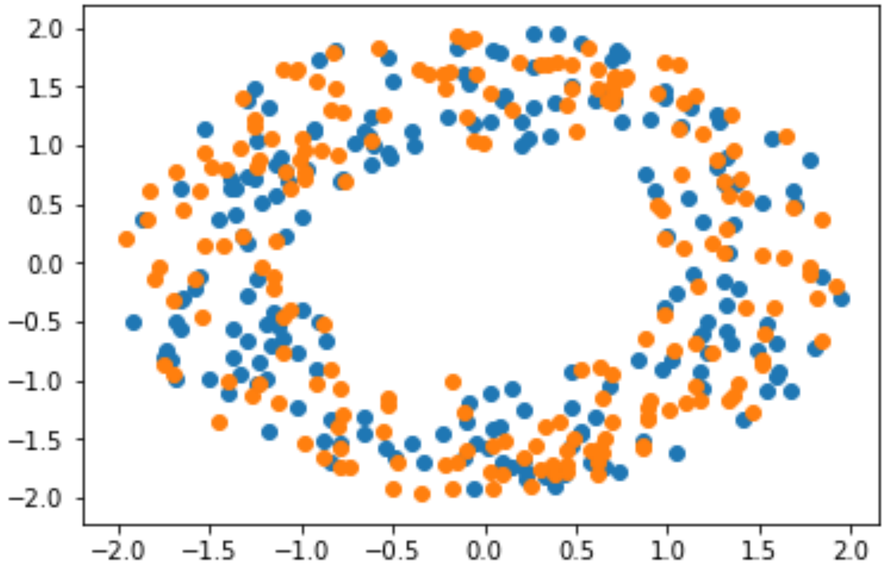
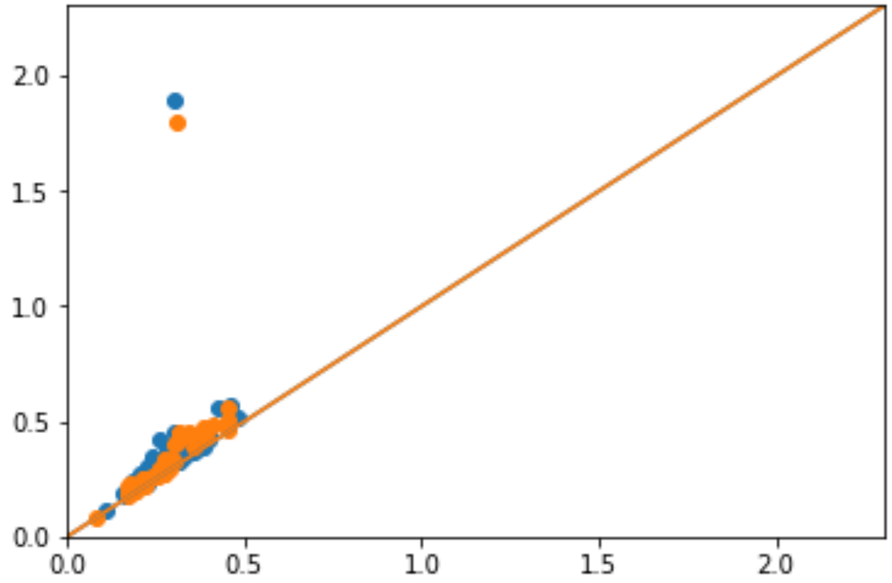
Distances, Stability, Vectorization, and Applications

Dr. Elena Xinyi Wang

November 4, 2026

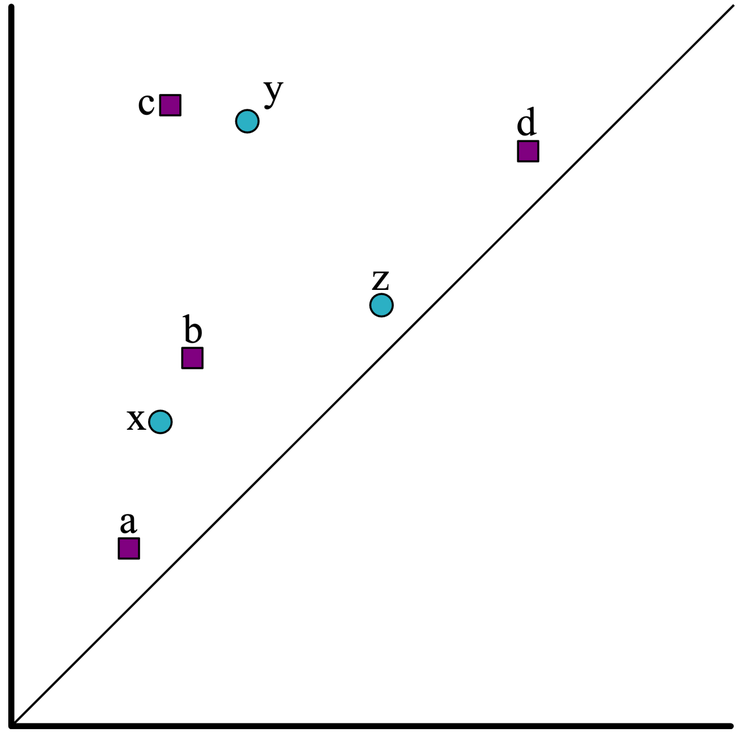
Similar Persistence Diagrams

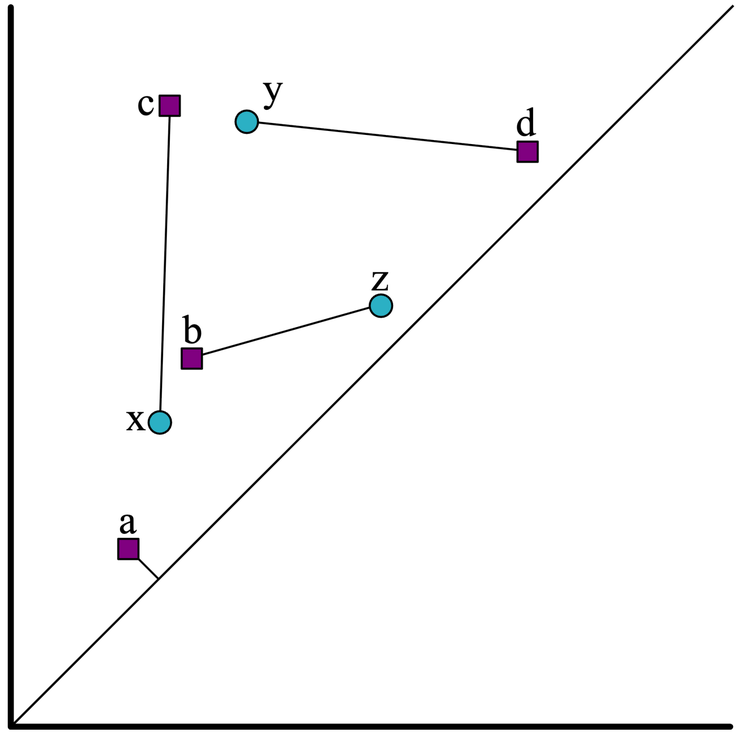


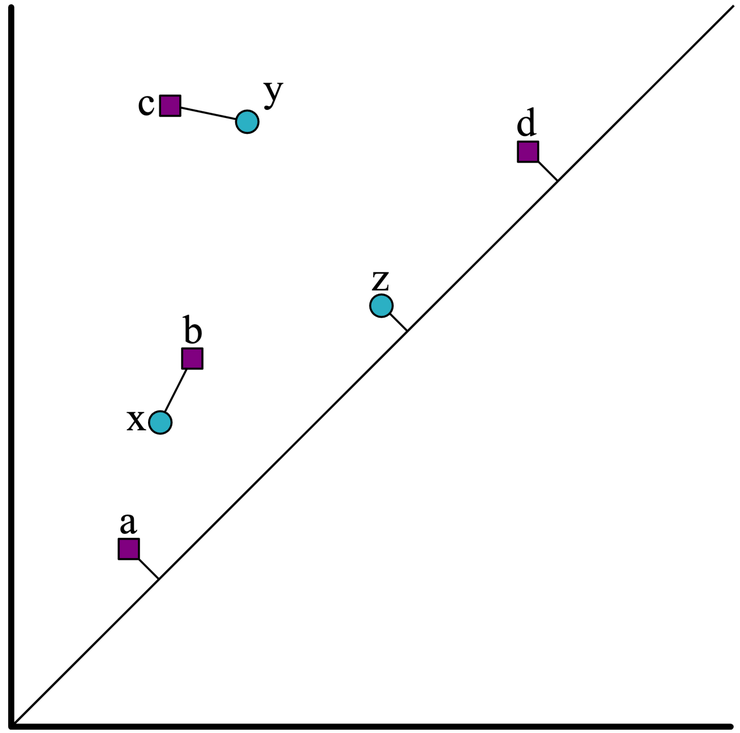


How different are they?

Find Matchings





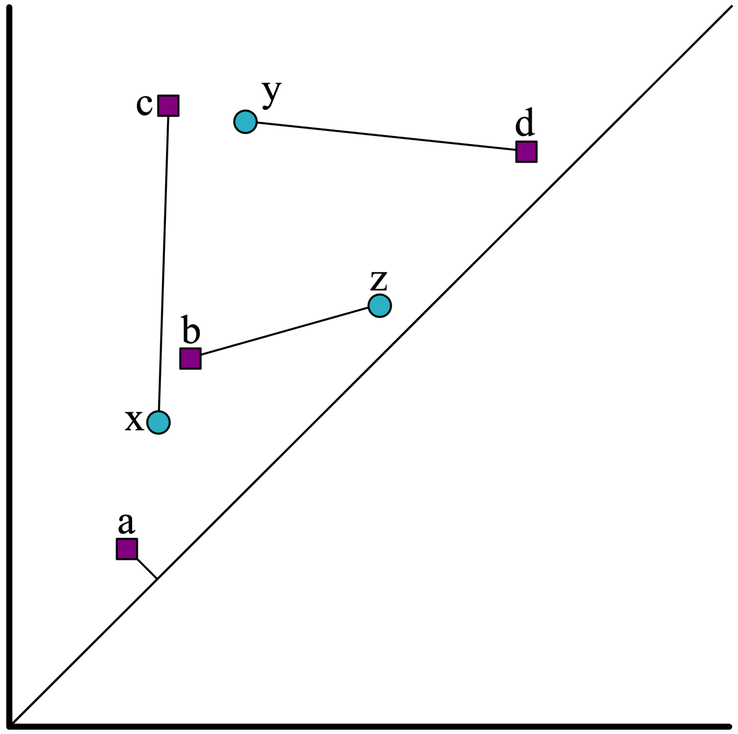


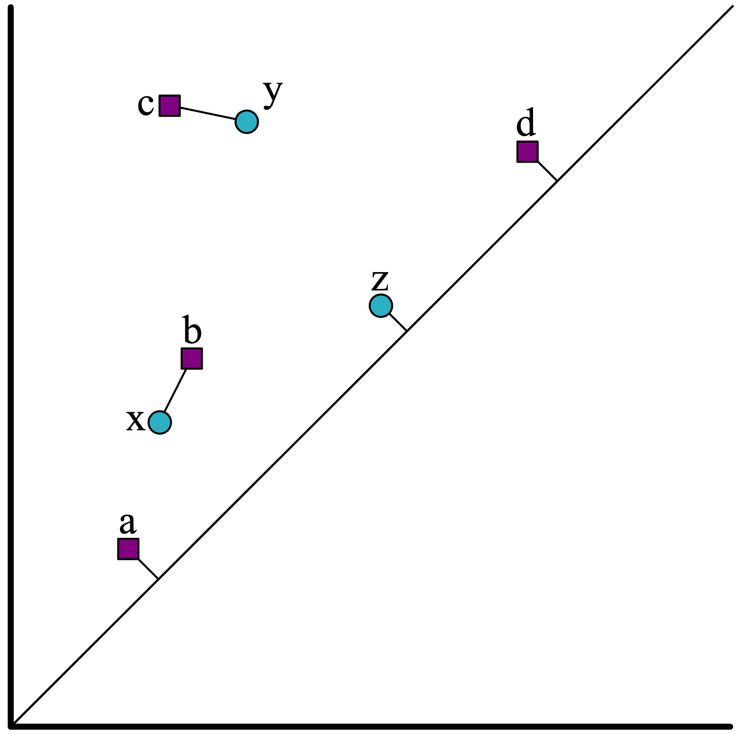
Bottleneck Distance, Version 1

- Given two persistence diagrams X, Y, the **bottleneck distance** is

$$d_B(X, Y) = \inf_{\eta: X \to Y} \sup_{x \in X} \|x - \eta(x)\|_\infty$$

- $\eta : X \to Y$ ranges over all bijections

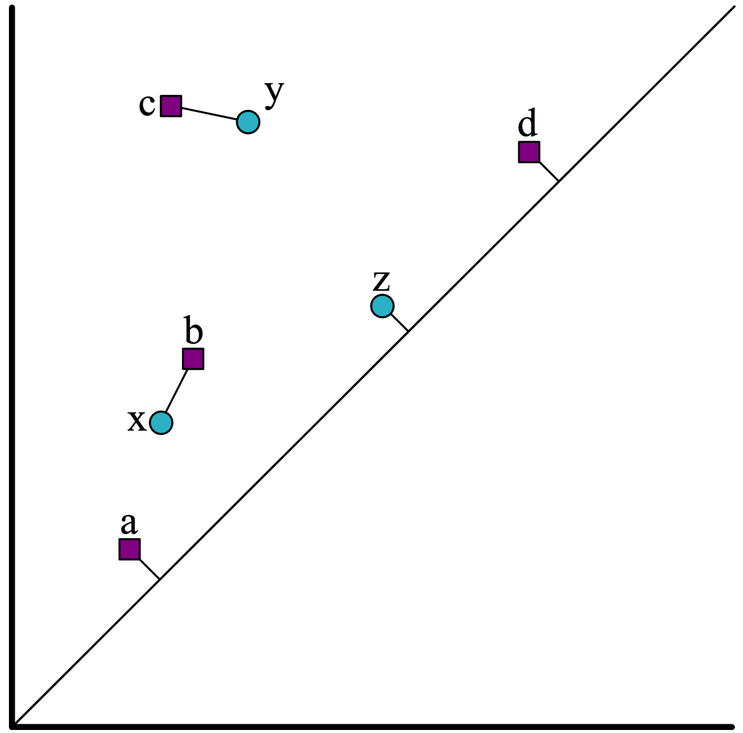




Bottleneck Distance, Version 2

Partial Matching

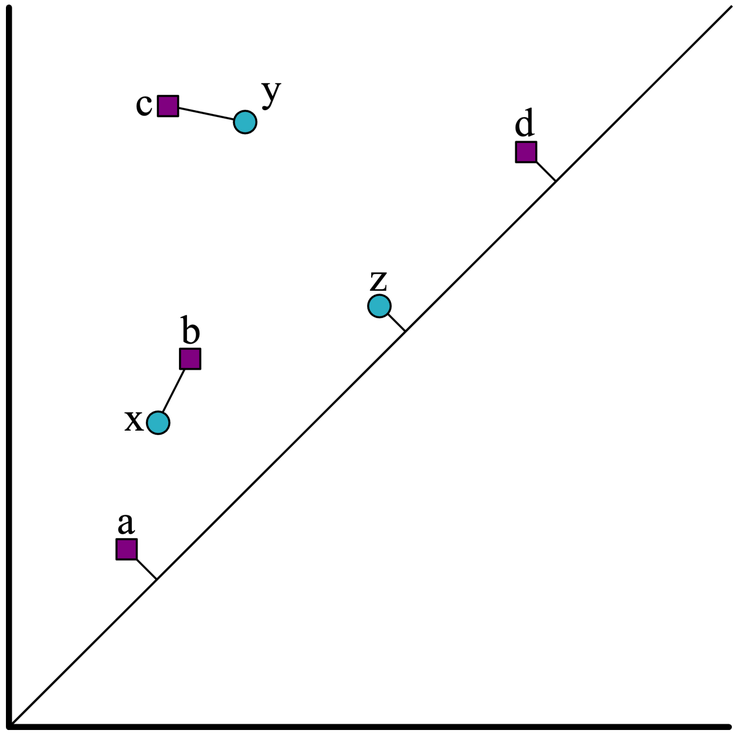
A partial matching between X and Y is a bijection on a subset of the off-diagonal points of the two diagrams $M : X' \to Y'$ for $X' \subseteq X$ and $Y' \subseteq Y$



Cost of a matching

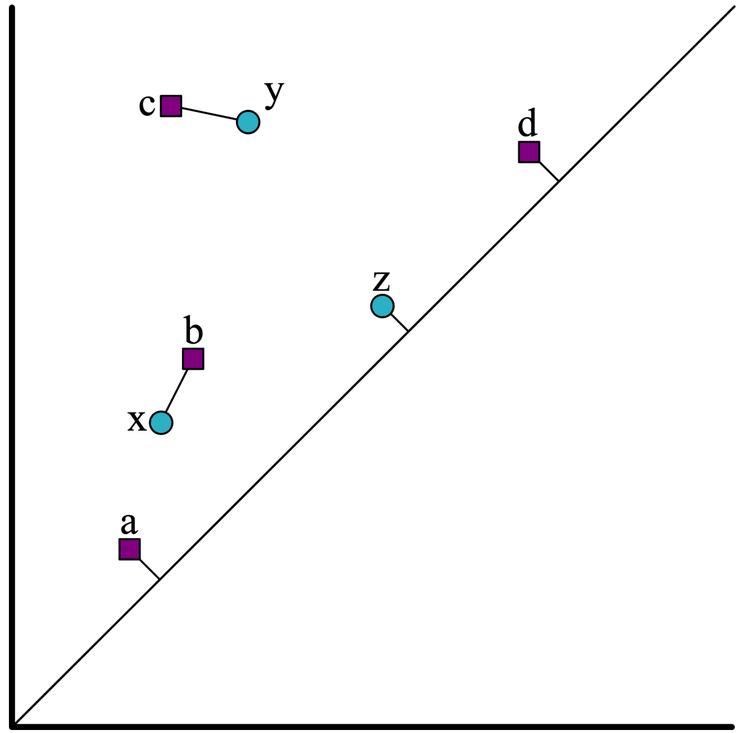
The cost of this partial matching is given by

$$c(M) = \sup\left(\{\|x - M(x)\|_\infty \mid x \in X'\} \cup \left\{\frac{1}{2}|x_1 - x_2| \mid (x_1, x_2) \in (X \setminus X') \cup (Y \setminus Y')\right\}\right)$$



Bottleneck Distance, V2

The bottleneck distance between X and Y is $d_B(X, Y) = \inf_M c(M)$, where the infinum runs over all possible partial matchings



Hausdorff distance

Let $X, Y \subset \mathbb{R}^d$ be finite point clouds. The Hausdorff distance is

$$d_H(X, Y) = \max \left\{ \max_{x \in X} \min_{y \in Y} \|x - y\|, \max_{y \in Y} \min_{x \in X} \|x - y\| \right\}.$$

Bottleneck Stability

Point Cloud Version

Let $X, Y \subset \mathbb{R}^d$ be finite point clouds. Let d_H be the Hausdorff distance. Let $\text{Dgm}(\mathbb{C}(X))$ be the persistence diagram of the filtration defined by the Čech complex on X. Then

$$0 \leq d_B(\text{Dgm}_p(\mathbb{C}(X), \text{Dgm}_p(\mathbb{C}(Y))) \leq d_H(X, Y)$$

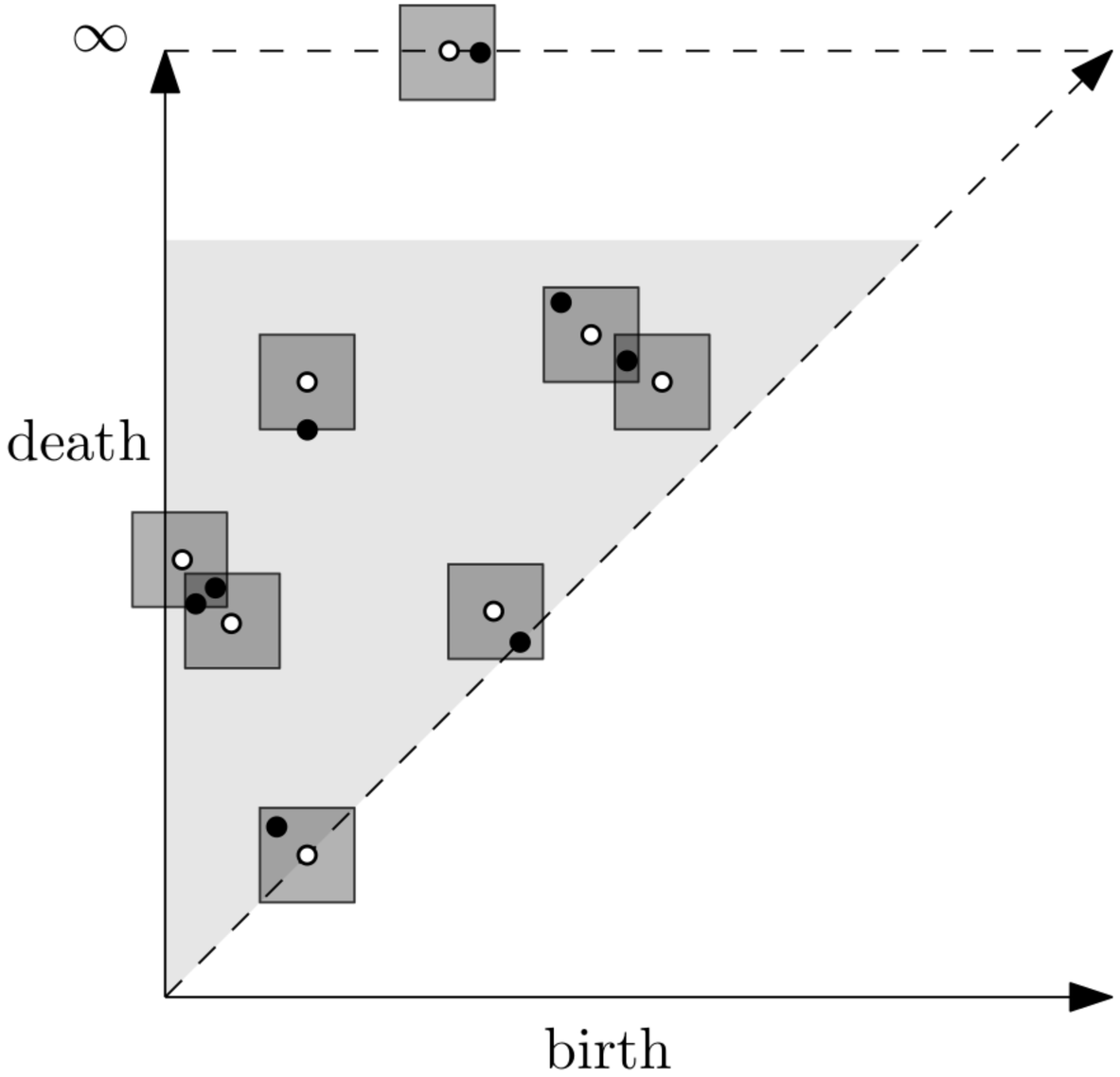
Bottleneck Stability

Function Version

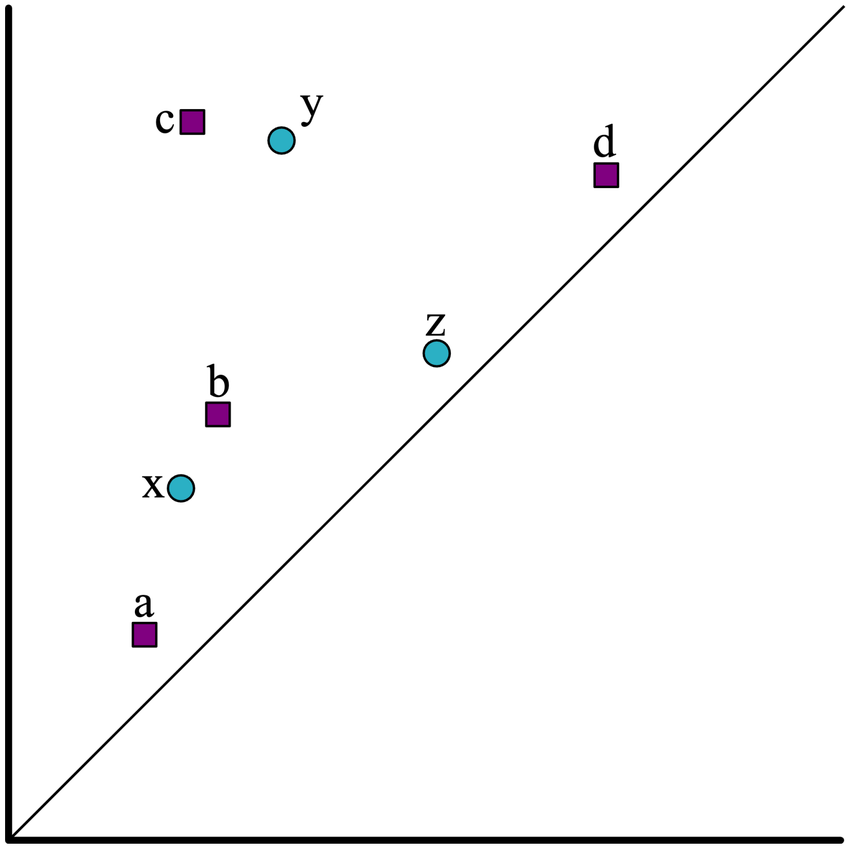
Let $f, g : K \to \mathbb{R}$ be two simplex-wise monotone functions giving rise to simplicial filtrations $\mathscr{F}_f$ and $\mathscr{F}_g$. Then, for every $p \geq 0$,

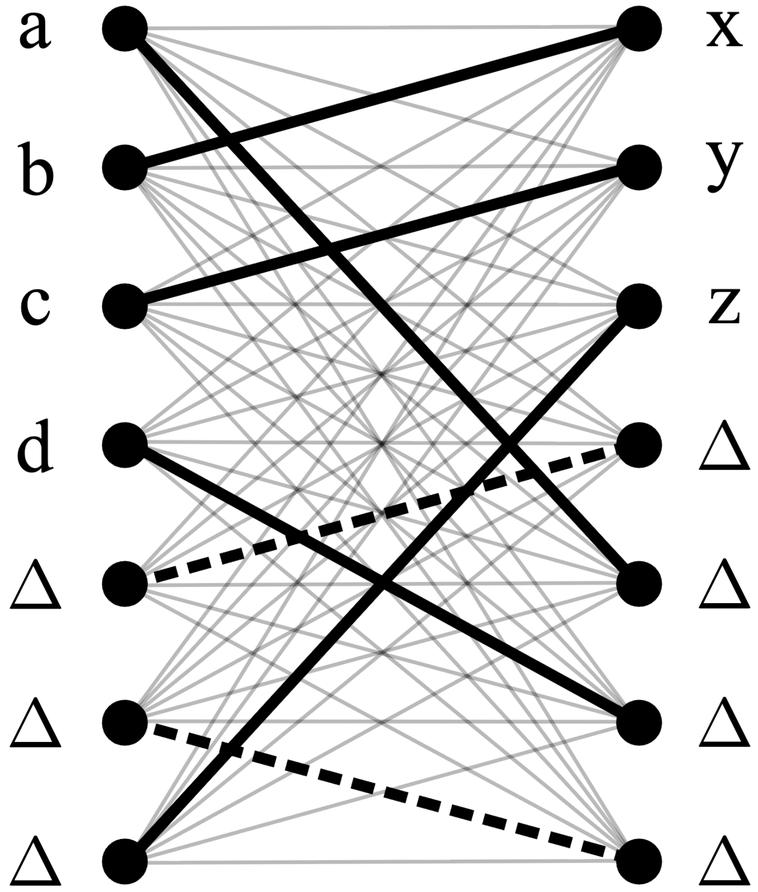
$$0 \leq d_B(\mathrm{Dgm}_p(\mathscr{F}_f), \mathrm{Dgm}_p(\mathscr{F}_g) \leq \|f - g\|_\infty$$

Another View: Nearby Diagrams



Computing Bottleneck Distance

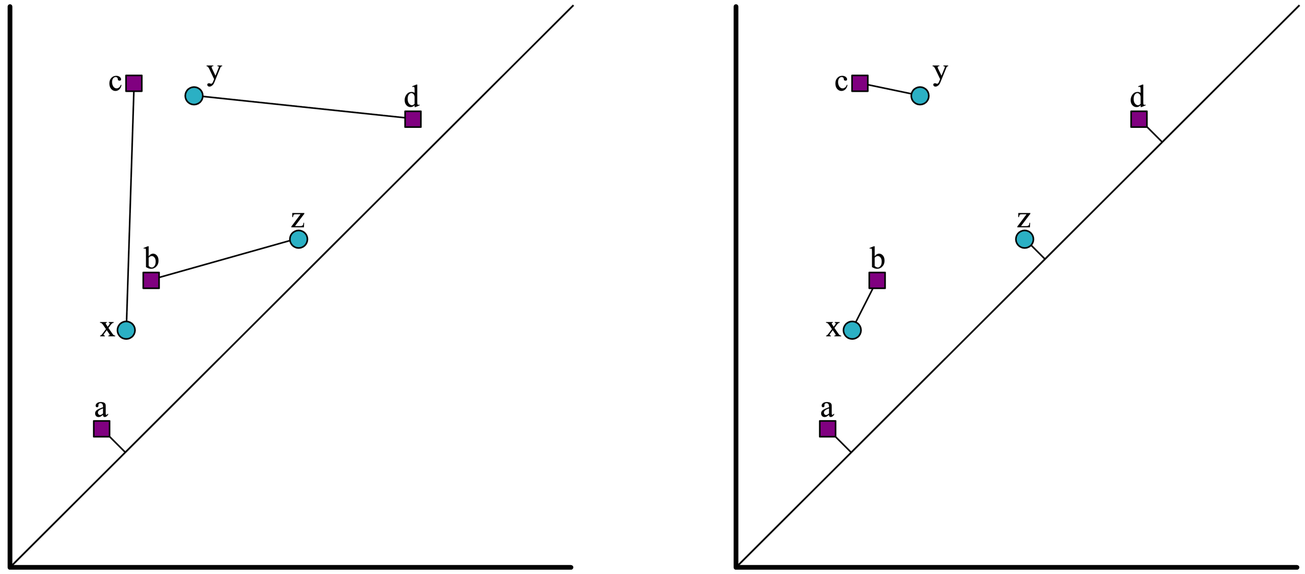




Wasserstein Distance

Given diagrams X and Y, the p-Wasserstein distance between them is

$$W_p^q(X,Y) = \inf_{\ell : X \to Y} \left(\sum_{x \in X} \|x - \ell(x)\|_p^q \right)^{1/q} = \inf_{\ell : X \to Y} \left(\sum_{x \in X} \left(\|x - \ell(x)\|_p \right)^q \right)^{1/q}.$$



Warning

Notation

- Textbook definition: $d_{W_q}\left(\mathrm{Dgm}_p(\mathscr{F}_f), \mathrm{Dgm}_p(\mathscr{F}_g)\right) = \inf_{\pi \in \Pi} \left[\sum_{x \in \mathrm{Dgm}_p(\mathscr{F}_f)} \|x - \pi(x)\|_q^q \right]^{1/q}.$
- Other reference: $W_q(X, Y) = \left[\inf_{\eta : X \to Y} \sum_{x \in X} \|x - \eta(x)\|_\infty^q \right]^{1/q}.$
- Different references use different conventions for Wasserstein distances.
- The ground metric may be ℓ^q or ℓ^∞

Wasserstein Stability

For simplicial filtrations

Let K be a finite simplicial complex, and let $f, g : K \to \mathbb{R}$ be simplex-wise monotone functions. Then, for any $p \geq 0$ and $q \geq 1$: $d_{W,q}\left(\mathrm{Dgm}_p(\mathscr{F}_f), \mathrm{Dgm}_p(\mathscr{F}_g)\right) \leq \|f - g\|_q$, where

$$\|f - g\|_q = \left(\sum_{\sigma \in K} |f(\sigma) - g(\sigma)|^q \right)^{1/q}.$$

Wasserstein Stability

For Lipschitz functions

A function $f : X \to \mathbb{R}$ is L-Lipschitz if $|f(x) - f(y)| \leq L\, d(x, y) \qquad \forall x, y \in X.$

Let X be a compact triangulable metric space with bounded degree-k total persistence.

Let $f, g : X \to \mathbb{R}$ be tame Lipschitz functions. Then there exists $C > 0$ such that, for any $p \geq 0$ and $q \geq k$: $d_{W,q}\left(\mathbf{Dgm}_p(\mathscr{F}_f), \mathbf{Dgm}_p(\mathscr{F}_g)\right) \leq C\, \|f - g\|_\infty^{1-k/q}.$

Stats and ML for Persistence

Questions and Tasks

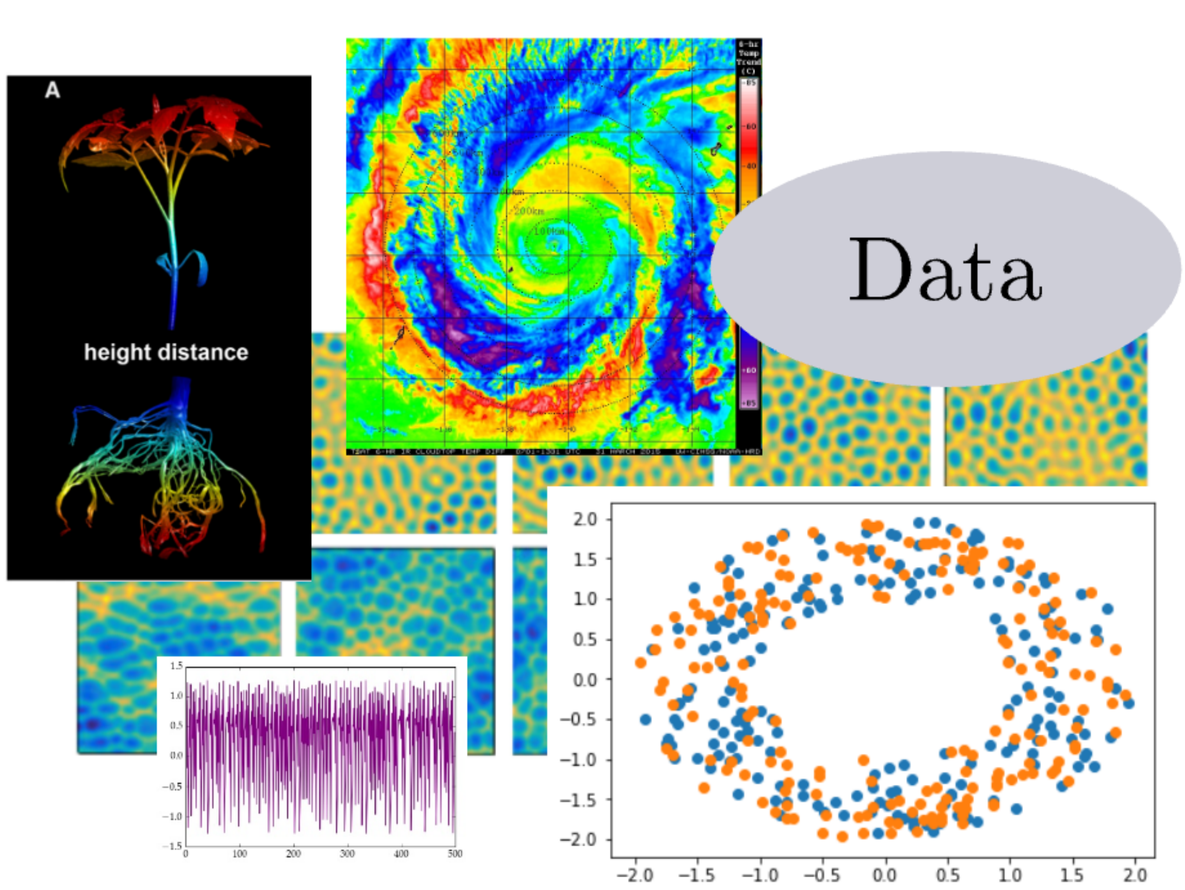
- **Tasks**
  - Classification
  - Regression
- **The Issues**
  - Most methods want vectors in $\mathbb{R}^n$
  - Implies sorting or ordering, which we don't want for persistence diagrams
  - Implies the same n for all data points
  - Usually need data points to come from a vector space

More Problems

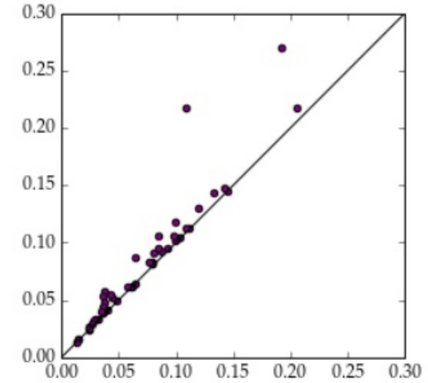
- Usually, we want ML inputs to come from a Banach space (complete normed vector space) or a Hilbert space (vector space with an inner product and limits).
- Note all Hilbert spaces are Banach spaces, but not all Banach spaces are Hilbert spaces
- Theorem (Bubenik Wagner 2020): The space $(\mathscr{D}, W_p)$ does not admit an isometric embedding into a Hilbert space for any $1 \leq p \leq \infty$

Featurization: the Idea

Here be dragons (Hic sunt dracones)



Data

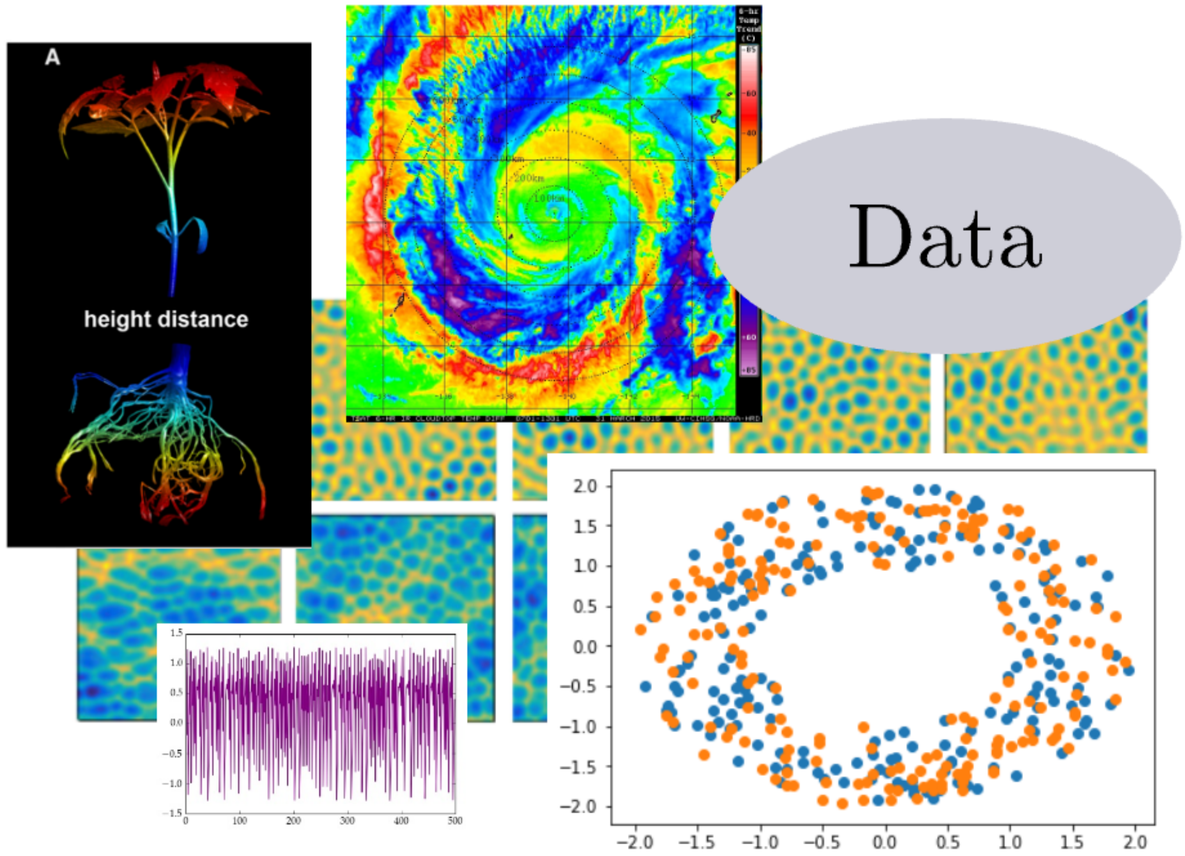


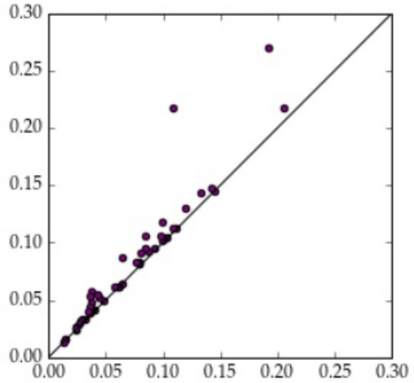
Pers

Featurization: the Idea

Here be dragons (Hic sunt dracones)

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

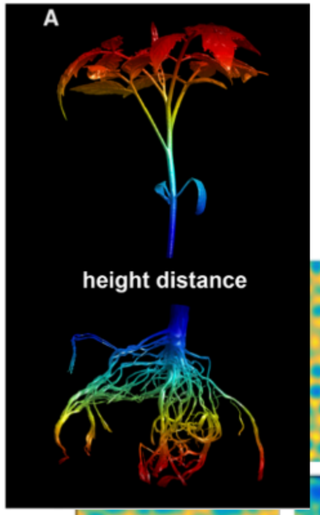


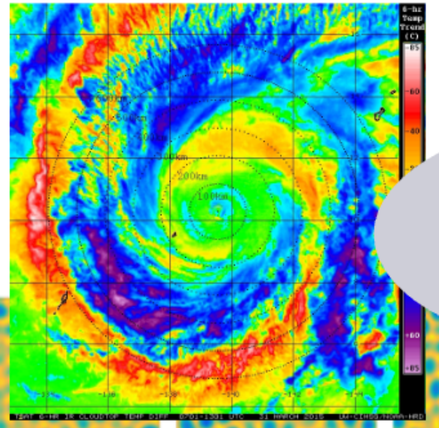


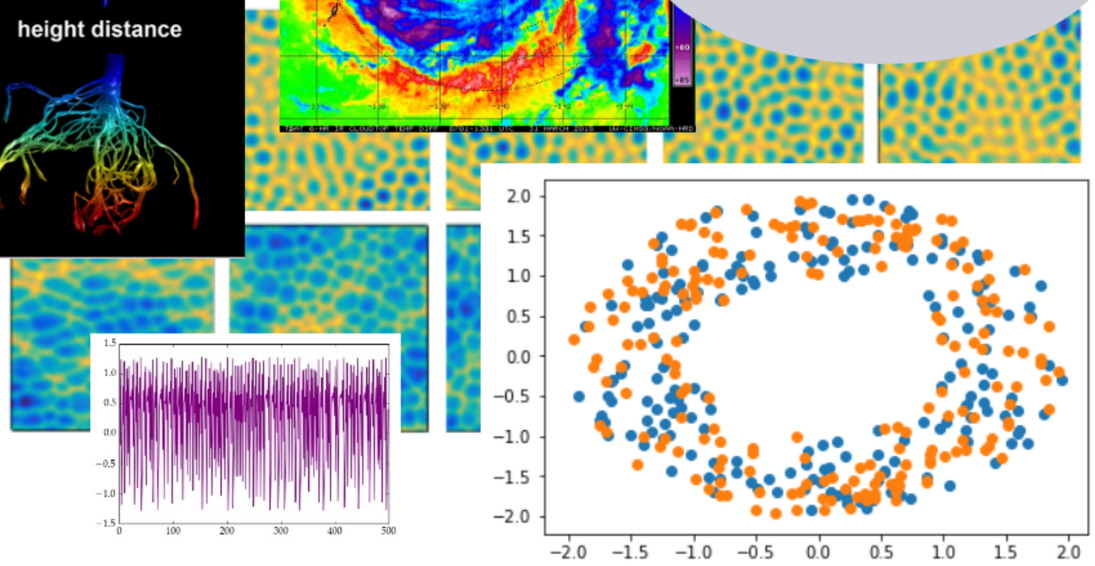
Pers

Featurization: the Idea

Here be dragons (Hic sunt dracones)

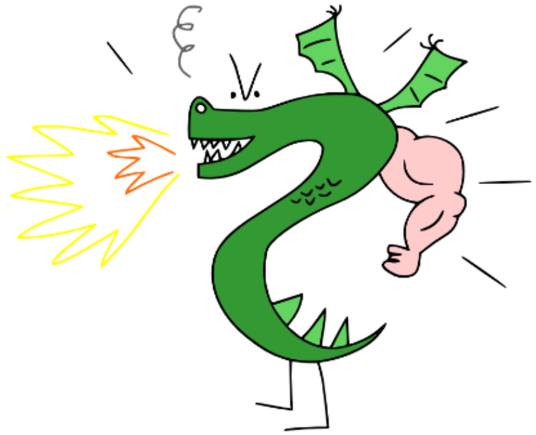






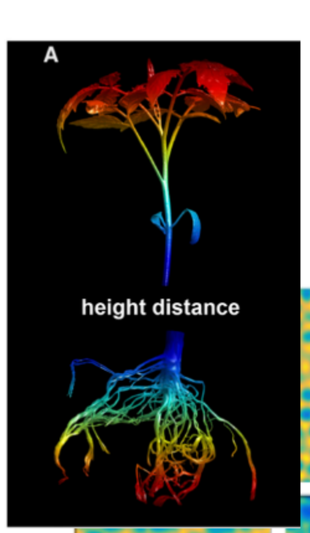
Data

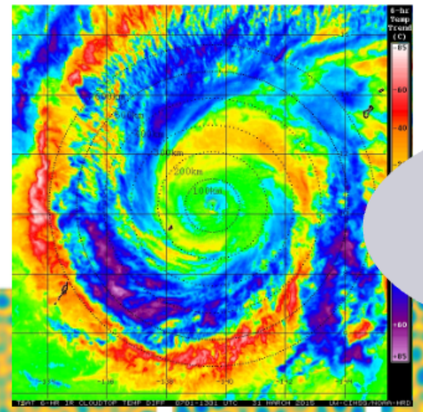
Pers

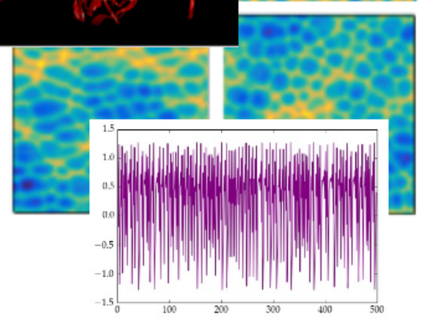


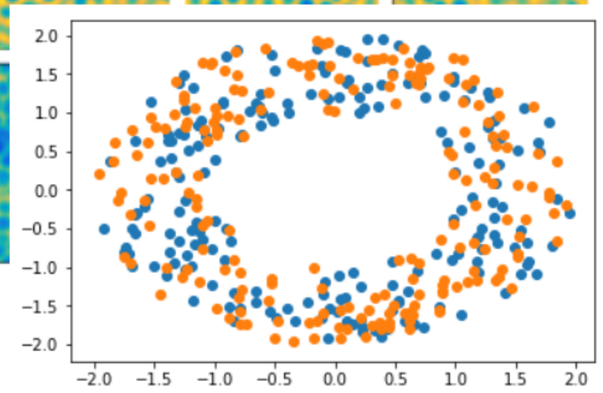
Featurization: the Idea

Here be dragons (Hic sunt dracones)

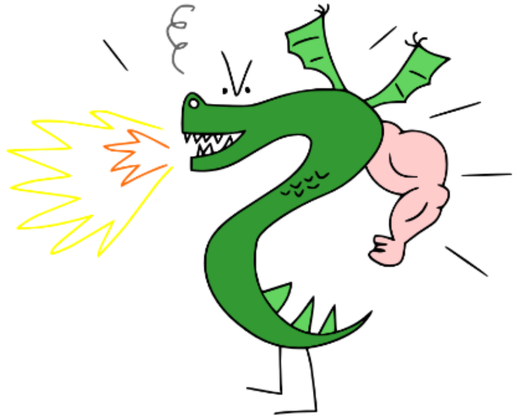








Data



Pers

$\mathbb{R}^d$

Turning PDs into something else

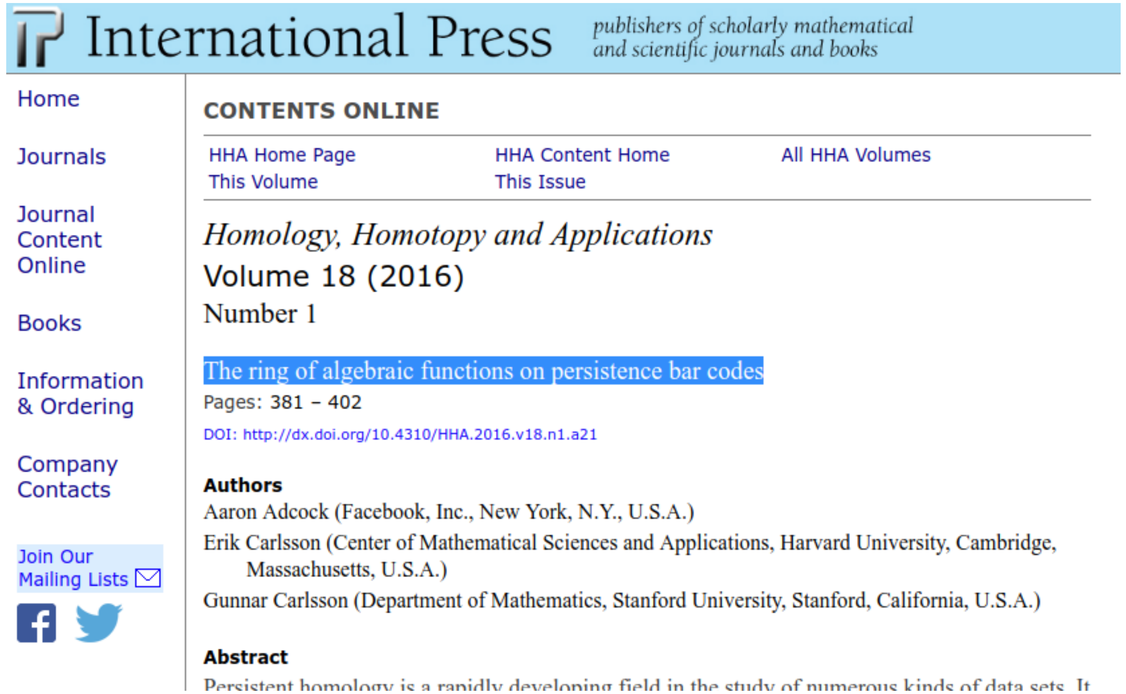
Algebraic Structures

Forget about the diagrams

- Just treat a persistence diagram as a set of points in $\mathbb{R}^2$
- **Problems:**
  - Order shouldn't matter
  - n isn't fixed

Algebraic Geometry to the Rescue

- Associate persistence diagrams to $k[x_1, y_1, x_2, y_2, \ldots]/\sim$
- 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

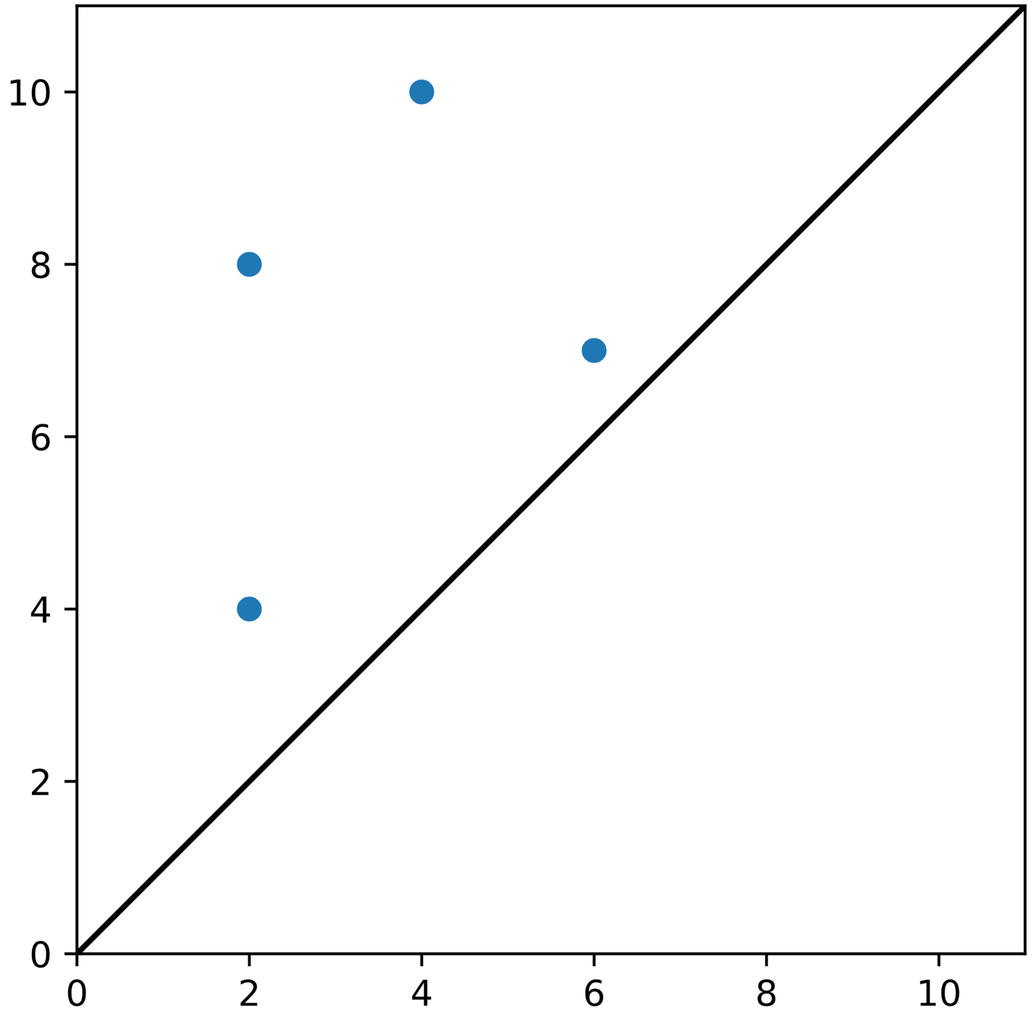
Homology, Homotopy and Applications
Volume 18 (2016)
Number 1

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



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

MNIST Data



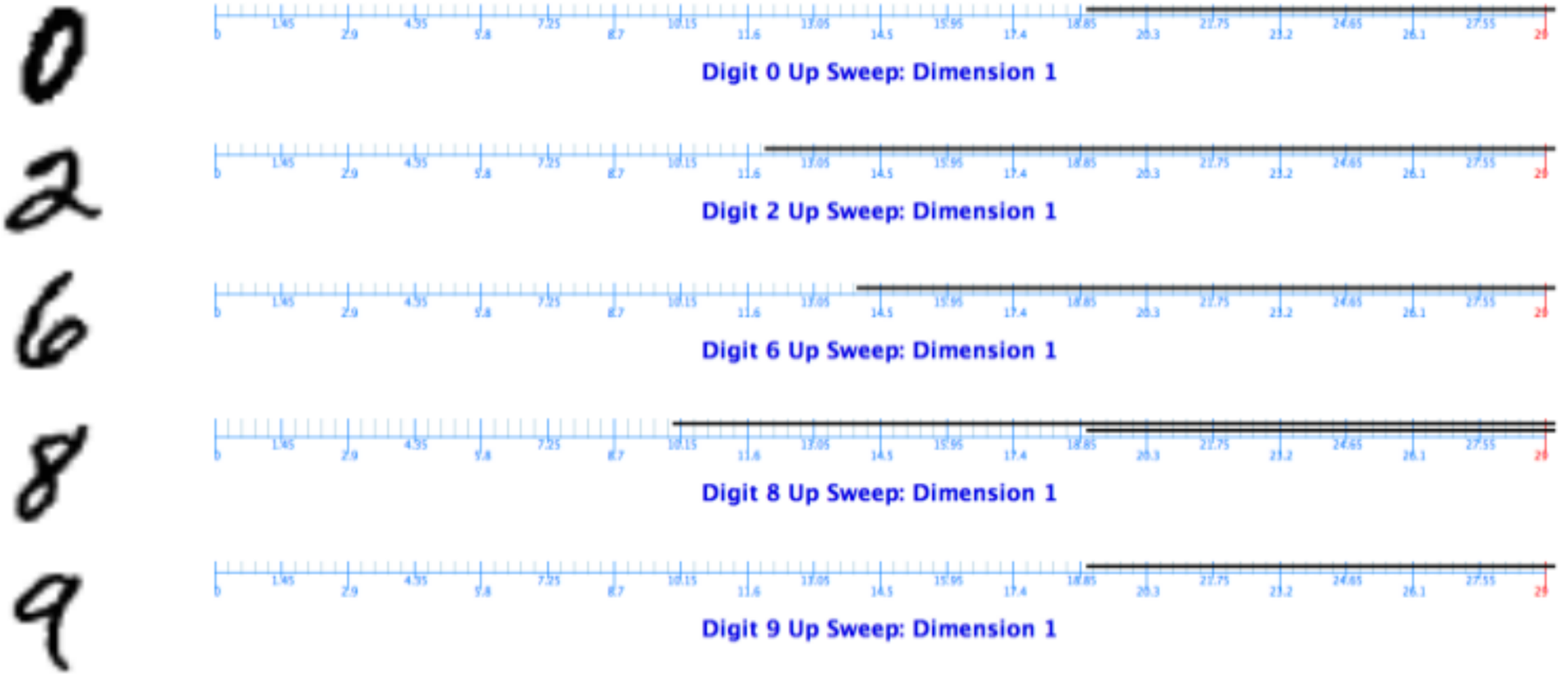
Build complex

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

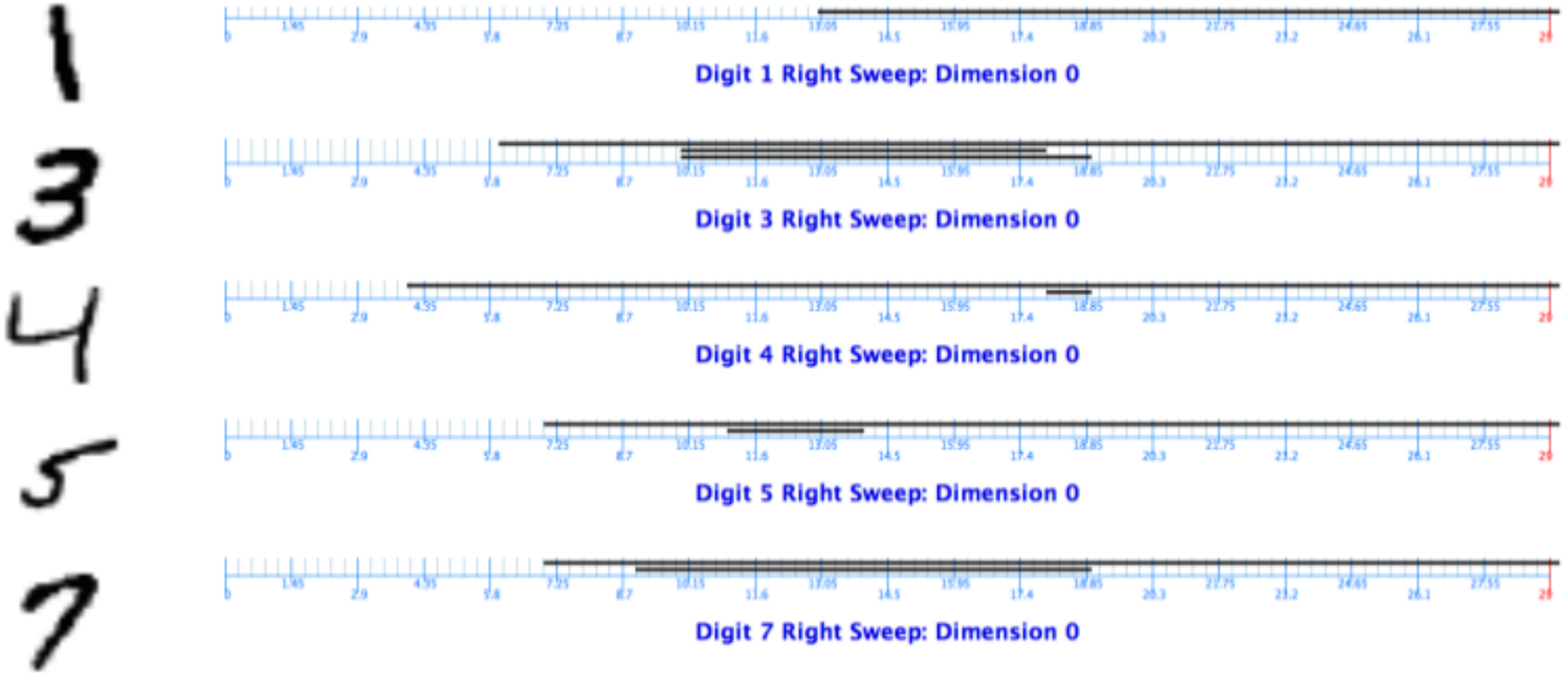
Filtration:

- 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

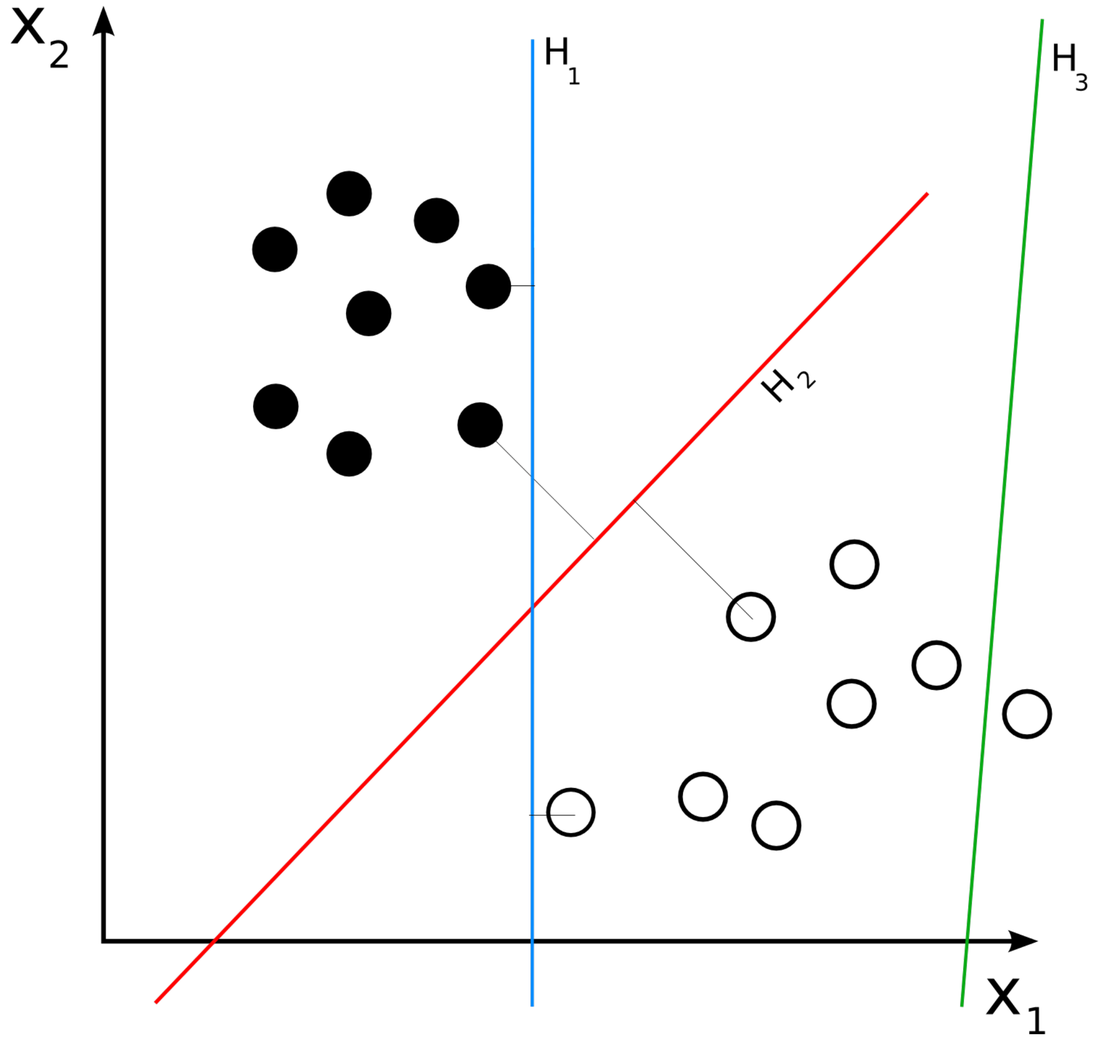
Features:

$y_{\max}$ = 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 Machines (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:

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

Cons:

- Not stable with respect to distances

Landscapes

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

Statistical Topological Data Analysis using Persistence Landscapes

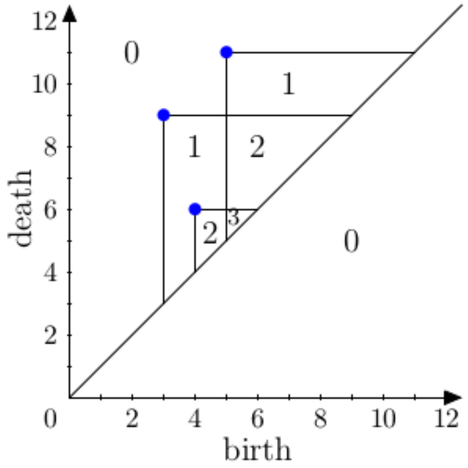
Peter Bubenik PETER.BUBENIK@GMAIL.COM
Department of Mathematics
Cleveland State University
Cleveland, OH 44115-2214, USA

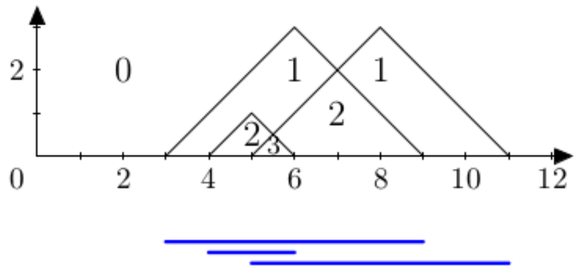
Editor: David Dunson

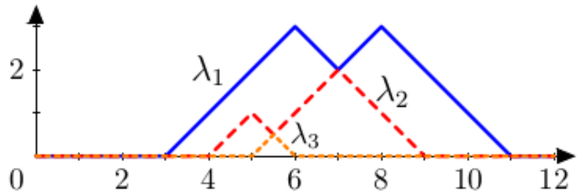
Abstract

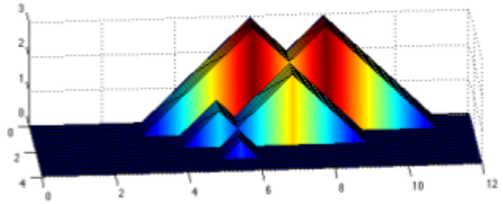
We define a new topological summary for data that we call the persistence landscape. Since this summary lies in a vector space, it is easy to combine with tools from statistics and machine learning, in contrast to the standard topological summaries. Viewed as a random variable with values in a Banach 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







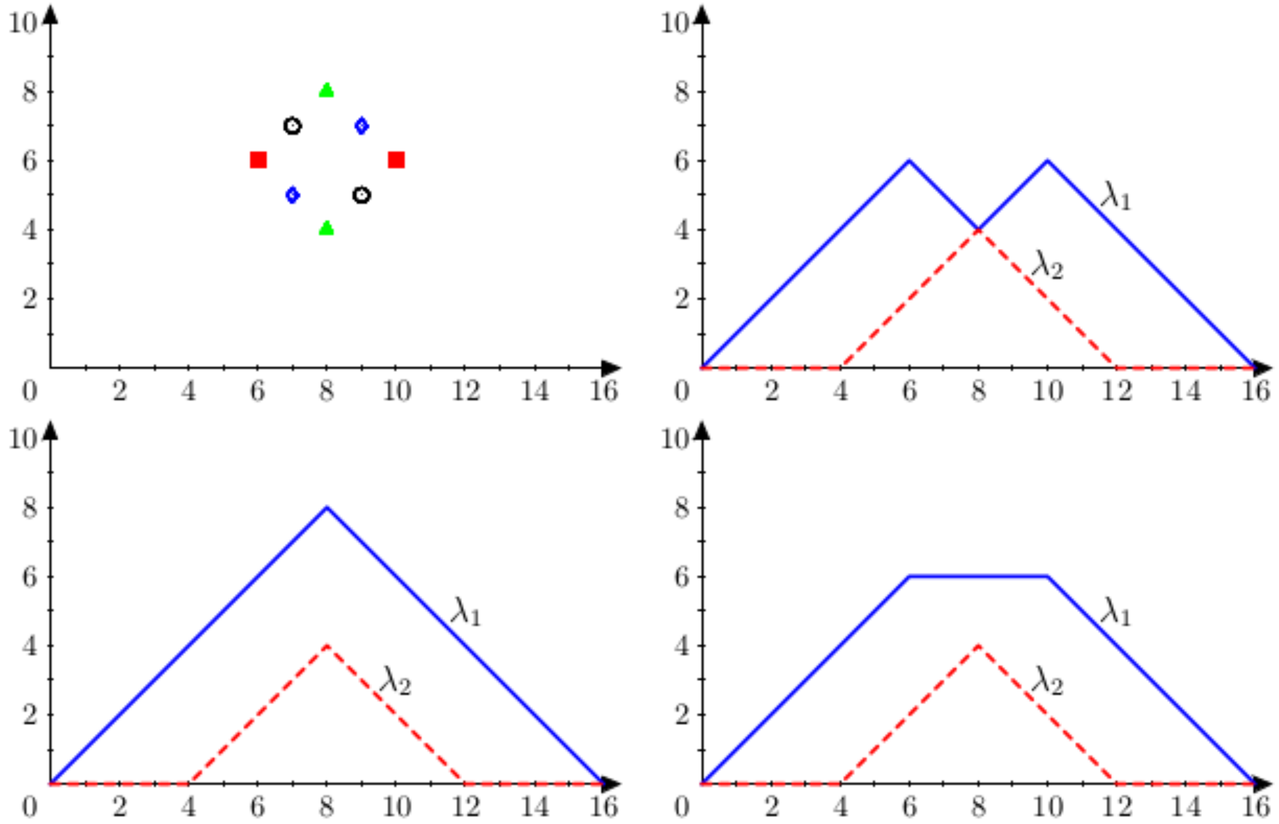


- Take diagram
- Rank function:

$$\beta^{a,b} = \dim\left(\mathrm{Im}\left(H_p(X_a) \to H_p(X_b)\right)\right)$$

- Rotate $(x, y) \mapsto \left(\frac{x+y}{2}, \frac{y-x}{2}\right)$
- $\lambda : \mathbb{N} \times \mathbb{R} \to \mathbb{R}, \quad (k, t) \mapsto \lambda_k(t)$
- $\lambda_k(t) = \sup\left\{m \geq 0 \mid \beta^{t-m,t+m} \geq k\right\}$

Example

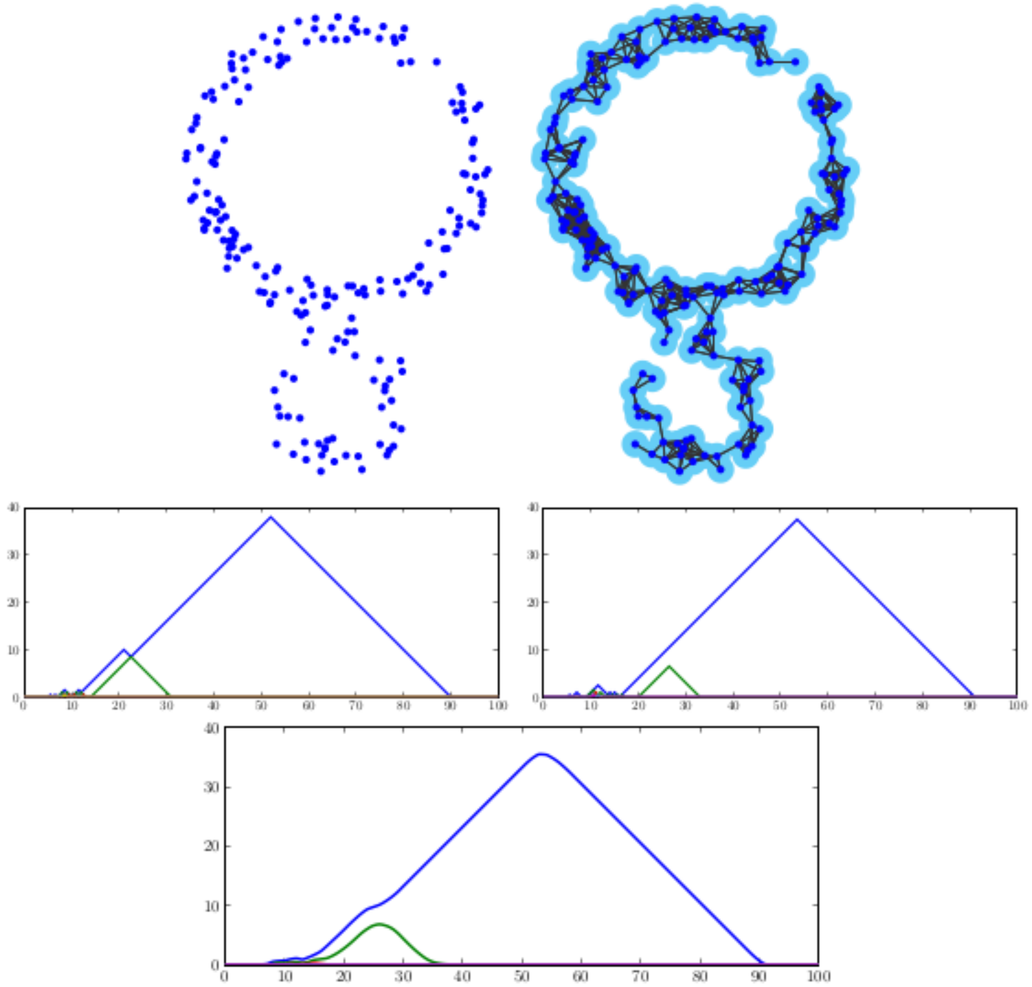

10
8
6
4
2
0
2
4
6
8
10
12
14
16

10
8
6
4
2
0
2
4
6
8
10
12
14
16
λ_1
λ_2

10
8
6
4
2
0
2
4
6
8
10
12
14
16
λ_1
λ_2

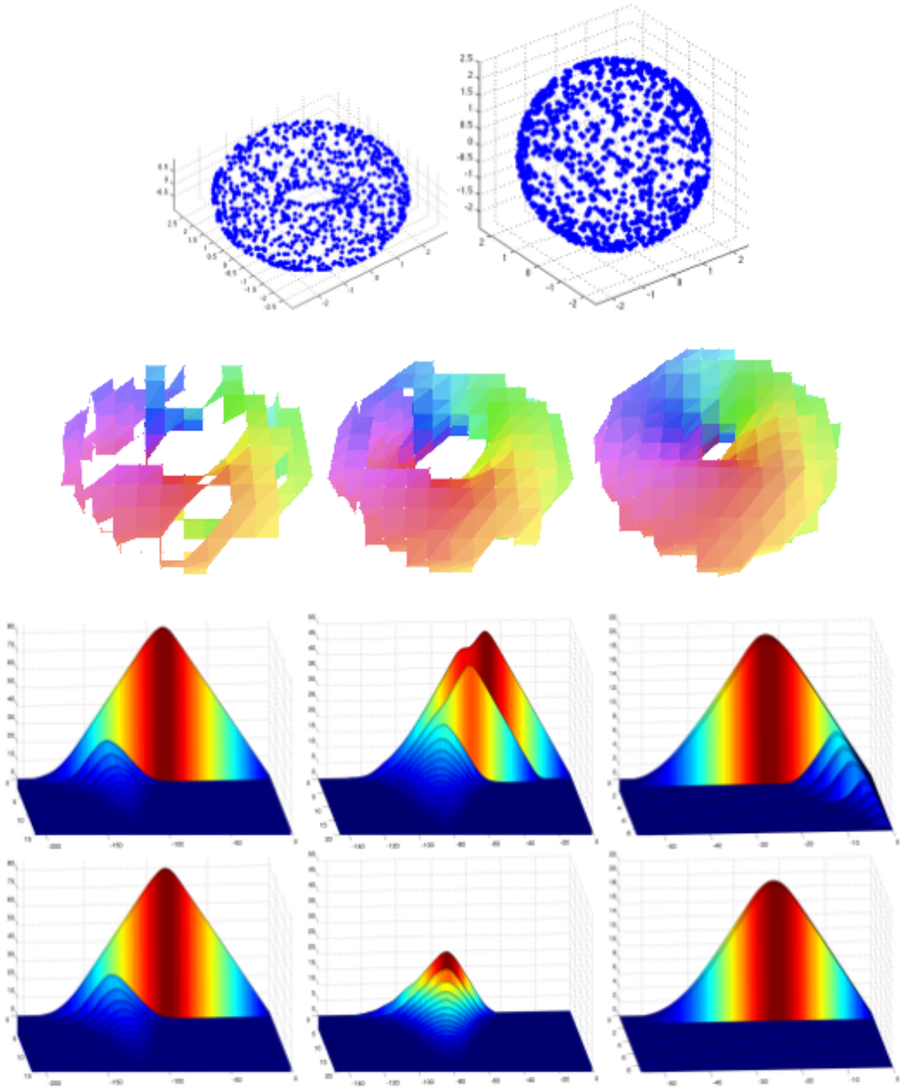
10
8
6
4
2
0
2
4
6
8
10
12
14
16
λ_1
λ_2

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
- Probably the first used ML featurization approach
- Satbility

Cons:

- Averages might not be something that comes from a diagram

Persistence Images

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

Their View on the Problem

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?

Setup

data

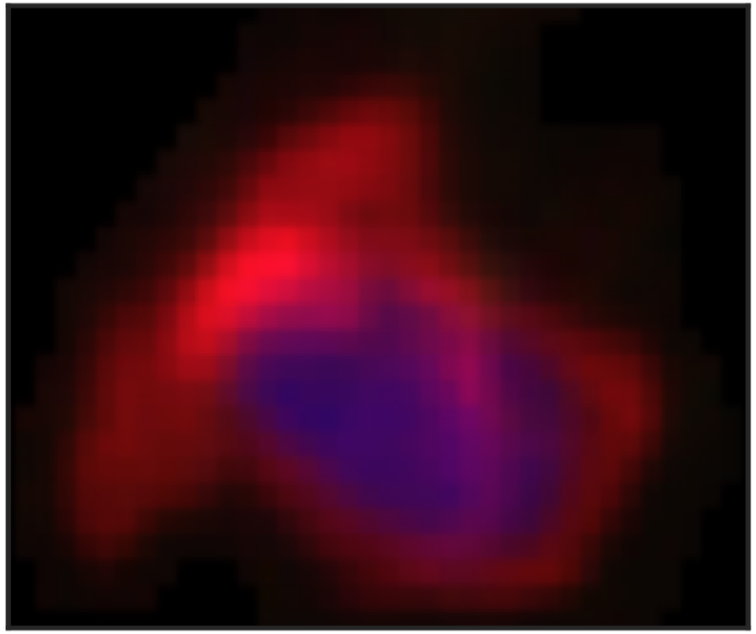


diagram B

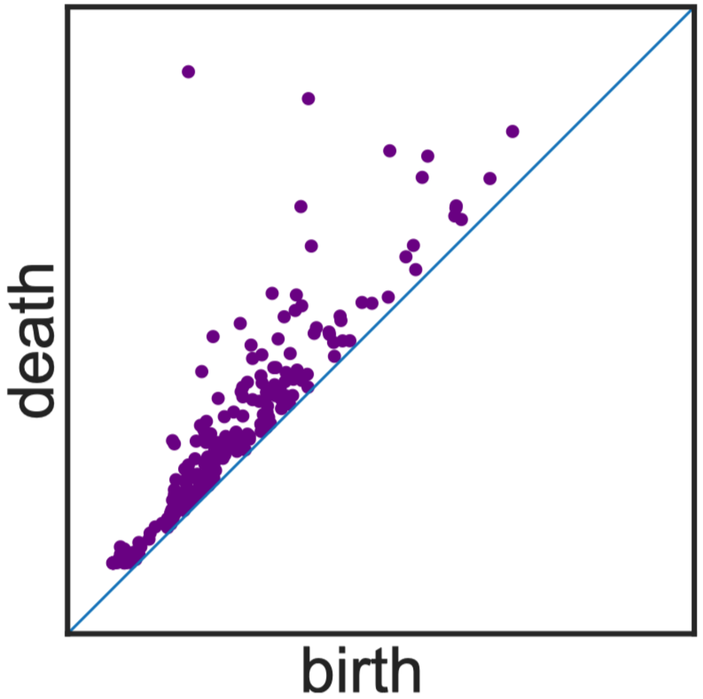
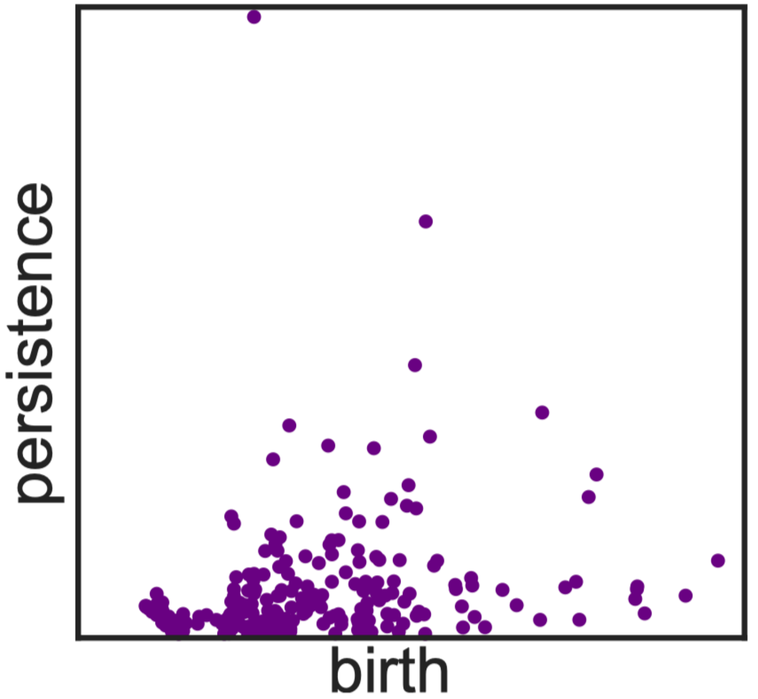
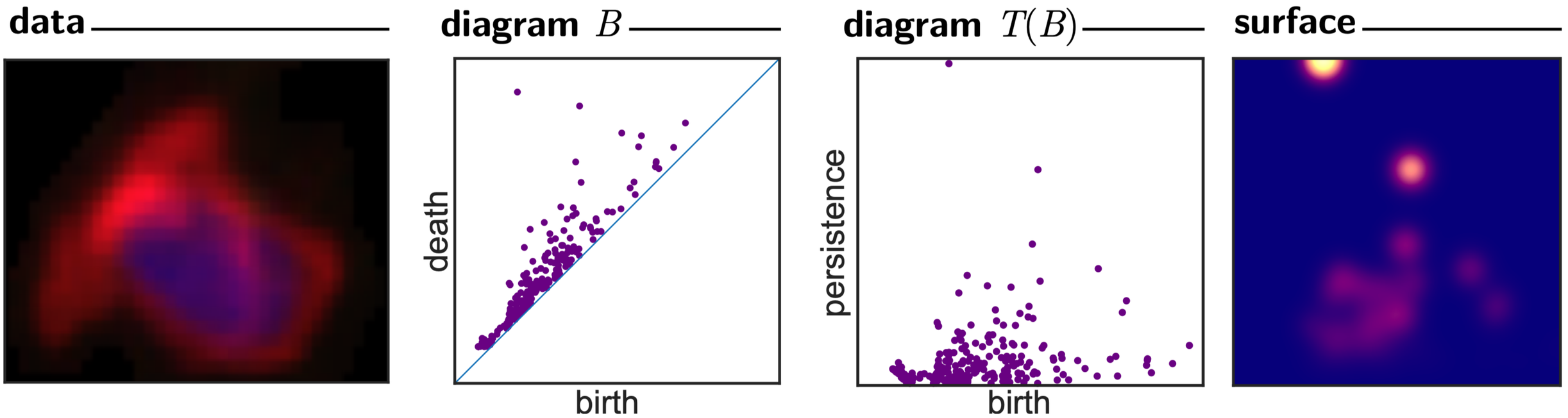


diagram $T(B)$



Persistence Surface



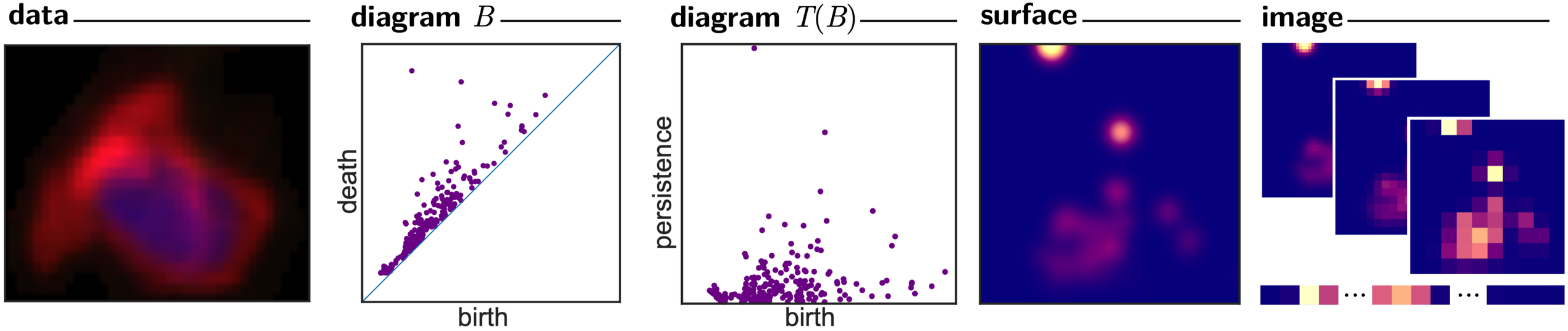
- Weight Function e.g., $f(u) = u_y/\text{maxPers}$
- Gaussian:
  $\phi_u(z) = \frac{1}{2\pi\sigma^2} \exp(-[(x-u_x)^2 + (y-u_y)^2]/(2\sigma^2))$

- For a Pers Dgm X, the persistence surface is

$$\mu_X : \quad \mathbb{R}^2 \longrightarrow \mathbb{R}$$

$$z \longmapsto \sum_{u \in T(X)} f(u)\varphi_u(z)$$

Persistence Images



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

Example: Linked Twist Map

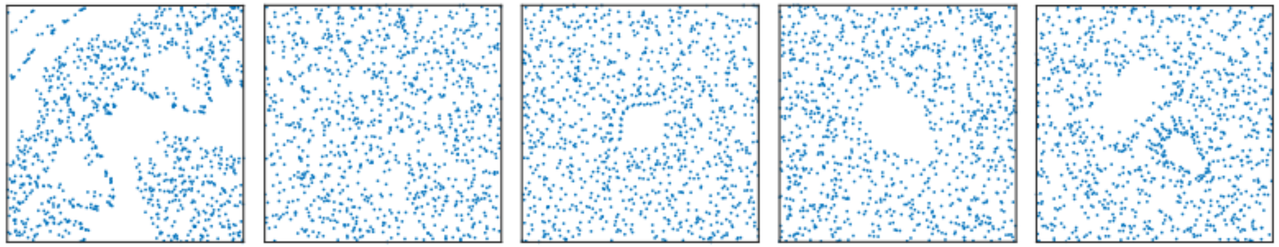


Figure 4: Examples of the first 1000 iterations, $\{(x_n, y_n) : n = 0, \ldots, 1000\}$, of the linked twist map with parameter values $r = 2$, 3.5, 4.0, 4.1 and 4.3, respectively.

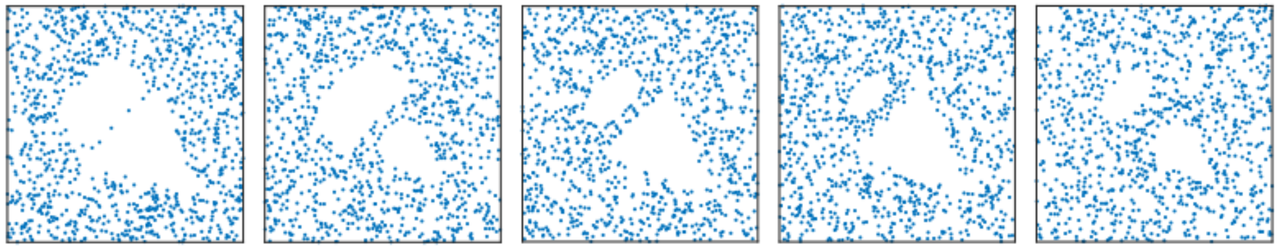


Figure 5: Truncated orbits, $\{(x_n, y_n) : n = 0, \ldots, 1000\}$, of the linked twist map with fixed $r = 4.3$ for different initial conditions (x_0, y_0).

- Generate point clouds from a discrete dynamical system

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

- 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:*

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

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.

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