

Learning Weakly Convex Sets in Metric Spaces

Eike Stadtländer¹, Tamás Horváth^{1,2,3} ✉, and Stefan Wrobel^{1,2,3}

¹ Dept. of Computer Science, University of Bonn, Bonn, Germany

² Fraunhofer IAIS, Schloss Birlinghoven, Sankt Augustin, Germany

³ Fraunhofer Center for Machine Learning, Sankt Augustin, Germany
{stadtlaender,horvath,wrobel}@cs.uni-bonn.de

Abstract. We introduce the notion of weak convexity in metric spaces, a generalization of ordinary convexity commonly used in machine learning. It is shown that weakly convex sets can be characterized by a closure operator and have a unique decomposition into a set of pairwise disjoint connected blocks. We give two generic efficient algorithms, an extensional and an intensional one for learning weakly convex concepts and study their formal properties. Our experimental results concerning vertex classification clearly demonstrate the excellent predictive performance of the extensional algorithm. Two non-trivial applications of the intensional algorithm to polynomial PAC-learnability are presented. The first one deals with learning k -convex Boolean functions, which are already known to be efficiently PAC-learnable. It is shown how to derive this positive result in a fairly easy way by the generic intensional algorithm. The second one is concerned with the Euclidean space equipped with the Manhattan distance. For this metric space, weakly convex sets form a union of pairwise disjoint axis-aligned hyperrectangles. We show that a weakly convex set that is consistent with a set of examples and contains a minimum number of hyperrectangles can be found in polynomial time. In contrast, this problem is known to be NP-complete if the hyperrectangles may be overlapping.

Keywords: abstract convexity · concept learning · vertex classification

1 Introduction

Several results in the theory of machine learning are concerned with concept classes defined by various forms of *convexity* (e.g., polygons formed by the intersection of a bounded number of half-spaces [2], conjunctions [14], or geodesic convexity in graphs [11]). In a broad sense, convex sets constitute *contiguous* subsets of the domain. This property can, however, be a drawback when the target concept cannot be represented by a single convex set. To overcome this problem, we relax the notion of convexity by introducing that of *weak convexity* for *metric spaces*. More precisely, a subset A of a metric space is *weakly convex* if for all $x, y \in A$ and for all points z in the ground set, z belongs to A whenever x and y are *near* to each other and the three points satisfy the triangle inequality with *equality*. This definition has been inspired by the following *relaxation* of

convexity in the Hamming metric space [6]: A Boolean function is *k-convex* for some positive integer k if for all true points x and y having a Hamming distance of at most k , all points on all shortest paths between x and y are also true. Our definition of weak convexity generalizes this notion to *arbitrary* metric spaces.

We present some properties of weakly convex sets of a metric space. In particular, we show that they form a *convexity space* [15] and hence, a *closure system*. Furthermore, they give rise to a *unique* decomposition into a set of “connected” blocks that have a pairwise minimum distance. We also study two scenarios for *learning* weakly convex sets. The first one considers the case that the metric space is *finite* and weakly convex sets are given *extensionally*. For this setting we define a *preclosure* operator and show that weakly convex sets can be characterized by a *closure* operator defined by the fixed points of the iterative applications of this preclosure operator. This characterization gives rise to an *efficient* algorithm computing the weakly convex hull for any set of points. We then prove that a weakly convex set that is *consistent* with a set of examples and has the *smallest* number of blocks can be found in polynomial time. This result makes use of the unique decomposition of weakly convex sets. As a proof of concept, we *experimentally* demonstrate on graph vertex classification that a remarkable accuracy can be obtained already with a relatively small training data set.

The second scenario deals with the case that the metric spaces are *not* necessarily finite and that weakly convex sets are given *intensionally*, using some compact representation. We present a simple generic algorithm, which iteratively “merges” weakly convex connected blocks and give sufficient conditions for the efficiency of a more sophisticated version of this naïve algorithm. Similarly to the extensional setting, we prove that a weakly convex set consistent with a set of examples and containing a *minimum* number of blocks can be found in *polynomial* time if certain conditions are fulfilled. We also present two non-trivial applications of this general result to polynomial PAC-learnability [14]. The first one deals with learning *k-convex Boolean functions*, for which there already exists a positive PAC result [6]. We still consider this problem because we show that the same result can be obtained in a very simple way by our intensional learning algorithm. Furthermore, our general purpose algorithm calculates the *k-convex* Boolean function for a set of examples in the same asymptotic time complexity as the domain specific one in [6]. The second application deals with the metric space defined by \mathbb{R}^d endowed with the Manhattan (or L_1) distance. Weakly convex sets for this case are the union of a set of pairwise disjoint axis-aligned closed *hyperrectangles*. Using our general learning algorithm, we prove in a very simple way that the concept class formed by weakly convex sets containing at most k hyperrectangles is *polynomially* PAC-learnable. To underline the strength and utility of our approach, we note that the consistent hypothesis finding problem for the related problem that the hyperrectangles are not required to be pairwise disjoint is NP-complete, even for $d = 2$ (see, e.g., [1]).

Related Work To the best of our knowledge, our notion of weak convexity in metric spaces is *new*. As mentioned above, it has been inspired by the definition of *k-convex Boolean functions* introduced in [6]. In fact, our notion generalizes

that for k -convex Boolean functions to a broad class of metric spaces, including infinite ones as well. We also mention the somewhat related, but fundamentally distinct notion of α -hulls (see, e.g., [5]). They are defined as the intersection of enclosing generalized disks, but only for *finite* subsets of \mathbb{R}^2 and \mathbb{R}^3 . Furthermore, it is known that the α -hull operator is *not* idempotent [8]. In contrast, our notion results in an abstract convexity structure in the sense of [15] and has therefore a corresponding closure operator defined for *arbitrary* subsets of a broad class of metric spaces. Last, but not least, even though our definitions resemble those of the density-based clustering approach [7], DBSCAN clusters are generally *not* weakly convex, except for very specific parameter values.

Outline The rest of the paper is organized as follows. We collect the necessary notions and fix the notation in Section 2. In Section 3 we define weakly convex sets in metric spaces and prove some of their basic properties. Sections 4 and 5 are devoted to learning weakly convex sets in the extensional and intensional problem settings. Finally, we conclude in Section 6 and mention some problems for future work. Most of the proofs are omitted for space limitations. They can be found in [13].

2 Preliminaries

In this section we collect the necessary notions and fix the notation. For any $n \in \mathbb{N}$, $[n]$ denotes the set $\{1, 2, \dots, n\}$. The family of all finite subsets of a set X is denoted by $[X]^{<\omega}$. A *metric space* is a pair (X, D) , where X is a set and D is a metric on X (i.e., (i) $D(x, y) = 0$ iff $x = y$, (ii) $D(x, y) = D(y, x)$, and (iii) $D(x, y) \leq D(x, z) + D(z, y)$ for all $x, y, z \in X$).

A *closure system* over some ground set X is a pair (X, \mathcal{C}) with $\mathcal{C} \subseteq 2^X$ such that \mathcal{C} is closed under arbitrary intersection, where 2^X denotes the *power set* of X . We require that $X \in \mathcal{C}$. The elements of \mathcal{C} are called *closed sets*. One of the elementary properties of closure systems is that they can be characterized in terms of closure operators (see, e.g., [3]). More precisely, a function $\rho : 2^X \rightarrow 2^X$ is a *closure operator* if it satisfies the following properties for all $A, B \subseteq X$: (i) $A \subseteq \rho(A)$ (*extensivity*), (ii) $\rho(A) \subseteq \rho(B)$ whenever $A \subseteq B$ (*monotonicity*), and (iii) $\rho(\rho(A)) = \rho(A)$ (*idempotency*). If ρ is extensive and monotone, but not necessarily idempotent, then it is a *preclosure operator*. The fixed points of a closure operator are called *closed sets* and the set system (X, \mathcal{C}_ρ) with $\mathcal{C}_\rho = \{A \subseteq X : \rho(A) = A\}$ is always a closure system. Conversely, for any closure system (X, \mathcal{C}) , the function $\rho : 2^X \rightarrow 2^X$ with $\rho : A \mapsto \bigcap \{C \in \mathcal{C} : A \subseteq C\}$ is a closure operator satisfying $\mathcal{C} = \{\rho(A) : A \subseteq X\}$. Finally, a *convexity space* [15] over a set X is a closure system (X, \mathcal{C}) such that (i) $\emptyset, X \in \mathcal{C}$ and (ii) \mathcal{C} is closed under *nested unions* (i.e., $\bigcup \mathcal{D} \in \mathcal{C}$ for any $\mathcal{D} \subseteq \mathcal{C}$ that is totally ordered w.r.t. set inclusion).

Our notion of *weak convexity* defined in the next section is inspired by that of *k-convexity* introduced in the seminal paper by Ekin, Hammer, and Kogan [6]. More precisely, consider the metric space (\mathbb{H}_d, D_H) , where $\mathbb{H}_d = \{0, 1\}^d$ is the

d -dimensional *Hamming cube* and D_H is the Hamming distance. A subset X of \mathbb{H}_d is *k-convex* for an integer $k \geq 1$ if for all $x, y \in X$ with $D_H(x, y) \leq k$ and for all $z \in \mathbb{H}_d$, $z \in X$ whenever the triangle inequality holds with equality (i.e., $D_H(x, y) = D_H(x, z) + D_H(z, y)$).

An (*undirected*) *graph* is a pair $G = (V, E)$, where V is a finite set of vertices and $E \subseteq \{e \subseteq V : |e| = 2\}$ is a set of edges. An edge $\{x, y\}$ will sometimes be denoted by xy . A graph $G' = (V', E')$ is a *subgraph* of G if $V' \subseteq V$ and $E' \subseteq E$. A *path* is a graph $P = (V_P, E_P)$ with $V_P = \{v_1, \dots, v_n\}$ and $E_P = \{v_i v_{i+1} : i \in [n-1]\}$. The *length* of a path P is the number of edges it contains. A graph is *connected* if all pairs of its vertices are connected by a path. If two vertices of a graph G are connected by a path, we define their *geodesic distance* by the length of a shortest path connecting them. Note that it is a metric on the set of vertices for connected graphs. A subset $X \subseteq V$ is called (*geodesically*) *convex* in a graph $G = (V, E)$ if for all $u, v \in V$ and for all shortest paths $P = (V_P, E_P)$ connecting u and v , we have $V_P \subseteq X$.

For the standard definitions of *concepts*, *concept classes*, *VC-dimension*, and *polynomial PAC-learnability* from computational learning theory, the reader is referred to some standard text book about learning theory (see, e.g., [9]). Let \mathcal{C} be a concept class over some domain X . The *k-fold union* of \mathcal{C} for some $k \geq 1$ integer is defined by $\mathcal{C}_{\cup}^k = \{C_1 \cup \dots \cup C_k : C_i \in \mathcal{C} \text{ for all } i \in [k]\}$. Note that the definition does not require the C_i s to be pairwise different. The following problem is central to concept learning:

Problem 1 (The Consistency Problem). Given a concept class $\mathcal{C} \subseteq 2^X$ over some domain X and disjoint sets $E^+, E^- \subseteq X$ of examples, return a concept $C \in \mathcal{C}$ that is consistent with E^+ and E^- , i.e., $E^+ \subseteq C$ and $E^- \cap C = \emptyset$ if such a concept exists; o/w return “NO”.

In order to prove polynomial PAC-learnability, we will use the following results from computational learning theory [2].

Theorem 1. Let $\mathcal{C} \subseteq 2^X$ be a concept class over some domain X with VC-dimension $d > 0$.

- (i) \mathcal{C} is polynomially PAC-learnable if d is bounded by a polynomial of its parameters and Problem 1 can be solved in polynomial time in the parameters.
- (ii) For all $k \geq 1$, the VC-dimension of \mathcal{C}_{\cup}^k is at most $2dk \log(3k)$.

3 Weak Convexity in Metric Spaces

In this section we relax the notion of convexity defined for Euclidean spaces to *weak convexity* in *metric spaces* and discuss some basic formal properties of weakly convex sets. The main result of this section is formulated in Thm. 2. It states that weakly convex sets have a *unique* decomposition into a set of weakly convex “connected” blocks that have a pairwise minimum distance from each other. To define weak convexity, recall that a subset $A \subseteq \mathbb{R}^d$ is *convex* if

$$D_2(x, z) + D_2(z, y) = D_2(x, y) \text{ implies } z \in A \quad (1)$$

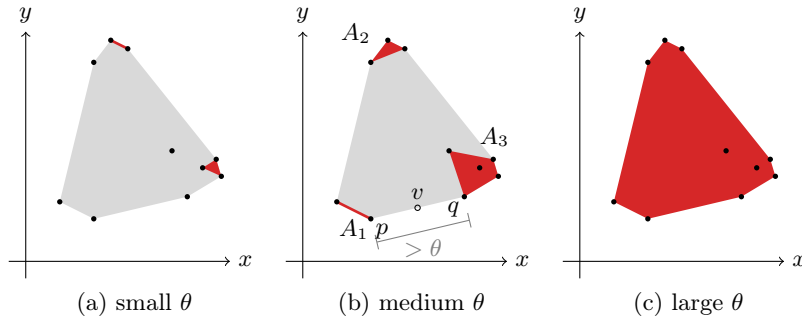


Fig. 1. Illustration of weakly convex sets in the Euclidean plane \mathbb{R}^2

for all $x, y \in A$ and for all $z \in \mathbb{R}^d$, where D_2 is the Euclidian distance. Our notion of weak convexity in metric spaces incorporates a relaxation of (1), motivated by the fact that convex sets defined by (1) are always “contiguous” and cannot therefore capture well-separated regions of the domain. We address this problem by adapting the idea of *k-convexity* over Hamming metric spaces [6] to *arbitrary* ones. Analogously to [6], we do not require (1) to hold for all points x and y , but only for such pairs which have a distance of at most a user-specified threshold. In other words, while ordinary convexity is based on a *global* condition resulting in a single “contiguous” region, our notion of weak convexity relies on a *local* one, resulting in potentially several isolated regions, where the spread of locality is controlled by the above mentioned user-specified threshold. This consideration yields the following formal definition of *weakly convex* sets in *metric spaces*:

Definition 1. Let (X, D) be a metric space and $\theta \geq 0$. A set $A \subseteq X$ is θ -convex (or simply, weakly convex) if for all $x, y \in A$ and $z \in X$ it holds that $z \in A$ whenever $D(x, y) \leq \theta$ and $z \in \Delta_{\leq}(x, y)$, where

$$\Delta_{\leq}(x, y) = \{z \in X : D(x, z) + D(z, y) = D(x, y)\} . \quad (2)$$

Notice that (2) does not require $x \neq y$. In particular, $\Delta_{\leq}(x, x) = \{x\}$ for all $x \in X$. The family of all weakly convex sets is denoted by $\mathcal{C}_{\theta, D}$; we omit D if it is clear from the context. It always holds that $\mathcal{C}_{0, D} = 2^X$.

In order to illustrate the notion of weak convexity, consider the finite set of points $A \subseteq \mathbb{R}^2$ depicted by filled dots in Fig. 1b. The strongly (i.e., ordinary) convex hull of A is indicated by the gray area. In contrast, the \subseteq -smallest θ -convex set containing A for some suitable $\theta \geq 0$ is drawn in red. The most obvious difference is that there are three separated regions A_1, A_2 , and A_3 , instead of a single contiguous area. In other words, weakly convex sets need not be connected despite that strongly convex sets in \mathbb{R}^2 do. This is a consequence of considering only such pairs for membership witnesses that have a distance of at most θ (see, also, Fig. 1a and 1c). For example, the points p and q in Fig. 1b have a distance strictly greater than θ , implying that they do not witness the membership of the point v . Notice that in the same way as strongly convex sets, (parts of) weakly

convex sets may degenerate. While A_2 and A_3 are regions with strictly positive area, A_1 is just a segment. We may even have isolated points as shown in Fig. 1a.

Despite this unconventional behavior of weakly convex sets, (X, \mathcal{C}_θ) forms a *convexity space*. To see this, note that $\emptyset, X \in \mathcal{C}_\theta$. Furthermore, \mathcal{C}_θ is stable for arbitrary intersections and nested unions. Indeed, if $\mathcal{F} \subseteq \mathcal{C}_\theta$ is a family of θ -convex sets, $x, y \in \bigcap \mathcal{F}$ with $D(x, y) \leq \theta$ then $\Delta_=(x, y) \subseteq F$ for all $F \in \mathcal{F}$ implying that $\bigcap \mathcal{F}$ is θ -convex. If, in addition, \mathcal{F} is totally ordered by inclusion and $x, y \in \bigcup \mathcal{F}$ with $D(x, y) \leq \theta$ then there are $F_x, F_y \in \mathcal{F}$, say $F_x \subseteq F_y$, such that $x \in F_x$ and $y \in F_y$. Then, according to (2), $\Delta_=(x, y) \subseteq F_y$ implying that $\bigcup \mathcal{F}$ is θ -convex. Hence, \mathcal{C}_θ is a convexity space as claimed.

Since \mathcal{C}_θ is stable for arbitrary intersections, it has an associated *closure operator* $\rho_\theta : 2^X \rightarrow 2^X$ with $A \mapsto \bigcap \{C \in \mathcal{C}_\theta : A \subseteq C\}$ for all $A \subseteq X$. That is, ρ_θ maps a set A to the \subseteq -smallest θ -convex set containing A . It is called the *weakly convex hull operator* and its fixed points (i.e., the ρ_θ -closed sets) form exactly \mathcal{C}_θ . Moreover, ρ_θ is *domain finite* [15], i.e., $\rho_\theta(A) = \bigcup \{\rho_\theta(F) : F \subseteq [A]^{<\omega}\}$.

3.1 Some Basic Properties of Weakly Convex Sets

We now present some basic properties of weakly convex sets that make this kind of closed sets interesting for machine learning from a practical as well as from a theoretical viewpoint. As already mentioned, weakly convex sets need *not* be contiguous (cf. Fig. 1), in contrast to, for instance, ordinary convex sets in the Euclidean space. Instead, one can observe regions that are separated from each other. This is again due to the fact that the notion of weak convexity utilizes a distance threshold θ . As a consequence, *separate* regions may arise with a pairwise distance of at least θ . In Thm. 2 below, which is one of our main technical results for this work, we formally state this property of weakly convex sets. We note that this result generalizes that stated in [6, Proposition 3.2] for the Hamming metric space to arbitrary metric spaces.

We first introduce some necessary notions. Let $\mathcal{M} = (X, D)$ be a metric space, $\theta \geq 0$, and $A \subseteq X$. Two points $a, b \in A$ are *θ -connected* w.r.t. A , denoted $a \sim_{\theta, A} b$, if there is a finite sequence $a = p_1, p_2, \dots, p_r = b \in A$ such that $D(p_i, p_{i+1}) \leq \theta$ for all $i \in [r - 1]$. A is *θ -connected* if $a \sim_{\theta, A} b$ for all $a, b \in A$. Note that $\sim_{\theta, A}$ is an equivalence relation on A ; its equivalence classes, denoted by $[a]_{\sim_{\theta, A}} = \{b \in A : a \sim_{\theta, A} b\}$ for all $a \in A$, will be referred to as *θ -connected components*.

Theorem 2. *Let (X, D) be a metric space and $\theta \geq 0$. Then $A \subseteq X$ is θ -convex iff there is a uniquely defined family of non-empty sets $(A_i \subseteq A)_{i \in I}$ for some index set I satisfying the following conditions:*

- (i) $A = \bigcup_{i \in I} A_i$,
- (ii) A_i is θ -convex for all $i \in I$,
- (iii) A_i is θ -connected for all $i \in I$,
- (iv) for all $i, j \in I$ with $i \neq j$, $D(a, b) > \theta$ for all $a \in A_i, b \in A_j$.

In what follows, the family $(A_i)_{i \in I}$ satisfying conditions (i)–(iv) in Thm. 2 will be referred to as the θ -decomposition of the θ -convex set A . Furthermore, the sets A_i in the θ -decomposition of A will be called θ -blocks or simply, *blocks*. The theorem above tells us that weakly convex sets can be partitioned *uniquely* into a family of non-empty blocks in a way that the distance between each pair of such weakly convex components is at least θ . The uniqueness of the θ -decomposition in Thm. 2 gives rise to a naïve algorithm for computing the weakly convex hull of a finite set intensionally (cf. Alg. 2 in Section 5). The idea is to start with the singletons and enforce conditions (i)–(iv) by repeatedly merging invalid pairs of blocks. However, that requires the strict inequality in condition (iv) not only to hold for pairs of points, but also between blocks. Cor. 1 below is concerned with metric spaces in which this property holds.

Corollary 1. *Let $\mathcal{M} = (X, D)$ be a metric space, $\theta \geq 0$, $A \subseteq X$, and $(A_i)_{i \in I}$ the θ -decomposition of $\rho_\theta(A)$.*

- (i) *We have $|I| \leq |A|$. In particular, if A is finite, then I is finite.*
- (ii) *If \mathcal{M} is complete and A_i and A_j are (topologically) closed for some $i \neq j$ then $D(A_i, A_j) = \inf_{a \in A_i, b \in A_j} D(a, b) > \theta$.*

Accordingly, Cor. 1 motivates the following definition of *well-behaved* metric spaces. A metric space $\mathcal{M} = (X, D)$ is *compatible* with the convexity space (X, \mathcal{C}_θ) if $\rho_\theta(A)$ is (topologically) closed for all $A \in [X]^{<\omega}$ [15]. If, in addition, \mathcal{M} is complete, we call it *well-behaved*.

Finally, we claim that the weakly convex hull operator is monotone w.r.t. θ . This property will be utilized by our consistent hypothesis finding algorithms.

Proposition 1. *Let (X, D) be a metric space and $0 \leq \theta \leq \theta'$. Then for all $A \subseteq X$, (i) $\rho_\theta(A) \subseteq \rho_{\theta'}(A)$ and (ii) for all $x, y \in \rho_\theta(A)$, x, y are in the same θ' -block of the θ' -decomposition of $\rho_{\theta'}(A)$ if they are in the same θ -block of the θ -decomposition of $\rho_\theta(A)$.*

4 Learning in the Extensional Problem Setting

In this section we consider the case that the underlying metric space is *finite* and weakly convex sets are represented *extensionally*, e.g. because they have *no* (natural) compact representation. Examples of this scenario include, among others, the case that the metric space is given by the set of vertices of a graph together with some distance on vertices. To formulate some basic properties of ρ_θ introduced in Section 3, we define a preclosure operator $\hat{\rho}_\theta$ over X . More precisely, let $\mathcal{M} = (X, D)$ be a finite metric space and $\theta \geq 0$. For all $x, y \in X$, let $W_\theta(x, y) = \Delta_{=}(x, y)$ if $D(x, y) \leq \theta$; o/w $W_\theta(x, y) = \emptyset$. Finally, define the function $\hat{\rho}_\theta : 2^X \rightarrow 2^X$ by $\hat{\rho}_\theta(A) = \bigcup_{x, y \in A} W_\theta(x, y)$ for all $A \subseteq X$.

Lemma 1. *The function $\hat{\rho}_\theta$ over \mathcal{M} is a preclosure operator.*

Algorithm 1 EXTENSIONAL WEAKLY CONVEX HULL $\hat{\rho}_\theta$

Require: finite metric space (X, D) and $S_x = \{(x', D(x, x')) : x' \in X \setminus \{x\}\}$ sorted in increasing order in the second component, for all $x \in X$

Input: $A \subseteq X$ and $\theta \geq 0$

Output: θ -decomposition of $\rho_\theta(A)$

```

1:  $C, E \leftarrow \emptyset$ , queue  $Q \leftarrow A$ 
2: mark all elements in  $A$ 
3: while  $Q \neq \emptyset$  do
4:    $x \leftarrow \text{DEQUEUE}(Q)$ ,  $C \leftarrow C \cup \{x\}$ 
5:   for all  $y \in N_\theta(x) \cap C$  do
6:      $E \leftarrow E \cup \{xy\}$ 
7:     for all  $z \in N_{D(x,y)}(x) \cap N_{D(x,y)}(y)$  do
8:       if  $z$  is unmarked and  $z \in \Delta_=(x, y)$  then
9:         mark  $z$ ,  $\text{ENQUEUE}(Q, z)$ 
10: return  $\mathcal{D} = \{V(Z) : Z \text{ is a connected component of } G_\theta = (C, E)\}$ 

```

Let $\hat{\rho}_\theta^0(A) = A$ and $\hat{\rho}_\theta^{i+1}(A) = \hat{\rho}_\theta(\hat{\rho}_\theta^i(A))$ for all $i \in \mathbb{N}$ and $A \subseteq X$. Since $\hat{\rho}_\theta$ is monotone by Lemma 1 and X is finite, for all $A \subseteq X$ there exists a positive integer $\gamma(A)$ such that $\hat{\rho}_\theta^{\gamma(A)}(A) = \hat{\rho}_\theta^{\gamma(A)+1}(A)$, implying

$$\hat{\rho}_\theta^{\gamma(A)}(A) = \hat{\rho}_\theta^{\gamma(A)+l}(A) \quad (3)$$

for all $l \geq 0$. Furthermore, $\Gamma = \max\{\gamma(A) : A \subseteq X\} < \infty$. In the theorem below we claim that $\hat{\rho}_\theta^\Gamma$ yields exactly ρ_θ .

Theorem 3. *Let (X, D) be a finite metric space and $\theta \geq 0$. Then*

- (i) $\rho : 2^X \rightarrow 2^X$ with $\rho(A) = \hat{\rho}_\theta^\Gamma(A)$ for all $A \subseteq X$ is a closure operator and
- (ii) for all $A \subseteq X$, $\rho_\theta(A) = A$ iff $\rho(A) = A$.

We now consider the problem of computing weakly convex sets for the case that the metric space is *finite* and weakly convex sets are represented *extensionally*. More precisely, we are interested in the following problem setting:

Problem 2 (The Extensional Weakly Convex Hull Problem). Given a finite metric space $\mathcal{M} = (X, D)$ with $|X| = n$, a set $A \subseteq X$, and a threshold $\theta \geq 0$, compute the θ -decomposition A_1, \dots, A_ℓ of $\rho_\theta(A)$, where the A_i s are given extensionally.

The algorithm solving Problem 2 is given in Alg. 1. Its input consists of a set $A \subseteq X$ for some finite metric space (X, D) and a non-negative real number θ . The algorithm assumes that the pairwise distances for (X, D) are given explicitly and that each element $x \in X$ is associated with a sorted sequence S_x of pairs $(x', D(x, x'))$, for all $x' \in X \setminus \{x\}$. We assume that these sequences are calculated and stored once in a preprocessing step. The reason behind this assumption is that in order to solve a related consistency problem defined below, Alg. 1 will

be called with different values of θ . For any $\delta \geq 0$, these sequences allow the δ -neighborhood $N_\delta(x) = \{y \in X : D(x, y) \leq \delta\}$ of a point $x \in X$ to be calculated in time $\mathcal{O}(|N_\delta(x)|)$ for all $\delta \geq 0$ (cf. lines 5 and 7 of Alg. 1).

Alg. 1 maintains three variables. In particular, one can show in the proof of the theorem below, the set $\rho_\theta(A)$ is calculated in C . All elements of C are added first to the queue Q , which is initialized with A (cf. line 1). The elements of Q are processed one by one (cf. lines 4–9). In particular, for the element x of Q considered in line 4, we move x from Q to C (line 4) and take all elements y in the θ -neighborhood of x that have already been added to $C \subseteq \rho_\theta(A)$ (line 5). In variable E we maintain the set of edges of the θ -neighborhood graph over C , i.e., two elements of C are connected by an edge iff their distance is at most θ . As x is a new element in C , in line 6 we connect it with all y considered in line 5. In lines 7–9 we take all $z \in W_\theta(x, y)$ that have not yet been considered, mark z , and add it to the queue Q . Regarding line 7, note that the triangle inequality implies that if $z \in W_\theta(x, y)$ then $D(x, z), D(y, z) \leq D(x, y)$. Finally, after we have processed all elements that have been added to Q , in line 10 we calculate the connected components of the θ -neighborhood graph $G_\theta = (C, E)$ and return the family formed by the sets of vertices of the connected components.

Theorem 4. *Alg. 1 is correct and solves Problem 2 in $\mathcal{O}(nd^2)$ time, where d is the degree of the θ -neighborhood graph $G_\theta = (C, E)$.*

In Section 4.1 we will be concerned with an application scenario of the following *consistent hypothesis finding (CHF)* problem:

Problem 3 (The CHF Problem for Extensional Weakly Convex Hulls). Given a finite metric space $\mathcal{M} = (X, D)$ with $|X| = n$, disjoint sets $E^+, E^- \subseteq X$ of positive and negative examples, and an integer $k > 0$, return “YES” and the θ -decomposition of a θ -convex set consistent with E^+ and E^- that consists of at most k blocks, if it exists for some θ ; o/w the answer “NO”.

Remark 1. Note that if Problem 3 can be solved in polynomial time then, as k cannot be greater than $|E^+|$ by (i) of Corollary 1, a consistent hypothesis with the *smallest* number of blocks can be found in polynomial time. It always exists, as $\rho_0(E^+) = E^+$ and $E^+ \cap E^- = \emptyset$ by assumption.

Theorem 5. *Problem 3 can be solved in $\mathcal{O}(T_P(\mathcal{M}) + n^3 \log n)$ time, where $T_P(\mathcal{M})$ is the time complexity of computing all pairwise distances for X .*

4.1 Application Scenario: Vertex Classification

As a proof of concept, in this section we *empirically* demonstrate the learnability of weakly convex concepts over *graphs*. More precisely, we consider the metric space $\mathcal{M} = (V, D_g)$ for some undirected graph $G = (V, E)$, where D_g is the *geodesic* distance on V . In the learning setting, V is partitioned into V^+ and V^- , such that V^+ is θ -convex for some $\theta \geq 0$. The *target concept* V^+ as well as θ are *unknown* to the learning algorithm. The problem we investigate empirically

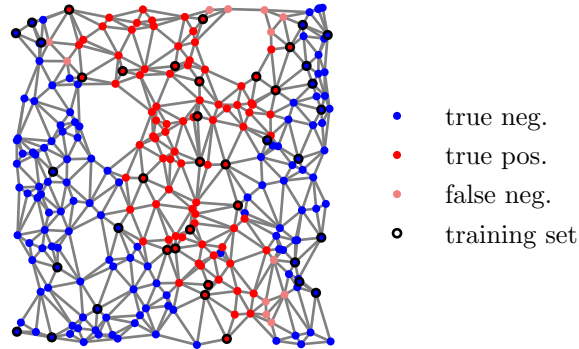


Fig. 2. Example of a graph with 250 vertices and 40 training examples.

is to approximate V^+ given some *small* labeled set $E = E^+ \cup E^-$ of positive and negative examples.

We solve this learning task by computing the hypothesis $C = \rho_{\theta'}(E^+)$ for the greatest $\theta' \leq \max_{u,v \in V} D(u,v)$ that is consistent with E . Such a θ' always exists (cf. Remark 1). Furthermore, C is computed by performing a binary search for θ' (see [13] for the details). To measure the predictive performance, we use *accuracy* (i.e., number of correctly classified vertices in $V \setminus E$ over $|V \setminus E|$) and compare it to the *baseline* $\max\{|V^+|/|V|, |V^-|/|V|\}$ defined by majority vote. We stress that the only purpose of these experiments is to empirically demonstrate that weakly convex concepts can be learned with a remarkable accuracy, *without* utilizing any domain specific properties and with using only a *few* training examples. An adaptation of our approach to the domain specific problem of learning on graphs and a rigorous empirical comparison of its predictive performance with state-of-the-art problem specific algorithms goes far beyond the scope of this paper.

We generated 50 random graphs for $|V| = 100, 250, 1000$, and 2500 for the experiments as follows: According to Prop. 1, the diameter of a graph is an upper bound on the parameter θ . In order to provide a diverse set of target concepts and possible values for θ , we generated random graphs based on *Delaunay triangulations* [4] as follows: After choosing the respective number of nodes $V \subset [0, 1]^2$ uniformly at random, we have computed the Delaunay triangulation. We then connected two nodes in V by an (undirected) edge iff they co-occur in at least one simplex of the triangulation. We considered the two cases that the edges are unweighted or they are weighted with the Euclidean distance between their endpoints. However, the resulting graph often contains a small number of very long edges (in terms of the Euclidean distance), especially near the “outline” of the chosen point set. Since such edges reduce the graph’s diameter substantially, we removed the longest 5% of the edges, i.e., those that are *not* contained in the 95th percentile w.r.t. the Euclidean distance of their endpoints.

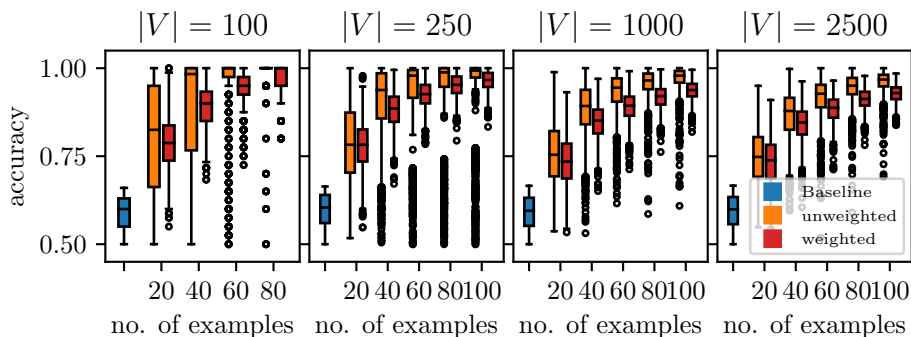


Fig. 3. Results for Delaunay-based graphs for varying number of vertices ($|V|$).

For each graph $G = (V, E)$ in the resulting dataset, we have generated random partitionings (V^+, V^-) of V in a way that V^+ and V^- are balanced (i.e., $|V^+| \approx |V^-|$) and V^+ is θ -convex. We note that for all random partitionings obtained, V^+ was not strongly (i.e., ordinary) convex. The training examples E^+ and E^- have been sampled uniformly at random from V^+ and V^- , respectively, such that $|E^+| \approx |E^-|$. The number of training examples (i.e., $|E^+ \cup E^-|$) was varied over 20, 40, 60, 80, 100. This overall procedure generates 5,000 learning tasks (50 graphs \times 20 random target concepts \times 5 training set sizes), for each graph size $|V| = 100, 250, 1000, 2500$. In Fig. 2 we give an example graph with $|V| = 250$, together with the node prediction using 40 training examples. The training examples are marked with black outline and the predictions are encoded by colors. In particular, dark red corresponds to true positive, dark blue to true negative, and light red to false negative nodes. In the example we have no false positive node, which was the case for most graphs.

Fig. 3 shows the accuracy (y -axes) of the baseline (blue box plots) and our learner (orange box plots for unweighted and red ones for weighted edges) grouped by the number of provided examples (x -axes) and the graph sizes $|V|$. In all cases, our learner outperforms the baseline significantly by noting that for $|V| = 100$, the high accuracy results obtained from 60 training examples are less interesting. For $|V| > 100$ it is remarkable that the learner does *not* require much more examples with increasing graph size. For example, for graphs with 2,500 vertices, already 80 examples are sufficient to achieve an average accuracy of 0.94 for unweighted graphs. Notice that the baseline is in all cases very close to 0.6. This is due to our construction of the target concepts: We chose θ maximal such that $2|V^+| < |V|$. Therefore, in almost all cases there are about 10% less positive nodes than negative. We have tested the generated weakly convex sets for strong convexity: almost all of them were *not* strongly convex. In summary, our experimental results clearly show that a remarkable predictive accuracy can be obtained already with relatively small training sets with our generic approach, without utilizing any domain specific knowledge.

Algorithm 2 INTENSIONAL WEAKLY CONVEX HULL (NAÏVE)

Require: well-behaved metric space $\mathcal{M} = (X, D)$ and representation scheme μ for \mathcal{M} **Input:** $A \in [X]^{<\omega}$ and $\theta \geq 0$ **Output:** $\mu(\theta, A_1), \dots, \mu(\theta, A_\ell)$, where A_1, \dots, A_ℓ is the θ -decomposition of $\rho_\theta(A)$

- 1: $\mathcal{D} \leftarrow \{\mu(\theta, \{x\}) : x \in A\}$
 - 2: **while** $\exists B_i, B_j \in \mathcal{D}$ such that $B_i \neq B_j$ and $\overline{D}(B_i, B_j) \leq \theta$ **do**
 - 3: $\mathcal{D} \leftarrow (\mathcal{D} \setminus \{B_i, B_j\}) \cup \{\text{MERGE}(\theta, A, B_i, B_j)\}$
 - 4: **return** \mathcal{D}
-

5 The Intensional Problem Setting

In this section we consider the *intensional* problem setting, that is, the scenario that weakly convex sets have some compact representation. In contrast to the extensional case, the metric spaces in this section are allowed to be *infinite*. They are, however, required to be *well-behaved* (see Sect. 3 for the definition). To formulate the problem setting considered in this section, we introduce the following notion for a metric space $\mathcal{M} = (X, D)$: A *representation scheme* for \mathcal{M} is a function $\mu : \mathbb{R}_{\geq 0} \times [X]^{<\omega} \rightarrow \{0, 1\}^*$ satisfying $\mu(\theta, A) = \mu(\theta, B)$ iff $\rho_\theta(A) = \rho_\theta(B)$ for all $A, B \in [X]^{<\omega}$ and $\theta \geq 0$. In other words, μ returns some representation of $\rho_\theta(A)$ for all finite subsets $A \subseteq X$. Note that $\rho_\theta(A)$ can be infinite. Analogously to Problem 2, we are interested in the following computational problem:

Problem 4 (The Intensional Weakly Convex Hull Problem). Given a well-behaved metric space $\mathcal{M} = (X, D)$, a representation scheme μ for \mathcal{M} , a set $A \subseteq [X]^{<\omega}$ with $|A| = m$, and $\theta \geq 0$, compute $\mu(\theta, A)$.

For page limitation, we give a very simple naïve algorithm for Problem 4 (see Alg. 2), by noting that it is not optimal. It assumes a well-behaved metric space $\mathcal{M} = (X, D)$ and some representation scheme μ for \mathcal{M} . The input to the algorithm consists of a finite subset $A \subseteq X$ and a distance threshold $\theta \geq 0$. Its output is the set $\{\mu(\theta, A_1), \dots, \mu(\theta, A_\ell)\}$ of binary strings representing the blocks A_1, \dots, A_ℓ in the θ -decomposition of $\rho_\theta(A)$. The algorithm first initializes the variable \mathcal{D} with the set of the representations of $\rho_\theta(\{x\}) = \{x\}$ for all $x \in A$ (cf. line 1). It then iteratively selects two different blocks $B_i, B_j \in \mathcal{D}$ such that $\overline{D}(B_i, B_j) = \min_{x \in B_i, y \in B_j} D(x, y) \geq \theta$. If there are no such B_i and B_j , then it terminates by returning \mathcal{D} ; o/w it updates \mathcal{D} by removing B_i, B_j and adding their merge defined by $\text{MERGE}(\theta, A, B_i, B_j) = \mu(\theta, (\text{ext}(B_i) \cup \text{ext}(B_j)) \cap A)$ if $\overline{D}(B_i, B_j) \leq \theta$; o/w $\text{MERGE}(\theta, A, B_i, B_j) = \perp$, where $\text{ext}(B_i), \text{ext}(B_j)$ denote the extensions of B_i, B_j , respectively. The proof of the proposition below follows by induction on $|\mathcal{D}|$ from Thm. 2 and Corollary 1.

Proposition 2. *Alg. 2 is correct.*

Let T_S, T_D , and T_M denote the time complexity of computing $\mu(\theta, \{x\})$, the distance between B_i and B_j , and the merge of B_i and B_j , respectively, for any

$x \in X$ and θ -blocks B_i and B_j . One can easily check that the time complexity of Alg. 2 is $\mathcal{O}(mT_S + m^3T_D + mT_M)$. Using a more sophisticated version of Alg. 2 not presented in this paper for space limitations, we have the following improved complexity result.

Theorem 6. *Problem 4 can be solved in time $\mathcal{O}(mT_S + m^2T_D + mT_M)$.*

We consider the consistency problem also for the intensional scenario.

Problem 5 (The Consistency Problem for Intensional Weakly Convex Hulls). Given a well-behaved metric space $\mathcal{M} = (X, D)$, representation scheme μ for \mathcal{M} , disjoint finite sets $E^+, E^- \subseteq X$ of labeled examples with $|E^+ \cup E^-| = m$, and $k > 0$, return “YES” and the representations of the blocks in the θ -decomposition of a θ -convex set that is consistent with E^+ and E^- and has at most k blocks, if such a decomposition exists for some $\theta > 0$; o/w the answer “NO”.

Note that Remark 1 applies also to the problem above. Using the same idea as for the solution of Problem 3 (i.e., to decide whether a desired θ exists, we perform a binary search on the sorted set of pairwise distances between the elements in A), we have the following result on the above problem:

Theorem 7. *Problem 5 can be solved in $\mathcal{O}((mT_S + m^2T_D + mT_M) \log m)$ time.*

In Sections 5.1 and 5.2 below we present two non-trivial applications of Thm. 7 to polynomial PAC-learnability.

5.1 Learning Weakly Convex Boolean Functions

As a first application of Thm. 7, we show that the concept class formed by *weakly convex Boolean functions* is efficiently PAC-learnable. This result is not new, it has been obtained with a *domain specific* algorithm in [6]. Still, we present it as an application because, as we show below, we can derive it in a very simple way by using Thm.7. Furthermore, our *general purpose* algorithm solving Problem 4 has the same asymptotic complexity on this problem as the *domain specific* one in [6].

Consider the metric space $\mathcal{M}_H = (\mathbb{H}_n, D_H)$ for some $n \in \mathbb{N}$ (see Section 2). Clearly, the finiteness of \mathbb{H}_n implies that \mathcal{M}_H is well-behaved for all $\theta \geq 0$. A Boolean function $f : \{0, 1\}^n \rightarrow \{0, 1\}$ ($n \in \mathbb{N}$) is θ -convex for some $\theta \geq 0$ if for all $x, y, z \in \mathbb{H}_n$, $f(z) = 1$ whenever $f(x) = f(y) = 1$, $D_H(x, y) \leq \theta$, and $z \in \Delta_{=} (x, y)$. Note that for \mathcal{M}_H it suffices to consider the values in $[n]$ for θ .

Throughout this section we will use the following notation: L_n denotes the set $\{x_1, \neg x_1, \dots, x_n, \neg x_n\}$ of Boolean literals. A *term* T is a conjunction of literals from L_n ; T is sometimes regarded as the set of literals it contains. A *conflict* between two terms T_i and T_j over L_n is an integer $p \in [n]$ such that $x_p \in T_i$ and $\neg x_p \in T_j$ or vice versa. Finally, for a Boolean function f , $\text{ext}(f)$ denotes the *extension* of f (i.e., $\text{ext}(f) = \{x \in \mathbb{H}_n : f(x) = 1\}$). We will use the following auxiliary result:

Lemma 2. *Let $A \subseteq \mathbb{H}_n$ be θ -convex and θ -connected for some $\theta \geq 2$. Then A is convex (i.e., n -convex) and can be represented by a term T over L_n .*

For any $n > 0$, the concept class $\mathcal{B}_{n,k} \subseteq 2^{\mathbb{H}_n}$ is defined as follows: For all $A \subseteq \mathbb{H}_n$, $A \in \mathcal{B}_{n,k}$ iff A is θ -convex for some $\theta \geq 0$ and its θ -decomposition has at most k blocks. Thm. 2 and Lemma 2 together imply that any such weakly convex set A can be represented *uniquely* by a k -term DNF F such that the extensions of the terms in F represent precisely the blocks in the θ -decomposition of A . Since the blocks are non-empty, no term contains a variable and its negation. This gives rise to the following definition of the representation scheme μ for \mathcal{M}_H : For all $S \subseteq \mathbb{H}_n$, define $\mu_H : \mathbb{R}_{\geq 0} \times \mathbb{H}_n \rightarrow \{0, 1\}^*$ by $\mu_H(\theta, S) = F$, where F is the *unique* DNF representation of $\rho_\theta(S)$, if $\rho_\theta(S)$ is θ -connected; o/w by \perp .

Lemma 3. *Problem 5 can be solved in $\mathcal{O}(nm^2 \log m)$ time for \mathcal{M}_H .*

Proof. Let μ in Problem 5 be defined by μ_H . We show that T_S, T_D, T_M in Thm. 7 are all in $\mathcal{O}(n)$. For T_S , the claim follows from $\mu_H(\theta, \{x\}) = \bigwedge_i l_i$, where $l_i = x_i$ if $x_i = 1$; o/w $l_i = \neg x_i$. Let T_i and T_j be terms over L_n . Their distance $\overline{D}_H(T_i, T_j)$ is equal to the number of conflicts between T_i and T_j , implying $T_D \in \mathcal{O}(n)$. Finally, if $\overline{D}_H(T_i, T_j) \leq \theta$ then $\text{MERGE}(T_i, T_j)$ is the term T with literals $T_i \cap T_j$. Thus, $T_{\text{MERGE}} = \mathcal{O}(n)$. The statement then follows by Thm. 7. \square

Theorem 8. *For all $d, k \geq 0$, $\mathcal{B}_{n,k}$ is polynomially PAC-learnable.*

Proof. Since $\mathcal{B}_{n,k} \subseteq (\mathcal{B}_{n,1})_{\cup}^k$, $\text{VC-dim}(\mathcal{B}_{n,k}) \leq \text{VC-dim}((\mathcal{B}_{n,1})_{\cup}^k) \leq 4nk \log(3k)$ by $\text{VC-dim}(\mathcal{B}_{n,1}) \leq 2n$ and by (ii) of Thm. 1. Hence, the VC-dimension of $\mathcal{B}_{n,k}$ is polynomial in n and k . Furthermore, by Lemma 3, the consistency problem for $\mathcal{B}_{n,k}$ can be solved in time polynomial in n , k , and $m = |E^+ \cup E^-|$. The theorem then follows by (i) of Thm. 1. \square

Note that if the extensions of the terms in the DNF are *not* required to be pairwise disjoint then, in contrast to the positive result in Thm. 8, k -term DNF formulas are *not* polynomially PAC-learnable for any $k \geq 2$ if $P \neq RP$ [10]. In [6] it is shown that the class of θ -convex Boolean functions is not polynomially PAC-learnable for $\theta > n/2 - 1$. The reason is that the number of terms having a pairwise distance greater than $n/2 - 1$ can be exponential in n . Notice that the number of terms in $\mathcal{B}_{n,k}$ is bounded by the *parameter* k in our problem setting. Finally we note that the time complexity of the *domain specific* algorithm in [6] that solves Problem 4 for \mathbb{H}_n is $\mathcal{O}(m^2n)$, which is the same as that of the sophisticated version of our *general purpose* Algorithm 2 in [13].

5.2 Learning Weakly Convex Axis-Aligned Hyperrectangles

Our second application of Thm. 7 is concerned with polynomial PAC-learnability of weakly convex sets in $\mathcal{M}_R = (\mathbb{R}^d, D_1)$, where D_1 is the *Manhattan* (or L_1) distance, i.e., $D_1(x, y) = \sum_i |x_i - y_i|$ for all $x = (x_1, \dots, x_d)$ and $y = (y_1, \dots, y_d) \in \mathbb{R}^d$. Note that \mathcal{M}_R can be regarded as a generalization of \mathcal{M}_H considered in the

previous section, as D_1 becomes equal to D_H over the domain $\mathbb{H}_d \subseteq \mathbb{R}^d$. Clearly, \mathcal{M}_R is complete. Furthermore, for all $x, y, z \in \mathbb{R}^d$, $D_1(x, z) + D_1(z, y) = D_1(x, y)$ iff z belongs to the smallest axis-aligned (topologically) closed hyperrectangle in \mathbb{R}^d that contains x and y . This implies that all axis-aligned closed hyperrectangles are θ -convex for all $\theta > 0$ and $\rho_\theta(A)$ is closed for all finite subsets $A \subseteq \mathbb{R}^d$.

All concepts in the concept class $\mathcal{R}_{d,k}$ considered in this section are defined by the union of at most k pairwise disjoint axis-aligned closed hyperrectangles in \mathbb{R}^d , for some $d, k > 0$. More precisely, for all $R \subseteq \mathbb{R}^d$, $R \in \mathcal{R}_{d,k}$ iff R is θ -convex for some $\theta > 0$ with respect to \mathcal{M}_R and the θ -decomposition of R consists of at most k blocks (i.e., axis-aligned closed hyperrectangles). For all finite sets $S \in [\mathbb{R}^d]^{<\omega}$, define $\mu_R : \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \times [\mathbb{R}^d]^{<\omega} \rightarrow \{0, 1\}^*$ by $\mu_R(\theta, S) = (S_{\min}, S_{\max})$ if $\rho_\theta(S)$ is θ -connected; o/w by \perp ,⁴ where S_{\min} (resp. S_{\max}) denotes the componentwise minimum (resp. maximum) of the points in S .

Lemma 4. *Problem 5 can be solved in $\mathcal{O}(dm^2 \log m)$ time for $\mathcal{M} = (\mathbb{R}^d, D_1)$.*

Proof. Let μ in Problem 5 be defined by μ_R . We prove the claim by showing that $T_S, T_D, T_{\text{MERGE}}$ in Thm. 7 are all in $\mathcal{O}(d)$. In particular, $T_S \in \mathcal{O}(d)$ follows from $\mu_R(\theta, \{x\}) = (x, x)$. Let B_i (resp. B_j) be an axis-aligned closed hyperrectangle, $u = \min B_i$, and $v = \max B_i$ (resp. $x = \min B_j$ and $y = \max B_j$). We have $T_D \in \mathcal{O}(d)$ by the fact that $\overline{D}_1(B_i, B_j) = \sum_{i=1}^d D'_1([u_i, v_i], [x_i, y_i])$, where $D'_1([u_i, v_i], [x_i, y_i]) = \min\{|x_i - v_i|, |u_i - y_i|\}$ if $[u_i, v_i] \cap [x_i, y_i] \neq \emptyset$; o/w $D'_1([u_i, v_i], [x_i, y_i]) = 0$. Finally, if $\overline{D}(B_i, B_j) \leq \theta$ then $\text{MERGE}(B_i, B_j)$ is the smallest axis-aligned closed hyperrectangle containing $\min\{u, x\}$ and $\max\{v, y\}$, implying $T_{\text{MERGE}} = \mathcal{O}(d)$. The claim then follows by Thm. 7. \square

Theorem 9. *For all $d, k \geq 0$, $\mathcal{R}_{d,k}$ is polynomially PAC-learnable.*

Proof. Since $\mathcal{R}_{d,k} \subseteq (\mathcal{R}_{d,1})_{\cup}^k$, $\text{VC-dim}(\mathcal{R}_{d,k}) \leq \text{VC-dim}((\mathcal{R}_{d,1})_{\cup}^k) \leq 4dk \log(3k)$ by $\text{VC-dim}(\mathcal{R}_{d,1}) = 2d$ and by (ii) of Thm. 1. Hence, the VC-dimension of $\mathcal{R}_{d,k}$ is polynomial in d and k . Furthermore, by Lemma 4, the consistency problem for $\mathcal{R}_{d,k}$ can be solved in time polynomial in d, k , and $|E^+ \cup E^-|$. Thus, the theorem follows by (i) of Theorem 1. \square

While Lemma 4 implies that a consistent hypothesis that has the smallest number of pairwise disjoint axis-aligned d -dimensional closed hyperrectangles can be found in polynomial time for all $d \geq 1$, this problem becomes NP-complete even for $d = 2$, if disjointness is not required (see, e.g., [1]).

6 Concluding Remarks

The theoretical and experimental results of this paper demonstrate the usefulness of weakly convex sets for machine learning. While our focus in this paper was solely on applications to *machine learning*, weakly convex sets seem to be useful for *data mining* applications (e.g., itemset mining, subgroup discovery) as well.

⁴ We assume that real numbers are represented in $\mathcal{O}(1)$ space upto a certain precision.

The notion of weak convexity can be uninteresting for certain metric spaces. For example, for finite subspaces of (\mathbb{R}^d, D_2) , $\Delta_=(x, y) = \{x, y\}$ holds almost surely for all points x and y . To overcome this problem, in the long version of this paper we allow the triangle inequality to hold up to some tolerance ε instead of equality. All results of Sect. 3 can be generalized to this relaxed definition.

There are several interesting questions for further research. Note, for example, that Alg. 2 is very similar to *single linkage clustering*, raising the following question: Can the time complexity in Thm. 7 be further improved by using techniques (e.g., in [12]) designed for single linkage clustering algorithms?

Acknowledgments

This work was partly supported by the Ministry of Education and Research of Germany (BMBF) under project ML2R (grant number 01/S18038C) and by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy - EXC 2070 - 390732324.

References

1. Bereg, S., Cabello, S., Díaz-Báñez, J., Pérez-Lantero, P., Seara, C., Ventura, I.: The class cover problem with boxes. *Computational Geometry* **45**(7), 294–304 (2012)
2. Blumer, A., Ehrenfeucht, A., Haussler, D., Warmuth, M.K.: Learnability and the vapnik-chervonenkis dimension. *J. ACM* **36**(4), 929–965 (1989)
3. Davey, B.A., Priestley, H.A.: *Introduction to Lattices and Order*. Cambridge University Press, 2nd edn. (2002)
4. Delaunay, B.N.: Sur la sphère vide. *Bull. Acad. Sci. URSS* (6), 793–800 (1934)
5. Edelsbrunner, H., Mücke, E.P.: Three-dimensional alpha shapes. *ACM Trans. Graph.* **13**(1), 43–72 (1994)
6. Ekin, O., Hammer, P.L., Kogan, A.: Convexity and logical analysis of data. *Theoretical Computer Science* **244**(1), 95 – 116 (2000)
7. Ester, M., Kriegel, H.P., Sander, J., Xu, X., et al.: A density-based algorithm for discovering clusters in large spatial databases with noise. In: *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD)*. vol. 96, pp. 226–231. AAAI Press (1996)
8. Hemmer, M., Portaneri, C., Alliez, P.: Alpha Hulls. Research report (Nov 2020)
9. Kearns, M.J., Vazirani, U.V.: *An Introduction to Computational Learning Theory*. MIT Press (1994)
10. Pitt, L., Valiant, L.G.: Computational limitations on learning from examples. *J. ACM* **35**(4), 965–984 (1988)
11. Seiffarth, F., Horváth, T., Wrobel, S.: Maximal closed set and half-space separations in finite closure systems. In: *Proc. of ECML PKDD 2019. Lecture Notes in Computer Science*, vol. 11906, pp. 21–37. Springer (2019)
12. Sibson, R.: SLINK: an optimally efficient algorithm for the single-link cluster method. *Comput. J.* **16**(1), 30–34 (1973)
13. Stadtländer, E., Horváth, T., Wrobel, S.: Learning weakly convex sets in metric spaces. *CoRR* **abs/2105.06251** (2021), <https://arxiv.org/abs/2105.06251>
14. Valiant, L.G.: A theory of the learnable. *Commun. ACM* **27**(11), 1134–1142 (1984)
15. van de Vel, M.: *Theory of Convex Structures*, North-Holland Mathematical Library, vol. 50. Elsevier (1993)