

위상학적 자료 분석(Topological Data Analysis)의 통계적 추정 및 기계학습에의 응용

김지수



가톨릭관동대학교
2021-11-10

위상학적 자료 분석(Topological Data Analysis) 소개

호몰로지(homology)를 통계적으로 추정하기

군집 나무(Cluster Tree)와 이를 통계적으로 추정하기

Persistent Homology를 통계적으로 추정하기

Persistent Homology의 신뢰띠를 격자 위에서 계산하기

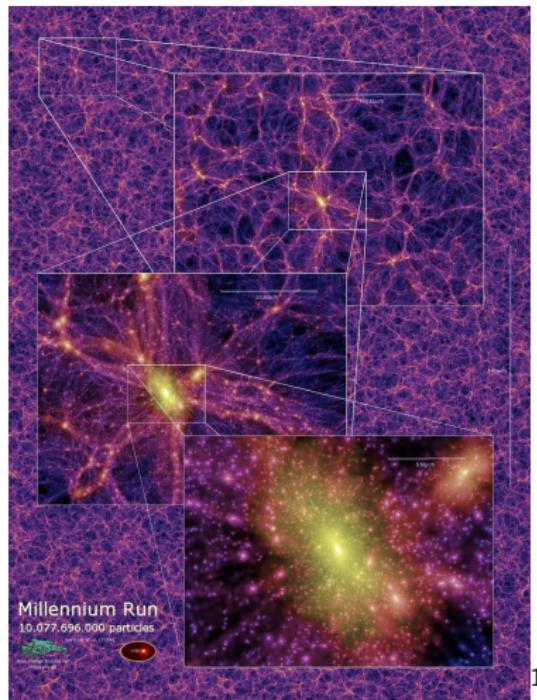
Persistent Homology의 신뢰띠를 자료 위에서 계산하기

Persistent Homology를 이용하여 특성(Feature) 만들기

R 패키지 TDA: 위상학적 자료 분석을 위한 통계 계산 도구

참조문헌

자료의 위상학적 구조로부터 정보를 얻을 수 있습니다.



1

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

위상학적 자료 분석(Topological Data Analysis)의 통계적 추정을 소개합니다.

- ▶ 호몰로지(homology)를 통계적으로 추정하기
 - ▶ Statistical Inference for Cluster Trees (Kim, Chen, Balakrishnan, Rinaldo, Wasserman, 2016)
- ▶ Persistent Homology를 통계적으로 추정하기
 - ▶ Confidence sets for persistence diagrams (Fasy, Lecci, Rinaldo, Wasserman, Balakrishnan, Singh, 2014b)
 - ▶ Persistent homology of KDE filtration on Vietoris-Rips complex (Shin, Kim, Rinaldo, Wasserman, 2021?)
- ▶ Persistent Homology를 이용하여 특성(Feature) 만들기
 - ▶ Time Series Featurization via Topological Data Analysis (Kim, Kim, Rinaldo, Chazal, 2020)
 - ▶ Efficient Topological Layer based on Persistence Landscapes (Kim, Kim, Zaheer, Kim, Chazal, Wasserman, 2020)
- ▶ R 패키지 TDA: 위상학적 자료 분석을 위한 통계 계산 도구
 - ▶ Introduction to the R package TDA (Fasy, Kim, Lecci, Maria, Millman, Rouvreau, 2014a)

위상학적 자료 분석(Topological Data Analysis) 소개

호몰로지(homology)를 통계적으로 추정하기

군집 나무(Cluster Tree)와 이를 통계적으로 추정하기

Persistent Homology를 통계적으로 추정하기

Persistent Homology의 신뢰띠를 격자 위에서 계산하기

Persistent Homology의 신뢰띠를 자료 위에서 계산하기

Persistent Homology를 이용하여 특성(Feature) 만들기

R 패키지 TDA: 위상학적 자료 분석을 위한 통계 계산 도구

참조문헌

구멍의 개수로 기하학적 대상들을 분류할 수 있습니다.

- ▶ 기하학적 대상들:
 - ▶ ㄱ, ㄴ, ㄷ, ㄹ, ㅁ, ㅂ, ㅅ, ㅇ, ㅈ, ㅊ, ㅋ, ㅌ, ㅍ, ㅎ
 - ▶ A, 字, あ
- ▶ 여러 차원에서 구멍들의 개수들을 각각 고려합니다.

1. β_0 =연결된 성분의 개수



2. β_1 =고리(1차원 구의 구멍)의 개수



3. β_2 =2차원 구의 구멍의 개수



예제: 대상들을 호몰로지에 따라 분류합니다.

1. β_0 =연결된 성분의 개수



2. β_1 =고리의 개수



| $\beta_0 \setminus \beta_1$ | 0 | 1 | 2 |
|-----------------------------|---------------------------|------------------|---|
| 1 | ㄱ, ㄴ, ㄷ, ㄹ, ㅅ, ㅈ, ㅌ, ㅊ | ㅁ, ㅇ, ㅂ, ㅍ, ㅏ | ㅓ |
| 2 | ㅊ, ㅈ | | |
| 3 | | ㅎ | |

호몰로지(homology)를 통계적으로 추정합니다.

- ▶ Statistical Inference for Cluster Trees (Kim, Chen, Balakrishnan, Rinaldo, Wasserman, 2016)

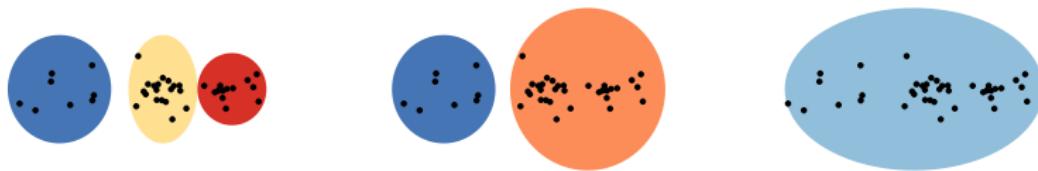
자료를 군집으로 묶고자 합니다.

- ▶ Statistical Inference for Cluster Trees (Kim, Chen, Balakrishnan, Rinaldo, Wasserman, 2016)



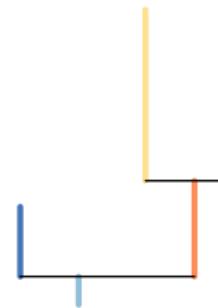
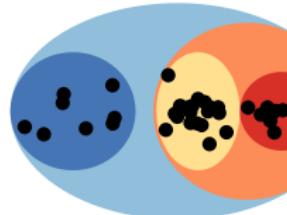
원하는 해상도에 따라 다른 군집이 생길 수 있습니다.

- ▶ Statistical Inference for Cluster Trees (Kim, Chen, Balakrishnan, Rinaldo, Wasserman, 2016)
- ▶ 국소적(local)이고 상세한 정보를 묘사하고 싶으면 (높은 해상도), 작은 규모의 많은 군집이 생깁니다.
- ▶ 대역적(global)이고 개략적인 정보를 묘사하고 싶으면 (낮은 해상도), 큰 규모의 적은 군집이 생깁니다.



군집들의 네트워크가 나무를 형성합니다: 군집 나무 (cluster tree)

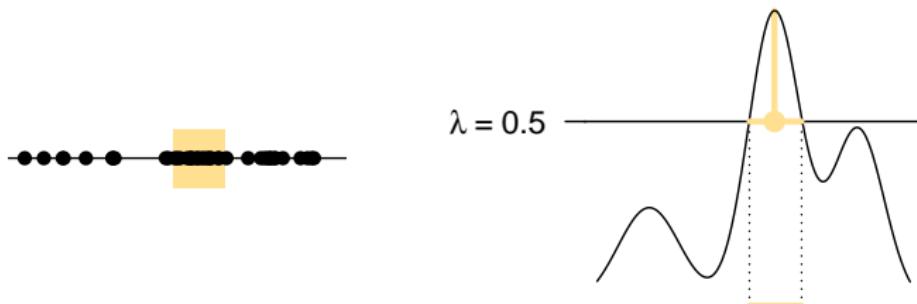
- ▶ Statistical Inference for Cluster Trees (Kim, Chen, Balakrishnan, Rinaldo, Wasserman, 2016)
- ▶ 다른 수준의 해상도로부터 얻어지는 군집들은 포함 관계에 의해 자연스러운 네트워크가 생깁니다.
- ▶ 포함 관계 네트워크는 나무로 표현될 수 있습니다: 군집 나무(cluster tree)



군집 나무(Cluster Tree)는 고밀도 군집들의 계층 구조입니다.

Definition

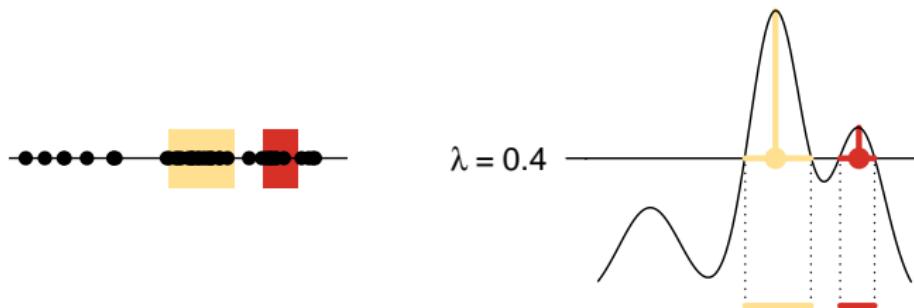
주어진 밀도함수 p 의 군집 나무(Cluster Tree) $T_p : \mathbb{R} \rightarrow \mathcal{P}(\mathcal{X})$ 는 각 실수값 λ 마다 윗레벨 집합 $\{x : p(x) \geq \lambda\}$ 의 연결부분들의 집합이 $T_p(\lambda)$ 로써 대응되는 함수입니다.



군집 나무(Cluster Tree)는 고밀도 군집들의 계층 구조입니다.

Definition

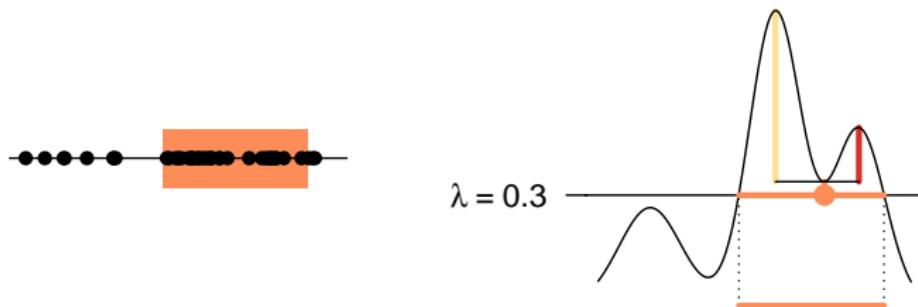
주어진 밀도함수 p 의 군집 나무(Cluster Tree) $T_p : \mathbb{R} \rightarrow \mathcal{P}(\mathcal{X})$ 는 각 실수값 λ 마다 윗레벨 집합 $\{x : p(x) \geq \lambda\}$ 의 연결부분들의 집합이 $T_p(\lambda)$ 로써 대응되는 함수입니다.



군집 나무(Cluster Tree)는 고밀도 군집들의 계층 구조입니다.

Definition

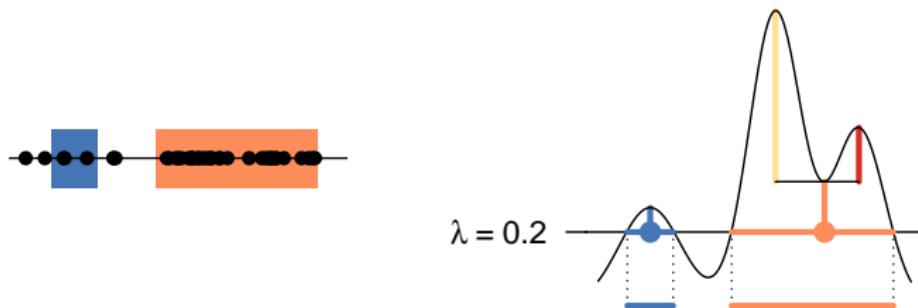
주어진 밀도함수 p 의 군집 나무(Cluster Tree) $T_p : \mathbb{R} \rightarrow \mathcal{P}(\mathcal{X})$ 는 각 실수값 λ 마다 윗레벨 집합 $\{x : p(x) \geq \lambda\}$ 의 연결부분들의 집합이 $T_p(\lambda)$ 로써 대응되는 함수입니다.



군집 나무(Cluster Tree)는 고밀도 군집들의 계층 구조입니다.

Definition

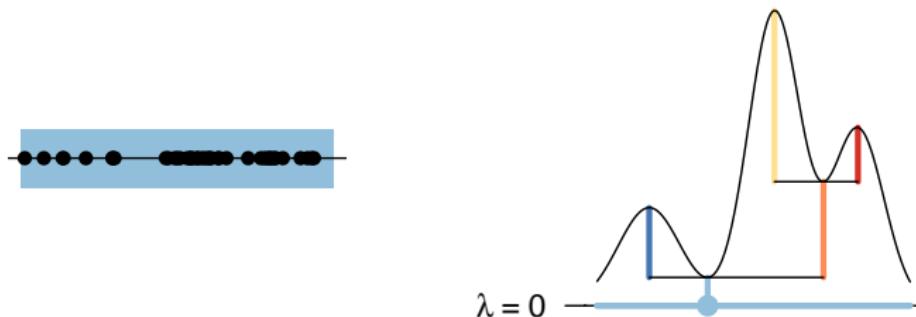
주어진 밀도함수 p 의 군집 나무(Cluster Tree) $T_p : \mathbb{R} \rightarrow \mathcal{P}(\mathcal{X})$ 는 각 실수값 λ 마다 윗레벨 집합 $\{x : p(x) \geq \lambda\}$ 의 연결부분들의 집합이 $T_p(\lambda)$ 로써 대응되는 함수입니다.



군집 나무(Cluster Tree)는 고밀도 군집들의 계층 구조입니다.

Definition

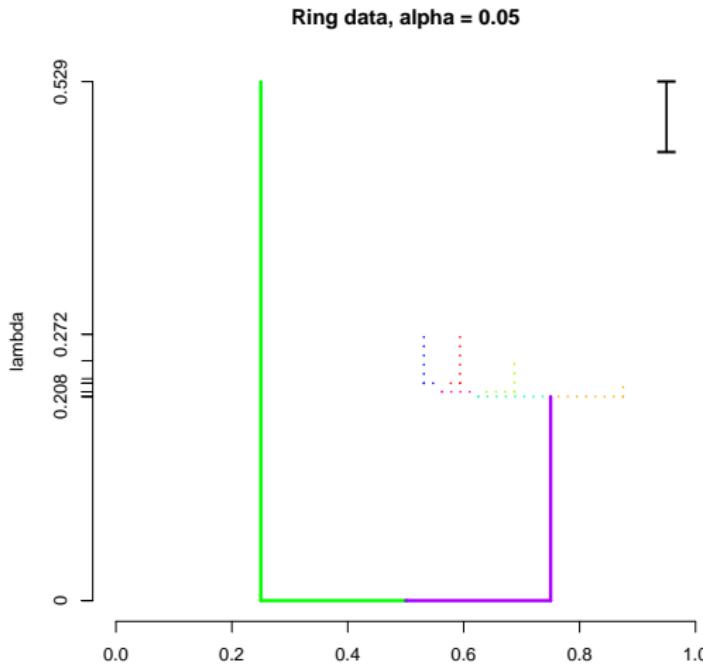
주어진 밀도함수 p 의 군집 나무(Cluster Tree) $T_p : \mathbb{R} \rightarrow \mathcal{P}(\mathcal{X})$ 는 각 실수값 λ 마다 윗레벨 집합 $\{x : p(x) \geq \lambda\}$ 의 연결부분들의 집합이 $T_p(\lambda)$ 로써 대응되는 함수입니다.



신뢰집합은 경험적 군집 나무에서 잡음을 줄이는 데에 도움을 줍니다.

- ▶ 점근적 $1 - \alpha$ 신뢰집합 \hat{C}_α 는 다음을 만족하는 군집 나무들의 집합입니다:

$$P(T_p \in \hat{C}_\alpha) = 1 - \alpha + o(1).$$



$1 - \alpha$ 신뢰집합 C_α 는 부트스트랩으로 계산할 수 있습니다.

- ▶ $T_{\hat{p}_h}$ 를 핵밀도추정(kernel density estimator) \hat{p}_h 에서 계산된 군집 나무로 놓습니다. 이 때,

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

그리고 신뢰집합을 $T_{\hat{p}_h}$ 을 중심으로 하고 반지름이 t_α 인 공으로 정의합니다. 즉,

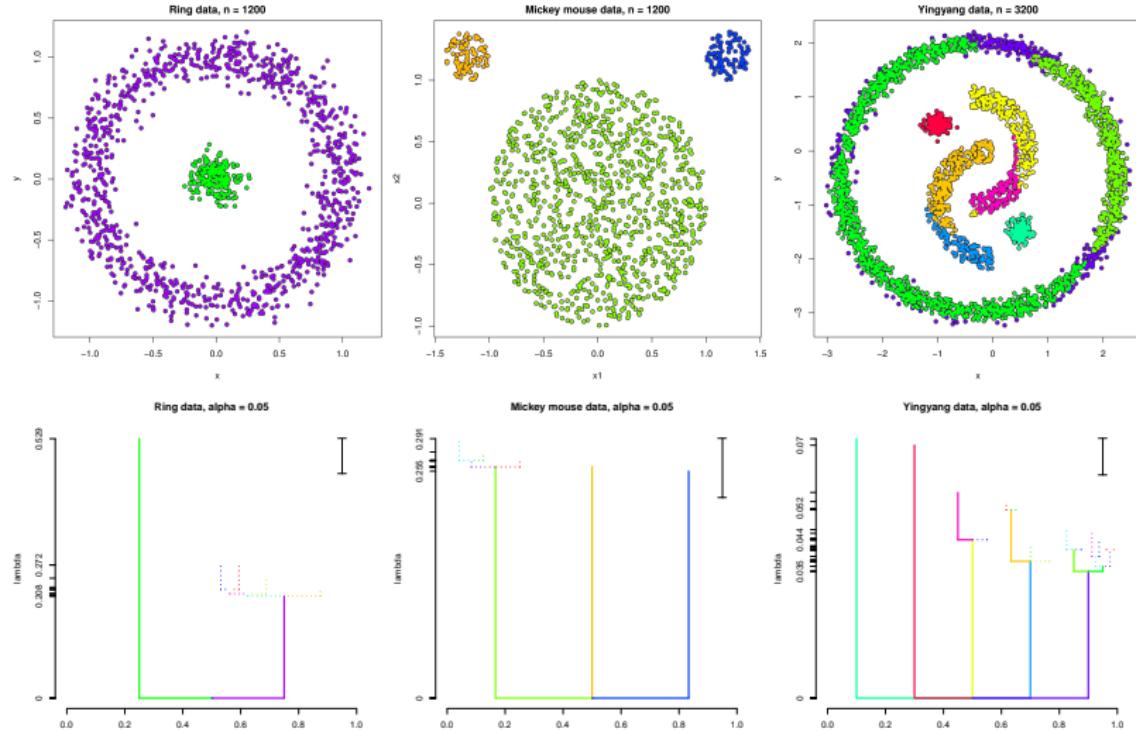
$$\hat{C}_\alpha = \{T : d_\infty(T, T_{\hat{p}_h}) \leq t_\alpha\}.$$

Theorem

(Theorem 3) 위의 신뢰집합 \hat{C}_α 은 다음을 만족합니다:

$$P\left(T_h \in \hat{C}_\alpha\right) = 1 - \alpha + O\left(\left(\frac{\log^7 n}{nh^m}\right)^{1/6}\right).$$

신뢰집합을 이용하여 가지치기한 군집나무로 실제 군집나무를 찾을 수 있습니다.



위상학적 자료 분석(Topological Data Analysis) 소개

호몰로지(homology)를 통계적으로 추정하기

군집 나무(Cluster Tree)와 이를 통계적으로 추정하기

Persistent Homology를 통계적으로 추정하기

Persistent Homology의 신뢰띠를 격자 위에서 계산하기

Persistent Homology의 신뢰띠를 자료 위에서 계산하기

Persistent Homology를 이용하여 특성(Feature) 만들기

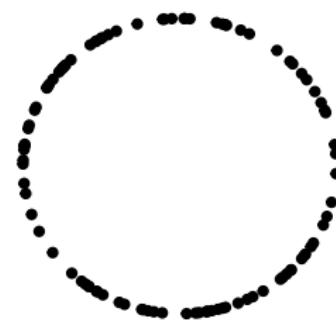
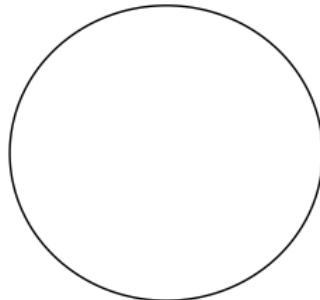
R 패키지 TDA: 위상학적 자료 분석을 위한 통계 계산 도구

참조문헌

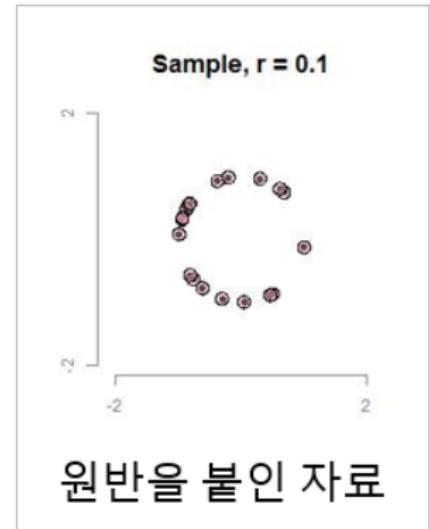
유한한 자료의 호몰로지는 기저 구조의 호몰로지와 다르기 때문에, 유한한 자료로 직접 기저 구조의 호몰로지를 추정할 수는 없습니다.

- ▶ 자료를 분석할 때, 기저 구조의 특성을 자료의 특성으로부터 추정할 수 있는 로버스트한 특성을 선호합니다.
- ▶ 호몰로지는 로버스트하지 않습니다:

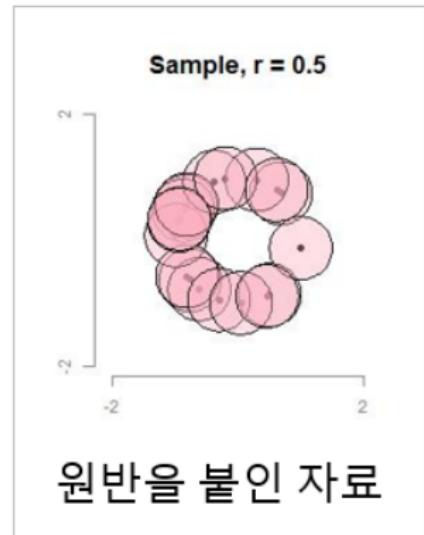
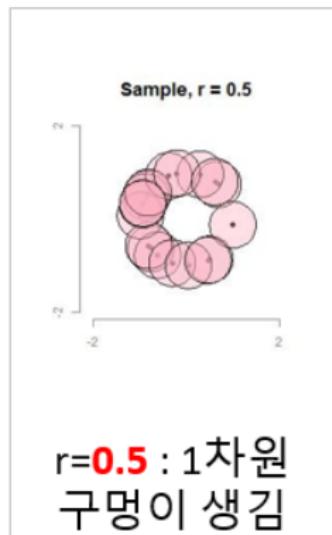
Underlying circle: $\beta_0 = 1$, $\beta_1 = 1$ 100 samples: $\beta_0 = 100$, $\beta_1 = 0$



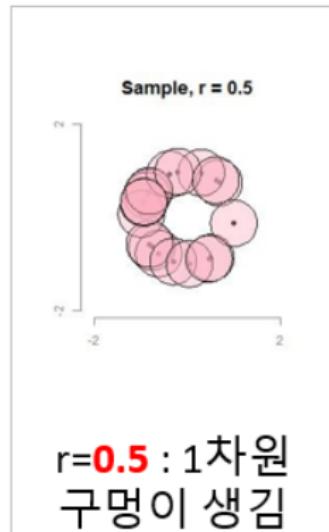
Persistent homology는 집합들의 모임에서 호몰로지를 계산하고, 호몰로지가 언제 나타나고 사라지는지 기록합니다.



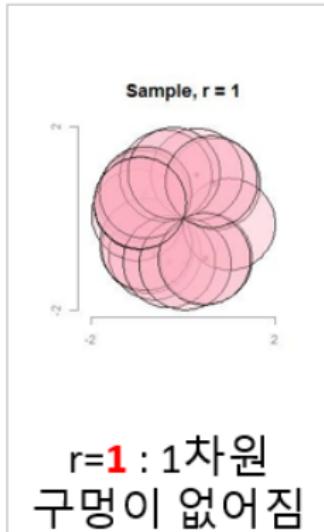
Persistent homology는 집합들의 모임에서 호몰로지를 계산하고, 호몰로지가 언제 나타나고 사라지는지 기록합니다.



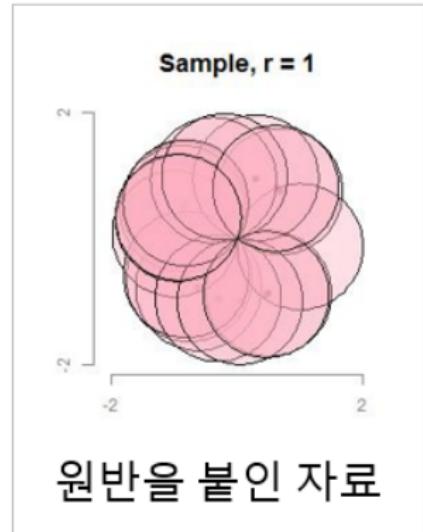
Persistent homology는 집합들의 모임에서 호몰로지를 계산하고, 호몰로지가 언제 나타나고 사라지는지 기록합니다.



$r=0.5$: 1차원
구멍이 생김

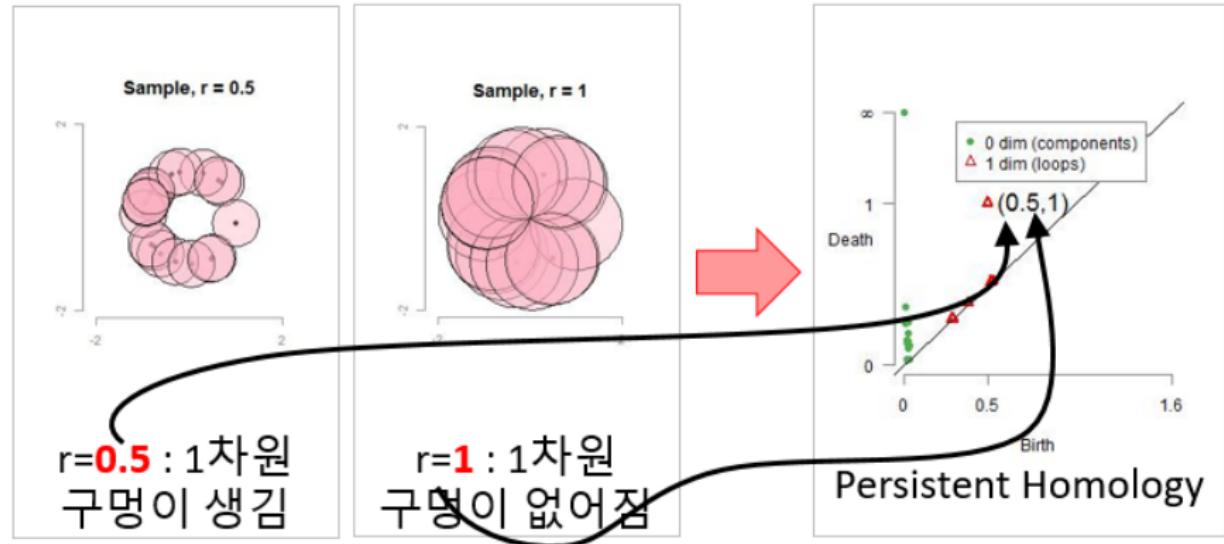


$r=1$: 1차원
구멍이 없어짐



원반을 붙인 자료

Persistent homology는 집합들의 모임에서 호몰로지를 계산하고, 호몰로지가 언제 나타나고 사라지는지 기록합니다.

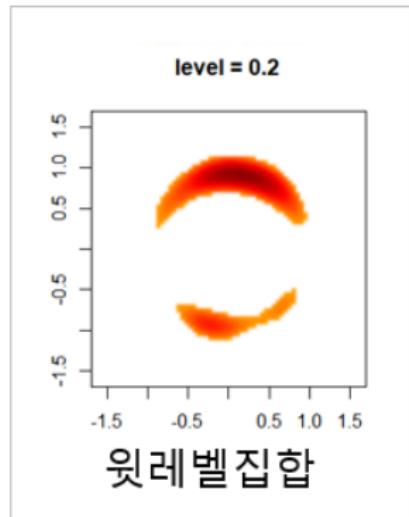


기저 구조의 위상학적 정보를 추출하는 데에 핵밀도추정(kernel density estimator)을 사용합니다.

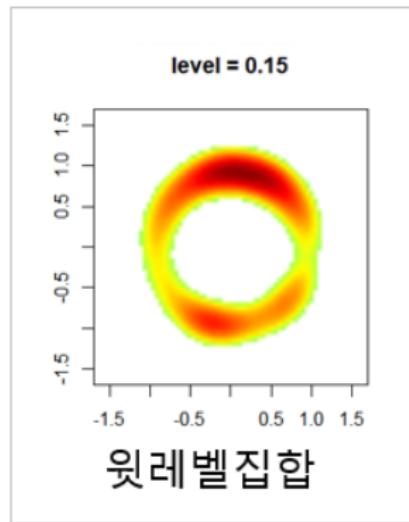
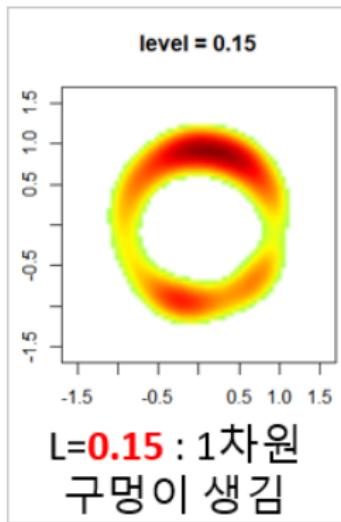
- ▶ 핵밀도추정(kernel density estimator)은 다음과 같습니다:

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

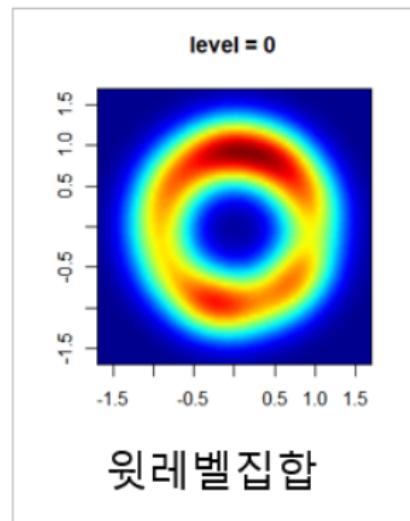
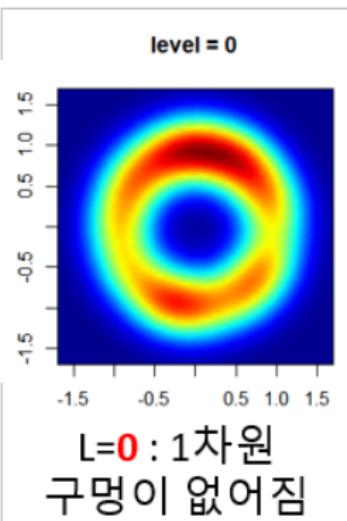
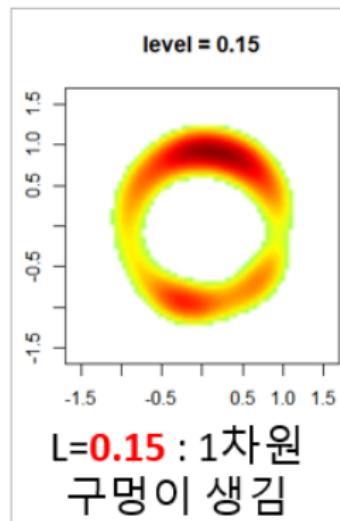
Persistent homology는 집합들의 모임에서 호몰로지를 계산하고, 호몰로지가 언제 나타나고 사라지는지 기록합니다.



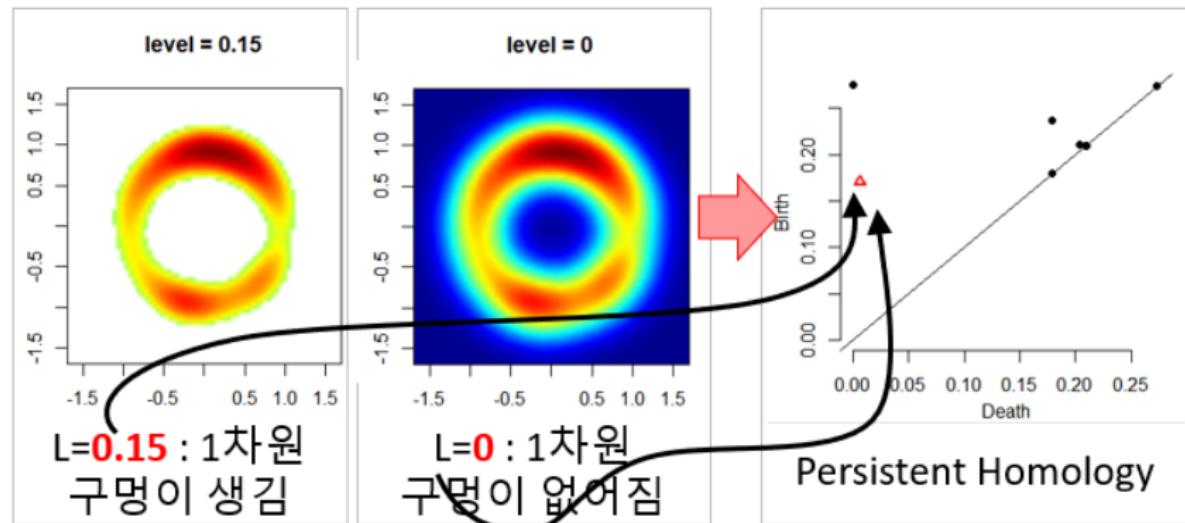
Persistent homology는 집합들의 모임에서 호몰로지를 계산하고, 호몰로지가 언제 나타나고 사라지는지 기록합니다.



Persistent homology는 집합들의 모임에서 호몰로지를 계산하고, 호몰로지가 언제 나타나고 사라지는지 기록합니다.

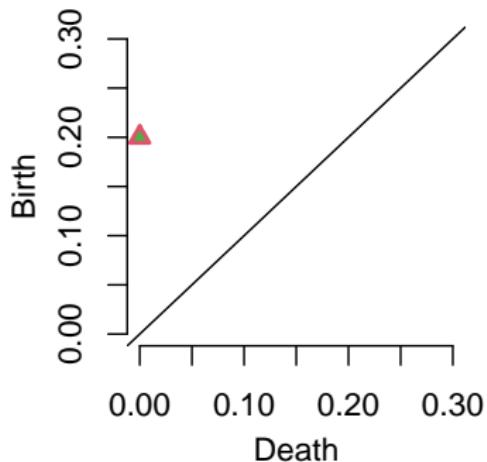


Persistent homology는 집합들의 모임에서 호몰로지를 계산하고, 호몰로지가 언제 나타나고 사라지는지 기록합니다.

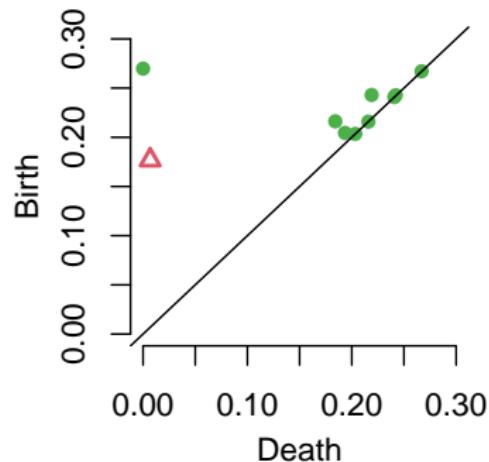


유한한 자료의 Persistent homology로부터 기저 구조의 Persistent homology를 추정할 수 있습니다.

Circle

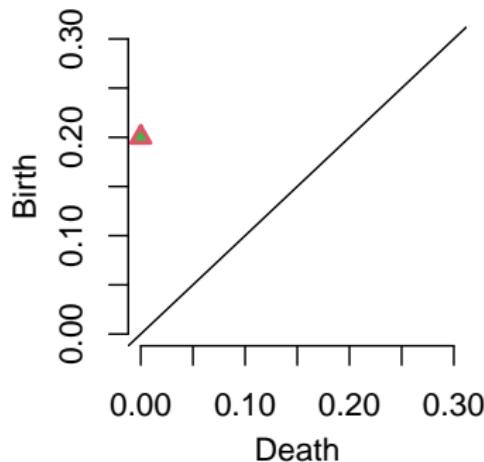


200 samples

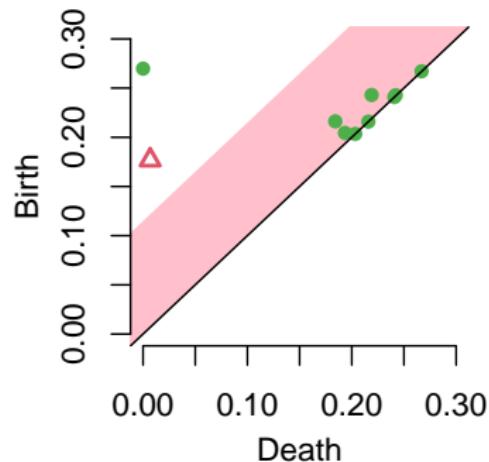


통계적으로 유의한 호몰로지 특성과 그렇지 않은
호몰로지 특성을 어떻게 구분할까요?

Circle



200 samples



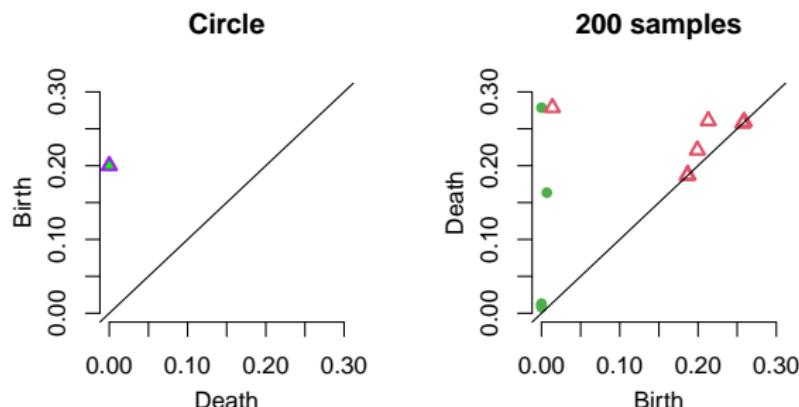
Bottleneck distance는 Persistent homology 공간에 거리 함수를 줍니다.

Definition

D_1, D_2 를 두 Persistent homology라고 하면, Bottleneck distance는 다음과 같이 정의됩니다:

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

이 때, γ 는 D_1 에서 D_2 로 가는 모든 일대일대응이 될 수 있습니다.



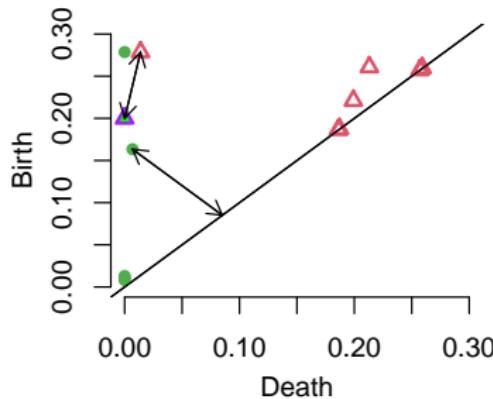
Bottleneck distance는 Persistent homology 공간에 거리 함수를 줍니다.

Definition

D_1, D_2 를 두 Persistent homology라고 하면, Bottleneck distance는 다음과 같이 정의됩니다:

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

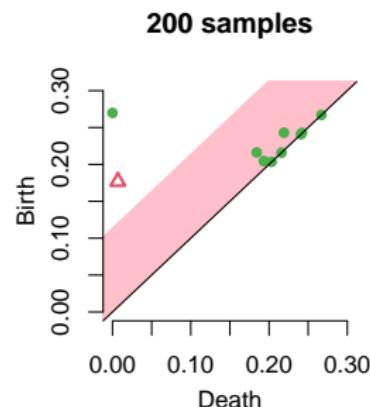
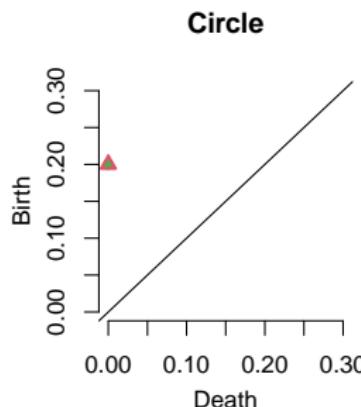
이 때, γ 는 D_1 에서 D_2 로 가는 모든 일대일대응이 될 수 있습니다.



Persistent homology의 신뢰띠는 Persistent homology 를 높을 확률로 포함하는 확률변수입니다.

기저 M 과 자료 X 의 Persistent homology를 각각 $Dgm(M)$ 과 $Dgm(X)$ 라고 놓습니다. 유의수준 $\alpha \in (0, 1)$ 가 주어졌을 때, $(1 - \alpha)$ 신뢰띠 $c_n = c_n(X)$ 는 다음을 만족하는 확률변수입니다:

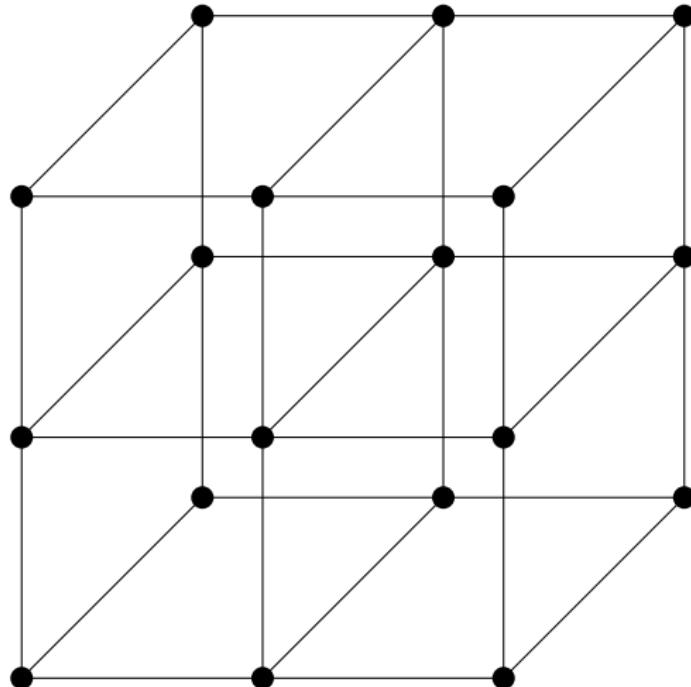
$$\mathbb{P}(W_\infty(Dgm(M), Dgm(X)) \leq c_n) \geq 1 - \alpha.$$



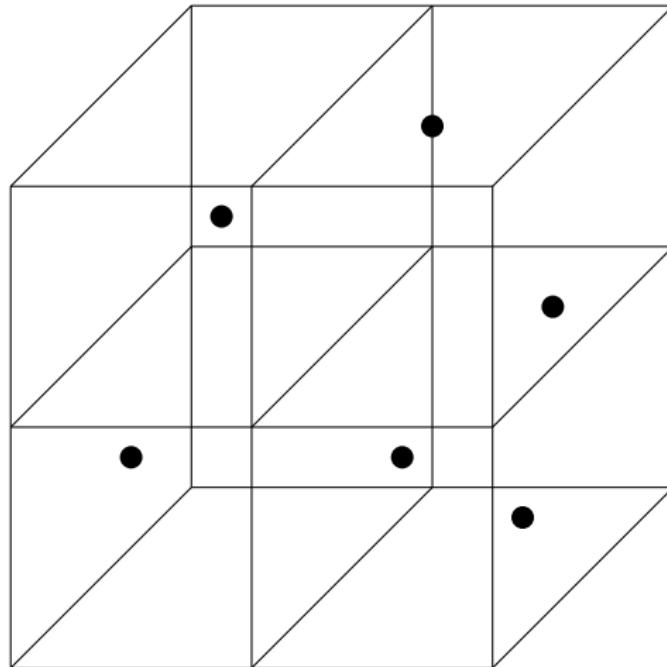
Persistent homology의 신뢰띠는 봇스트랩으로 계산할 수 있습니다.

1. 주어진 자료 $X = \{x_1, \dots, x_n\}$ 에서 핵밀도추정(kernel density estimator) \hat{p}_h 를 계산합니다.
2. $X = \{x_1, \dots, x_n\}$ 로부터 $X^* = \{x_1^*, \dots, x_n^*\}$ 를 복원추출하고, X^* 의 핵밀도추정 \hat{p}_h^* 을 계산한 후, $\theta^* = \sqrt{nh^d} \|\hat{p}_h^*(x) - \hat{p}_h(x)\|_\infty$ 를 계산합니다.
3. 전단계를 B 번 반복하여 $\theta_1^*, \dots, \theta_B^*$ 를 얻습니다.
4. 분위수 $\hat{z}_\alpha = \inf \left\{ q : \frac{1}{B} \sum_{j=1}^B I(\theta_j^* \geq q) \leq \alpha \right\}$ 를 계산합니다.
5. $\mathbb{E}[\hat{p}_h]$ 의 $(1 - \alpha)$ 신뢰띠는 $\left[\hat{p}_h - \frac{\hat{z}_\alpha}{\sqrt{nh^d}}, \hat{p}_h + \frac{\hat{z}_\alpha}{\sqrt{nh^d}} \right]$ 이 됩니다.

Peresistent homology의 신뢰띠을 계산하기 위해서는 격자 위에서 계산을 해야 하는데, 고차원 공간에서는 계산이 힘듭니다.

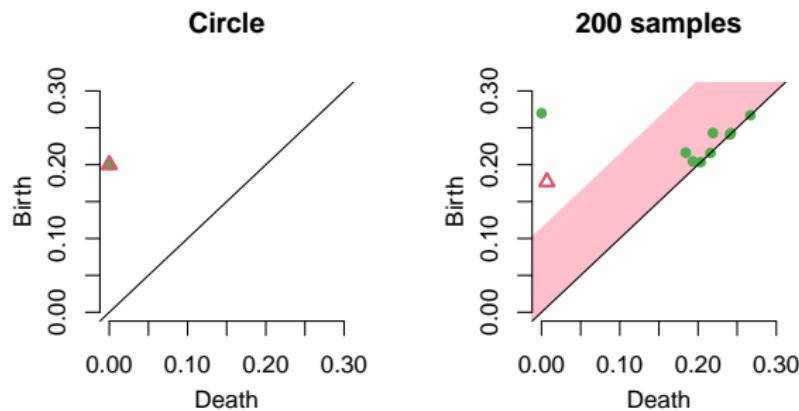


Persistent homology를 자료 위에서 계산할 수 있으면 계산복잡도가 줄어듭니다.



Persistent homology의 신뢰띠을 자료 위에서 계산할 수 있을까요?

- ▶ (Shin, Kim, Rinaldo, Wasserman, 2021?)

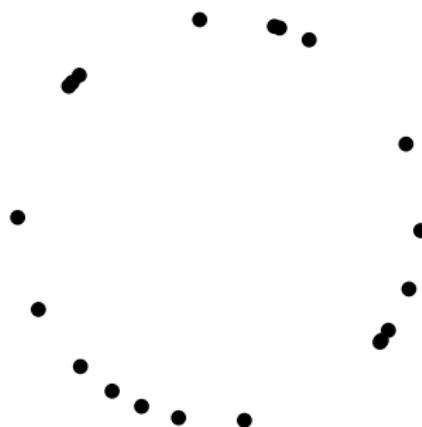


목표 persistent homology를 추정하기 위해 Vietoris-Rips complex를 사용합니다.

- ▶ 주어진 $\mathcal{X} \subset \mathbb{R}^d$ 와 $r > 0$ 에 대해, Vietoris-Rips complex $\text{Rips}(\mathcal{X}, r)$ 는 다음과 같이 정의됩니다:

$$\text{Rips}(\mathcal{X}, r) = \{\{x_1, \dots, x_k\} \subset \mathcal{X} : d(x_i, x_j) < 2r, \text{ for all } 1 \leq i, j \leq k\}.$$

Vietoris-Rips complex



목표 persistent homology를 추정하기 위해 Vietoris-Rips complex를 사용합니다.

- ▶ 주어진 $\mathcal{X} \subset \mathbb{R}^d$ 와 $r > 0$ 에 대해, Vietoris-Rips complex $\text{Rips}(\mathcal{X}, r)$ 는 다음과 같이 정의됩니다:

$$\text{Rips}(\mathcal{X}, r) = \{\{x_1, \dots, x_k\} \subset \mathcal{X} : d(x_i, x_j) < 2r, \text{ for all } 1 \leq i, j \leq k\}.$$

Vietoris-Rips complex

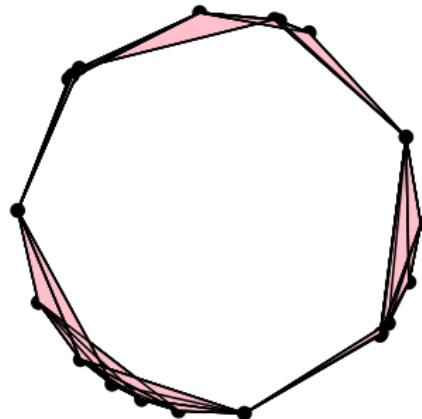


목표 persistent homology를 추정하기 위해 Vietoris-Rips complex를 사용합니다.

- ▶ 주어진 $\mathcal{X} \subset \mathbb{R}^d$ 와 $r > 0$ 에 대해, Vietoris-Rips complex $\text{Rips}(\mathcal{X}, r)$ 는 다음과 같이 정의됩니다:

$$\text{Rips}(\mathcal{X}, r) = \{\{x_1, \dots, x_k\} \subset \mathcal{X} : d(x_i, x_j) < 2r, \text{ for all } 1 \leq i, j \leq k\}.$$

Vietoris-Rips complex



핵밀도추정 레벨집합으로 만든 Vietoris-Rips complex를 사용하여 목표 레벨집합을 추정합니다.

- ▶ 주어진 $\mathcal{X} \subset \mathbb{R}^d$ 와 $r > 0$ 에 대해, Vietoris-Rips complex $\text{Rips}(\mathcal{X}, r)$ 는 다음과 같이 정의됩니다:

$$\text{Rips}(\mathcal{X}, r) = \{\{x_1, \dots, x_k\} \subset \mathcal{X} : d(x_i, x_j) < 2r, \text{ for all } 1 \leq i, j \leq k\}.$$

- ▶ 핵밀도추정(kernel density estimator)은 다음과 같습니다:

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

- ▶ 주어진 $\mathcal{X}_n = \{X_1, \dots, X_n\}$ 에 대해, 핵밀도추정 레벨집합으로 만든 Vietoris-Rips complex를 사용합니다.

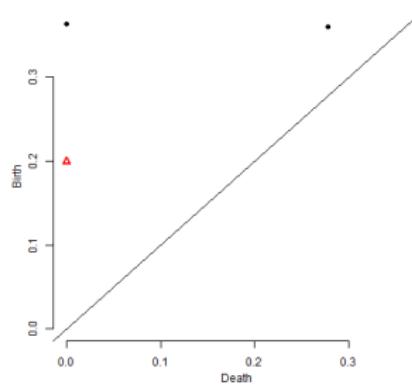
$$\left\{ \text{Rips} \left(\mathcal{X}_{n,L}^{\hat{p}_h}, r \right) \right\}_{L>0}, \text{ where } \mathcal{X}_{n,L}^{\hat{p}_h} = \{X_i \in \mathcal{X}_n : \hat{p}_h(X_i) \geq L\}.$$

Vietoris-Rips complex 위에서의 핵밀도추정 persistent homology를 사용하여 목표 persistent homology를 추정합니다.

- ▶ 목표 persistent homology를 추정하기 위해 핵밀도추정 레벨집합으로 만든 Vietoris-Rips complex의 persistent homology를 사용합니다:

$$\left\{ \text{Rips} \left(\mathcal{X}_{n,L}^{\hat{p}_h}, r \right) \right\}_{L>0}, \text{ where } \mathcal{X}_{n,L}^{\hat{p}_h} = \{ X_i \in \mathcal{X}_n : \hat{p}_h(X_i) \geq L \}.$$

그리고 그 persistent homology를 $PH_*^R(\hat{p}_h, r)$ 로 씁니다.



Vietoris-Rips complex 위에서의 핵밀도추정 persistent homology는 일치성(consistency)이 있습니다.

Theorem

(Theorem 16, Corollary 17) $\{r_n\}_{n \in \mathbb{N}}$ 과 $\{h_n\}_{n \in \mathbb{N}}$ 에 대하여 $r_n = \Omega\left(\left(\frac{\log n}{n}\right)^{1/d}\right)$, $r_n = o(1)$, $\frac{\log(1/h_n)}{nh_n^d} = O(1)$ 을 만족한다고 가정합니다. 그러면

$$W_\infty(PH_*(p_{h_n}), PH_*^R(\hat{p}_{h_n}, r_n)) = O_P\left(\sqrt{\frac{\log(1/h_n)}{nh_n^d}} + \|r_n\|_\infty\right).$$

신뢰띠

- ▶ 점근적 $1 - \alpha$ 신뢰띠 \hat{b}_α 는 다음을 만족하는 확률변수입니다.

$$\mathbb{P} \left(W_\infty \left(PH_*(p_{h_n}), PH_*^R(\hat{p}_{h_n}, r_n) \right) \leq \hat{b}_\alpha \right) \geq 1 - \alpha + o(1).$$

Vietoris-Rips complex 위에서의 핵밀도추정 persistent homology의 신뢰띠

- ▶ 븁스트랩으로 계산한 \hat{b}_α 는 다음 정리에 의해 유효한 신뢰띠입니다.

Theorem

(Theorem 20)

$$\mathbb{P} \left(W_\infty \left(PH_*(p_{h_n}), PH_*^R(\hat{p}_{h_n}, r_n) \right) \leq \hat{b}_\alpha \right) \geq 1 - \alpha + o(1).$$

위상학적 자료 분석(Topological Data Analysis) 소개

호몰로지(homology)를 통계적으로 추정하기

군집 나무(Cluster Tree)와 이를 통계적으로 추정하기

Persistent Homology를 통계적으로 추정하기

Persistent Homology의 신뢰띠를 격자 위에서 계산하기

Persistent Homology의 신뢰띠를 자료 위에서 계산하기

Persistent Homology를 이용하여 특성(Feature) 만들기

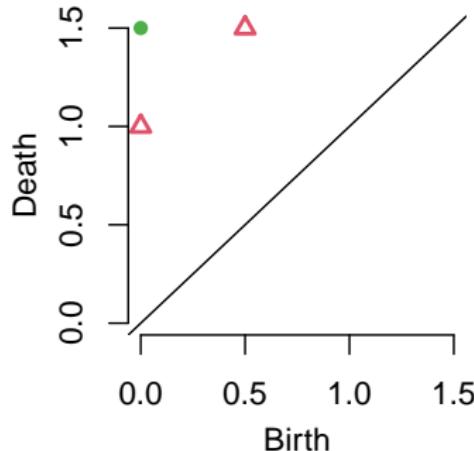
R 패키지 TDA: 위상학적 자료 분석을 위한 통계 계산 도구

참조문헌

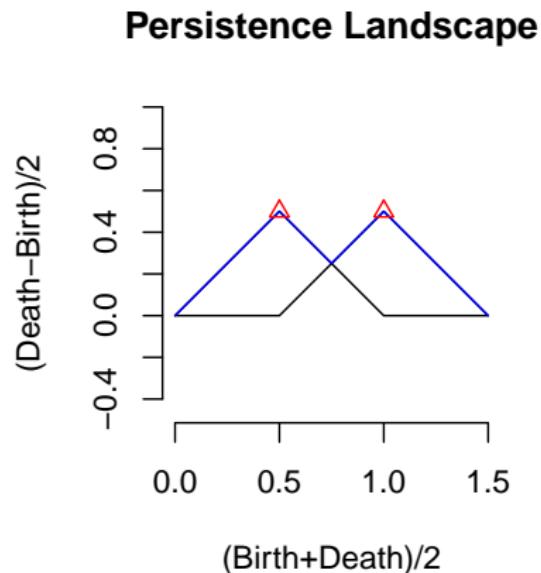
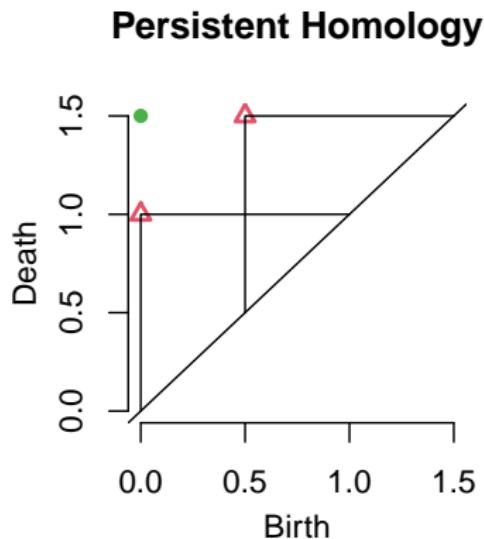
Persistent homology를 한 번 더 요약해서 유클리드 공간 또는 함수 공간에 넣습니다.

- ▶ Persistent homology의 공간은 구조적으로 복잡하여 기계학습 알고리즘과 같이 사용하기는 힘듭니다.
- ▶ Persistent homology를 한 번 더 요약해서 유클리드 공간 또는 함수 공간에 넣으면 기계학습의 알고리즘에 사용하기 편합니다.
 - ▶ Persistence Landscape, Persistence Silhouette, Persistence Image 등 여러 방법이 있습니다.

Persistent Homology



Persistence Landscape은 Persistent homology의 함수
요약입니다.

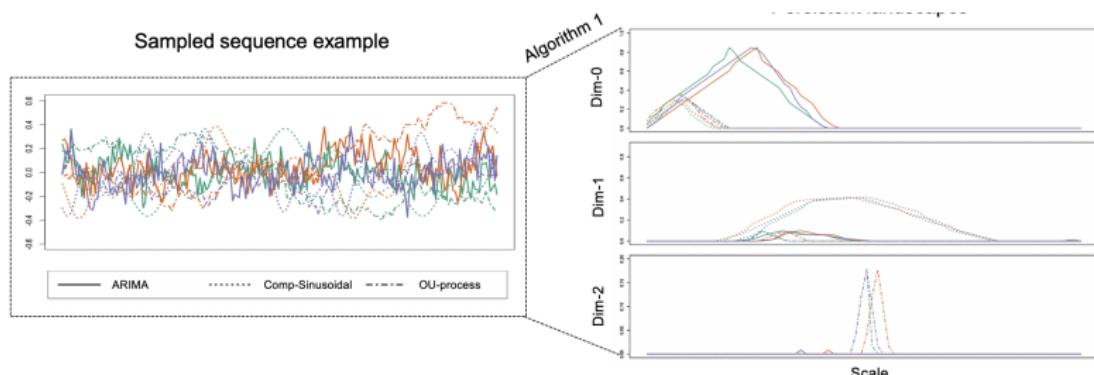


Persistence Landscape을 이용하여 특성(feature)을 만들 수 있습니다.

- ▶ 시간지체법(time-delayed embedding)과 Persistence Landscape 으로 특성(feature) 만들기
 - ▶ Time Series Featurization via Topological Data Analysis (Kim, Kim, Rinaldo, Chazal, 2020)
- ▶ Persistence Landscape으로 위상학적 층(topological layer) 만들기
 - ▶ PLLay: Efficient Topological Layer based on Persistence Landscapes (Kim, Kim, Zaheer, Kim, Chazal, Wasserman, 2020)

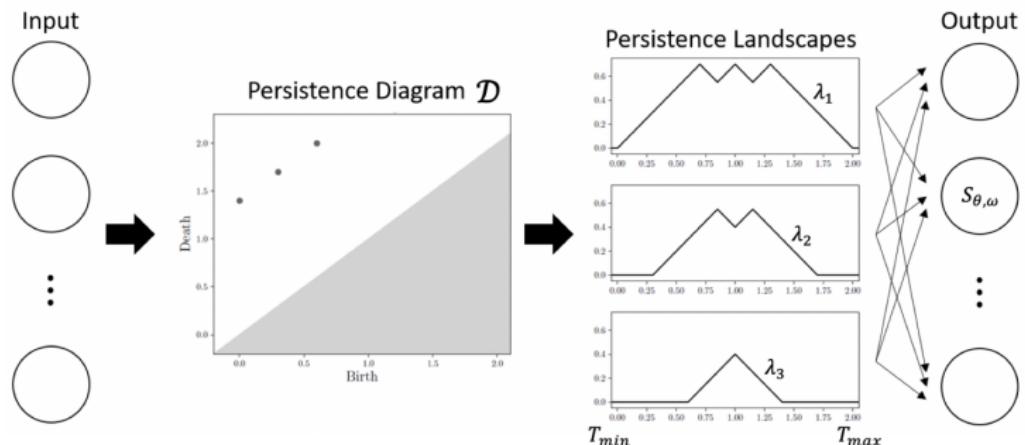
시간지체법(time-delayed embedding)과 Persistence Landscape 으로 특성(feature) 만들기

1. 시계열자료에 시간지체법(time-delayed embedding)을 사용하여 점구름 자료(point cloud) $X \subset \mathbb{R}^m$ 을 만듭니다.
2. X 에 주성분분석(PCA, Principal Component Analysis)을 사용하여 $X' \subset \mathbb{R}^l$ 을 만듭니다.
3. Vietoris-Rips filtration $R_{X'}$ 을 만들고 그의 Persistence diagram $Dgm(X')$ 을 계산합니다.
4. $Dgm(X')$ 으로부터 landscape $\lambda : \mathbb{N} \times \mathbb{R} \rightarrow \mathbb{R}$ 을 계산하고, 이를 벡터화하여 $\lambda^K \in \mathbb{R}^K$ 를 만듭니다.
5. λ^K 에 주성분분석을 사용하여 $\lambda^k \in \mathbb{R}^k$ 를 만듭니다.

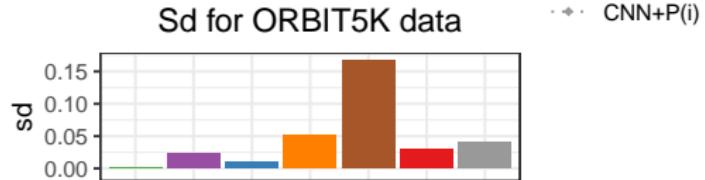
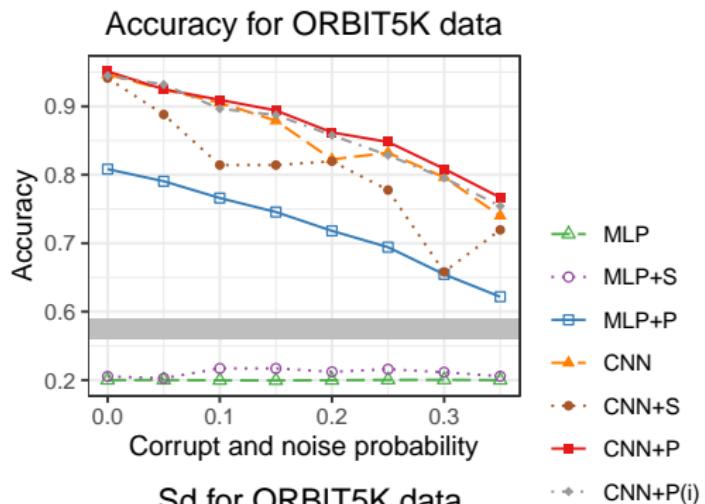
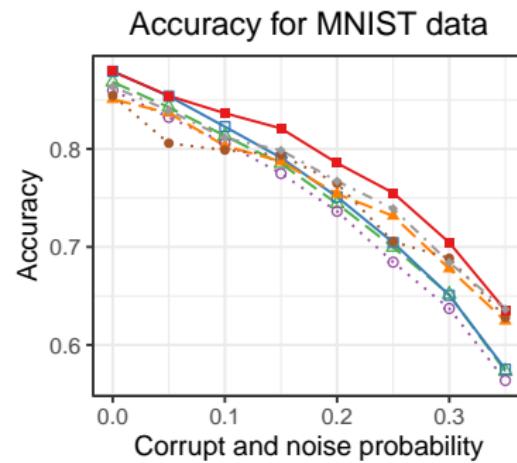


Persistence Landscape으로 위상학적 층(topological layer) 만들기

1. 자료 X 에 적당한 simplicial complex K 와 함수 f 를 선택하여 Persistent Homology \mathcal{D} 를 계산합니다.
2. \mathcal{D} 로부터 Landscape $\lambda : \mathbb{N} \times \mathbb{R} \rightarrow \mathbb{R}$ 을 계산합니다.
3. 매개변수 $\omega \in \mathbb{R}^{K_{\max}}$ 를 이용하여 가중평균함수 $\bar{\lambda}_{\omega}(t) := \sum_{k=1}^{K_{\max}} \omega_k \lambda_k(t)$ 을 계산합니다.
4. $\bar{\lambda}_{\omega}$ 를 벡터화하여 $\bar{\Lambda}_{\omega} \in \mathbb{R}^m$ 을 만듭니다.
5. 매개화된 미분가능함 함수 $g_{\theta} : \mathbb{R}^m \rightarrow \mathbb{R}$ 을 사용하여, $S_{\theta, \omega}(\mathcal{D}) := g_{\theta}(\bar{\Lambda}_{\omega})$ 를 계산합니다.



Persistence Landscape으로 위상학적 층(topological layer) 만들기



위상학적 자료 분석(Topological Data Analysis) 소개

호몰로지(homology)를 통계적으로 추정하기

군집 나무(Cluster Tree)와 이를 통계적으로 추정하기

Persistent Homology를 통계적으로 추정하기

Persistent Homology의 신뢰띠를 격자 위에서 계산하기

Persistent Homology의 신뢰띠를 자료 위에서 계산하기

Persistent Homology를 이용하여 특성(Feature) 만들기

R 패키지 TDA: 위상학적 자료 분석을 위한 통계 계산 도구

참조문헌

위상학적 자료 분석(Topological Data Analysis)를 해주는
많은 프로그램들이 있습니다.

- ▶ 위상학적 자료 분석을 해주는 프로그램들 예시: Dionysus, DIPHA, GUDHI, javaPlex, Perseus, PHAT, Ripser, TDA, TDAsstats

R 패키지 TDA는 위상학적 자료 분석을 해주는 C++ 라이브러리의 R 인터페이스를 제공합니다.

- ▶ 웹사이트:
<https://cran.r-project.org/web/packages/TDA/index.html>
- ▶ 저자: Brittany Terese Fasy, Jisu Kim, Fabrizio Lecci, Clément Maria, David Milman, and Vincent Rouvreau.
- ▶ R은 통계 계산과 시각화를 위한 프로그래밍 언어입니다.
- ▶ R은 개발시간이 짧고, C/C++는 실행시간이 짧습니다.
- ▶ R package TDA는 위상학적 자료 분석을 해주는 C++ 라이브러리인 GUDHI/Dionysus/PHAT의 R 인터페이스를 제공합니다.

위상학적 자료 분석(Topological Data Analysis) 소개

호몰로지(homology)를 통계적으로 추정하기

군집 나무(Cluster Tree)와 이를 통계적으로 추정하기

Persistent Homology를 통계적으로 추정하기

Persistent Homology의 신뢰띠를 격자 위에서 계산하기

Persistent Homology의 신뢰띠를 자료 위에서 계산하기

Persistent Homology를 이용하여 특성(Feature) 만들기

R 패키지 TDA: 위상학적 자료 분석을 위한 통계 계산 도구

참조문헌

참조문헌

- Frédéric Chazal, Vin de Silva, Marc Glisse, and Steve Oudot. The structure and stability of persistence modules. *arXiv preprint arXiv:1207.3674*, 2012.
- Frédéric Chazal, Brittany Terese Fasy, Fabrizio Lecci, Bertrand Michel, Alessandro Rinaldo, and Larry Wasserman. Robust topological inference: Distance-to-a-measure and kernel distance. *Technical Report*, 2014.
- Herbert Edelsbrunner and John Harer. *Computational topology: an introduction*. American Mathematical Society, 2010.
- Brittany T. Fasy, Jisu Kim, Fabrizio Lecci, Clément Maria, David L. Millman, and Vincent Rouvreau. Introduction to the R package TDA. *CoRR*, abs/1411.1830, 2014a. URL
<http://arxiv.org/abs/1411.1830>.
- Brittany Terese Fasy, Fabrizio Lecci, Alessandro Rinaldo, Larry Wasserman, Sivaraman Balakrishnan, and Aarti Singh. Confidence sets for persistence diagrams. *Ann. Statist.*, 42(6):2301–2339, 2014b. ISSN 0090-5364. doi: 10.1214/14-AOS1252. URL
<https://doi.org/10.1214/14-AOS1252>.
- Jisu KIM, Yen-Chi Chen, Sivaraman Balakrishnan, Alessandro Rinaldo, and Larry Wasserman. Statistical inference for cluster trees. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, editors,

감사합니다!

군집 나무(Cluster Tree)와 이를 통계적으로 추정하기

Persistent Homology를 통계적으로 추정하기

Persistent Homology의 신뢰띠를 자료 위에서 계산하기

Persistent Homology의 신뢰띠를 자료 위에서 계산하기

Persistent Homology를 이용하여 특성(Feature) 만들기

R 패키지 TDA: 위상학적 자료 분석을 위한 통계 계산 도구

다양체(manifold)에서의 표본 추출, 거리 함수, 밀도 함수

Persistent Homology와 Persistence Landscape

Persistence Homology와 Persistence Landscape의 통계적 추정

ℓ_∞ 거리로 군집 나무간에 거리를 쟈 수 있습니다.

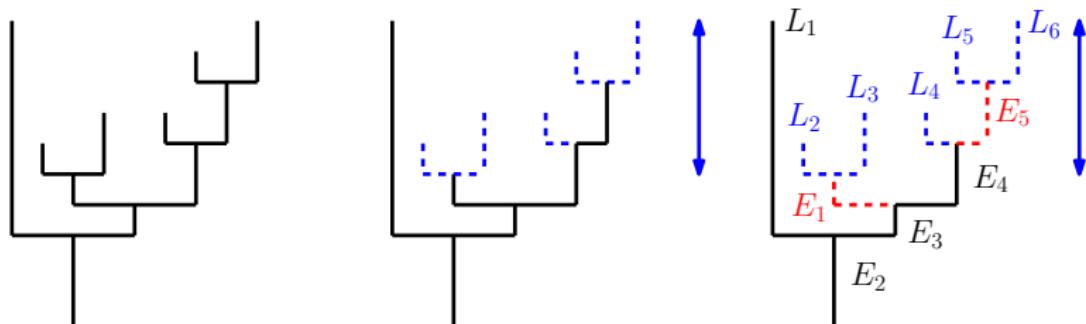
Definition

군집 나무간의 ℓ_∞ 거리는 다음과 같이 정의됩니다.

$$d_\infty(T_p, T_q) = \sup |p(x) - q(x)|.$$

경험적 군집 나무를 가지치기함으로써 신뢰집합을 탐색할 수 있습니다.

- ▶ 경험적 군집 나무 $T_{\hat{p}_h}$ 보다 단순하고 신뢰집합에 속해 있는 군집 나무를 다음의 두 가지 가지치기 방법으로 찾을 수 있습니다.
 - ▶ 잎사귀만 가지치기: 길이가 $2\hat{t}_\alpha$ 보다 작은 모든 잎사귀를 없앱니다.
 - ▶ 잎사귀와 내부가지를 가지치기: 말단으로부터의 누적길이가 $2\hat{t}_\alpha$ 보다 작은 모든 가지들을 제거합니다.



군집 나무(Cluster Tree)와 이를 통계적으로 추정하기

Persistent Homology를 통계적으로 추정하기

Persistent Homology의 신뢰띠를 자료 위에서 계산하기

Persistent Homology의 신뢰띠를 자료 위에서 계산하기

Persistent Homology를 이용하여 특성(Feature) 만들기

R 패키지 TDA: 위상학적 자료 분석을 위한 통계 계산 도구

다양체(manifold)에서의 표본 추출, 거리 함수, 밀도 함수

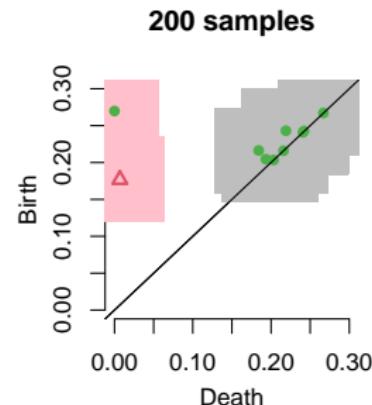
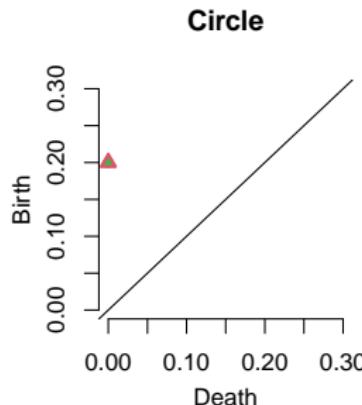
Persistent Homology와 Persistence Landscape

Persistence Homology와 Persistence Landscape의 통계적 추정

Persistent homology의 신뢰집합은 Persistent homology를 높을 확률로 포함하는 랜덤집합입니다.

기저 M 과 자료 X 의 Persistent homology를 각각 $Dgm(M)$ 과 $Dgm(X)$ 라고 놓습니다. 유의수준 $\alpha \in (0, 1)$ 가 주어졌을 때, $(1 - \alpha)$ 신뢰집합 $\{D \in Dgm : W_\infty(Dgm(X), D) \leq c_n\}$ 은 다음을 만족하는 랜덤집합입니다:

$$\mathbb{P}(Dgm(M) \in \{D \in Dgm : W_\infty(Dgm(X), D) \leq c_n\}) \geq 1 - \alpha.$$



Bottleneck distance는 그에 상응하는 함수간의 거리로 조정할 수 있습니다: 안정성 정리

Theorem

[Edelsbrunner and Harer, 2010][Chazal, de Silva, Glisse, and Oudot, 2012] \mathbb{X} 를 유한히 삼각분할(finitely triangulable)할 수 있는 공간이라 하고 $f, g : \mathbb{X} \rightarrow \mathbb{R}$ 를 두 연속함수라 합니다. $Dgm(f)$ 와 $Dgm(g)$ 를 그에 상응하는 persistent homology라고 할 때, 다음이 성립합니다:

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

Persistent homology의 신뢰띠는 그에 상응하는 함수의 신뢰띠로 계산할 수 있습니다.

안정성 정리로부터, $\mathbb{P}(||f_M - f_X|| \leq c_n) \geq 1 - \alpha$ 는 다음을 유도합니다:

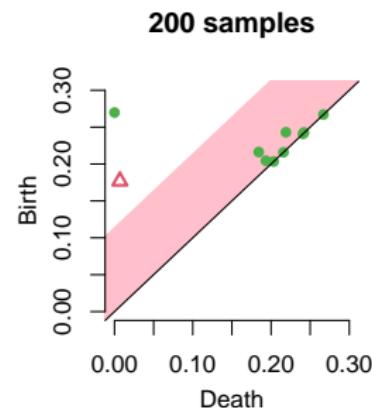
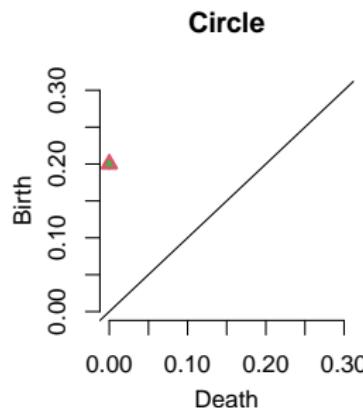
$$\mathbb{P}(W_\infty(Dgm(f_M), Dgm(f_X)) \leq c_n) \geq \mathbb{P}(||f_M - f_X||_\infty \leq c_n) \geq 1 - \alpha,$$

따라서 f_M 의 신뢰띠를 persistent homology $Dgm(f_M)$ 의 신뢰띠로 이용할 수 있습니다.

Persistent homology의 신뢰띠는 봇스트랩으로 계산할 수 있습니다.

봇스트랩 알고리즘을 persistent homology에 적용할 수 있다는 것이 증명되었습니다.

- ▶ Fasy et al. [2014b] 이 핵밀도추정(kernel density estimator)에서 보였고,
- ▶ Chazal et al. [2014] 이 distance to measure와 kernel distance에서 보였습니다.



받침(support) 위에서 평균 핵밀도추정(average kernel density estimator)의 윗레벨집합(upper level set)을 고려합니다.

- ▶ 자료를 $X_1, \dots, X_n \sim P$ 로 관찰할 때, 평균 핵밀도추정(average kernel density estimator)은 다음과 같습니다:

$$p_h(x) = \mathbb{E} [\hat{p}_h(x)] = \frac{1}{h^d} \mathbb{E} \left[K \left(\frac{x - X}{h} \right) \right].$$

- ▶ 받침(support) 위에서 평균 핵밀도추정의 윗레벨집합(upper level set)들을 고려합니다.

$$\{D_L\}_{L>0}, \text{ where } D_L := \{x \in \text{supp}(P) : p_h(x) \geq L\}.$$

받침(support) 위에서 평균 핵밀도추정(average kernel density estimator)의 윗레벨집합(upper level set)을 고려합니다.

- ▶ 받침(support) 위에서 평균 핵밀도추정의 윗레벨집합(upper level set)들을 고려합니다.

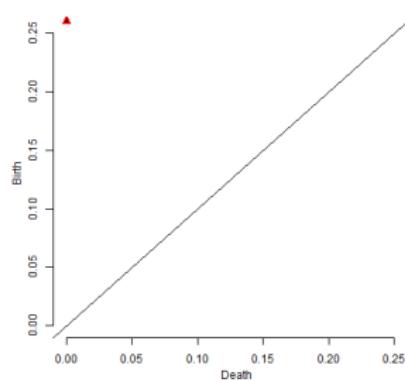
$$\{D_L\}_{L>0}, \text{ where } D_L := \{x \in \text{supp}(P) : p_h(x) \geq L\}.$$

받침 위에서 평균 핵밀도추정의 윗레벨집합의 persistent homology를 목표로 합니다.

- ▶ 받침(support) 위에서 평균 핵밀도추정의 윗레벨집합(upper level set)들을 고려합니다:

$$\{D_L\}_{L>0}, \text{ where } D_L := \{x \in \text{supp}(P) : p_h(x) \geq L\},$$

그리고 그것의 persistent homology $PH_*^{\text{supp}(P)}(p_h)$ 을 목표로 합니다.



핵밀도추정 레벨집합으로 만든 Vietoris-Rips complex를 사용하여 목표 레벨집합을 추정합니다.

- ▶ 주어진 $\mathcal{X}_n = \{X_1, \dots, X_n\}$ 에 대해, 핵밀도추정 레벨집합으로 만든 Vietoris-Rips complex를 사용합니다.

$$\left\{ \text{Rips} \left(\mathcal{X}_{n,L}^{\hat{p}_h}, r \right) \right\}_{L>0}, \text{ where } \mathcal{X}_{n,L}^{\hat{p}_h} = \{X_i \in \mathcal{X}_n : \hat{p}_h(X_i) \geq L\}.$$

핵밀도추정 레벨집합으로 만든 Vietoris-Rips complex으로
목표 레벨집합을 추정합니다.

- ▶ 목표 레벨집합을 근사하기 위해:

$$\{D_L\}_{L>0}, \text{ where } D_L := \{x \in \text{supp}(P) : p_h(x) \geq L\},$$

핵밀도추정 레벨집합으로 만든 Vietoris-Rips complex를 사용합니다.

$$\left\{ R \left(\mathcal{X}_{n,L}^{\hat{p}_h}, r \right) \right\}_{L>0}, \text{ where } \mathcal{X}_{n,L}^{\hat{p}_h} = \{X_i \in \mathcal{X}_n : \hat{p}_h(X_i) \geq L\}.$$

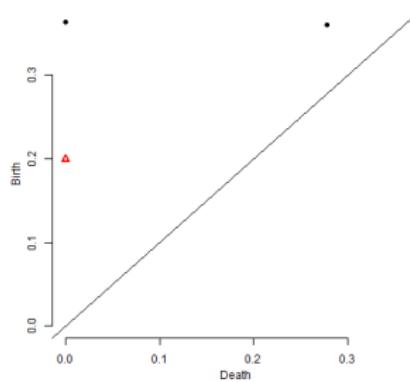
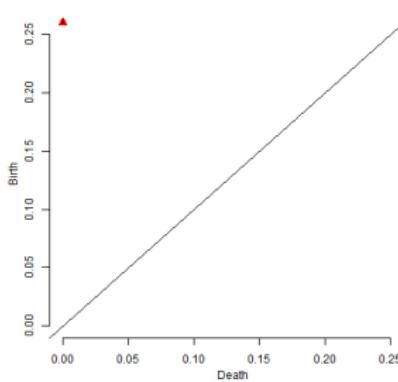
Vietoris-Rips complex 위에서의 핵밀도추정 persistent homology로 목표 persistent homology를 추정합니다.

- ▶ 목표 persistent homology를 근사하기 위해:

$$PH_*^{\text{supp}(P)}(p_h),$$

Rips complex 위에서의 핵밀도추정 persistent homology를 사용합니다.

$$PH_*^R(\hat{p}_h, r).$$



군집 나무(Cluster Tree)와 이를 통계적으로 추정하기

Persistent Homology를 통계적으로 추정하기

Persistent Homology의 신뢰띠를 자료 위에서 계산하기

Persistent Homology의 신뢰띠를 자료 위에서 계산하기

Persistent Homology를 이용하여 특성(Feature) 만들기

R 패키지 TDA: 위상학적 자료 분석을 위한 통계 계산 도구

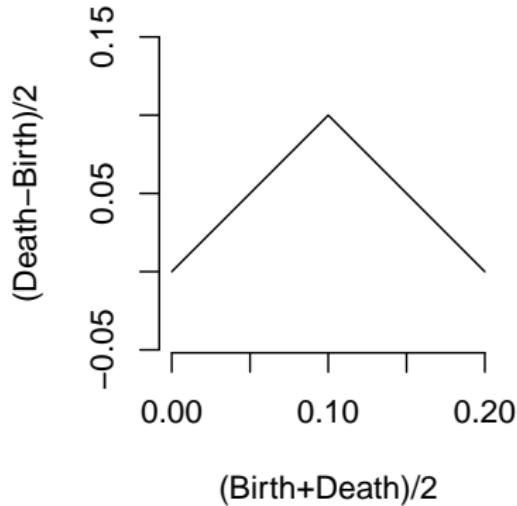
다양체(manifold)에서의 표본 추출, 거리 함수, 밀도 함수

Persistent Homology와 Persistence Landscape

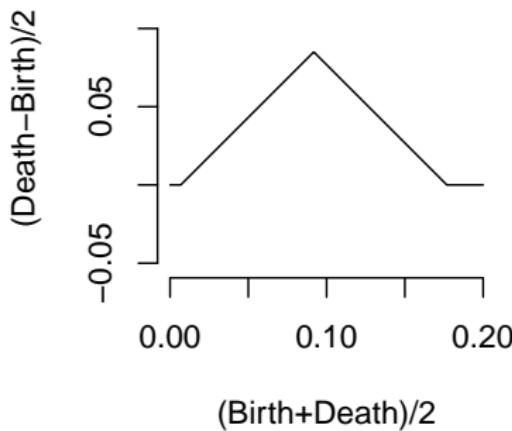
Persistence Homology와 Persistence Landscape의 통계적 추정

유한한 자료의 Persistence Landscape으로부터 기저 구조의 Persistence Landscape를 추정할 수 있습니다.

Circle

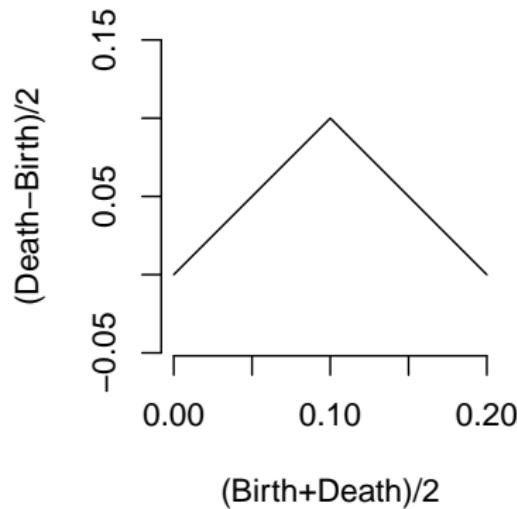


200 samples

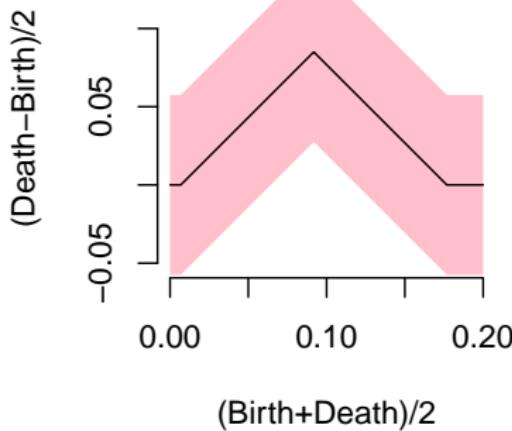


Persistent homology의 신뢰띠로 Persistence Landscape의 랜덤성을 정량화할 수 있습니다.

Circle



200 samples

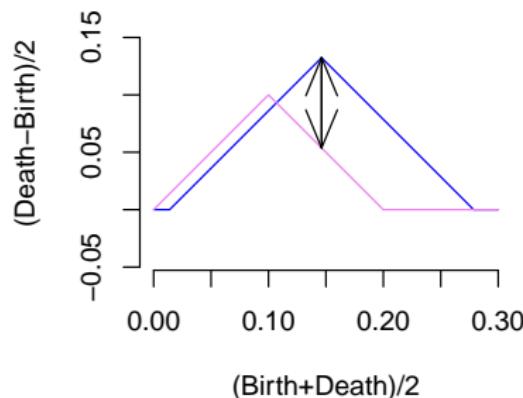


∞ -landscape 거리는 persistence landscape 공간에 거리를 줍니다.

Definition

[?] D_1, D_2 을 점들의 multiset이라 하고, 그에 해당하는 persistence landscape를 λ_1, λ_2 라고 놓습니다. ∞ -landscape 거리는 다음과 같이 정의합니다:

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



∞ -landscape 거리는 그에 대응되는 함수 간의 거리로 조정할 수 있습니다: 안정성 정리(stability theorem).

Theorem

$f, g : \mathbb{X} \rightarrow \mathbb{R}$ 를 두 함수로 놓고, 그에 해당하는 persistence landscape를 $\lambda(f)$ 과 $\lambda(g)$ 로 놓습니다. 그러면,

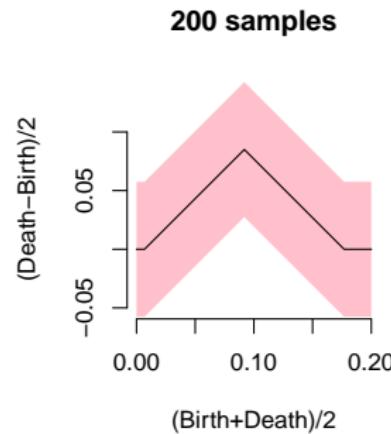
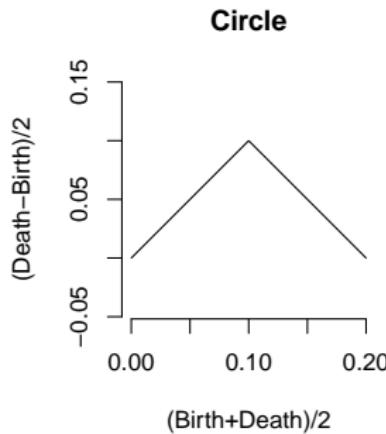
$$\Lambda_\infty(\lambda(f), \lambda(g)) \leq \|f - g\|_\infty.$$

persistence landscape의 신뢰띠는 븋스트랩으로 계산할 수 있습니다.

- ▶ 기저 M 과 표본 X 의 persistence landscape를 각각 λ_M 과 λ_X 로 놓습니다. 안정성 정리(stability theorem)으로부터, $\mathbb{P}(\|f_M - f_X\| \leq c_n) \geq 1 - \alpha$ 는 다음을 유도합니다:

$$\mathbb{P}(\lambda_X(t) - c_n \leq \lambda_M(t) \leq \lambda_X(t) + c_n \forall t) \geq \mathbb{P}(\|f_M - f_X\| \leq c_n) \geq 1 - \alpha,$$

따라서 대응되는 함수인 f_M 의 신뢰띠를 persistence landscape λ_M 의 신뢰띠로 사용할 수 있습니다.



persistence landscape의 신뢰띠는 븁스트랩으로 계산할 수 있습니다.

- ▶ persistence landscape의 신뢰띠는 multiplier bootstrap으로도 계산할 수 있습니다; [?].

군집 나무(Cluster Tree)와 이를 통계적으로 추정하기

Persistent Homology를 통계적으로 추정하기

Persistent Homology의 신뢰띠를 자료 위에서 계산하기

Persistent Homology의 신뢰띠를 자료 위에서 계산하기

Persistent Homology를 이용하여 특성(Feature) 만들기

R 패키지 TDA: 위상학적 자료 분석을 위한 통계 계산 도구

다양체(manifold)에서의 표본 추출, 거리 함수, 밀도 함수

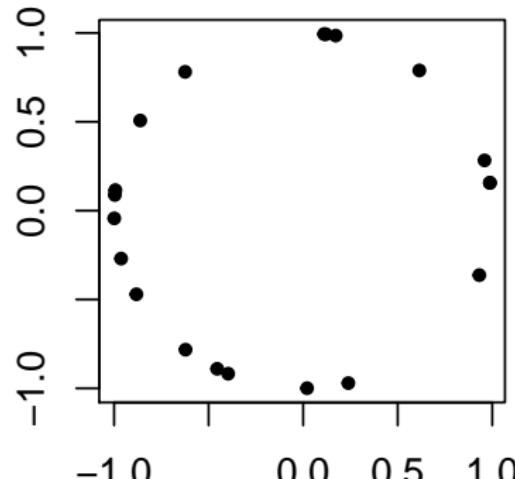
Persistent Homology와 Persistence Landscape

Persistence Homology와 Persistence Landscape의 통계적 추정

R 패키지 TDA는 원 위에서 표본 추출할 수 있는 함수를 제공합니다.

The function `circleUnif()` generates n sample from the uniform distribution on the circle in \mathbb{R}^2 with radius r .

```
circleSample <- circleUnif(n = 20, r = 1)
plot(circleSample, xlab = "", ylab = "", pch = 20)
```



R 패키지 TDA는 격자 위에서의 거리 함수와 밀도 함수를 제공합니다.

Suppose $n = 400$ points are generated from the unit circle, and grid of points are generated.

```
X <- circleUnif(n = 400, r = 1)

lim <- c(-1.7, 1.7)
by <- 0.05
margin <- seq(from = lim[1], to = lim[2], by = by)
Grid <- expand.grid(margin, margin)
```

R 패키지 TDA는 격자 위에서의 핵밀도추정(KDE)을 제공합니다.

The Gaussian Kernel Density Estimator (KDE) $\hat{p}_h : \mathbb{R}^d \rightarrow [0, \infty)$ is defined as

$$\hat{p}_h(y) = \frac{1}{n(\sqrt{2\pi}h)^d} \sum_{i=1}^n \exp\left(\frac{-\|y - x_i\|_2^2}{2h^2}\right),$$

where h is a smoothing parameter.

The function `kde()` computes the KDE function \hat{p}_h on a grid of points.

```
h <- 0.3
KDE <- kde(X = X, Grid = Grid, h = h)

par(mfrow = c(1,2))
plot(X, xlab = "", ylab = "", main = "Sample X", pch = 20)
persp(x = margin, y = margin,
      z = matrix(KDE, nrow = length(margin), ncol = length(margin)),
      xlab = "", ylab = "", zlab = "", theta = -20, phi = 35, scale = FALSE,
      expand = 3, col = "red", border = NA, ltheta = 50, shade = 0.5,
      main = "KDE")
```

R 패키지 TDA는 격자 위에서의 핵밀도추정(KDE)을 제공합니다.

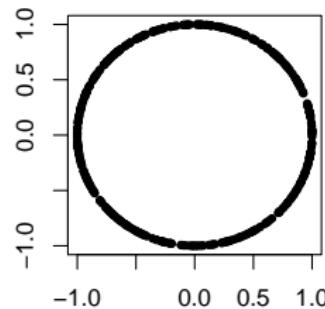
The Gaussian Kernel Density Estimator (KDE) $\hat{p}_h : \mathbb{R}^d \rightarrow [0, \infty)$ is defined as

$$\hat{p}_h(y) = \frac{1}{n(\sqrt{2\pi}h)^d} \sum_{i=1}^n \exp\left(\frac{-\|y - x_i\|_2^2}{2h^2}\right),$$

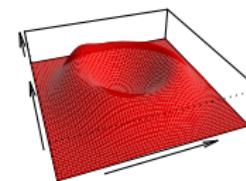
where h is a smoothing parameter.

The function `kde()` computes the KDE function \hat{p}_h on a grid of points.

Sample X



KDE



R 패키지 TDA는 격자 위에서의 Persistent Homology를 계산합니다.

- ▶ The function `gridDiag()` computes the persistence diagram of sublevel (and superlevel) sets of the input function.
 - ▶ `gridDiag()` evaluates the real valued input function over a grid.
 - ▶ `gridDiag()` constructs a filtration of simplices using the values of the input function.
 - ▶ `gridDiag()` computes the persistent homology of the filtration.
- ▶ The user can choose to compute persistent homology using either C++ library GUDHI, Dionysus, or PHAT.

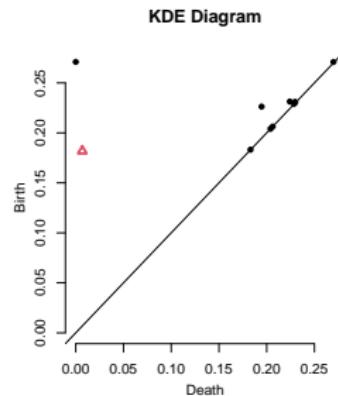
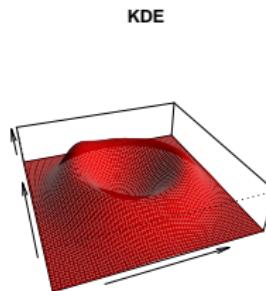
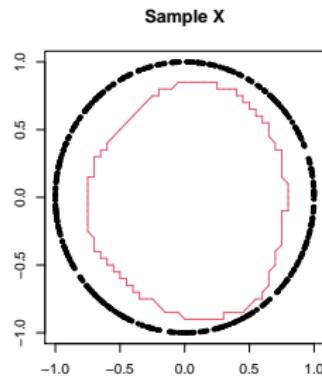
R 패키지 TDA는 격자 위에서의 Persistent Homology를 계산합니다.

```
DiagGrid <- gridDiag(X = X, FUN = kde, lim = c(lim, lim), by = by,
  sublevel = FALSE, library = "Dionysus", location = TRUE,
  printProgress = FALSE, h = h)

par(mfrow = c(1,3))
plot(X, xlab = "", ylab = "", main = "Sample X", pch = 20)
one <- which(DiagGrid[["diagram"]][, 1] == 1)
for (i in seq(along = one)) {
  for (j in seq_len(dim(DiagGrid[["cycleLocation"]][[one[i]]])[1])) {
    lines(DiagGrid[["cycleLocation"]][[one[i]]][j, , ], pch = 19, cex = 1,
      col = i + 1)
  }
}
persp(x = margin, y = margin,
  z = matrix(KDE, nrow = length(margin), ncol = length(margin)),
  xlab = "", ylab = "", zlab = "", theta = -20, phi = 35, scale = FALSE,
  expand = 3, col = "red", border = NA, ltheta = 50, shade = 0.9,
  main = "KDE")
plot(x = DiagGrid[["diagram"]], main = "KDE Diagram")
```

R 패키지 TDA는 격자 위에서의 Persistent Homology를 계산합니다.

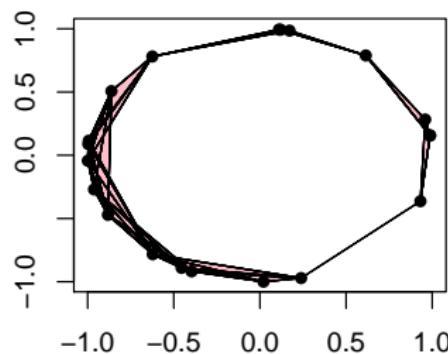
- ▶ The function `gridDiag()` computes the persistent homology of sublevel (and superlevel) sets of the input function.
 - ▶ `gridDiag()` evaluates the real valued input function over a grid.
 - ▶ `gridDiag()` constructs a filtration of simplices using the values of the input function.
 - ▶ `gridDiag()` computes the persistent homology of the filtration.
- ▶ The user can choose to compute persistent homology using either GUDHI, Dionysus, or PHAT.



R 패키지 TDA는 Vietoris-Rips Persistent Homology를 계산합니다.

- ▶ Vietoris-Rips complex consists of simplices whose pairwise distances of vertices are at most ϵ apart, i.e.

$$R(X, \epsilon) = \{[X_{n_1}, \dots, X_{n_r}] : d(X_{n_i}, X_{n_j}) \leq \epsilon\}.$$



- ▶ Rips filtration is formed by Rips complexes with gradually increasing ϵ .

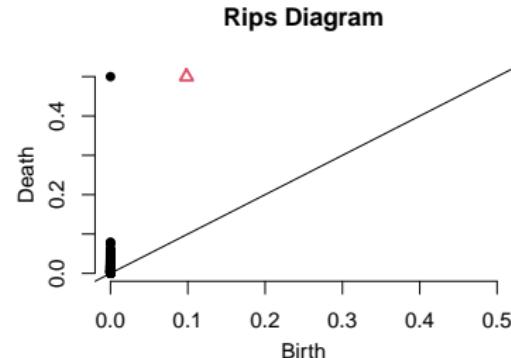
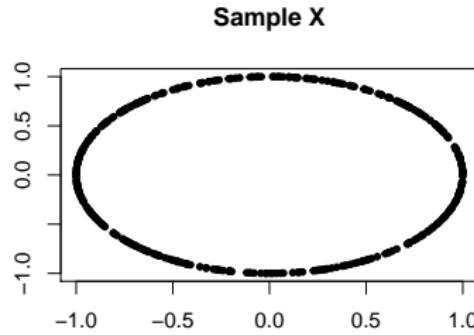
R 패키지 TDA는 Vietoris-Rips Persistent Homology를 계산합니다.

- ▶ The function `ripsDiag()` computes the persistence diagram of the Rips filtration built on top of a point cloud.
 - ▶ `ripsDiag()` constructs the Vietoris-Rips filtration using the data points.
 - ▶ `ripsDiag()` computes the persistent homology of the Vietoris-Rips filtration.
- ▶ The user can choose to compute persistent homology using either C++ library GUDHI, Dionysus, or PHAT.

```
DiagRips <- ripsDiag(X = X, maxdimension = 1, maxscale = 0.5,  
library = c("GUDHI", "Dionysus"), location = TRUE)  
  
par(mfrow = c(1,2))  
plot(X, xlab = "", ylab = "", main = "Sample X", pch = 20)  
plot(x = DiagRips[["diagram"]], main = "Rips Diagram")
```

R 패키지 TDA는 Vietoris-Rips Persistent Homology를 계산합니다.

- ▶ The function `ripsDiag()` computes the persistence diagram of the Rips filtration built on top of a point cloud.
 - ▶ `ripsDiag()` constructs the Vietoris-Rips filtration using the data points.
 - ▶ `ripsDiag()` computes the persistent homology of the Vietoris-Rips filtration.
- ▶ The user can choose to compute persistent homology using either C++ library GUDHI, Dionysus, or PHAT.



R 패키지 TDA는 Persistence Landscape를 계산합니다.

- ▶ Let Λ_p be created by tenting each point $p = (x, y) = (\frac{b+d}{2}, \frac{d-b}{2})$ representing a birth-death pair (b, d) in the persistence diagram D .
- ▶ The persistence landscape of D is the collection of functions

$$\lambda_k(t) = k \max_p \Lambda_p(t), \quad t \in [0, T], k \in \mathbb{N},$$

where $k \max$ is the k th largest value in the set.

- ▶ The function `landscape()` evaluates the persistence landscape function $\lambda_k(t)$.

```
tseq <- seq(0, 0.2, length = 1000)
Land <- landscape(DiagGrid[["diagram"]], dimension = 1, KK = 1, tseq = tseq)

par(mfrow = c(1,2))
plot(x = DiagGrid[["diagram"]], main = "KDE Diagram")
plot(tseq, Land, type = "l", xlab = "(Birth+Death)/2",
      ylab = "(Death-Birth)/2", asp = 1, axes = FALSE, main = "Landscape")
axis(1); axis(2)
```

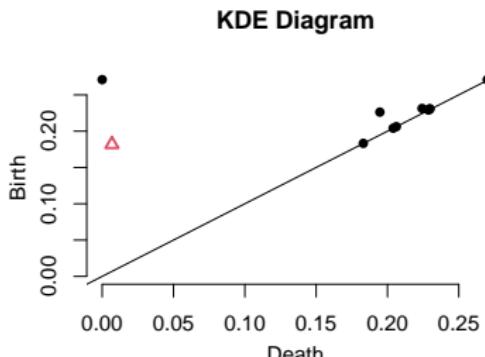
R 패키지 TDA는 Persistence Landscape를 계산합니다.

- ▶ Let Λ_p be created by tenting each point $p = (x, y) = (\frac{b+d}{2}, \frac{d-b}{2})$ representing a birth-death pair (b, d) in the persistence diagram D .
- ▶ The persistence landscape of D is the collection of functions

$$\lambda_k(t) = k \max_p \Lambda_p(t), \quad t \in [0, T], k \in \mathbb{N},$$

where $k \max$ is the k th largest value in the set.

- ▶ The function `landscape()` evaluates the persistence landscape function $\lambda_k(t)$.



R 패키지 TDA는 함수의 븋스트랩 신뢰띠를 계산합니다.

The function `bootstrapBand()` computes $(1 - \alpha)$ bootstrap confidence band for $\mathbb{E}[\hat{p}_h]$.

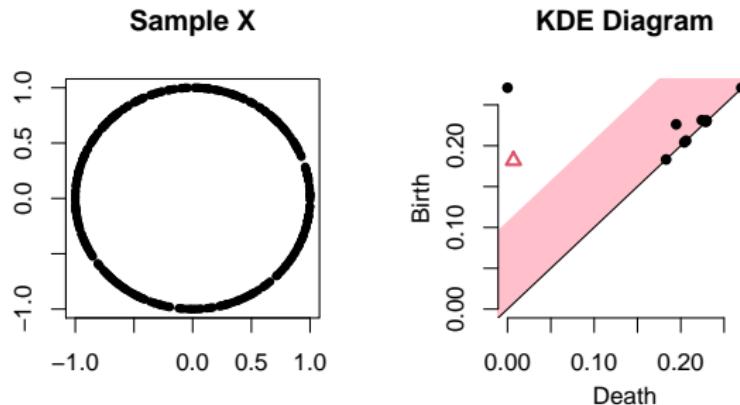
```
bandKDE <- bootstrapBand(X = X, FUN = kde, Grid = Grid, B = 20,
    parallel = FALSE, alpha = 0.1, h = h)
print(bandKDE[["width"]])

##         90%
## 0.0537502
```

R 패키지 TDA는 persistent homology의 블스트랩 신뢰띠를 계산합니다.

The $(1 - \alpha)$ bootstrap confidence band for $\mathbb{E}[\hat{p}_h]$ is used as the confidence band for the persistent homology.

```
par(mfrow = c(1,2))
plot(X, xlab = "", ylab = "", main = "Sample X", pch = 20)
plot(x = DiagGrid[["diagram"]], band = 2 * bandKDE[["width"]],
     main = "KDE Diagram")
```



R 패키지 TDA는 persistence landscape의 블스트랩 신뢰띠를 계산합니다.

The $(1 - \alpha)$ bootstrap confidence band for $\mathbb{E}[\hat{p}_h]$ is used as the confidence band for the persistence landscape.

```
par(mfrow = c(1,2))
plot(X, xlab = "", ylab = "", main = "Sample X", pch = 20)
plot(tseq, Land, type = "l", xlab = "(Birth+Death)/2",
      ylab = "(Death-Birth)/2", asp = 1, axes = FALSE, main = "200 samples")
axis(1); axis(2)
polygon(c(tseq, rev(tseq)), c(Land - bandKDE[["width"]],
      rev(Land + bandKDE[["width"]])), col = "pink", lwd = 1.5,
      border = NA)
lines(tseq, Land)
```

R 패키지 TDA는 persistence landscape의 블스트랩 신뢰띠를 계산합니다.

The $(1 - \alpha)$ bootstrap confidence band for $\mathbb{E}[\hat{p}_h]$ is used as the confidence band for the persistence landscape.

