Topological Data Analysis (Spring 2018)
Persistence Landscapes

Instructor: Mehmet Aktas

March 27, 2018
This presentation is inspired by the following paper and presentation from the same author:

- Peter Bubenik, Statistical Topological Data Analysis Using Persistence Landscapes, 2015.
- Peter Bubenik, Topology for Data Science, 2017.
Outline

1. Introduction
2. Persistence Landscape
3. Mathematical Background
4. Examples
Outline

1. Introduction
2. Persistence Landscape
3. Mathematical Background
4. Examples
TDA Summary

1. Raw Data
2. Preprocess
3. Clean data
4. Filtered simplicial complex
5. Transform
6. Homology
7. Persistence module
8. Topological summary
9. Statistics and Machine Learning
The standard paradigm for TDA is as follows:

- Take a data set as a finite set of points in $\mathbb{R}^n$ or more generally in some metric space
- Apply some geometric construction to which one applies tools from algebraic topology
- The end result is a topological summary of the data.
- The standard topological descriptors are the **barcode** and the **persistence diagram**
- **Question**: How can one combine the main tool of the subject, the barcode or persistence diagram with statistics and machine learning?
Statistics with barcodes/persistence diagrams

Statistics and machine learning tasks

- clustering
- certain hypothesis test
- calculating averages
- understanding variances
- classification
One solution: Persistence Landscapes

One way to turn a barcode or persistence diagram into a vector is the persistence landscape.
Advantages:
- it does not lose information
- it is stable
- it has a discrete and a continuous version
Outline

1. Introduction
2. Persistence Landscape
3. Mathematical Background
4. Examples
Persistence landscape from a barcode

Replace

with
Persistence landscape from a barcode

Barcode:

Persistence Landscape:

\[ \lambda_k = 0, \quad \text{for } k \geq 4 \]
Persistence landscape from a persistence diagram
Persistence landscape from a persistence diagram
Persistence landscape from a persistence diagram
Barcodes and persistence diagrams

Time | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9
---|---|---|---|---|---|---|---|---|---|---
Betti number effect | $\beta_0$ | +

Birth–Death pairs for $H_0$:

Birth–Death pairs for $H_1$:
### Barcodes and persistence diagrams

**Persistence Landscape**

- **Time** | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9
---|---|---|---|---|---|---|---|---|---|---
**Betti number effect** | $\beta_0$ | $\beta_0$ | $\beta_0$ | $\beta_0$ | $\beta_0$ | $\beta_0$ | $\beta_0$ | $\beta_1$ | $\beta_1$ | $\beta_1$

| Birth–Death pairs for $H_0$: | $(0, \infty)$, $(1, 3)$, $(2, 6)$, $(4, 5)$ |
| Birth–Death pairs for $H_1$: | $(7, \infty)$, $(8, 9)$ |
Barcodes and persistence diagrams

Birth–Death pairs for $H_0$: (0, $\infty$), (1, 3), (2, 6), (4, 5)
Birth–Death pairs for $H_1$: (7, $\infty$), (8, 9)

Barcode

$H_1$

$H_0$
Exercise 2: Graphing the persistence landscape

Graph the persistence landscape of $H_0$
Graph the persistence landscape of $H_0$
Graph the persistence landscape of $H_0$
Persistence Landscape as a Vector

\[
\lambda_1 \quad 0 \quad 1 \quad 2 \quad 3 \quad 4 \quad 3 \quad 2 \quad 3 \quad 2 \quad 1 \quad 0 \\
\lambda_2 \quad 0 \quad 0 \quad 0 \quad 1 \quad 2 \quad 1 \quad 2 \quad 1 \quad 0 \quad 0 \quad 0 \\
\lambda_3 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0 \\

(0, 1, 2, 3, 4, 3, 2, 3, 2, 1, 0, 0, 0, 0, 1, 2, 1, 2, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0)
\]
Outline

1. Introduction
2. Persistence Landscape
3. Mathematical Background
4. Examples
Let $M$ be a *persistence module* i.e. it consists of a vector space $M_a$ for all $a \in \mathbb{R}$ and linear maps $M(a \leq b) : M_a \to M_b$ for all $a \leq b$ s.t. $M(a \leq a)$ is the identity map and for all $a \leq b \leq c$, 

$$M(b \leq c) \circ M(a \leq b) = M(a \leq c).$$

For $a \leq b$, the corresponding Betti number of $M$ is given by the dimension of the image of the corresponding linear map. That is,

$$\beta^{a,b} = \dim(\text{im}(M(a \leq b))).$$
Let $M$ be a persistence module i.e. it consists of a vector space $M_a$ for all $a \in \mathbb{R}$ and linear maps $M(a \leq b) : M_a \to M_b$ for all $a \leq b$ s.t. $M(a \leq a)$ is the identity map and for all $a \leq b \leq c$, $M(b \leq c) \circ M(a \leq b) = M(a \leq c)$.

For $a \leq b$, the corresponding Betti number of $M$ is given by the dimension of the image of the corresponding linear map. That is,

$$\beta^{a,b} = \dim(\text{im}(M(a \leq b))).$$
The Rank Function

Definition

The *rank function* is the function \( \lambda : \mathbb{R}^2 \to \mathbb{R} \) given by

\[
\lambda(b, d) = \begin{cases} 
\beta^{b,d}, & \text{if } b \leq d \\
0, & \text{otherwise}
\end{cases}
\]

Definition

The *rescaled rank function* is the function \( \lambda : \mathbb{R}^2 \to \mathbb{R} \) given by

\[
\lambda(m, h) = \begin{cases} 
\beta^{m-h,m+h}, & \text{if } h \leq 0 \\
0, & \text{otherwise}
\end{cases}
\]

with \( m = \frac{b+d}{2} \), and \( h = \frac{d-b}{2} \).
The Rank Function

Definition

The rank function is the function $\lambda : \mathbb{R}^2 \rightarrow \mathbb{R}$ given by

$$\lambda(b, d) = \begin{cases} \beta^{b,d}, & \text{if } b \leq d \\ 0, & \text{otherwise} \end{cases}$$

Definition

The rescaled rank function is the function $\lambda : \mathbb{R}^2 \rightarrow \mathbb{R}$ given by

$$\lambda(m, h) = \begin{cases} \beta^{m-h, m+h}, & \text{if } h \leq 0 \\ 0, & \text{otherwise} \end{cases}$$

with $m = \frac{b+d}{2}$, and $h = \frac{d-b}{2}$.
Definition

The *persistence landscape* is a sequence of functions $\lambda_k : \mathbb{R} \to \bar{\mathbb{R}}$ where $\lambda_k(t) = \lambda(k, t)$.

Lemma

- $\lambda_k(t) \geq 0$,
- $\lambda_k(t) \geq \lambda_{k+1}(t)$,
- $\lambda_k$ is 1-Lipshitz, i.e. $|\lambda_k(t) - \lambda_k(s)| \geq |t - s|$.

Landscape Stability Theorem

Let $f, g : X \to \mathbb{R}$ be functions. Then

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

The *persistence landscape* is a sequence of functions $\lambda_k : \mathbb{R} \rightarrow \mathbb{R}$ where $\lambda_k(t) = \lambda(k, t)$.

Lemma

- $\lambda_k(t) \geq 0$,
- $\lambda_k(t) \geq \lambda_{k+1}(t)$,
- $\lambda_k$ is 1-Lipshitz, i.e. $|\lambda_k(t) - \lambda_k(s)| \geq |t - s|$.

Landscape Stability Theorem

Let $f, g : X \rightarrow \mathbb{R}$ be functions. Then

$$\|\lambda(f) - \lambda(g)\|_\infty \leq \|f - g\|_\infty.$$
Persisteance Landscape

Definition

The *persistence landscape* is a sequence of functions $\lambda_k : \mathbb{R} \to \bar{\mathbb{R}}$ where $\lambda_k(t) = \lambda(k, t)$.

Lemma

- $\lambda_k(t) \geq 0$,
- $\lambda_k(t) \geq \lambda_{k+1}(t)$,
- $\lambda_k$ is 1-Lipschitz, i.e. $|\lambda_k(t) - \lambda_k(s)| \geq |t - s|$.

Landscape Stability Theorem

Let $f, g : X \to \mathbb{R}$ be functions. Then

$$\|\lambda(f) - \lambda(g)\|_\infty \leq \|f - g\|_\infty.$$
Average landscapes

Persistence landscapes, $\lambda^{(1)}, \ldots, \lambda^{(n)}$, have a pointwise average,

$$\bar{\lambda}(k, t) = \frac{1}{n} \sum_{i=1}^{n} \lambda^{(i)}(k, t)$$
Average diagram vs average landscape

Mathematical Background

Persistence Landscapes
Outline

1. Introduction
2. Persistence Landscape
3. Mathematical Background
4. Examples
Linked Annuli

Examples
Torus and Sphere
Examples

Torus and Sphere