Topological Data Analysis

2022–2023

Lecture 13

Persistence Descriptors

19 December 2022

Descriptors

A **persistence descriptor** is a numerical summary or a vectorized summary from persistence diagrams.

Numerical summaries

- Average life
- Average midlife
- Entropy
- Complex polynomials

Vectorized summaries

- Betti curves
- Landscapes
- Persistence images

Numerical Summaries

Average life:

$$\frac{1}{n} \sum_{i=1}^{n} (d_i - b_i)$$

Average midlife:

$$\frac{1}{n} \sum_{i=1}^{n} \frac{b_i + d_i}{2}$$

Entropy:

$$-\sum_{i=1}^{n} \frac{d_i - b_i}{L} \log_2\left(\frac{d_i - b_i}{L}\right), \quad \text{where} \quad L = \sum_{i=1}^{n} (d_i - b_i).$$

The **entropy** of a random variable is the average level of uncertainty inherent in its outcomes (Shannon, 1948).

Numerical Summaries

Complex polynomials

Let

$$p(x) = x^n + a_1 x^{n-1} + \cdots + a_{n-1} x + a_n$$

be a monic polynomial with coefficients in $\mathbb{C}$ whose roots are the points (b, d) in a given persistence diagram. Then the collection $a_1, \ldots, a_n$ of coefficients of $p(x)$, or a subset of this collection, can be used as a persistence descriptor.

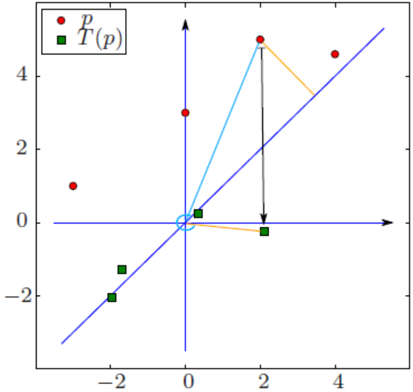
It is convenient to first transform the points as follows:

$$T(b, d) = \frac{d - b}{2} \left(\cos\alpha - \sin\alpha + i\,(\cos\alpha + \sin\alpha)\right),$$

where $\alpha = \sqrt{b^2 + d^2}$.

Numerical Summaries

This transformation T brings close to the origin the points (b, d) that are close to the diagonal in the persistence diagram, at an angle proportional to their distance to the origin:



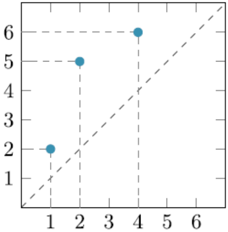
Vectorized Summaries

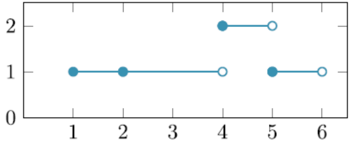
Betti curves

For each $k \geq 0$, let $\beta_k \colon \mathbb{R} \to \mathbb{R}$ be defined as

$$\beta_k(t) = \#\{(b, d) \mid b \leq t \leq d\},$$

where (b, d) ranges over the points in a given persistence diagram for homological dimension k.





Vectorized Summaries

Persistence images

For a given persistence diagram, consider a function

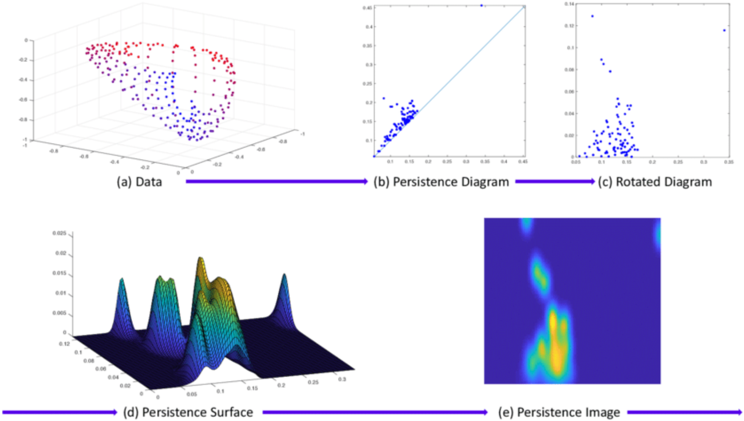
$$\Phi(s,t) = \sum_{i=1}^{n} w_i \, G_i(s,t)$$

for (s,t) in a square, where each w_i is a weight and G_i is a 2-dimensional Gaussian function centered at (b_i, d_i).

This yields a smoothing of the persistence diagram called a **persistence surface.**

A **persistence image** is a discretization of Φ on a grid overlay.

Vectorized Summaries



Generate a surface by centering 2D Gaussian distributions at each point, and generate a **persistence image** by summing the volume under the Gaussian distributions over the area of each pixel.

References

B. T. Fasy, Y. Qin, B. Summa, C. Wenk, *Comparing distance metrics on vectorized persistence summaries*, Topological Data Analysis and Beyond, 34th Conference on Neural Information Processing Systems (NeurIPS 2020)

B. Di Fabio, M. Ferri, *Comparing persistence diagrams through complex vectors*, Image Analysis and Processing (ICIAP 2015), Lecture Notes in Computer Science, vol. 9279, Springer, 2015

X. Arnal, R. Ballester, C. Casacuberta, C. Corneanu, S. Escalera, M. Madadi, *Towards explaining the generalization gap in neural networks using topological data analysis* (2021)

Kernels

Let X be any set. A **kernel** is a function $K\colon X \times X \to \mathbb{R}$ which is

- **symmetric:** $K(x, y) = K(y, x)$ for all $x, y \in X$, and
- **positive definite:**

  $$\sum_{i,j=1}^{n} c_i c_j \, K(x_i, x_j) \geq 0$$

  for all n and $c_1, \ldots, c_n \in \mathbb{R}$ and $x_1, \ldots, x_n \in X$, and moreover equality holds if and only if $c_i = 0$ for all i.

Example: The **linear kernel** in $\mathbb{R}^d$ is given by

$$K(x, y) = x^T y.$$

In this case $\sum_{i,j} c_i c_j \, K(x_i, x_j) = K\Big(\sum_i c_i x_i, \sum_i c_i x_i\Big)$ by bilinearity.

Kernels

Alternative definitions

- A function $K \colon X \times X \to \mathbb{R}$ is a kernel if and only if for each finite ordered subset $\{x_1, \dots, x_n\}$ of X the matrix $(K(x_i, x_j))$ is symmetric and positive definite.

- A function $K \colon X \times X \to \mathbb{R}$ is a kernel if and only if there exist a Hilbert space H and a map $\Phi \colon X \to H$ such that

  $$K(x, y) = \langle \Phi(x), \Phi(y) \rangle$$

  for all x, y. The Hilbert space H is called **feature space** and the map Φ is called **feature map.**

Kernels

Further examples

The following are kernels in Euclidean space $\mathbb{R}^d$:

- **Polynomial:** $K(x, y) = (1 + x^T y)^n$ with $n \geq 1$.
- **Gaussian:** $K(x, y) = \exp\Big(-\dfrac{\|x - y\|^2}{2\sigma^2}\Big)$ with $\sigma > 0$.
- **Laplacian:** $K(x, y) = \exp\big(-\alpha\|x - y\|\big)$ with $\alpha > 0$.

A **radial basis function** (RBF) is a real-valued function whose value depends only on the distance between the input and some fixed point. The Gaussian kernel is also called **RBF kernel.**

Kernels

The **heat kernel**

$$K_t(x, y) = \frac{1}{(4\pi t)^{d/2}} e^{-\|x-y\|^2/4t}$$

solves the **heat equation**

$$\frac{\partial K_t}{\partial t}(x, y) = \Delta_x K_t(x, y)$$

for $t > 0$ and $x, y \in \mathbb{R}^d$, with the initial condition

$$\lim_{t \to 0} K_t(x, y) = \delta_x(y),$$

where δ_x is a Dirac delta distribution centered at x.

Kernels

Every kernel $K\colon X \times X \to \mathbb{R}$ induces a **pseudometric** on X corresponding to the norm distance on the feature space:

$$d_K(x,y) = \sqrt{K(x,x) - 2K(x,y) + K(y,y)} = \|\Phi(x) - \Phi(y)\|.$$

Here it is possible that $d_K(x,y) = 0$ with $x \neq y$ since the feature map Φ need not be injective.

Kernels

For a set X, a Hilbert space H of functions $f\colon X \to \mathbb{R}$ is a **reproducing kernel Hilbert space (RKHS)** if the evaluation map $H \to \mathbb{R}$ given by $f \mapsto f(x)$ is continuous for all $x \in X$, i.e., if $\|f - g\|$ is small then $|f(x) - g(x)|$ is small for all x.

Every RKHS determines a unique kernel $K\colon X \times X \to \mathbb{R}$ with

- $K(x, -) \in H$ for all $x \in X$;
- $\langle f, K(x, -)\rangle = f(x)$ for all $x \in X$ and all $f \in H$.

This is called a **reproducing kernel.**

Conversely, every kernel K determines a unique RKHS inducing K as its reproducing kernel.

Kernels

Scale-space kernel (Reininghaus et al., 2015)

$K\colon \mathcal{D} \times \mathcal{D} \to \mathbb{R}$ where $\mathcal{D}$ is the set of all persistence diagrams.

It is defined via a feature map $\Phi\colon \mathcal{D} \to L^2(\Omega)$, where $\Omega = \{(x, y) \in \mathbb{R}^2 \mid y \geq x\}$ is the half plane above the diagonal.

To each persistence diagram $D \in \mathcal{D}$ we could assign the sum $\sum_{p \in D} \delta_p$ of Dirac delta distributions. Here δ_p is viewed as a functional that evaluates each smooth function at $p = (b, d)$.

However, the induced metric on $\mathcal{D}$ does not take into account the distance to the diagonal and hence it is not robust against noise.

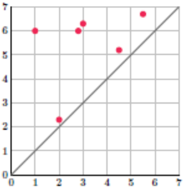
Kernels

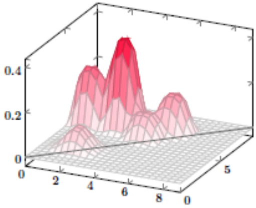
Instead, take the sum of Dirac deltas as initial condition for a heat diffusion problem with a boundary condition on the diagonal.

Find a solution $u\colon \Omega \times \mathbb{R}_+ \to \mathbb{R}$ of the Dirichlet problem

$$\Delta_x u = \partial_t u \ \text{ in } \Omega \times \mathbb{R}_+, \qquad u = 0 \ \text{ on } \partial\Omega \times \mathbb{R}_+,$$

$$u = \sum_{p \in D} \delta_p \ \text{ on } \Omega \times \{0\}.$$





Kernels

Then define $\boldsymbol{\Phi_\sigma(D) = u|_{t=\sigma}}$ for each $D \in \mathcal{D}$ and each scale parameter $\sigma > 0$. Thus,

$$K_\sigma(D_1, D_2) = \langle \Phi_\sigma(D_1), \Phi_\sigma(D_2) \rangle.$$

In this case the feature map Φ_σ is injective, so K_σ yields a metric.

Explicitly, one obtains that

$$u(x,t) = \frac{1}{4\pi t} \sum_{p \in D} e^{-\|x-p\|^2/4t} - e^{-\|x-\bar{p}\|^2/4t}$$

where $\bar{p} = (d, b)$ if $p = (b, d)$. Therefore

$$K_\sigma(D_1, D_2) = \frac{1}{8\pi\sigma} \sum_{p \in D_1,\, q \in D_2} e^{-\|p-q\|^2/8\sigma} - e^{-\|p-\bar{q}\|^2/8\sigma}$$

Kernels

Stability

This kernel is stable with respect to the 1-Wasserstein distance:

$$\|\Phi_\sigma(D_1) - \Phi_\sigma(D_2)\| \leq \frac{1}{\sigma\sqrt{8\pi}}\, W_1(D_1, D_2),$$

but not with respect to p-Wasserstein distances with $p > 1$.

Kernels

Landscape kernel

Landscapes represent persistence diagrams as functions in $L^p(\mathbb{N} \times \mathbb{R})$ for any p. For $p = 2$, we can use the Hilbert space structure of $L^2(\mathbb{N} \times \mathbb{R})$ to define a kernel K^L with feature map

$$\Phi^L \colon \mathcal{D} \longrightarrow L^2(\mathbb{N} \times \mathbb{R}),$$

and a corresponding distance d^L.

This kernel is stable with respect to a weighted version of the 2-Wasserstein distance.

Kernels

Classification performance

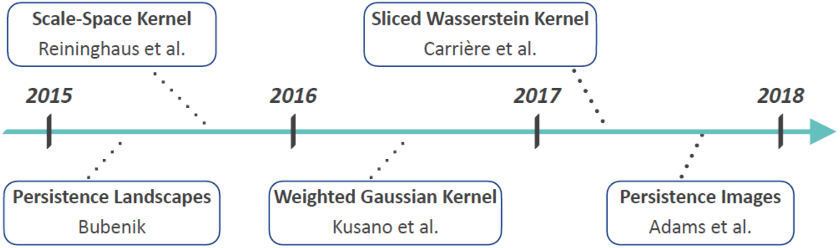
The following percentages were obtained over a range of 10 time parameters t_i using the kernels K^L and K_σ with an SVM classifier (support vector machine) on SHREC 2014:

HKS t_i	k^L	k_σ	Δ
t_1	68.0 ± 3.2	94.7 ± 5.1	+26.7
t_2	$\mathbf{88.3} \pm 3.3$	$\mathbf{99.3} \pm 0.9$	+11.0
t_3	61.7 ± 3.1	96.3 ± 2.2	+34.7
t_4	81.0 ± 6.5	97.3 ± 1.9	+16.3
t_5	84.7 ± 1.8	96.3 ± 2.5	+11.7
t_6	70.0 ± 7.0	93.7 ± 3.2	+23.7
t_7	73.0 ± 9.5	88.0 ± 4.5	+15.0
t_8	81.0 ± 3.8	88.3 ± 6.0	+7.3
t_9	67.3 ± 7.4	88.0 ± 5.8	+20.7
t_{10}	55.3 ± 3.6	91.0 ± 4.0	+35.7

Source: Reininghaus et al. (2015)

Kernels

Other kernels



Source: U. Fugacci, CNR-IMATI, Genova

Reference

J. Reininghaus, S. Huber, U. Bauer, R. Kwitt,
A stable multi-scale kernel for topological machine learning,
2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, 4741–4748