

Variably Scaled Persistence Kernels (VSPKs) for persistent homology applications

S. De Marchi*, F. Lot**, F. Marchetti*, D. Poggiali***

**Dipartimento di Matematica “Tullio Levi-Civita”, Università di Padova, 35121 Padova, Italy;*

***Fachbereich Mathematik und Informatik, Philipps-Universität Marburg, 35032 Marburg, Germany;*

****FAR Networks srl, 20063 Cernusco S.N., Milan, Italy*

Abstract

In recent years, various kernels have been proposed in the context of *persistent homology* to deal with *persistence diagrams* in supervised learning approaches. In this paper, we consider the idea of variably scaled kernels, for approximating functions and data, and we interpret them in the framework of persistent homology. We call them *Variably Scaled Persistence Kernels (VSPKs)*. These new kernels are then tested in different classification experiments. The obtained results show that they can improve the performance and the efficiency of existing standard kernels.

Keywords: kernel-based learning; variably scaled persistence kernel; persistence diagrams; persistent homology

2020 MSC: : 41A05, 55N31, 65D05, 65D15.

1. Introduction

Let $\Omega \subset \mathbb{R}^v$ and let $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\} \subset \Omega$ be a set of input data, $v, n \in \mathbb{N}$. Assume that each element $\mathbf{x}_i$ is uniquely associated to a *label* (or *class*) $y_i \in \mathcal{T}$, with $\mathcal{T} = \{c_1, \dots, c_t\}$, so that we can define the dataset $\mathcal{Z} = \{(\mathbf{x}_i, y_i) \mid \mathbf{x}_i \in \mathcal{X}, y_i \in \mathcal{T}\}$. The supervised learning task consists in finding a decision function $s : \Omega \rightarrow \mathcal{T}$ such that:

- (i) it models the input-output relation in $\mathcal{Z}$;
- (ii) it models the input-output relation of unseen labeled instances $\{(\xi_i, y_i)\}_{i=1, \dots, m} \subset (\Omega \setminus \mathcal{X}) \times \mathcal{T}$, $m \geq 1$.

The *generalization* capability required in (ii) is fundamental in the learning problem, since it is trivial to find a decision function that satisfies (i); for an

Email addresses: demarchi@math.unipd.it (S. De Marchi*), lotf@mathematik.uni-marburg.de (F. Lot**), francesco.marchetti@math.unipd.it (F. Marchetti*), dr.davide.poggiali@gmail.com (D. Poggiali***)

introduction concerning (statistical) learning theory, refer to e.g. [1, §1.3] and [2].

Kernel methods are well-established tools in supervised machine learning, as well as in a variety of research and applied fields [3, 4]. The flexibility provided by kernel-based schemes allows the handling of different kinds of possible structured data, e.g. graphs and words, which are encoded in some dot product space where even complex patterns may be distinguished [5, 6, 7, 8].

In the following, our data consist of *persistence diagrams*, which represent an output of the so-called *persistent homology*, in the framework of Topological Data Analysis (TDA). The usage of topological methods and analysis techniques to extract significant features and patterns from data is receiving more and more interest and, in particular, persistent homology captures the evolution of topological features in the data; for a complete overview concerning TDA, see e.g. [9, 10, 11]. In the last years, several kernels specifically devoted to deal with the peculiar structure of persistence diagrams have been proposed [12, 13, 14, 15], therefore the construction of suitable kernels is a very active research line in this context.

Here, we introduce what we call the Variably Scaled Persistence Kernels (VSPKs), which are inspired by the variably scaled kernels introduced in the context of approximation theory in [16] and employed as a feature augmentation strategy in [17, 18] for kernel-based learning. Indeed, our aim is to build a bridge between classical variably scaled kernels and kernels defined in the context of persistent homology. After defining VSPKs, we design some scaling functions and we test the performance of the resulting kernels by experimenting with three different datasets. The obtained results show that the variably scaled setting may yields to better classification outcomes with respect to the standard setting, and thus can be considered for further investigations and applications involving persistence diagrams.

The paper is organized as follows. In Sections 2 and 3, we introduce kernel-based learning with Support Vector Machines (SVMs) and we recall some notions concerning persistent homology, also presenting some kernels of recent introduction that are dedicated to persistent diagrams. VSPKs are proposed and analysed in Section 4, and related numerical experiments are exposed in Section 5. Finally, in Section 6 we present some conclusions and final remarks.

2. Positive definite and variably scaled kernels

Let $\kappa : \Omega \times \Omega \longrightarrow \mathbb{R}$ be a kernel. Given a set of data $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\} \subset \Omega$, the $n \times n$ matrix K with elements $K_{ij} := \kappa(\mathbf{x}_i, \mathbf{x}_j)$, $i, j = 1, \dots, n$, is the *Gram matrix* of the kernel κ with respect to $\mathcal{X}$. If κ is positive definite (strictly positive definite) on $\Omega \times \Omega$, i.e., K is positive semi-definite (definite) for all possible datasets in Ω , then it is possible to decompose the kernel according to Mercer theorem [19], and to interpret such decomposition as an inner product in a *Reproducing Kernel Hilbert Space* (RKHS) $\mathcal{H}$. Indeed, there exists a (non-

unique) *feature map* $\Phi : \Omega \longrightarrow \mathcal{F}$ such that

$$\kappa(\mathbf{x}, \mathbf{y}) = \langle \Phi(\mathbf{x}), \Phi(\mathbf{y}) \rangle_{\mathcal{F}} \quad \mathbf{x}, \mathbf{y} \in \Omega,$$

being $\langle \cdot, \cdot \rangle_{\mathcal{F}}$ the bilinear form related to the RKHS (e.g. $\Phi(\mathbf{x}) = \kappa(\cdot, \mathbf{x})$). Moreover, the kernel κ induces a distance d_κ on Ω

$$d_\kappa(\mathbf{x}, \mathbf{y}) := \kappa(\mathbf{x}, \mathbf{x}) + \kappa(\mathbf{y}, \mathbf{y}) - 2\kappa(\mathbf{x}, \mathbf{y}). \quad (1)$$

Variably Scaled Kernels (VSKs) have been introduced in [16] in the context of kernel-based approximation, with the aim of overcoming instability issues. Then, they have been extended to work in a more general setting in [17], as presented in the following form. Let $\Lambda \subseteq \mathbb{R}^\nu$, $\nu > 0 \in \mathbb{N}$ and let $\kappa : \Omega \times \Omega \longrightarrow \mathbb{R}$ be a continuous (strictly) positive definite kernel, where $\tilde{\Omega} = \Omega \times \Lambda \subseteq \mathbb{R}^{\nu+\nu}$. Given a *scaling* function $\Psi : \Omega \longrightarrow \Lambda$, a VSK $\kappa_\Psi : \Omega \times \Omega \longrightarrow \mathbb{R}$ is defined as

$$\kappa_\Psi(\mathbf{x}, \mathbf{y}) = \kappa((\mathbf{x}, \Psi(\mathbf{x})), (\mathbf{y}, \Psi(\mathbf{y}))) \quad (2)$$

for $\mathbf{x}, \mathbf{y} \in \Omega$. The function Ψ can be interpreted as a *feature augmentation* map, which adds ν coordinates (features) to the original sample. In this view, the VSK setting has been analysed in [17] as a *stacking* technique, which is capable of enhancing the prediction performances of classical kernel-based classifiers such as, e.g., SVMs.

Letting $\mathbf{x} = (x_1, \dots, x_d)^\top \in \Omega$, we recall that a binary (i.e. $\mathcal{T} = \{-1, +1\}$) SVMs classifier is characterised by the decision function

$$s(\mathbf{x}) = \text{sign}(h(\mathbf{x})) = \text{sign}(\langle \Phi(\mathbf{x}), \mathbf{w} \rangle_{\mathcal{F}} + b),$$

where

$$\mathbf{w} = \sum_{i=1}^n \alpha_i y_i \Phi(\mathbf{x}_i) \in \mathcal{F}.$$

The coefficients $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_n) \in \mathbb{R}^n$ are the solution of the following *soft margin* problem [20, §18, p. 346–347]

$$\begin{cases} \min_{\boldsymbol{\alpha} \in \mathbb{R}^n} \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \kappa(\mathbf{x}_i, \mathbf{x}_j) - \sum_{i=1}^n \alpha_i, \\ \text{s.t. } \sum_{i=1}^n \alpha_i y_i = 0, \\ 0 \leq \alpha_i \leq \zeta, \quad i = 1, \dots, n, \end{cases}$$

where $[0, \zeta]^n$ is the *bounding box*, with $\zeta \in [0, +\infty)$. Usually, a binary SVMs classifier is extended to the multiclass setting by considering a *one-vs-rest* approach.

3. Persistent homology and kernels

3.1. Basics on persistent homology

In the following, we recall some basic ideas about some tools of persistent homology. For a more detailed treatment, especially concerning the algebraic aspects of the construction, we refer e.g. to [10, 21].

Let our data $\mathcal{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\} \subset \Omega$ be interpreted as a set of vertices sampled from some manifold M , and suppose that we wish to highlight some of its intrinsic homological properties. Letting $\varepsilon > 0$, a possible concrete way to proceed consists of studying

$$M^{\mathcal{X}, \varepsilon} = \bigcup_{i=1}^n B(\mathbf{x}_i, \varepsilon)$$

as an approximation of M , where $B(\mathbf{x}_i, \varepsilon)$ is the ball of radius ε and centre $\mathbf{x}_i$. We point out that we can associate to $M^{\mathcal{X}, \varepsilon}$ a *Vietoris-Rips simplicial complex* $K^{\mathcal{X}, \varepsilon}$, where two distinct vertices $\mathbf{x}_i, \mathbf{x}_j$ are connected by an edge if and only if $\|\mathbf{x}_i - \mathbf{x}_j\|_2 \leq \varepsilon$. Formal linear combinations of r -dimensional faces in $K^{\mathcal{X}, \varepsilon}$ form the r -chains group $C_r^{\mathcal{X}, \varepsilon}$. Moreover, letting $[\mathbf{x}_{i_0}, \dots, \mathbf{x}_{i_r}]$ be the face constructed upon distinct vertices $\mathbf{x}_{i_0}, \dots, \mathbf{x}_{i_r}$ in $\mathcal{X}$, $i_0, \dots, i_r \in \{1, \dots, n\}$, we define the linear *boundary operator* $\partial_r : C_r^{\mathcal{X}, \varepsilon} \longrightarrow C_{r-1}^{\mathcal{X}, \varepsilon}$ as

$$\partial_r[\mathbf{x}_{i_0}, \dots, \mathbf{x}_{i_r}] = \sum_{j=0}^r (-1)^j [\mathbf{x}_{i_0}, \dots, \mathbf{x}_{i_{j-1}}, \mathbf{x}_{i_{j+1}}, \dots, \mathbf{x}_{i_r}].$$

The r -cycles and r -boundaries groups are defined as $Z_r^{\mathcal{X}, \varepsilon} = \ker \partial_r$ and $B_r^{\mathcal{X}, \varepsilon} = \text{im } \partial_{r+1}$, respectively. Furthermore, the rank of the r -homology group $H_r^{\mathcal{X}, \varepsilon} = Z_r^{\mathcal{X}, \varepsilon} / B_r^{\mathcal{X}, \varepsilon}$ expresses the concept of r -dimensional holes in $K^{\mathcal{X}, \varepsilon}$.

Finding an optimal ε^* that represents the intrinsic geometric properties of M is a tough and unstable process. Instead, one may analyse the whole *filtration* $\{M^{\mathcal{X}, \varepsilon} \mid \varepsilon > 0\}$. In particular, letting $\varepsilon_1 < \dots < \varepsilon_u$ be increasing real numbers, the nested sequence $K^{\mathcal{X}, \varepsilon_1} \subseteq \dots \subseteq K^{\mathcal{X}, \varepsilon_u}$ is obtained. Then, for $r \geq 0$ and $i \in \{1, \dots, u\}$, we consider the ℓ -persistent homology group

$$H_{r, \ell}^{\mathcal{X}, \varepsilon_i} = Z_r^{\mathcal{X}, \varepsilon_i} / (Z_r^{\mathcal{X}, \varepsilon_i} \cap B_r^{\mathcal{X}, \varepsilon_{i+\ell}}).$$

The group $H_{r, \ell}^{\mathcal{X}, \varepsilon_i}$ contains the homology classes that persist in the *time interval* $[i, i + \ell]$, i.e., they are *born* with $K^{\mathcal{X}, \varepsilon_j}$ for some $j < i$ and they are *alive* with $K^{\mathcal{X}, \varepsilon_{i+\ell}}$. We point out that such homology classes might persist *indefinitely* (we will denote this case as an ∞ level), or they might *die* with a certain ε_j , $i < j \leq u$.

Therefore, each element of the persistent homology groups obtained by considering the whole filtration can be represented by a birth-death pair $(b, d) \in \mathbb{R}_+^2$, $b = \varepsilon_h$, $d = \varepsilon_k$ for some $h \in \{1, \dots, m\}$, $k \in \{1, \dots, m\} \cup \{\infty\}$, $h < k$. We say that a birth-death pair is r -dimensional if it is related to r -dimensional homology groups. Moreover, letting $\varepsilon = (\varepsilon_1, \dots, \varepsilon_u)$, we define a *persistence diagram* $D_r(\mathcal{X}, \varepsilon)$ related to the filtration $K^{\mathcal{X}, \varepsilon_1} \subseteq \dots \subseteq K^{\mathcal{X}, \varepsilon_m}$ as

$$D_r(\mathcal{X}, \varepsilon) = \{(b, d) \mid (b, d) \in P_r(\mathcal{X}, \varepsilon)\} \cup \mathcal{B}, \quad r \geq 0, \quad (3)$$

where $P_r(\mathcal{X}, \varepsilon)$ denotes the set of r -dimensional birth-death pairs obtained with the filtration and $\mathcal{B} = \{(z, z) \mid z \geq 0\}$. We remark that $D_r(\mathcal{X}, \varepsilon)$ is a multiset,

since a couple (b, d) might appear more than once, i.e., might have multiplicity greater than one. Furthermore, the bisector $\mathcal{B}$ is composed by an infinite number of elements characterized by infinite multiplicity, and it is added in order to achieve some *uniformity* among different persistence diagrams and facilitate the formulation of proper metrics, as we present below. In Figure 1, we display the steps and the properties of the discussed analysis by means of a two-dimensional example. We observe that such analysis captures some intrinsic geometrical properties of $\mathcal{X}$. In particular, the unique most persistent 0-dimensional pair, which is significantly far from the others, suggests that a unique connected component underlies $\mathcal{X}$. Moreover, the persistent 1-dimensional pair indicates the presence of a 1-dimensional hole in the structure of the dataset, which is highlighted by the simplicial complex depicted in the bottom left figure.

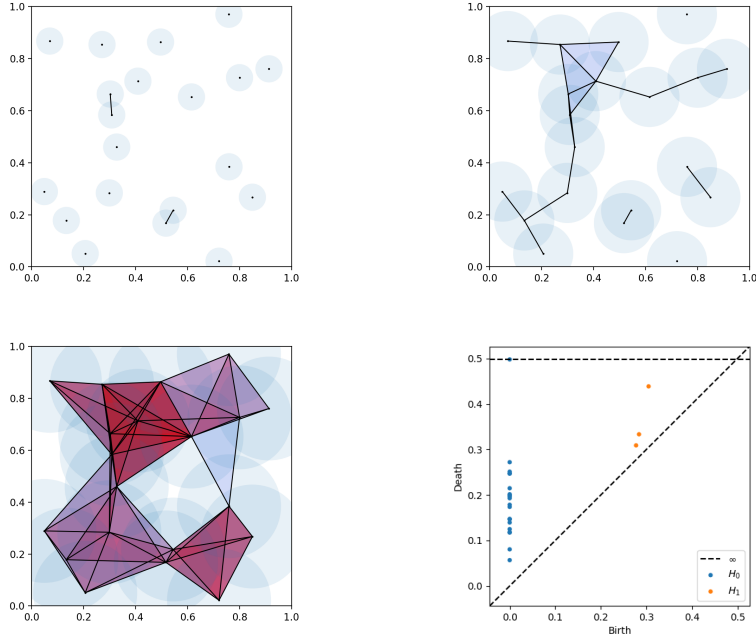


Figure 1: Given a random set of points $\mathcal{X}$, the construction of the Vietoris-Rips complex along the filtration is depicted on top and bottom left figures. The persistence diagrams $D_0(\mathcal{X}, \varepsilon)$ and $D_1(\mathcal{X}, \varepsilon)$ are overlapped in the bottom right figure.

Persistence diagrams show some stability properties with respect to perturbations of the involved dataset [22]. To better clarify this aspect, let us recall some useful metrics. Let $\mathcal{X}, \mathcal{Y} \subset \Omega$ be two non-empty datasets. The *Hausdorff distance* is defined as

$$d_H(\mathcal{X}, \mathcal{Y}) = \max \left\{ \sup_{x \in \mathcal{X}} \inf_{y \in \mathcal{Y}} \|x - y\|_\infty, \sup_{y \in \mathcal{Y}} \inf_{x \in \mathcal{X}} \|x - y\|_\infty \right\}.$$

Moreover, letting $D_r(\mathcal{X}, \epsilon)$ and $D_r(\mathcal{Y}, \epsilon)$ be persistence diagrams for some $r \geq 0$ and filtration radii vector ϵ , we recall the p -Wasserstein distance

$$d_{W,p}(D_r(\mathcal{X}, \epsilon), D_r(\mathcal{Y}, \epsilon)) = \left(\inf_{\gamma \in \Gamma} \sum_{x \in D_r(\mathcal{X}, \epsilon)} \|x - \gamma(x)\|_\infty^p \right)^{\frac{1}{p}},$$

where $\Gamma = \{\gamma : D_r(\mathcal{X}, \epsilon) \rightarrow D_r(\mathcal{Y}, \epsilon) \mid \gamma \text{ is a bijection}\}$. In particular, letting $p \rightarrow \infty$, we obtain the *bottleneck distance*

$$d_{W,\infty}(D_r(\mathcal{X}, \epsilon), D_r(\mathcal{Y}, \epsilon)) = d_B(D_r(\mathcal{X}, \epsilon), D_r(\mathcal{Y}, \epsilon)) = \inf_{\gamma \in \Gamma} \sup_{x \in D_r(\mathcal{X}, \epsilon)} \|x - \gamma(x)\|_\infty.$$

We have the stability result [13]

$$d_B(D_r(\mathcal{X}, \epsilon), D_r(\mathcal{Y}, \epsilon)) \leq d_H(\mathcal{X}, \mathcal{Y}),$$

i.e., the bottleneck distance between persistence diagrams is controlled as long as the underlying datasets are close in the Hausdorff metric.

3.2. Kernels for persistence diagrams

In order to better measure similarities between persistence diagrams, various positive definite kernels that are suitable for dealing with the peculiar structure of persistence diagrams have been introduced and studied in the recent literature. In our applications, we will consider the following ones.

- The **Persistence Scale Space (PSS) kernel** [15]

$$\kappa_\sigma(D_1, D_2) = \frac{1}{8\pi\sigma} \sum_{\substack{\mathbf{y} \in D_1 \\ \mathbf{z} \in D_2}} \exp\left(-\frac{\|\mathbf{y} - \mathbf{z}\|^2}{8\sigma}\right) - \exp\left(-\frac{\|\mathbf{y} - \bar{\mathbf{z}}\|^2}{8\sigma}\right),$$

which is 1-Wasserstein stable, i.e.:

$$\|\kappa_\sigma(D_1, D_2)\|_{L_2(\Omega)} \leq \frac{1}{2\sqrt{\pi}\sigma} d_{W,1}(D_1, D_2)$$

- The **Persistence Weighted Gaussian (PWG) kernel** [13]

$$\kappa_G(D_1, D_2; \kappa, \omega) := \exp\left(-\frac{1}{2\tau^2} \|E_\kappa(\mu_{D_1}^\omega) - E_\kappa(\mu_{D_2}^\omega)\|_{\mathcal{F}}^2\right) \quad \tau > 0,$$

which is built upon a *standard* kernel κ and a weight function ω , where

$$E_\kappa(\mu_{D_1}^\omega) := \sum_{\mathbf{x} \in D_1} \omega(\mathbf{x}) \kappa(\cdot, \mathbf{x}).$$

It is both 1-Wasserstein and bottleneck stable, if we choose as weight function

$$\omega_{\text{arc}}(\mathbf{x}) = \arctan(C(d-b)^\delta) \quad \mathbf{x} = (b, d), C > 0, \delta \in \mathbb{Z}_{>0}.$$

Indeed, there exist $\delta \in \mathbb{Z}_{>0}$ and $L > 0$ such that

$$\|E_\kappa(\mu_{D_1}^{\omega_{\text{arc}}}) - E_\kappa(\mu_{D_2}^{\omega_{\text{arc}}})\| \leq Ld_B(D_1, D_2).$$

- The **Sliced Wasserstein (SW) kernel** [12]

$$\kappa_{SW}(D_1, D_2) := \exp\left(-\frac{SW(D_1, D_2)}{2\sigma^2}\right).$$

which is based on the so called Sliced Wasserstein distance, which is equivalent to the 1-Wasserstein distance, i.e.,

$$\frac{1}{2M}d_{W,1}(D_1, D_2) \leq d_{SW}(D_1, D_2) \leq 2\sqrt{2}d_{W,1}(D_1, D_2)$$

for some positive constant M .

Remark 1. We observe that the PWG and SW kernels are in the form

$$\kappa(D_r(\mathcal{X}, \varepsilon), D_r(\mathcal{Y}, \varepsilon)) = \exp(-\beta d(D_r(\mathcal{X}, \varepsilon), D_r(\mathcal{Y}, \varepsilon)))$$

for some $\beta > 0$, where $d(\cdot, \cdot)$ is the distance induced by the underlying metric in the case of the SW (for more details, we refer to the cited seminal papers), while for the PWG kernel the distance is induced by the kernel (see (1)).

4. Variably Scaled Persistence Kernels

In the following, our purpose is to interpret the idea underlying VSKs in the context of persistent homology. The main difference between standard kernels and kernels for persistence diagrams is the structure of the input data. Since persistent diagrams consists of a collection of topological features, i.e. birth-death couples in $\mathbb{R}_+^2$, introducing a scaling function whose output lies outside $\mathbb{R}_+^2$ would be meaningless. Hence, letting $\mathfrak{D}_r(\varepsilon) = \{D_r(\mathcal{X}, \varepsilon) \mid \mathcal{X} \subset \Omega\}$, we propose the following definition.

Definition 1. Let $\kappa : \mathfrak{D}_r(\varepsilon) \times \mathfrak{D}_r(\varepsilon) \longrightarrow \mathbb{R}$ be a kernel for persistence diagrams and let $\Psi : \mathfrak{D}_r(\varepsilon) \rightarrow \mathfrak{D}_r(\varepsilon)$. A variably scaled persistence kernel κ_Ψ on $\mathfrak{D}_r(\varepsilon) \times \mathfrak{D}_r(\varepsilon)$ is defined as

$$\kappa_\Psi(D_r(\mathcal{X}, \varepsilon), D_r(\mathcal{Y}, \varepsilon)) := \kappa(\Psi(D_r(\mathcal{X}, \varepsilon)), \Psi(D_r(\mathcal{Y}, \varepsilon)))$$

for $D_r(\mathcal{X}, \varepsilon), D_r(\mathcal{Y}, \varepsilon) \in \mathfrak{D}_r(\varepsilon)$.

As in other contexts, a proper *scaling* function Ψ needs to be designed. Let $\widetilde{D}_r(\mathcal{X}, \varepsilon) = D_r(\mathcal{X}, \varepsilon) \setminus \mathcal{B}$ and let $\psi : \mathfrak{D}_r(\varepsilon) \rightarrow \mathbb{R}_+^2$ be an auxiliary function defined as

$$\psi(D_r(\mathcal{X}, \varepsilon)) = \frac{1}{W} \sum_{\mathbf{x} \in \widetilde{D}_r(\mathcal{X}, \varepsilon)} w(\mathbf{x}) \mathbf{x} \quad (4)$$

where $w : \mathbb{R}_+^2 \rightarrow \mathbb{R}_+$ is a weight function and $W = \sum_{\mathbf{x} \in \widetilde{D}_r(\mathcal{X}, \varepsilon)} w(\mathbf{x})$. We observe that $\widetilde{D}_r(\mathcal{X}, \varepsilon)$ contains a finite number of elements (generators), therefore the sum in (4) is always defined (see also Remark 4). We propose the following alternative scaling functions Ψ .

1. We define

$$\Psi_a(D_r(\mathcal{X}, \varepsilon)) = D_r(\mathcal{X}, \varepsilon) \cup \psi(D_r(\mathcal{X}, \varepsilon)).$$

2. Letting $\rho \in \mathbb{N}$, we first define the set $\widetilde{D}_r(\mathcal{X}, \varepsilon, \rho)$ which consists of the ρ most persistent elements in $\widetilde{D}_r(\mathcal{X}, \varepsilon)$. Then, we define the function

$$\Psi_\rho(D_r(\mathcal{X}, \varepsilon)) = \widetilde{D}_r(\mathcal{X}, \varepsilon, \rho) \cup \psi(D_r(\mathcal{X}, \varepsilon) \setminus \widetilde{D}_r(\mathcal{X}, \varepsilon, \rho)) \cup \mathcal{B}.$$

We remark that $(D_r(\mathcal{X}, \varepsilon) \setminus \widetilde{D}_r(\mathcal{X}, \varepsilon, \rho)) \in \mathfrak{D}_r(\varepsilon)$, therefore $\Psi_\rho(D_r(\mathcal{X}, \varepsilon))$ is well defined.

We observe that Ψ_a plays the role of a *feature augmenting* map, since an additional generator is added in the persistence diagram. On the other hand, Ψ_ρ performs a *feature extraction* procedure. Indeed, the resulting persistence diagram consists of the most ρ persistent elements, $\mathcal{B}$ and the remaining elements, which are possible large in numbers, are *compressed* into a single generator.

Remark 2. Since $D_r(\mathcal{X}, \varepsilon)$ contains an infinite number of elements for all $\mathcal{X} \subset \Omega$ by definition, the VSPK κ_Ψ is still well defined on $\mathfrak{D}_r(\varepsilon) \times \mathfrak{D}_r(\varepsilon)$. Moreover, if κ is (strictly) positive definite, so it is κ_Ψ .

Remark 3. Referring to Remark 1, if κ is a PWG or SW kernel, then κ_Ψ can be directly expressed in terms of the distance $d_\Psi(\cdot, \cdot)$ induced in the variably scaled setting.

For the auxiliary function ψ , we propose the following weights.

1. Let $w_1(\mathbf{x}) = 1/|\widetilde{D}_r(\mathcal{X}, \varepsilon)|$, being $|\widetilde{D}_r(\mathcal{X}, \varepsilon)|$ the cardinality of the multiset, i.e., each element is counted with its multiplicity. We denote as *centre of uniform mass* the resulting auxiliary function

$$\psi_1(D_r(\mathcal{X}, \varepsilon)) = \frac{1}{|\widetilde{D}_r(\mathcal{X}, \varepsilon)|} \sum_{\mathbf{x} \in \widetilde{D}_r(\mathcal{X}, \varepsilon)} \mathbf{x}.$$

2. Let $\mathbf{x} = (b, d) \in \widetilde{D}_r(\mathcal{X}, \varepsilon)$, where $b, d \in \mathbb{R}_+$ are the birth-death *time* of the element $\mathbf{x}$ (see (3)), and let $w_2(\mathbf{x}) = d - b$ be the length of the persistence interval of $\mathbf{x}$. We denote as *centre of persistence* the auxiliary function

$$\psi_2(D_r(\mathcal{X}, \varepsilon)) = \frac{1}{\sum_{\mathbf{x}=(b,d) \in \widetilde{D}_r(\mathcal{X}, \varepsilon)} (d-b)} \sum_{\mathbf{x}=(b,d) \in \widetilde{D}_r(\mathcal{X}, \varepsilon)} (d-b) \mathbf{x}.$$

While the centre of uniform mass is the *barycentre* of the elements of the multiset, the centre of persistence assigns different weights according to the persistence of the elements. This is a natural choice to be analysed, since elements with low persistence are more likely to be related to *noise* structures resulting in the filtration, while elements of large persistence are linked to more representative geometrical features of the dataset (see the example in Figure 1).

Remark 4. *If we take $\mathbf{x} \in D_r(\mathcal{X}, \varepsilon)$ in (4), additional conditions on the weight function w are needed to guarantee the convergence of the sum. However, such infinite setting is not meaningful to be analysed, since elements in the bisector carry no topological information concerning the dataset. As a further observation, the centre of persistence might be formally computed by summing over $D_r(\mathcal{X}, \varepsilon)$, as in this case $w(\mathbf{x}) = 0$ for $\mathbf{x} \in \mathcal{B}$.*

5. Experiments

Our aim is to show how the SVM classifier may benefit of the introduced VSPKs.

In the experiments the kernels are handled using Python 3.8 and the modulus scikit-learn [23] on a 2.6 Ghz Dual-Core Intel Core i5. Persistence diagrams are constructed via the moduli *persim* [27], *ripser* [28] and *GUDHI* [30]. Free and open source PYTHON software is available at

https://github.com/reevost/vspk_paper.

We validate the following hyperparameters. About the considered kernels, we follow the guidelines provided by the authors of the seminal papers.

- Concerning the SVM classifier, we validate $\zeta \in \{10^j \mid j = -3, \dots, 3\}$.
- Concerning the PSS kernel, we take $\sigma \in \{10^j \mid j = -3, \dots, 3\} \cup \{5 \cdot 10^j \mid j = -3, \dots, 2\}$.
- Concerning the PWG kernel, the parameters C and τ of the PWGK are chosen in $\{10^j \mid j = -2, \dots, 2\}$, while δ is set to 10 (see [24, Theorem 3.2]). Moreover, as underlying standard kernel we use the Gaussian.
- Concerning the SW kernel, σ is obtained following the procedure carried out in [12, §4].

To assess the performance of the classifiers, we consider the following scores.

- The accuracy score

$$\text{accuracy} = \frac{\text{true positives} + \text{true negatives}}{\text{true positives} + \text{true negatives} + \text{false positives} + \text{false negatives}}.$$

- The f_1 -score

$$f_1\text{-score} = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}},$$

where

$$\text{precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}},$$

$$\text{recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}.$$

5.1. Alzheimer’s Disease diagnosis

The Open Access Series of Imaging Studies (OASIS) is a project aimed at making neuroimaging data sets of the brain freely available for the scientific community. In particular, OASIS-3 is a compilation of MRI and PET imaging and related clinical data for 1098 participants who were collected across several ongoing studies in the Washington University Knight Alzheimer Disease Research Center over the course of 15 years. Imaging data is accompanied by dementia and APOE status and longitudinal clinical and cognitive outcomes [25].

We consider a subset of the full study group, in order to have a balanced set of data.

A summary of demographic and neuropsychological details of the subjects considered in our study is presented in Table 1.

	AD (mean)	AD (st.dev.)	Control (mean)	Control (st.dev.)
No. of subjects	225	-	248	-
Gender (F/M)	114/111	-	126/122	-
Hand preference (A/L/R)	5/23/197	-	6/26/216	-
Age at entry	74.41	7.60	65.21	9.62
Education (years)	14.77	3.08	16.04	2.51
MMSE	20.33	6.38	29.27	1.30

Table 1: Demographic details and baseline cognitive status measures of the study population.

For each subject, we build the persistence diagrams using the estimation of cortical thickness on 34 points in both right and left hemisphere of the brain, for a total of 64 values. For simplicity, in the study we consider the same coordinates of the above mentioned points for all subjects. The coordinates are computed with the *scipy* toolbox [26]. From this coordinates we build the persistence diagrams and we extract 1 and 2-dimensional topological features, i.e., we obtain the generators associated with H_1 and H_2 homological groups.

In Figure 2 we show two examples of persistence diagrams, and in Figure 3 we highlight the generator added as centre of persistence.

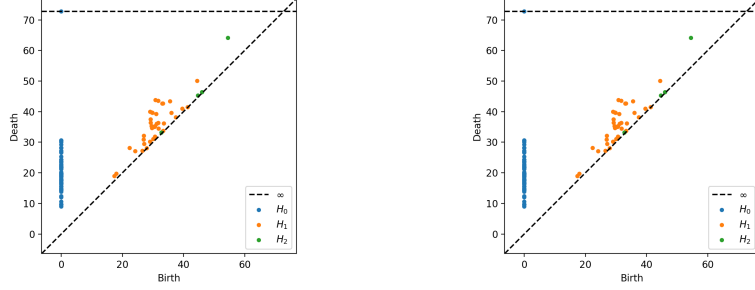


Figure 2: Persistence diagrams of an AD subject with a MMSE of 30 (left) and the persistence diagram of a control subject with a MMSE of 7 (right).

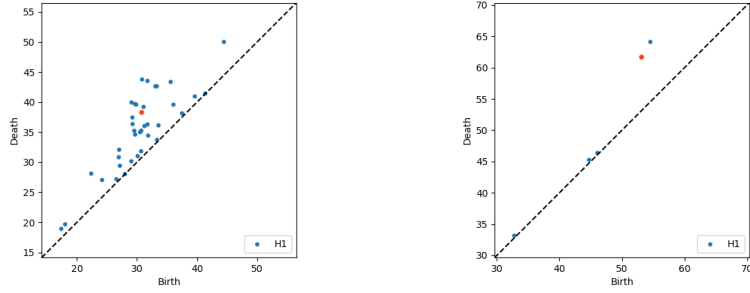


Figure 3: 1-dimensional (left) and 2-dimensional (right) persistence diagram of an AD subject. The red dot is the added centre of persistence via Psi_a and ψ_2 .

We evaluate the performance achieved by a SVMs classifier that makes use of the presented PSS, PWG and SW kernels, both in the classical and in the variably scaled settings.

In each test, we perform a random 70%/30% splitting of the dataset for training and testing, and we consider 5-fold cross validation on the training set for the tuning of the hyperparameters. The results displayed in Tables 2 3 4 have been averaged over 10 runs of tests.

For the variably scaled setting, we consider Ψ_a for both H_1 and H_2 diagrams, while Ψ_ρ , with $\rho = 10$, is employed with H_1 only, since H_2 diagrams are limited in the number of generators, and therefore compressing features is not meaningful.

Furthermore, we use ψ_2 as auxiliary function. Indeed, as highlighted in Table 5, ψ_2 definitely outperforms ψ_1 in our setting. Moreover, we observe that the performances achieved by the auxiliary function alone, i.e., taking the centres of mass or persistence in place of the persistence diagram, are definitely not competitive with respect to the classical and variably scaled settings.

	Ψ	ψ	Accuracy	f ₁ -score	validation time (s)
SW (H_1)	-	-	0.741	0.716	220
VSP-SW (H_1)	Ψ_a	ψ_2	0.732	0.700	223
VSP-SW (H_1)	Ψ_ρ	ψ_2	0.731	0.693	184
SW (H_2)	-	-	0.741	0.712	183
VSP-SW (H_2)	Ψ_a	ψ_2	0.753	0.720	300

Table 2: OASIS-3 dataset. Results of SVMs classification obtained considering H_1 and H_2 persistence diagrams and using the SW kernel.

	Ψ	ψ	Accuracy	f ₁ -score	validation time (s)
PWGG (H_1)	-	-	0.749	0.723	18165
VSP-PWG (H_1)	Ψ_a	ψ_2	0.750	0.726	18082
VSP-PWG (H_1)	Ψ_ρ	ψ_2	0.759	0.735	8731
PWG (H_2)	-	-	0.716	0.699	4422
VSP-PWG (H_2)	Ψ_a	ψ_2	0.709	0.683	4713

Table 3: OASIS-3 dataset. Results of SVMs classification obtained considering H_1 and H_2 persistence diagrams and using the PWG kernel.

	Ψ	ψ	Accuracy	f ₁ -score	validation time (s)
PSS (H_1)	-	-	0.743	0.721	9238
VSP-PSS (H_1)	Ψ_a	ψ_2	0.752	0.728	9045
VSP-PSS (H_1)	Ψ_ρ	ψ_2	0.750	0.723	3330
PSS (H_2)	-	-	0.781	0.762	2825
VSP-PSS (H_2)	Ψ_a	ψ_2	0.775	0.755	3120

Table 4: OASIS-3 dataset. Results of SVMs classification obtained considering H_1 and H_2 persistence diagrams and using the PSS kernel.

	accuracy	f ₁ -score
ψ_1	0.56	0.55
ψ_2	0.65	0.72

Table 5: OASIS-3 dataset. Results of SVMs classification obtained by using the centre of mass and persistence alone in place of the persistence diagrams.

We observe that VSPKs are competitive with respect to the classical setting, improving the performance in some cases. Moreover, the usage of Ψ_ρ leads to a consistent saving in validation time.

5.2. Orbit recognition

As second experiment, we follow the idea proposed in [29] and we analyse the linked twisted map, which models fluid flows. The corresponding orbits are computed via the discrete system

$$\begin{cases} x_{n+1} = x_n + ry_n(1 - y_n) \mod 1 \\ y_{n+1} = y_n + rx_{n+1}(1 - x_{n+1}) \mod 1 \end{cases}$$

where $(x_0, y_0) \in [0, 1] \times [0, 1]$ is the initial position and $r > 0$ is a real parameter that influences the orbit. The topological structure of the orbit changes with the initial position and r , as displayed in Figures 4 and 5, where we depict the first 1000 iterations $\{(x_n, y_n) : n = 0, \dots, 1000\}$.

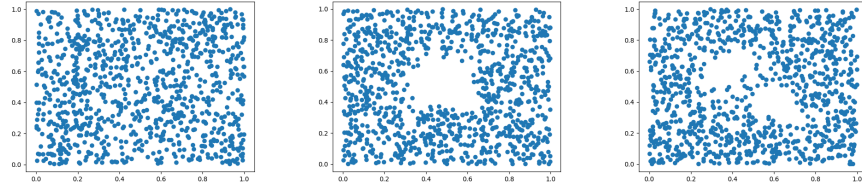


Figure 4: Fixed $(x_0, y_0) \in [0, 1]^2$, the orbits resulting from the linked twisted map taking $r = 2.5, 4.1, 4.3$, from left to right, respectively.

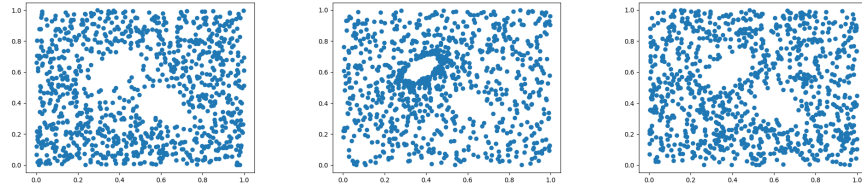


Figure 5: Fixed $r = 4.3$, the orbits resulting from the linked twisted map taking different starting points $(x_0, y_0) \in [0, 1]^2$.

In the following tests, accordingly to [29], we choose a set of five parameters $r = 2.5, 3.5, 4, 4.1, 4.3$ as set of classification labels. For each label, we compute the first 1000 points of 50 orbits, with random starting point. Therefore, the dataset consists of 250 elements. Then, we compute the persistence diagram related to each orbit.

Here, since each persistence diagram has a huge number of generators ($\approx 10^5$), we empirically reduce the persistence diagrams by restricting to the most 10 persistent elements. More precisely, here Ψ_a is computed with respect to such 10 elements, while in the case of Ψ_ρ the discarded less persistent generators are compressed in a unique element via ψ_2 . Moreover, since we are dealing with 2-dimensional orbits, we compute the H_1 homology group only.

As in Section 5.1, we consider a 5-fold cross validation on the training set, and the results displayed in Tables 6, 7 and 8 are averaged over 10 runs with 70%/30% training-test split of the data.

	Ψ	ψ	Accuracy	f ₁ -score
SW (H_1)	-	-	0.832	0.831
VSP-SW (H_1)	Ψ_a	ψ_2	0.814	0.812
VSP-SW (H_1)	Ψ_ρ	ψ_2	0.833	0.832

Table 6: Orbit Recognition. Results of SVMs classification on H_1 persistence diagrams using the SW kernel.

	Ψ	ψ	Accuracy	f ₁ -score
PWG (H_1)	-	-	0.858	0.866
VSP-PWG (H_1)	Ψ_a	ψ_2	0.858	0.866
VSP-PWG (H_1)	Ψ_ρ	ψ_2	0.846	0.853

Table 7: Orbit Recognition. Results of SVMs classification on H_1 persistence diagrams using the PWG kernel.

	Ψ	ψ	Accuracy	f ₁ -score
PSS (H_1)	-	-	0.806	0.803
VSP-PSS (H_1)	Ψ_a	ψ_2	0.824	0.821
VSP-PSS (H_1)	Ψ_ρ	ψ_2	0.826	0.823

Table 8: Orbit Recognition. Results of SVMs classification on H_1 persistence diagrams using the PSS kernel.

5.3. 3D shape segmentation

Here, we follow an experiment proposed also in [12]. We consider some categories of the mesh segmentation benchmark introduced in [31], which contains different 3D shapes of several categories. In each sample, every face is represented as a triplet of 3D points, and it is labeled with respect to a corresponding segmentation index. In our dataset, we consider 72546 faces in the category *Ant*, 59546 faces in the category *Airplane*, 50902 faces in the category *Bird* and 66248 faces in the category *Octopus*. A H_1 persistence diagram is then computed for each face by using the geodesic distance on the 3D shape; for more details concerning the dataset, we refer to [32], while in Figure 6 we display some examples of shape segmentation.

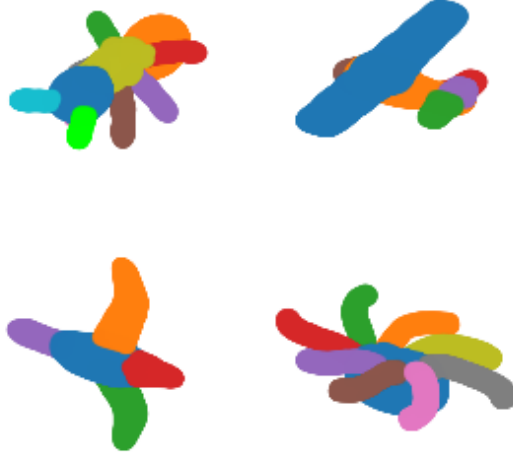


Figure 6: Examples of 3D shapes and corresponding segmentation labels. From top left, clock-wise, Ant, Airplane, Bird and Octopus.

For each category, the classification task consists in predicting the segmentation index corresponding to the persistence diagram. As in the previous subsections, we consider a 5-fold cross validation on the training set, and the results displayed in Tables 9 and 10 are averaged over 10 runs with 70%/30% training-test split of the data.

We remark that the elements of the persistence diagrams in this experiment are limited in numbers and free of noisy generators, because of the particular construction of the diagrams obtained via the geodesic metric. Therefore, the usage of Ψ_ρ is not significant.

	PSS (H_1)	VSP-PSS (H_1)	SW (H_1)	VSP-SW (H_1)
Ant	0.781	0.793	0.753	0.746
Airplane	0.677	0.688	0.717	0.704
Bird	0.612	0.598	0.636	0.629
Octopus	0.771	0.773	0.755	0.746

Table 9: Accuracy achieved by the SVMs classifiers in the carried out tests. In the variably scaled setting, we used Ψ_a and ψ_2 functions.

	PSS (H_1)	VSP-PSS (H_1)	SW (H_1)	VSP-SW (H_1)
Ant	0.744	0.751	0.710	0.700
Airplane	0.580	0.582	0.580	0.563
Bird	0.619	0.616	0.624	0.617
Octopus	0.704	0.707	0.679	0.666

Table 10: f1-score achieved by the SVMs classifiers in the carried out tests. In the variably scaled setting, we used Ψ_a and ψ_2 functions.

6. Conclusions and future work

In this paper, we proposed VSPKs for dealing with persistence diagrams in the context of persistent homology. The proposed framework, which is directly inspired by the variably scaled setting explored in kernel-based approximation and learning, may enhance the performance and the efficiency of existing kernels for persistence diagrams, as suggested by the obtained results. Future work consists of investigating more on the design of the scaling function, which plays a key role in the construction of the kernel. In this view, the analysis of VSPKs in the context of algebraic topology may provide useful insights.

Acknowledgements

This research has been accomplished within the Rete Italiana di Approssimazione (RITA) and the thematic group on Approximation Theory and Applications of the Italian Mathematical Union. We also received the support of GNCS-INdAM.

References

- [1] B. Schölkopf, A. J. Smola, F. Bach, et al., Learning with kernels: support vector machines, regularization, optimization, and beyond, MIT press, 2002.
- [2] V. N. Vapnik, An overview of statistical learning theory, IEEE transactions on neural networks 10 (5) (1999) 988–999.
- [3] R. Schaback, H. Wendland, Kernel techniques: From machine learning to meshless methods, Acta Numerica 15 (2006) 543–639. doi:10.1017/S0962492906270016.
- [4] J. Shawe-Taylor, N. Cristianini, et al., Kernel methods for pattern analysis, Cambridge university press, 2004.
- [5] K. M. Borgwardt, C. S. Ong, S. Schönauer, S. Vishwanathan, A. J. Smola, H.-P. Kriegel, Protein function prediction via graph kernels, Bioinformatics 21 (suppl_1) (2005) i47–i56.

- [6] M. Collins, N. Duffy, Convolution kernels for natural language, in: *Advances in neural information processing systems*, 2001, pp. 625–632.
- [7] C. H. Lampert, *Kernel methods in computer vision*, Now Publishers Inc, 2009.
- [8] S. V. N. Vishwanathan, N. N. Schraudolph, R. Kondor, K. M. Borgwardt, Graph kernels, *Journal of Machine Learning Research* 11 (2010) 1201–1242.
- [9] G. Carlsson, Topology and data, *Bulletin of the American Mathematical Society* 46 (2) (2009) 255–308.
- [10] H. Edelsbrunner, D. Letscher, A. Zomorodian, Topological persistence and simplification, *Proceedings 41st Annual Symposium on Foundations of Computer Science* (2000) 454–463.
- [11] A. T. Fomenko, *Visual geometry and topology*, Springer Science & Business Media, 2012.
- [12] M. Carriere, M. Cuturi, S. Oudot, Sliced wasserstein kernel for persistence diagrams, in: *International Conference on Machine Learning*, PMLR, 2017, pp. 664–673.
- [13] G. Kusano, K. Fukumizu, Y. Hiraoka, Kernel method for persistence diagrams via kernel embedding and weight factor, *The Journal of Machine Learning Research* 18 (1) (2017) 6947–6987.
- [14] T. Le, M. Yamada, Persistence fisher kernel: A riemannian manifold kernel for persistence diagrams, in: *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, NIPS’18, Curran Associates Inc., Red Hook, NY, USA, 2018, p. 10028–10039.
- [15] J. Reininghaus, S. Huber, U. Bauer, R. Kwitt, A stable multi-scale kernel for topological machine learning, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 4741–4748.
- [16] M. Bozzini, L. Lenarduzzi, M. Rossini, R. Schaback, Interpolation with variably scaled kernels, *IMA Journal of Numerical Analysis* 35 (1) (2015) 199–219.
- [17] C. Campi, F. Marchetti, E. Perracchione, Learning via variably scaled kernels, *Adv. Comput. Math.* 47 (4) (2021) 51. doi:10.1007/s10444-021-09875-6.
URL <https://doi.org/10.1007/s10444-021-09875-6>
- [18] F. Marchetti, F. De Martino, M. Shamseddin, S. De Marchi, C. Briskin, Variably scaled kernels improve classification of hormonally-treated patient-derived xenografts, in: *2020 IEEE Conference on Evolving and Adaptive Intelligent Systems (EAIS)*, 2020, pp. 1–6. doi:10.1109/EAIS48028.2020.9122767.

- [19] J. Mercer, Functions of positive and negative type, and their connection with the theory of integral equations, *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 209 (1909) 415–446.
URL <http://www.jstor.org/stable/91043>
- [20] G. Fasshauer, M. McCourt, *Kernel-based Approximation Methods using MATLAB*, WORLD SCIENTIFIC, 2015. **arXiv:**<https://www.worldscientific.com/doi/pdf/10.1142/9335>, doi:10.1142/9335.
URL <https://www.worldscientific.com/doi/abs/10.1142/9335>
- [21] M. Guillemand, A. Iske, *Interactions Between Kernels, Frames, and Persistent Homology*, Springer International Publishing, Cham, 2017, pp. 861–888. doi:10.1007/978-3-319-55556-0_14.
URL https://doi.org/10.1007/978-3-319-55556-0_14
- [22] D. Cohen-Steiner, H. Edelsbrunner, J. Harer, Stability of persistence diagrams, *Discrete & computational geometry* 37 (1) (2007) 103–120.
- [23] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in Python, *Journal of Machine Learning Research* 12 (2011) 2825–2830.
- [24] G. Kusano, Y. Hiraoka, K. Fukumizu, Persistence weighted gaussian kernel for topological data analysis, in: *International Conference on Machine Learning*, PMLR, 2016, pp. 2004–2013.
- [25] P. J. LaMontagne, T. L. Benzinger, J. C. Morris, S. Keefe, R. Hornbeck, C. Xiong, E. Grant, J. Hassenstab, K. Moulder, A. Vlassenko, et al., Oasis-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and alzheimer disease, *MedRxiv* (2019).
- [26] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, Í. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, SciPy 1.0 Contributors, *SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python*, *Nature Methods* 17 (2020) 261–272. doi:10.1038/s41592-019-0686-2.
- [27] N. Saul, C. Tralie, *Scikit-tda: Topological data analysis for python* (2019). doi:10.5281/zenodo.2533369.
URL <https://doi.org/10.5281/zenodo.2533369>

- [28] C. Tralie, N. Saul, R. Bar-On, Ripser.py: A lean persistent homology library for python, *The Journal of Open Source Software* 3 (29) (2018) 925. doi:10.21105/joss.00925.
URL <https://doi.org/10.21105/joss.00925>
- [29] H. Adams, T. Emerson, M. Kirby, R. Neville, C. Peterson, P. Shipman, S. Chepushtanova, E. Hanson, F. Motta, L. Ziegelmeier, Persistence images: A stable vector representation of persistent homology, *Journal of Machine Learning Research* 18 (2017).
- [30] The GUDHI Project, GUDHI User and Reference Manual, 3.5.0 Edition, GUDHI Editorial Board, 2022.
URL <https://gudhi.inria.fr/doc/3.5.0/>
- [31] X. Chen, A. Golovinskiy, T. Funkhouser, A benchmark for 3D mesh segmentation, *ACM Transactions on Graphics (Proc. SIGGRAPH)* 28 (3) (Aug. 2009).
- [32] M. Carrière, S. Y. Oudot, M. Ovsjanikov, Stable topological signatures for points on 3d shapes, in: *Computer graphics forum*, Vol. 34, Wiley Online Library, 2015, pp. 1–12.