infty} \|\lambda_k^D\|_p^p \right)^{1/p}$$

Then

$$\|\lambda^{D_1} - \lambda^{D_2}\|_\infty \leq d_B(D_1, D_2).$$

Pros & Cons

Pros

- Can take averages, do ML etc
- Probably the #1 used ML featurization approach
- Stability

Cons

- Average might not be something that comes from a diagram

Section 3

Persistence Images

Images paper

Journal of Machine Learning Research 18 (2017) 1-35 Submitted 7/16; Published 2/17

Persistence Images: A Stable Vector Representation of Persistent Homology

Henry Adams ADAMS@MATH.COLOSTATE.EDU
Tegan Emerson EMERSON@MATH.COLOSTATE.EDU
Michael Kirby KIRBY@MATH.COLOSTATE.EDU
Rachel Neville NEVILLE@MATH.COLOSTATE.EDU
Chris Peterson PETERSON@MATH.COLOSTATE.EDU
Patrick Shipman SHIPMAN@MATH.COLOSTATE.EDU
Department of Mathematics
Colorado State University
1874 Campus Delivery
Fort Collins, CO 80523-1874

Sofya Chepushtanova SOFYA.CHEPUSHTANOVA@WILKES.EDU
Department of Mathematics and Computer Science
Wilkes University
84 West South Street
Wilkes-Barre, PA 18766, USA

Eric Hanson ERIC.HANSON@TCU.EDU
Department of Mathematics
Texas Christian University
Box 298900
Fort Worth, TX 76129

Francis Motta MOTTA@MATH.DUKE.EDU
Department of Mathematics
Duke University
Durham, NC 27708, USA

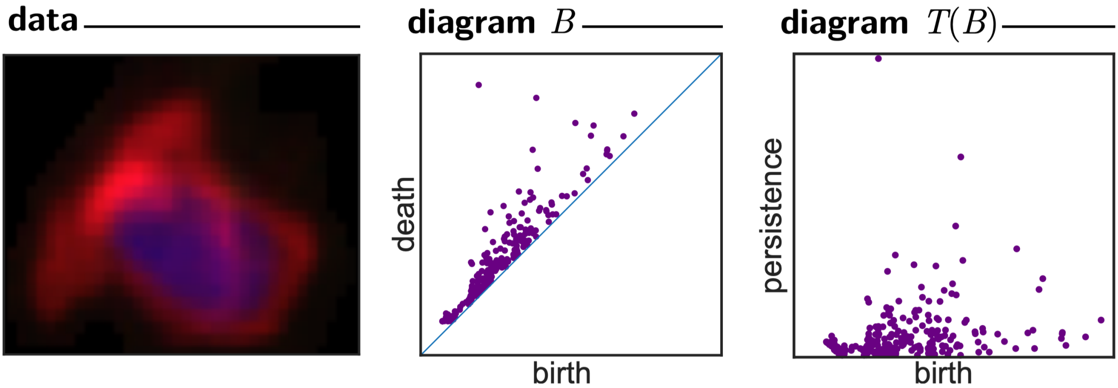
Lori Ziegelmeier LZIEGEL1@MACALESTER.EDU
Department of Mathematics, Statistics, and Computer Science
Macalester College
1600 Grand Avenue
Saint Paul, MN 55105, USA

Editor: Michael Mahoney

Their view on the problem

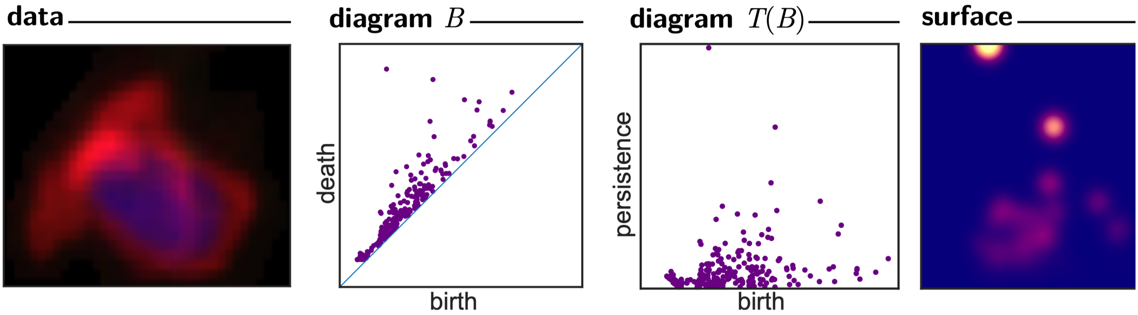
Problem Statement: How can we represent a persistence diagram so that

- the output of the representation is a vector in $\mathbb{R}^n$,
- the representation is stable with respect to input noise,
- the representation is efficient to compute,
- the representation maintains an interpretable connection to the original PD, and
- the representation allows one to adjust the relative importance of points in different regions of the PD?



- Transform: $T(X) = \{(x, y - x) \mid (x, y) \in X\}$

Persistence surface

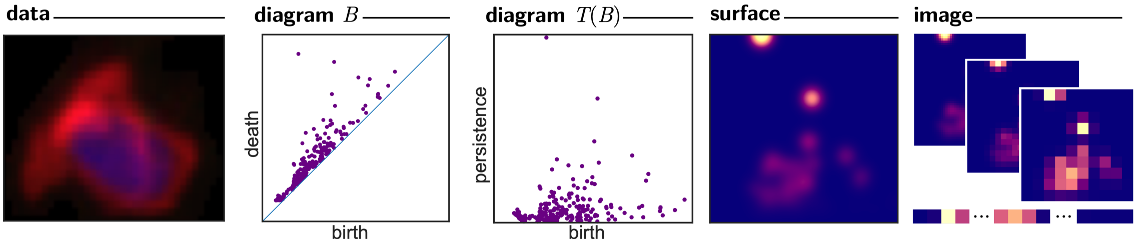


- Weight function e.g., $f(u) = u_y/\text{maxPers}$
- Gaussian: $\varphi_u(z) = \frac{1}{2\pi\sigma^2}\exp(-[(x-u_x)^2 + (y-u_y)^2]/(2\sigma^2))$

For a Pers Dgm X, the persistence surface is

$$\begin{aligned} \mu_X : \quad \mathbb{R}^2 &\to \mathbb{R} \\ x &\mapsto \sum_{u \in T(X)} f(u)\varphi_u(z) \end{aligned}$$

Persistence Images



- Grid up box
- Integrate the function in each box
- Treat the outputs as a vector

Example: linked twist map

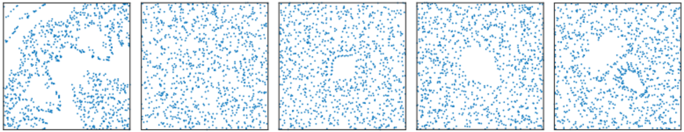


Figure 4: Examples of the first 1000 iterations, $\{(x_n, y_n) \; : \; n = 0, \ldots, 1000\}$, of the linked twist map with parameter values $r = 2$, 3.5, 4.0, 4.1 and 4.3, respectively.

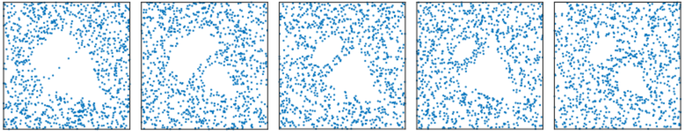


Figure 5: Truncated orbits, $\{(x_n, y_n) \; : \; n = 0, \ldots, 1000\}$, of the linked twist map with fixed $r = 4.3$ for different initial conditions (x_0, y_0).

- Generate point clouds from a discrete dynamical system

$$x_{n+1} = x_n + r y_n (1 - y_n) \quad \bmod 1$$
$$y_{n+1} = y_n + r x_n (1 - x_n) \quad \bmod 1$$

- Goal: Classify trials by r
- Scores
    - Both H_0 and H_1: 82.5%
    - H_0: 49.8%
    - H_1: 65.7%

Stability

Theorem 13.3. *Suppose persistence images are computed with the normalized Gaussian distribution with variance σ^2 and weight function $\omega : \mathbb{R}^2 \to \mathbb{R}$. Then the persistence images are stable w.r.t. the* 1*-Wasserstein distance between persistence diagrams. More precisely, given two finite and bounded persistence diagrams D and E, we have:*

$$\| \mathrm{I}_D - \mathrm{I}_E \|_1 \leq \left(\sqrt{5}|\nabla\omega| + \sqrt{\frac{10}{\pi}} \frac{\|\omega\|_\infty}{\sigma} \right) \cdot \mathsf{d}_{W,1}(D, E).$$

Here, $\nabla\omega$ stands for the gradient of ω, and $|\nabla\omega| = \sup_{z\in\mathbb{R}^2} \|\nabla\omega\|_2$ is the maximum norm of the gradient vector of ω at any point in $\mathbb{R}^2$. The same upper bound holds for $\| \mathrm{I}_D - \mathrm{I}_E \|_2$ and $\| \mathrm{I}_D - \mathrm{I}_E \|_\infty$ as well.

Section 4

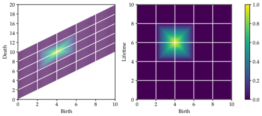
Tent Functions

Paper

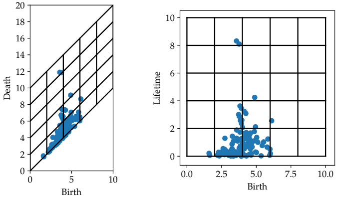
Template functions

Step 1: Choose collection of functions: $\{f_i\}$

- f_i :  $\to \mathbb{R}$
- compact support



Step 2: Evaluate each function f at each point in each diagram: $f(p)$ for $p \in D$.



Step 3: For fixed diagram D and function f, sum up for all points in diagram: $\nu_f(D) = \sum_{p \in D} f(p)$

Result: $\{f_i\}_{i=1}^k \quad \rightsquigarrow \quad D \mapsto (\nu_{f_1}(D), \nu_{f_2}(D), \cdots, \nu_{f_k}(D))$

Template function definition

- $\mathcal{D}$: Space of persistence diagrams
- $C_c(\mathbb{W})$: functions from $\mathbb{W} =$  to $\mathbb{R}$ with compact support

Definition

A **coordinate system** for $\mathcal{D}$ is a collection $\mathcal{F} \subset C(\mathcal{D}, \mathbb{R})$ which *separates points*.

$D \neq D' \in \mathcal{D}$, then there exists $F \in \mathcal{F}$ for which $F(D) \neq F(D')$.

Definition

A **template system** for $\mathcal{D}$ is a collection $\mathcal{T} \subset C_c(\mathbb{W})$ so that

$$\mathcal{F}_{\mathcal{T}} = \{\nu_f : f \in \mathcal{T}\}$$

is a coordinate system for $\mathcal{D}$.

The elements of $\mathcal{T}$ are called **template functions**.

$$\nu_f(D) = \sum_{p \in D} f(p)$$

In theory, it works...

Theorem (Perea, Munch, Khasawneh, 2019)

- *Let* $\mathcal{T} \subset C_c(\mathbb{W})$ *be a template system for* $\mathcal{D}$,
- $\mathcal{C} \subset \mathcal{D}$ *compact, and*
- $F : \mathcal{C} \longrightarrow \mathbb{R}$ *be continuous.*

Then for every $\varepsilon > 0$ *there exist*

- $N \in \mathbb{N}$,
- *a polynomial* $p \in \mathbb{R}[x_1, \ldots, x_N]$ *and*
- *template functions* $f_1, \ldots, f_N \in \mathcal{T}$

so that

$$\left| p(\nu_{f_1}(D), \nu_{f_2}(D), \cdots, \nu_{f_k}(D)) - F(D) \right| < \varepsilon$$

for every $D \in \mathcal{C}$.

... what about in practice?

Questions

- What choice of template functions?
- What polynomial?

Unsatisfying answers

- Any collection of functions on $\mathbb{W}$ with compact support that separate points should work.
- Machine learning to the rescue!

Tent functions

Parameters

- d: number of subdivisions
- δ: partition scale
- ε: shift away from the diagonal

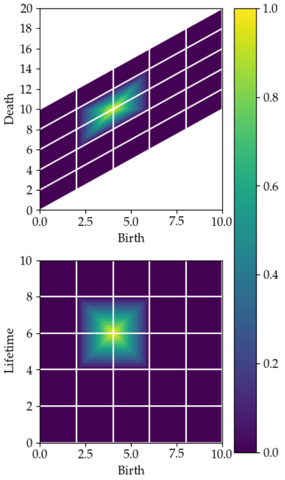
Tent function (i, j) in the birth-lifetime plane

$$i \in 0, \cdots, d; \; j \in 1, \cdots, d$$

$$\tilde{g}_{i,j}(x, \tilde{y}) = \left|1 - \frac{1}{\delta} \max\left\{|x - \delta i|, |\tilde{y} - (\delta j + \varepsilon)|\right\}\right|_+$$

Given a persistence diagram $D = (S, \mu)$,

$$G_{i,j}(D) = \sum_{\tilde{\mathbf{x}} \in S} \mu(\tilde{\mathbf{x}}) \cdot \tilde{g}_{i,j}(\tilde{\mathbf{x}})$$



Parameters

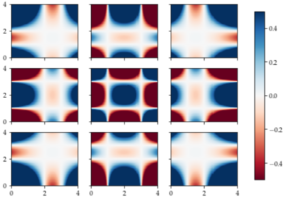
- $\mathcal{A} = \{a_1 < a_2 < \cdots < a_n\}$: partition of x-axis
- $\mathcal{B} = \{b_1 < b_2 < \cdots < b_n\}$: partition of y-axis

Interplating polynomial (i, j) in the birth-lifetime plane

$$f(x, y) = \ell_i^{\mathcal{A}}(x) \cdot \ell_j^{\mathcal{B}}(y)$$

where

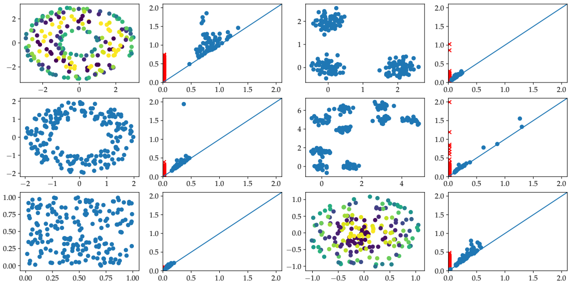
$$\ell_j^{\mathcal{A}}(x) = \prod_{i \neq j} \frac{x - a_i}{a_j - a_i} \qquad \ell_j^{\mathcal{B}}(x) = \prod_{i \neq j} \frac{x - b_i}{b_j - b_i}$$

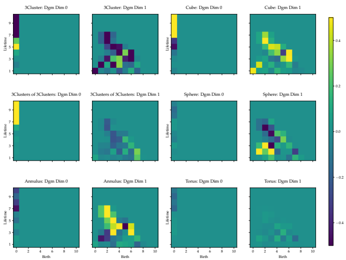


Given a persistence diagram $D = (S, \mu)$,

$$F_{i,j}^{\mathcal{A},\mathcal{B},K,\varepsilon}(D) := \sum_{\tilde{\mathbf{x}} \in \tilde{S}} \mu(\tilde{\mathbf{x}}) \cdot f_{i,j}^{\mathcal{A},\mathcal{B}}(\tilde{\mathbf{x}}) \cdot h_{K,\varepsilon}(\tilde{\mathbf{x}}).$$

Manifold experiment

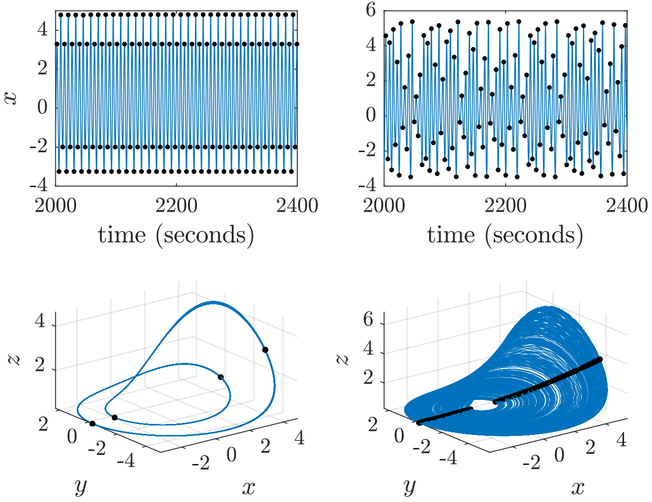




	Tents		Polynomials	
No. Dgms	Train	Test	Train	Test
10	$99.8\% \pm 0.9$	$96.5\% \pm 3.2$	$99.8\% \pm 0.9$	$95.0\% \pm 3.9$
25	$99.9\% \pm 0.3$	$99.0\% \pm 1.0$	$99.7\% \pm 0.5$	$97.6\% \pm 1.5$
50	$99.9\% \pm 0.2$	$99.9\% \pm 0.3$	$100\% \pm 0$	$99.2\% \pm 0.9$
100	$99.8\% \pm 0.1$	$99.7\% \pm 0.4$	$99.6\% \pm 0.2$	$99.3\% \pm 0.5$
200	$99.5\% \pm 0.1$	$99.5\% \pm 0.3$	$99.2\% \pm 0.2$	$98.9\% \pm 0.5$

Rössler

Classification of chaotic vs periodic



Section 5

The point

The point

- Can either work with persistence diagrams directly, or map somewhere else
- Working with persistence diagrams directly is hard
- Mapping somewhere else means you need to decide on that map.