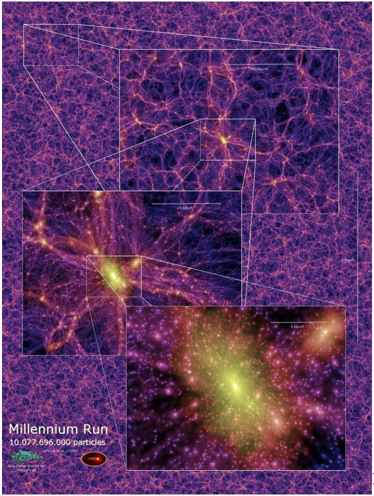
PLLay: Efficient Topological Layer based on Persistence Landscapes

Jisu KIM

The 24th Korea–Japan Joint Workshop on Algorithms and Computation (WAAC 2024)
2024-08-03

Topological structures in the data provide information.

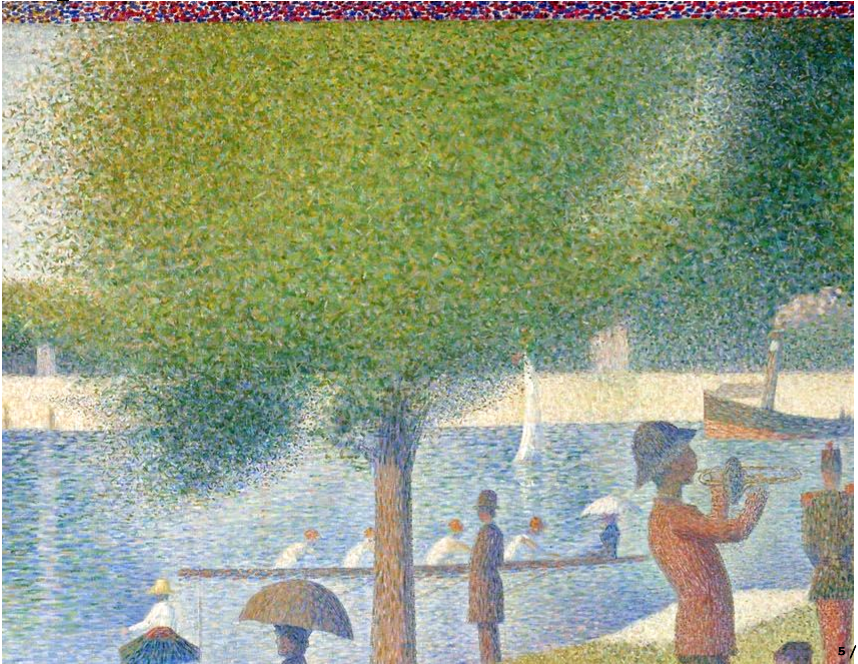


[1]

[1] http://www.mpa-garching.mpg.de/galform/virgo/millennium/poster_half.jpg

Topological structures are observed in different scales.

Topological structures are observed in different scales.



Topological structures are observed in different scales.

- Georges Seurat, A Sunday afternoon on the island of La Grande Jatte (Un dimanche après-midi à l'Île de la Grande Jatte)

Reference for Topological Data Analysis

- Introduction to Topological Data Analysis
  - Computational Topology: An Introduction (Edelsbrunner, Harer, 2010)
  - Topological Data Analysis (Wasserman, 2016)
  - An Introduction to Topological Data Analysis: Fundamental and Practical Aspects for Data Scientists (Chazal, Michel, 2021)
- Application of Topological Data Analysis to machine learning
  - A Survey of Topological Machine Learning Methods (Hensel, Moor, Rieck, 2021)

The number of holes is used to summarize topological features.

- Geometrical objects:
  - ㄱ, ㄴ, ㄷ, ㄹ, ㅁ, ㅂ, ㅅ, ㅇ, ㅈ, ㅊ, ㅋ, ㅌ, ㅍ, ㅎ
  - あ, い, う, え, お, ア, イ, ウ, エ, オ
- The number of holes of different dimensions is considered.
  1. $\beta_0 = \#$ of connected components 
  2. $\beta_1 = \#$ of loops (holes inside 1-dim sphere) 
  3. $\beta_2 = \#$ of voids (holes inside 2-dim sphere) 

Example : Objects are classified by homologies.

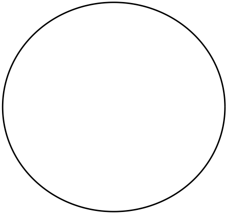
1. β_0 =# of connected components
2. β_1 =# of loops (holes inside 1-dim sphere)

$\beta_0 \setminus \beta_1$	0	1	2
1	ㄱ,ㄴ,ㄷ,ㄹ, ㅅ,ㅈ,ㅋ, ㅌ,ア, イ, ウ, エ, オ	ㅁ,ㅇ, ㅂ,ㅍ	あ
2	ㅊ, い, う, え	お	
3		ㅎ	

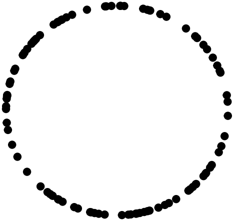
Homology of finite sample is different from homology of underlying manifold, hence it cannot be directly used for the inference.

- When analyzing data, we prefer robust features where features of the underlying manifold can be inferred from features of finite samples.
- Homology is not robust:

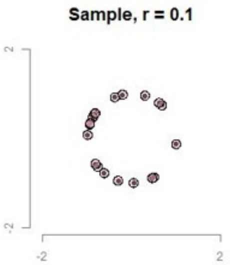
Underlying circle: $\beta_0 = 1$, $\beta_1 = 1$



100 samples: $\beta_0 = 100$, $\beta_1 = 0$

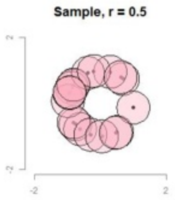


Persistent homology computes homologies on collection of sets, and tracks when topological features are born and when they die.

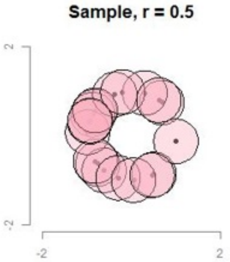


Augmented Set

Persistent homology computes homologies on collection of sets, and tracks when topological features are born and when they die.

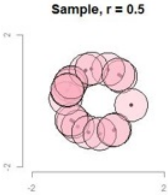


r=**0.5** : 1-dim hole is formed

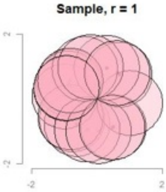


Augmented Set

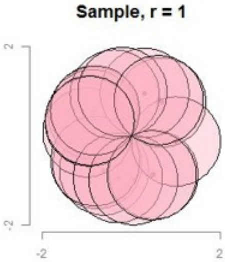
Persistent homology computes homologies on collection of sets, and tracks when topological features are born and when they die.



r=**0.5** : 1-dim hole is formed

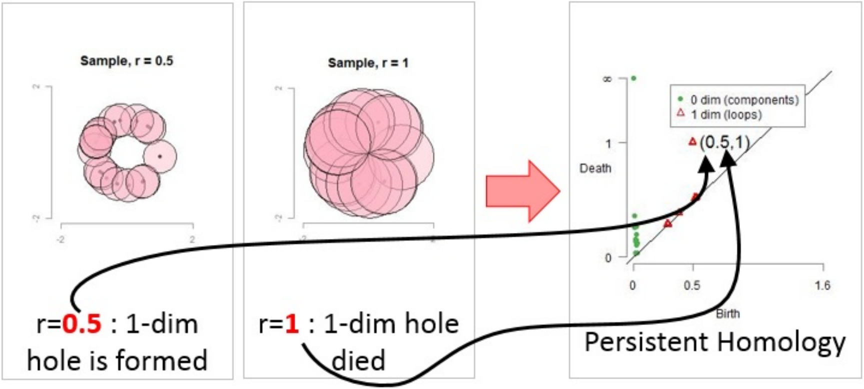


r=**1** : 1-dim hole died

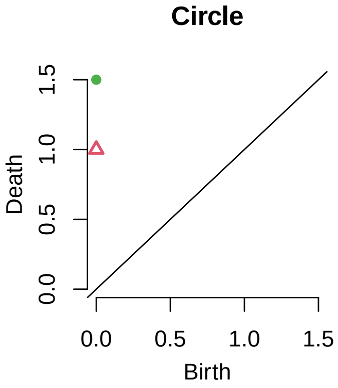


Augmented Set

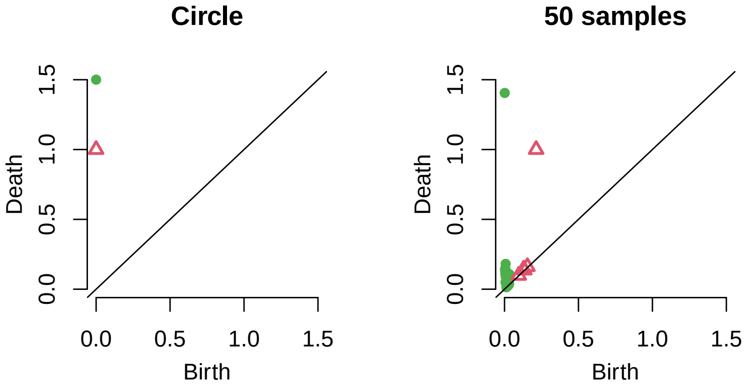
Persistent homology computes homologies on collection of sets, and tracks when topological features are born and when they die.



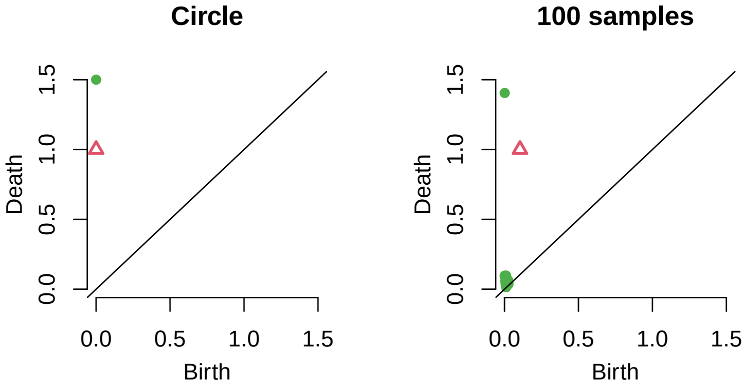
Persistent homology of the underlying manifold can be inferred from persistent homology of finite samples.



Persistent homology of the underlying manifold can be inferred from persistent homology of finite samples.



Persistent homology of the underlying manifold can be inferred from persistent homology of finite samples.



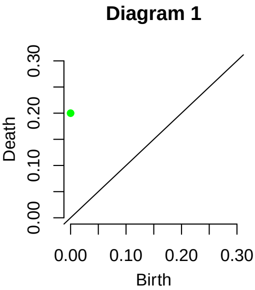
Bottleneck distance gives a metric on the space of persistent homology.

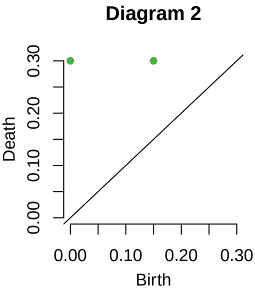
Definition

Let $D_1,\ D_2$ be multiset of points. Bottleneck distance is defined as

$$W_\infty(D_1,\ D_2) = \inf_{\gamma} \sup_{x \in D_1} \|x - \gamma(x)\|_\infty,$$

where γ ranges over all bijections from D_1 to D_2.





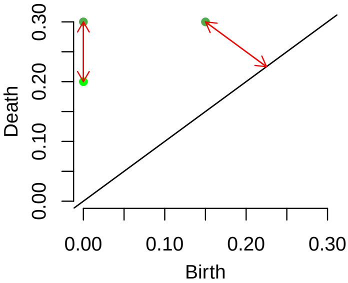
Bottleneck distance gives a metric on the space of persistent homology.

Definition

Let $D_1,\ D_2$ be multiset of points. Bottleneck distance is defined as

$$W_\infty(D_1,\ D_2) = \inf_{\gamma} \sup_{x \in D_1} \|x - \gamma(x)\|_\infty,$$

where γ ranges over all bijections from D_1 to D_2.



$$\sup_{x \in D_1} \|x - \gamma_1(x)\|_\infty = 0.1$$

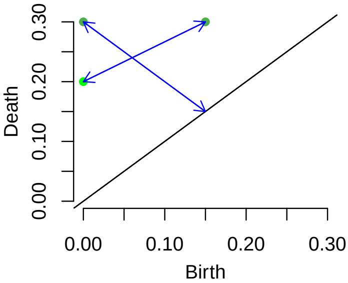
Bottleneck distance gives a metric on the space of persistent homology.

Definition

Let D_1, D_2 be multiset of points. Bottleneck distance is defined as

$$W_\infty(D_1, D_2) = \inf_{\gamma} \sup_{x \in D_1} \|x - \gamma(x)\|_\infty,$$

where γ ranges over all bijections from D_1 to D_2.



$$\sup_{x \in D_1} \|x - \gamma_2(x)\|_\infty = 0.15$$

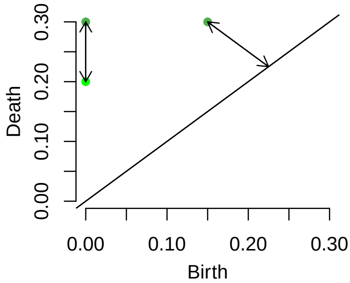
Bottleneck distance gives a metric on the space of persistent homology.

Definition

Let D_1, D_2 be multiset of points. Bottleneck distance is defined as

$$W_\infty(D_1,\, D_2) = \inf_{\gamma} \sup_{x \in D_1} \|x - \gamma(x)\|_\infty,$$

where γ ranges over all bijections from D_1 to D_2.



$$\inf_{\gamma} \sup_{x \in D_1} \|x - \gamma(x)\|_\infty = 0.1$$

Introduction

Persistent Homology

Application of Topological Data Analysis to Machine Learning

PLLay: Efficient Topological Layer based on Persistence Landscapes

Reference

(Very rough) sketch to Machine Learning

- For a given task and data, Machine Learning / Deep Learning fits a parametrized model.
  - Given data X,
  - Parametrized model f_θ,
  - Loss function $\mathcal{L}$ tailored to the task,
  - Machine Learning minimizes $\arg\min_\theta \mathcal{L}(f_\theta, \mathcal{X})$.
- Many cases, getting explicit formula for $\arg\min_\theta \mathcal{L}(f_\theta, \mathcal{X})$ is impossible or too costly (e.g., inverting a large scale matrix). So, gradient descent is used with the $\nabla_\theta \mathcal{L}(f_\theta, \mathcal{X})$:

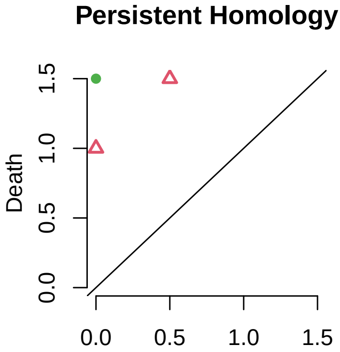
$$\theta_{n+1} = \theta_n - \lambda \nabla_\theta \mathcal{L}(f_\theta, \mathcal{X}).$$

Application of Topological Data Analysis to Machine Learning

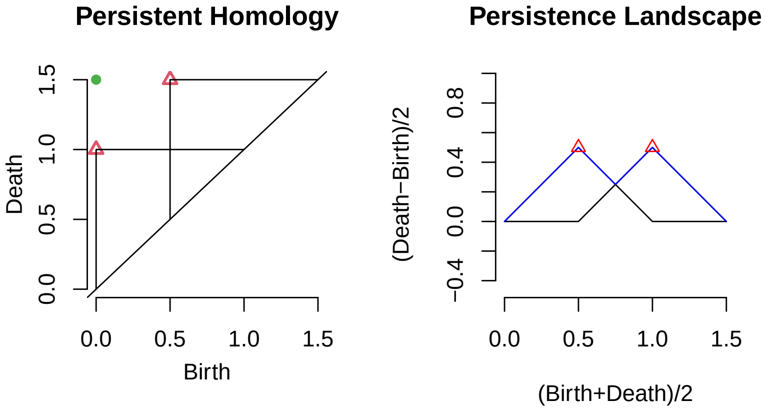
- A Survey of Topological Machine Learning Methods (Hensel, Moor, Rieck, 2021)
- Topological Data Analysis (TDA) is applied to Machine Learning in usually two directions:
  - Make features from TDA, so that the data X is augmented with extra TDA features: more common
    - PLLay: Efficient Topological Layer based on Persistence Landscapes (Kim, Kim, Zaheer, Kim, Chazal, Wasserman, 2020)
  - Evaluate quality of data $\mathcal{X}$ or model f_θ using TDA: recently received attentions

Persistent homology is further summarized and embedded into a Euclidean space or a functional space.

- The space of the persistent homology is complex, so directly applying in machine learning is difficult.
- If the persistent homology is further summarized and embedded into a Euclidean space or a functional space, then applying in machine learning becomes much more convenient.
  - e.g., Persistence Landscape, Persistence Silhouette, Persistence Image



Persistence Landscape is a functional summary of the persistent homology.

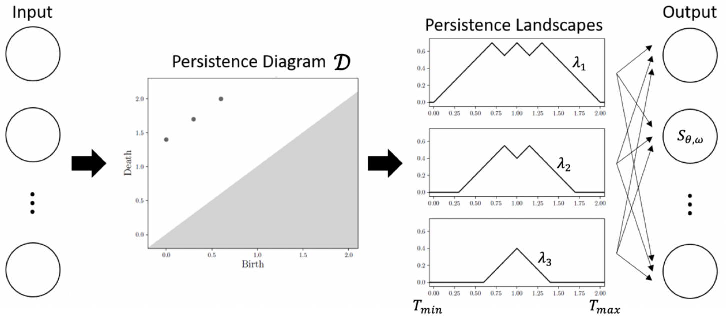


Featurizing using Persistence Landscape

- Featurization using time-delayed embedding and Persistence Landscape
  - Time Series Featurization via Topological Data Analysis (Kim, Kim, Rinaldo, Chazal, 2020)
- Build topological layer using Persistence Landscape
  - PLLay: Efficient Topological Layer based on Persistence Landscapes (Kim, Kim, Zaheer, Kim, Chazal, Wasserman, 2020)

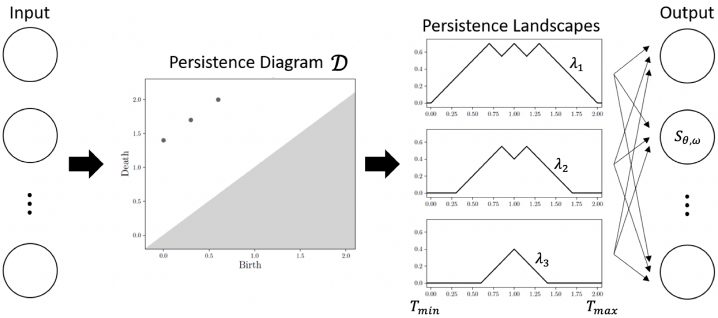
We build topological layer using persistence diagrams and persistence landscapes

- Given input X, we would like to build a topological layer that reflects topological properties of X.
- Computation of diagrams $X \to \mathcal{D}_X$: user chooses appropriate ways to compute persistence diagrams
- Construction of topological layer $\mathcal{D}_X \to h_{\text{top}}$: persistence landscapes and parametrized differentiable map



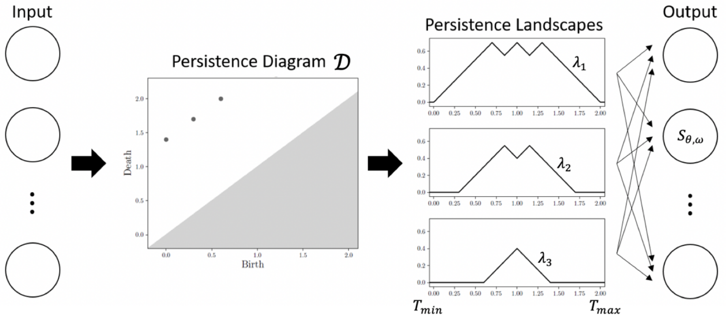
PLLay: Build topological layer using Persistence Landscape

1. From data X, choose an appropriate simplicial complex K and a function f to compute the Persistece diagram $\mathcal{D}$.
2. From the persistence diagram $\mathcal{D}$, compute the persistence landscape $\lambda : \mathbb{N} \times \mathbb{R} \to \mathbb{R}$.
3. Compute the weighted average function $\bar{\lambda}_\omega(t) := \sum_{k=1}^{K_{\max}} \omega_k \lambda_k(t)$, and vectorize to get $\bar{\Lambda}_\omega \in \mathbb{R}^m$.
4. For a parametrized differentiable map $g_\theta : \mathbb{R}^m \to \mathbb{R}$, compute $S_{\theta,\omega}(\mathcal{D}) := g_\theta(\bar{\Lambda}_\omega)$, then $h_{\mathrm{top}}(X) = \{S_{\theta,\omega}(\mathcal{D})\}_{i=1}^{n_h}$.



Computational complexity of PLLay

- Computation of diagrams $X \to \mathcal{D}_X$: $O(N^3)$, where N is the number of simplices in the simplicial complex
- Construction of topological layer $\mathcal{D}_X \to h_{\text{top}}$: $O(|\mathcal{D}_X|)$ where $|\mathcal{D}_X|$ is the number of points in $\mathcal{D}_X$, and $|\mathcal{D}_X| = O(N)$.



PLLay is differentiable.

- A deep learning model learns its parameters by back propagation, which is to apply gradient descent layer-wise.
- For a deep learning layer to be learnable, it should be differentiable:

Theorem (Theorem 3.1 in Kim et al. [2020])

When persistence diagrams are appropriately computed from the input data, the PLLay function h_{top} is differentiable with respect to the input data X.

PLLay is stable.

- PLLay is stable with respect to changes in persistence diagrams:

Theorem (Theorem 4.1 in Kim et al. [2020])

For two persistence diagrams $\mathcal{D}, \mathcal{D}'$,

$$|S_{\theta,\omega}(\mathcal{D}) - S_{\theta,\omega}(\mathcal{D}')| = O(W_\infty(\mathcal{D}, \mathcal{D}')),$$

where W_∞ is the bottleneck distance.

PLLay is stable.

- PLLay is stable with respect to perturbations in input X:

Theorem (Theorem 4.2 in Kim et al. [2020])

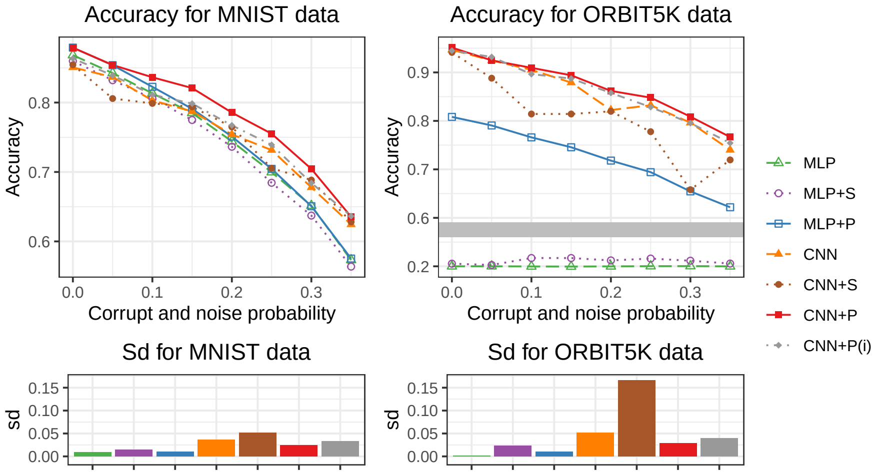
Let $X \sim P$ and P_n be the empirical distribution of X. Further, let $\mathcal{D}_P, \mathcal{D}_X$ be the persistence diagrams of P, X, respectively. Then

$$|S_{\theta,\omega}(\mathcal{D}_X) - S_{\theta,\omega}(\mathcal{D}_P)| = O(W_2(P_n, P)),$$

where W_2 is 2-Wasserstein distance.

PLLay is robust to noise

- PLLay is robust to noise and stabilizes the model.



Reference I

Frédéric Chazal and Bertrand Michel. An introduction to topological data analysis: Fundamental and practical aspects for data scientists. *Frontiers Artif. Intell.*, 4:667963, 2021. doi: 10.3389/frai.2021.667963. URL `https://doi.org/10.3389/frai.2021.667963`.

Frédéric Chazal, Vin de Silva, Marc Glisse, and Steve Oudot. The structure and stability of persistence modules. *arXiv preprint arXiv:1207.3674*, 2012.

Herbert Edelsbrunner and John L. Harer. *Computational topology*. American Mathematical Society, Providence, RI, 2010. ISBN 978-0-8218-4925-5. doi: 10.1090/mbk/069. URL `https://doi.org/10.1090/mbk/069`. An introduction.

Felix Hensel, Michael Moor, and Bastian Rieck. A survey of topological machine learning methods. *Frontiers Artif. Intell.*, 4:681108, 2021. doi: 10.3389/frai.2021.681108. URL `https://doi.org/10.3389/frai.2021.681108`.

Reference II

Kwangho Kim, Jisu Kim, Alessandro Rinaldo, and Frédéric Chazal. Time series featurization via topological data analysis: an application to cryptocurrency trend forecasting. *CoRR*, abs/1812.02987, 2020. URL `http://arxiv.org/abs/1812.02987`.

Kwangho Kim, Jisu Kim, Manzil Zaheer, Joon Sik Kim, Frédéric Chazal, and Larry Wasserman. PLLay: Efficient Topological Layer based on Persistent Landscapes. *arXiv e-prints*, art. arXiv:2002.02778, February 2020.

Larry Wasserman. Topological data analysis, 2016.

Thank you!

Persistent Homology and Persistence Landscape

We rely on the superlevel sets of the kernel density estimator to extract topological information of the underlying distribution.

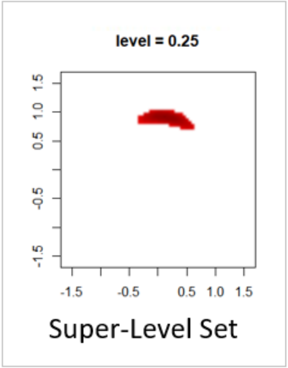
- The kernel density estimator is

$$\hat{p}_h(x) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right).$$

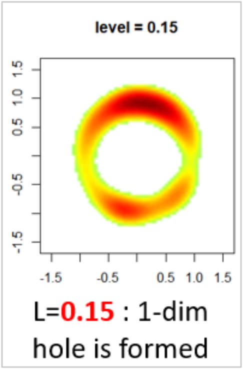
- We look at superlevel sets of the kernel density estimator as

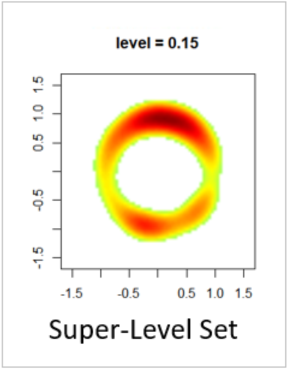
$$\left\{x \in \mathbb{R}^d : \hat{p}_h(x) \geq L\right\}_{L>0}.$$

Persistent homology computes homologies on collection of sets, and tracks when topological features are born and when they die.

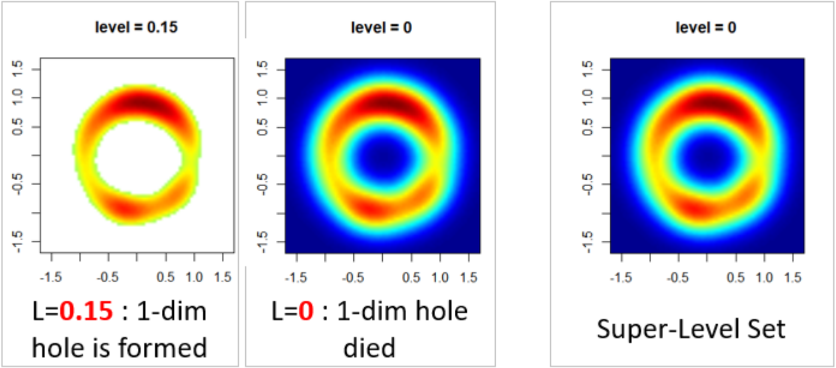


Persistent homology computes homologies on collection of sets, and tracks when topological features are born and when they die.

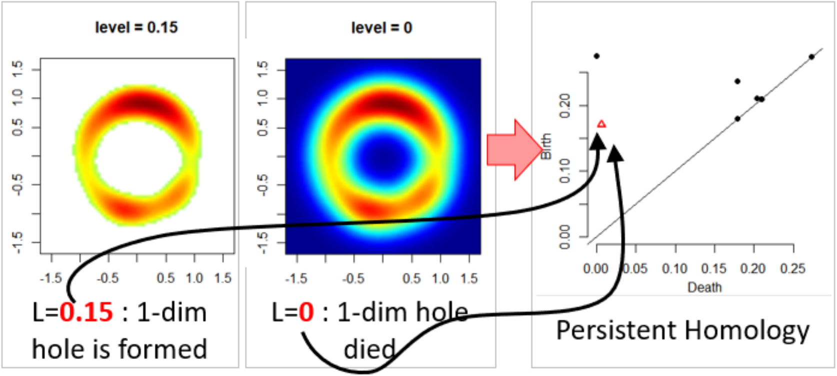




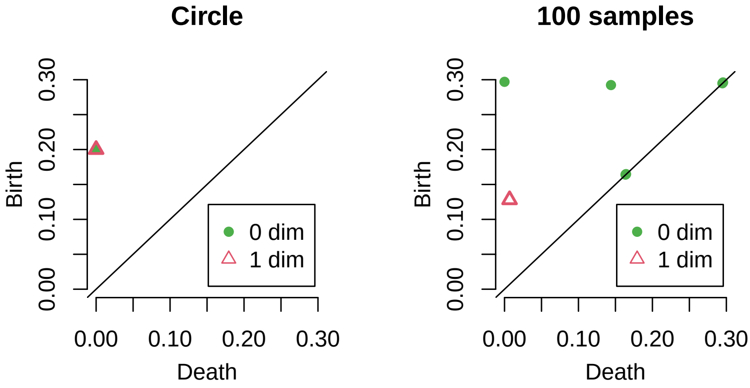
Persistent homology computes homologies on collection of sets, and tracks when topological features are born and when they die.



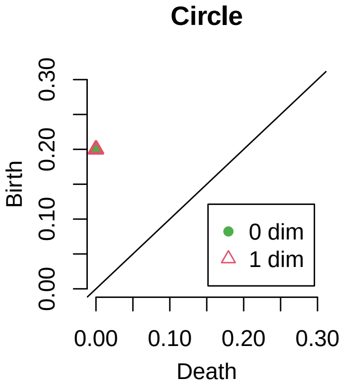
Persistent homology computes homologies on collection of sets, and tracks when topological features are born and when they die.



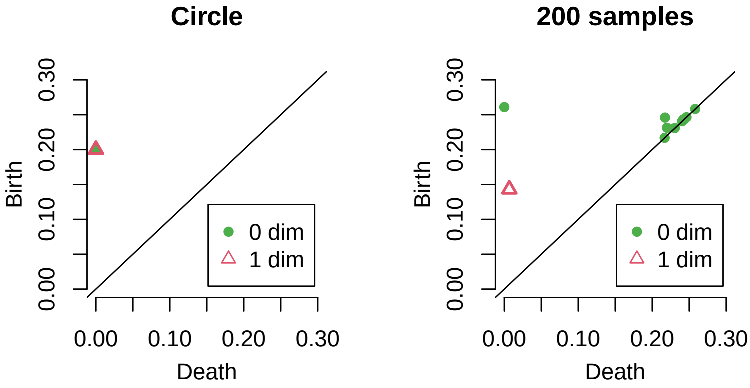
Persistent homology of the underlying manifold can be inferred from persistent homology of finite samples.



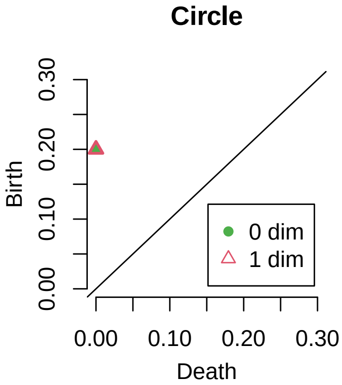
Persistent homology of the underlying manifold can be inferred from persistent homology of finite samples.



Persistent homology of the underlying manifold can be inferred from persistent homology of finite samples.



Persistent homology of the underlying manifold can be inferred from persistent homology of finite samples.



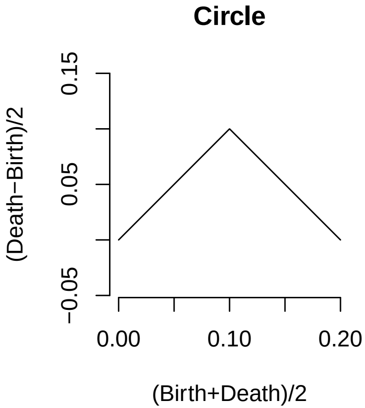
Bottleneck distance can be controlled by the corresponding distance on functions: Stability Theorem.

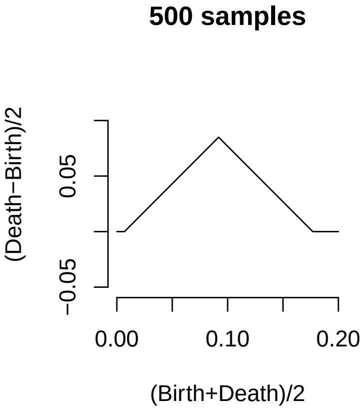
Theorem

[Edelsbrunner and Harer, 2010][Chazal, de Silva, Glisse, and Oudot, 2012] Let $\mathbb{X}$ be finitely triangulable space and $f, g : \mathbb{X} \to \mathbb{R}$ be two continuous functions. Let $Dgm(f)$ and $Dgm(g)$ be corresponding persistence diagrams. Then

$$W_\infty(Dgm(f), Dgm(g)) \leq \|f - g\|_\infty.$$

Persistence Landscape of the underlying manifold can be inferred from Persistence Landscape of finite samples.



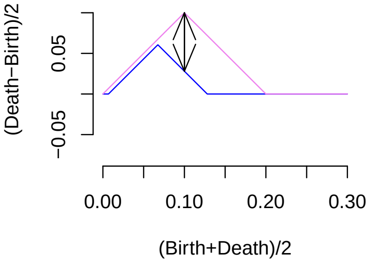


∞-landscape distance gives a metric on the space of persistence landscapes.

Definition

[?]Let D_1, D_2 be multiset of points, and λ_1 , λ_2 be corresponding persistence landscapes. ∞-landscape distance is defined as

$$\Lambda_\infty(D_1, D_2) = \|\lambda_1 - \lambda_2\|_\infty .$$



∞-landscape distance can be controlled by the corresponding distance on functions: Stability Theorem.

Theorem

Let $f, g : \mathbb{X} \to \mathbb{R}$ be two functions, and let $Dgm(f)$ and $Dgm(g)$ be corresponding persistent homologies. Then

$$\Lambda_\infty(\lambda(f), \lambda(g)) \leq \|f - g\|_\infty.$$