CS 6170: Computational Topology, Spring 2019
Lecture 16

Topological Data Analysis for Data Scientists

Dr. Bei Wang

School of Computing
Scientific Computing and Imaging Institute (SCI)
University of Utah
www.sci.utah.edu/~beiwang
beiwang@sci.utah.edu

Feb 28, 2019

Review on Kernel Methods

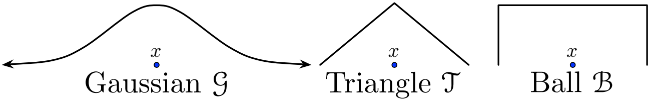
An information introduction to kernel

- Informally, a *kernel* K is a (nonnegative) similarity measure between a pair of points in $\mathbb{R}^d$, where more similar points have higher value:

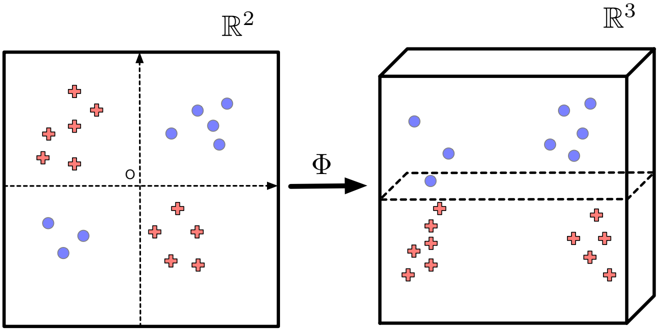
$$K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^+.$$

- Example: Gaussian kernel (positive definite)

$$K(p, x) = \exp(-||p - x||^2 / 2\sigma^2)$$



Kernels and feature space: the kernel trick

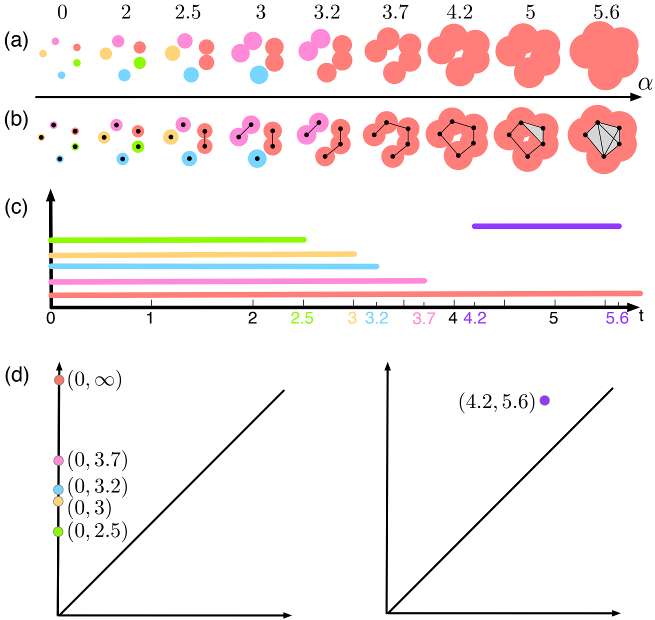


- There is no linear classifier in $\mathbb{R}^2$
- Map points to a higher dimensional feature space such that there is a linear classifier.
- $\Phi(x) = [x_1 x_2 x_1 x_2] \in \mathbb{R}^3$

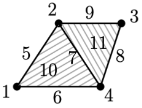
Stability of Persistence Diagrams

Edelsbrunner and Harer (2010), C.VIII

Barcode and persistence diagram



Review: computing persistent homology



$\delta =$

	1	2	3	4	5	6	7	8	9	10	11
1					1	1					
2					1		1		1		
3								1	1		
4						1	1	1			
5										1	
6										1	
7										1	1
8											1
9											1
10											
11											

$\Rightarrow$

	1	2	3	4	5	6	7(+6+5)	8(+6)	9(+8+5)	10	11
1					1	1	~~1~~	**1**	~~1~~		
2					[1]		~~1~~		~~1~~		
3								[1]	~~1~~		
4						[1]	~~1~~	~~1~~			
5										1	
6										1	
7										[1]	1
8											1
9											[1]
10											
11											

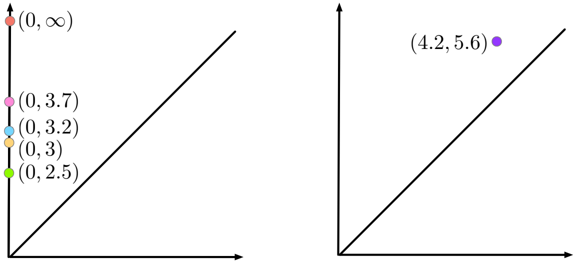
Now: imagine perturbing the function values at some particular simplifies: from 7 to 7.5 and from 11 to 10.5. How will the persistence diagram change?

Persistence diagram

Definition

A *persistence diagram* is a finite multisite of points in the extended plane. To simplify the definitions and results, we add infinitely many points on the diagonal, each with infinite multiplicity.

- The extended plane: $\bar{\mathbb{R}}^2 = (\mathbb{R} \cup \pm\infty)^2$.
- Let δ denote diagonal of the extended plane.
- For example see persistence diagrams in dimension 0 and dimension 1 below (left, Dgm_0; right Dgm_1).



Bottleneck distance

- Let X, Y be two persistence diagrams
- L_∞ norm: let $x = (x_1, x_2) \in X$ and $y = (y_1, y_2) \in Y$, then

$$||x - y||_\infty := \max\{|x_1 - y_1|, |x_2 - y_2|\}.$$

- Let $\eta : X \to Y$ be a bijection.

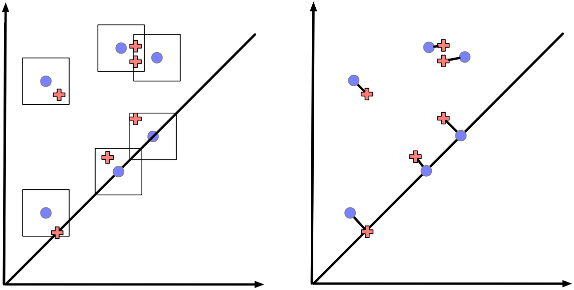
Definition

The *bottleneck distance* between two persistence diagrams X and Y is defined to be

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

Bottleneck distance: an illustration

- Let $\epsilon = W_\infty(X, Y)$
- Draw squares centered at $x \in Y$ with 2ϵ sides such that each square contains its corresponding point $\eta(x) \in Y$.



Bottleneck Stability for Filtrations

- K: simplicial complex
- $f : K \to \mathbb{R}$ is a *monotonic function* on K, if for every $\sigma < \tau$, $f(\sigma) \leq f(\tau)$.
- Let $f, g : K \to \mathbb{R}$ be two monotonic functions, define

$$||f - g||_\infty = \sup_{\sigma \in K} |f(v) - g(v)|.$$

Theorem (Stability Theorem for Filtrations)

Let K be a simplicial complex and $f, g : K \to \mathbb{R}$ two monotonic functions. For each dimension p, the bottleneck distance between the diagrams $X = \mathrm{Dgm}_p(f)$ and $Y = \mathrm{Dgm}_p(g)$ is bounded from above by the L_∞ distance between the functions (Edelsbrunner and Harer, 2010, Page 182), that is,

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

References I

Edelsbrunner, H. and Harer, J. (2010). *Computational Topology: An Introduction*. American Mathematical Society, Providence, RI, USA.