

Preference-based Teaching of Unions of Geometric Objects

Ziyuan Gao

Department of Computer Science, University of Regina

GAO257@CS.UREGINA.CA

David Kirkpatrick

Department of Computer Science, University of British Columbia

KIRK@CS.UBC.CA

Christoph Ries

Department of Mathematics, Ruhr-University Bochum

CHRISTOPH.RIES@RUB.DE

Hans Simon

Department of Mathematics, Ruhr-University Bochum

HANS.SIMON@RUB.DE

Sandra Zilles

Department of Computer Science, University of Regina

ZILLES@CS.UREGINA.CA

Editors: Steve Hanneke and Lev Reyzin

Abstract

This paper studies exact learning of unions of non-discretized geometric concepts in the model of preference-based teaching. In particular, it focuses on upper and lower bounds of the corresponding sample complexity parameter, the preference-based teaching dimension (PBTd), when learning disjoint unions of a bounded number of geometric concepts of various types – for instance balls, axis-aligned cubes, or axis-aligned boxes – in arbitrary dimensions. It is shown that the PBTd of disjoint unions of some such types of concepts grows linearly with the number of concepts in the union, independent of the dimensionality. Teaching the union of potentially overlapping objects turns out to be more involved and is hence considered here only for unions of up to two objects.

Keywords: preference-based teaching dimension, geometric concept classes, learning unions of concepts

1. Introduction

When designing machine learning algorithms, a natural task is to combine approaches for learning certain types of concepts into methods for learning unions of such concepts. Likewise, when analyzing machine learning models, a typical objective is to compare the complexity of learning certain types of concepts to the complexity of learning their unions. Already in the early days of research on computational learning theory, learning unions of learnable concepts was a topic of interest, as witnessed for example by publications on PAC-learning (Blumer et al., 1989; Long and Warmuth, 1994), on online learning (Maass and Turán, 1994), on learning from equivalence and membership queries (Goldberg et al., 1994), on learning in the limit (Wright, 1989), and on learning from teachers (Frazier et al., 1994). Not only various learning models, but also concept classes of numerous kinds were studied in this context, for instance geometric concept classes (Maass and Turán, 1994; Blumer et al., 1989), or, more recently, classes of formal languages (Ng and Shinohara, 2005), specific classes of recursively enumerable sets (Jain et al., 2007), and classes of recursive sets (Ouchi et al., 2017). Taking the union is a simple and natural way of forming a composite concept from ba-

sic ones, and yet learning unions typically requires more than just a straightforward generalization of a basic learning algorithm.

This paper studies the complexity of *exactly teaching unions of geometric concepts*. In the past, research on exact learning of geometric concepts typically dealt with finite classes of geometric concepts, by discretizing the input space. In particular, most existing models of algorithmic teaching are not suited to classes of non-discretized geometric objects. For example, the unique (exact) identification of a rectangle among the set of all axis-aligned rectangles over the rational numbers is not possible with a finite number of examples, either in the classical teaching dimension model (Goldman and Kearns, 1995; Shinohara and Miyano, 1991) or in the subsequently introduced model of recursive teaching (Zilles et al., 2011). By contrast though, the very recent model of *preference-based teaching* makes natural classes of non-discretized geometric concepts learnable. In preference-based teaching, the teacher and learner agree on a preference relation over the concepts in the class. After the teacher presents a set T of labeled examples to the learner, the latter selects the most preferred hypothesis that is consistent with T (Gao et al., 2017). For instance, to teach axis-aligned rectangles in \mathbb{R}^2 or in \mathbb{Q}^2 , the teacher and learner agree to prefer concepts of smaller area over those of larger area. It is now sufficient for the teacher to present two positively labeled points that are diagonally opposite in the target rectangle, as the learner will then correctly hypothesize the unique smallest axis-aligned rectangle containing these two points.

The worst-case number of examples a teacher needs to present to the learner when teaching any concept in a concept class \mathcal{L} is called the preference-based teaching dimension (PBTd) of \mathcal{L} . The PBTd of the class of all axis-aligned rectangles in \mathbb{R}^2 , for example, is 2. The focus of this paper is on the following question: assuming the PBTd of \mathcal{L} is finite, can we determine interesting upper and lower bounds on the PBTd of the class of unions of up to k concepts from \mathcal{L} , for a given finite number k ? We address this question for cases when \mathcal{L} is a class of geometric objects over the Euclidean domain, including, for instance, balls in \mathbb{R}^d , axis-aligned cubes in \mathbb{R}^d , or axis-aligned boxes in \mathbb{R}^d .

The problem of determining upper bounds on the PBTd of the class of unions of objects taken from a base class turns out to be non-trivial even when we focus on unions of mutually disjoint objects: given positive and negative examples for a union of up to k objects, the learner must find the right clustering of the examples (one cluster for each of the $\ell \leq k$ objects). For general (possibly non-disjoint) unions, this problem becomes quite involved even when the union of only two objects is considered. We present some general results on the PBTd of disjoint unions of objects, as well as some preliminary results on the non-disjoint case. In particular, it is shown that the PBTd of disjoint unions of objects grows linearly with the number of objects in the union, independent of the dimensionality, if each object is a uniformly scaled and translated version of a fixed convex body.

As is not uncommon in the field of computational learning theory, we also address the question of learning in situations in which negative examples are not available. Here we obtain that teaching with $k + 1$ positive examples is optimal for the admittedly very constrained case in which exactly k translated, mutually disjoint, and equally scaled copies of the same convex body must be learned.

The fact that preference-based teaching puts no restrictions on the preference relation over concepts often makes reasoning about lower bounds complicated (Gao et al., 2017). For the same reason, the freedom of choosing arbitrary preference relations makes the PBTd-model very powerful. Using a cardinality argument, we establish a new general upper bound on the PBTd which yields the surprising result that the PBTd of any family of plane algebraic curves of bounded degree (including lines and conics) is one.

2. Preliminaries

A *concept class* \mathcal{L} is a family of subsets over a domain \mathcal{X} , i.e., $\mathcal{L} \subseteq 2^{\mathcal{X}}$ where $2^{\mathcal{X}}$ denotes the powerset of \mathcal{X} . We write $\mathcal{L}' \subseteq_f \mathcal{L}$, if \mathcal{L}' is a finite subclass of \mathcal{L} . The elements of \mathcal{L} are called *concepts*. A *labeled example* is an element of $\mathcal{X} \times \{0, 1\}$. Elements of \mathcal{X} are called *examples*. For any set of labeled examples $S \subseteq \mathcal{X} \times \{0, 1\}$, we use the notation $S_{\mathcal{X}}$ to refer to the set of examples occurring in S , i.e., $S_{\mathcal{X}} = \{x \in \mathcal{X} \mid (x, \ell) \in S \text{ for some } \ell \in \{0, 1\}\}$. A set $L \subseteq \mathcal{X}$ is *consistent with* S if it includes all examples in S that are labeled “1” and excludes all examples in S that are labeled “0”. A set S of labeled examples such that L but not L' is consistent with S is said to *distinguish* L from L' . The classical model of teaching is then defined as follows.

Definition 1 (Shinohara and Miyano (1991); Goldman and Kearns (1995)) A teaching set for a concept $L \in \mathcal{L}$ w.r.t. \mathcal{L} is a set S of labeled examples such that L is the only concept in \mathcal{L} that is consistent with S , i.e., S distinguishes L from any other concept in \mathcal{L} . Define $\text{TD}(L, \mathcal{L}) = \inf\{|S| : S \text{ is a teaching set for } L \text{ w.r.t. } \mathcal{L}\}$. i.e., $\text{TD}(L, \mathcal{L})$ is the smallest possible size of a teaching set for L w.r.t. \mathcal{L} . If L has no finite teaching set w.r.t. \mathcal{L} , then $\text{TD}(L, \mathcal{L}) = \infty$. The number $\text{TD}(\mathcal{L}) = \sup_{L \in \mathcal{L}} \text{TD}(L, \mathcal{L}) \in \mathbb{N}_0 \cup \{\infty\}$ is called the teaching dimension of \mathcal{L} .

For technical reasons, we will occasionally deal with the number $\text{TD}_{\min}(\mathcal{L}) = \inf_{L \in \mathcal{L}} \text{TD}(L, \mathcal{L})$, i.e., the number of examples needed to teach the concept from \mathcal{L} that is easiest to teach. Let \mathcal{L}_1 and \mathcal{L}_2 be two concept classes whose domains are disjoint. The *direct sum* of \mathcal{L}_1 and \mathcal{L}_2 is defined as follows:

$$\mathcal{L}_1 \oplus \mathcal{L}_2 = \{L_1 \cup L_2 \mid L_1 \in \mathcal{L}_1 \wedge L_2 \in \mathcal{L}_2\} .$$

It is well known (Doliwa et al., 2014) that TD_{\min} acts additively on direct sums, i.e.,

$$\text{TD}_{\min}(\mathcal{L}_1 \oplus \mathcal{L}_2) = \text{TD}_{\min}(\mathcal{L}_1) + \text{TD}_{\min}(\mathcal{L}_2) . \quad (1)$$

In this paper, by “disjoint union” we mean “a union of mutually disjoint objects that are taken from the *same domain*”. This is quite different from “direct sum” where the objects participating in the union are taken from *mutually disjoint domains*. In the latter case, the student can form k clusters from the given labeled examples in a trivial manner so that we may invoke teaching protocols for the k individual classes participating in the direct sum. In the former case (with disjoint objects taken from the same domain), the right clustering of the labeled examples has to be taught as well.

We will be concerned with the teaching model suggested by Gao et al. (2017) in which the teacher and the student agree not only on a classification-rule system \mathcal{L} but also on a preference relation, denoted as \prec , imposed on \mathcal{L} . We assume that \prec is a *strict partial order* on \mathcal{L} , i.e., \prec is asymmetric and transitive. The partial order that makes every pair $L \neq L' \in \mathcal{L}$ incomparable is denoted by \prec_{\emptyset} . For every $L \in \mathcal{L}$, let

$$\mathcal{L}_{\prec L} = \{L' \in \mathcal{L} : L' \prec L\}$$

be the set of concepts over which L is strictly preferred. Note that $\mathcal{L}_{\prec_{\emptyset} L} = \emptyset$ for every $L \in \mathcal{L}$. In the sequel, we briefly repeat the basic definitions for this model of teaching.

As already noted above, a teaching set S of L w.r.t. \mathcal{L} distinguishes L from any other concept in \mathcal{L} . If a preference relation comes into play, then it is no longer necessary for S to distinguish L from the concepts in $\mathcal{L}_{\prec L}$ because L is strictly preferred over them anyway.

Definition 2 A teaching set for $L \in \mathcal{L}$ w.r.t. (\mathcal{L}, \prec) is defined as a teaching set for L w.r.t. $\mathcal{L} \setminus \mathcal{L}_{\prec L}$. Furthermore define

$$\text{PBTd}(L, \mathcal{L}, \prec) = \inf\{|S| : S \text{ is a teaching set for } L \text{ w.r.t. } (\mathcal{L}, \prec)\} \in \mathbb{N}_0 \cup \{\infty\} .$$

The number $\text{PBTd}(\mathcal{L}, \prec) = \sup_{L \in \mathcal{L}} \text{PBTd}(L, \mathcal{L}, \prec) \in \mathbb{N}_0 \cup \{\infty\}$ is called the teaching dimension of (\mathcal{L}, \prec) .

Definition 2 implies that

$$\text{PBTd}(L, \mathcal{L}, \prec) = \text{TD}(L, \mathcal{L} \setminus \mathcal{L}_{\prec L}) . \quad (2)$$

The classical model of teaching is obtained from the model described in Definition 2 when we plug in the empty preference relation \prec_{\emptyset} for \prec .

We are interested in finding the partial order that is optimal for the purpose of teaching and we aim at determining the corresponding teaching dimension. This motivates the following notion:

Definition 3 The preference-based teaching dimension of \mathcal{L} is given by

$$\text{PBTd}(\mathcal{L}) = \inf\{\text{PBTd}(\mathcal{L}, \prec) : \prec \text{ is a strict partial order on } \mathcal{L}\} . \quad (3)$$

We briefly note that i) $\text{PBTd}(\mathcal{L})$ is left unchanged when the infimum in (3) ranges over all strict linear orders (Gao et al., 2017), and ii) PBTd is sub-additive on direct sums of concept classes. The question of whether PBTd is additive on direct sums is open.

$\text{TD}^+(\mathcal{L})$ and $\text{PBTd}^+(\mathcal{L})$, called the *positive teaching dimension* of \mathcal{L} and *positive preference-based teaching dimension* of \mathcal{L} , are defined analogously for the setting where only positive examples can be employed by the teacher.

Classes over the Euclidean Domain. For $\mathcal{X} = \mathbb{R}^d$, we introduce the following classes of basic geometric objects:

- BOXES^d denotes the class of all d -dimensional boxes, i.e., the class of all d -dimensional axis-aligned hyper-rectangles.
- CUBES^d denotes the class of all d -dimensional axis-aligned cubes.
- BALLS^d denotes the class of all d -dimensional balls.

We are primarily interested in determining the PBTd of various classes containing unions of these basic objects.

Let $a = (a_1, a_2, \dots, a_d)$ and $b = (b_1, b_2, \dots, b_d)$ be points in \mathbb{R}^d . We say that a is *northern* of b (and that b is *southern* of a) if $a_d > b_d$. Let C denote a convex body in \mathbb{R}^d (i.e., a compact and convex subset of \mathbb{R}^d with a non-empty interior). A point (necessarily unique) in C that is northern of all other points in C is called the *north pole* of C . The notion *south pole* is understood analogously. Note that some convex bodies, for example elements of BOXES^d , do not have poles, but others, for example elements of BALLS^d do.

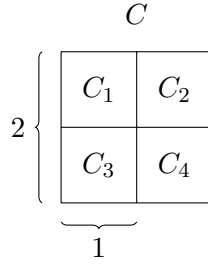


Figure 1: A finite subclass of CUBES² whose TD_{min} is at least 2; the class contains four small cubes, denoted C₁, . . . , C₄, and the larger cube C that equals their union.

3. A Simple General Lower Bound Technique

To lower-bound PBTD, it is often convenient to use the following simple observation from (Gao et al., 2017):

$$\text{PBTD}(\mathcal{L}) \geq \sup_{\mathcal{L}' \subseteq_f \mathcal{L}} \text{TD}_{\min}(\mathcal{L}') = \sup_{\mathcal{L}' \subseteq_f \mathcal{L}} \text{PBTD}(\mathcal{L}') . \quad (4)$$

This bound can be deployed to show that preference-based teaching of boxes or cubes in d dimensions is not possible using just a single example:

Example 1 For all $d \geq 1$, we have

$$\text{PBTD}(\text{BOXES}^d) = \text{PBTD}^+(\text{BOXES}^d) = \text{PBTD}(\text{CUBES}^d) = \text{PBTD}^+(\text{CUBES}^d) = 2 .$$

Proof $\text{PBTD}^+(\text{BOXES}^d) = 2$ was already shown by Gao et al. (2017). It now suffices to show that $\sup_{\mathcal{L} \subseteq_f \text{CUBES}^d} \text{TD}_{\min}(\mathcal{L}) > 1$. Suppose first that $d \geq 2$. Consider a finite class \mathcal{L} consisting of $2^d + 1$ cubes, one of which has side length 2 and forms the union of the remaining 2^d cubes, the latter having a side length of 1 each. See Figure 1 for an illustration in \mathbb{R}^2 . It is easy to see that $\text{TD}_{\min}(\mathcal{L}) > 1$: each point outside some cube in \mathcal{L} is also outside another cube in \mathcal{L} , and each point inside some cube in \mathcal{L} is also inside some other cube in \mathcal{L} . Hence, no singleton can be a teaching set for any concept in \mathcal{L} . For $d = 1$, boxes and cubes are intervals, and a similar argument as above works for the following finite subclass: $\{[0, 0], [1, 1], [2, 2], [0, 1], [1, 2], [0, 2]\}$. ■

The class of d -dimensional balls witnesses that the bound in Equation (4) is not tight:

Example 2 For every $d \geq 2$, we have $\text{PBTD}(\text{BALLS}^d) = \text{PBTD}^+(\text{BALLS}^d) = 2$ while

$$\sup_{\mathcal{L} \subseteq_f \text{BALLS}^d} \text{PBTD}(\mathcal{L}) = \sup_{\mathcal{L} \subseteq_f \text{BALLS}^d} \text{PBTD}^+(\mathcal{L}) = 1 .$$

Proof Let \mathcal{L} be a finite subclass of BALLS^d . Let us prefer ball B_1 over ball B_2 if the north pole of B_1 is northern of the north pole of B_2 . Among balls with the same north pole, let us prefer the smaller one. It is easy to see that, given this preference relation, we can teach each ball of a strictly positive radius by a single positive example that is sufficiently close, albeit not identical, to its north

pole. Each ball of radius 0 forms a singleton set, say $\{z\}$, and is taught by the example $(z, +)$. Thus $\sup_{\mathcal{L} \subseteq_f \text{BALLS}^d} \text{PBTD}^+(\mathcal{L}) = 1$.

It was shown by [Gao et al. \(2017\)](#) that a class containing all singletons can only then have a PBTD of 1, if it contains at most one additional concept on top of the singletons. Hence $\text{PBTD}(\text{BALLS}^d) \geq 2$. Finally, $\text{PBTD}(\text{BALLS}^d) = \text{PBTD}^+(\text{BALLS}^d) = 2$ is witnessed by a teaching strategy that assigns two diametrically opposed positive examples to any ball; the learner need only prefer smaller balls over larger ones in order to successfully identify any target in BALLS^d . ■

So, for the class of boxes and the class of cubes, the preference-based teaching dimension coincides with the supremum of the PBTD taken over all finite subclasses, while this is not true for the class of balls.

4. Upper Bounds Based on Cardinality Arguments

As we have seen in Section 3, the condition $\text{PBTD}(\mathcal{L}) \leq \sup_{\mathcal{L}' \subseteq_f \mathcal{L}} \text{PBTD}(\mathcal{L}')$ is not satisfied in general. The condition becomes weaker when we replace PBTD by TD. The weaker condition actually is satisfied by PBTD on all countably infinite families:

Theorem 4 *If \mathcal{L} is countably infinite, then $\text{PBTD}(\mathcal{L}) \leq \sup_{\mathcal{L}' \subseteq_f \mathcal{L}} \text{TD}(\mathcal{L}')$ and $\text{PBTD}^+(\mathcal{L}) \leq \sup_{\mathcal{L}' \subseteq_f \mathcal{L}} \text{TD}^+(\mathcal{L}')$.*

Proof Let L_1, L_2, \dots be any one-one enumeration of the family \mathcal{L} . Define the following preference relation:

$$\forall i, j \in \mathbb{N} : (L_i \prec L_j) \Leftrightarrow j < i ,$$

i.e., we prefer sets with a smaller index. So to teach L_i , we only need to distinguish L_i from all sets L_j such that $j < i$. The size of such a teaching set for L_i is upper-bounded by $\text{TD}(\{L_1, \dots, L_i\}) \leq \sup\{\text{TD}(\mathcal{L}') : \mathcal{L}' \text{ is a finite subfamily of } \mathcal{L}\}$. ■

This result is perhaps more powerful than it appears at first glance, since many concept classes will become countable if one restricts the underlying domain to be the set of computable reals. Moreover, it can be generalized in the following way.

Theorem 5 *For every infinite family of sets \mathcal{L} the following holds:*

$$\begin{aligned} \text{PBTD}(\mathcal{L}) &\leq \sup\{\text{TD}(\mathcal{L}') : \mathcal{L}' \subset \mathcal{L} \wedge \text{card}(\mathcal{L}') < \text{card}(\mathcal{L})\} , \\ \text{PBTD}^+(\mathcal{L}) &\leq \sup\{\text{TD}^+(\mathcal{L}') : \mathcal{L}' \subset \mathcal{L} \wedge \text{card}(\mathcal{L}') < \text{card}(\mathcal{L})\} . \end{aligned}$$

Proof The *initial segment* of $L \in \mathcal{L}$ w.r.t. a given linear order on \mathcal{L} is defined as the family of all sets in \mathcal{L} which are strict predecessors of L . This subfamily of \mathcal{L} is denoted as I_L .

Recall from set theory that a linear ordering on a set M is said to be a *well-ordering* if any non-empty subset of M has a smallest element. The following result (based on the Well-Ordering Principle) is well known ([Moschovakis, 1994](#)): every set \mathcal{L} can be well-ordered such that, for every $L \in \mathcal{L}$, the cardinality of the initial segment I_L is smaller than the cardinality of \mathcal{L} .

Now, fix a well-ordering on \mathcal{L} with this property. Choose the preference relation on \mathcal{L} such that, for every $L \in \mathcal{L}$, precisely the sets in I_L are strictly preferred over L . It follows that

$$\begin{aligned} \text{PBSD}(\mathcal{L}) &\leq \sup_{L \in \mathcal{L}} \text{TD}(L, I_L \cup \{L\}) \leq \sup_{L \in \mathcal{L}} \text{TD}(I_L \cup \{L\}) \\ &\leq \sup\{\text{TD}(\mathcal{L}') : \mathcal{L}' \subset \mathcal{L} \wedge \text{card}(\mathcal{L}') < \text{card}(\mathcal{L})\} , \end{aligned}$$

as desired. ■

As usual, \aleph_0 denotes the cardinality of a countably infinite set so that 2^{\aleph_0} denotes the cardinality of the reals. For simplicity, we set $|S| = \text{card}(S)$ in the sequel. Cardinal multiplication is denoted as $\nu \otimes \mu$, i.e., if $\mu = |S|$ and $\nu = |T|$, then $\nu \otimes \mu = |S \times T|$. With this notation, the following holds:

Theorem 6 *Let \mathcal{L} be a class of cardinality $\lambda = |\mathcal{L}| > \aleph_0$. Suppose that every element $L \in \mathcal{L}$ satisfies $|L| \geq \lambda$. Furthermore suppose that*

$$\nu := \sup\{|L \cap L'| : L \neq L' \in \mathcal{L}\} < \lambda .$$

Then $\text{PBSD}^+(\mathcal{L}) = 1$.

Proof We first note that ν , as a supremum of cardinalities, is a cardinality itself (Schimmerling, 2011). According to Theorem 5, it suffices to show that, for every $\mathcal{L}' \subset \mathcal{L}$ with $\nu' := |\mathcal{L}'| < \lambda$, we have that $\text{TD}^+(\mathcal{L}') \leq 1$. Let $\mu = \max\{\nu, \nu', \aleph_0\}$. Fix some $L \in \mathcal{L}'$. Let $S = \cup_{L' \in \mathcal{L}'} (L \cap L')$, i.e., S is the set of points in L that L shares with some other set in \mathcal{L}' . It clearly suffices to show that S is a proper subset of L . From the rules of cardinal arithmetic (Schimmerling, 2011), we conclude that

$$|S| \leq \nu \otimes \nu' \leq \mu \otimes \mu = \mu < \lambda \leq |L| .$$

Thus S must be a proper subset of L , indeed. ■

Example 3 *By Bezout's Theorem, any pair of plane algebraic curves intersect in a number of points bounded by the product of their degrees. Thus, any family of such curves, each with uncountably many points, satisfies the assumptions made in Theorem 6 and can therefore be taught in the PBSD-model at the expense of a single positive example.*

5. Upper Bounds on the PBSD of Disjoint Unions of Basic Objects

In this subsection, we let C denote a convex body in \mathbb{R}^d (i.e., a compact and convex subset of \mathbb{R}^d with a non-empty interior). Consider now the class

$$\mathcal{L}^C = \{s \cdot C + u : s \in \mathbb{R}_0^+ \wedge u \in \mathbb{R}^d\} .$$

It consists of all (non-negatively) scaled and translated versions of C . If $C' = s \cdot C + u$ for $s \in \mathbb{R}_0^+$ and $u \in \mathbb{R}^d$, then s is called the scaling factor of C' . Since the scaling factor 0 is among the possible choices, the class \mathcal{L}^C contains the singletons in \mathbb{R}^d as a subclass.

We denote by \mathcal{L}_k^C the class of unions of k or fewer objects from \mathcal{L}^C . The subclass consisting of unions of k or fewer mutually disjoint objects from \mathcal{L}^C is denoted as $\mathcal{L}_k^C[\text{disj}]$. For instance, if

B denotes the unit ball in \mathbb{R}^d , then \mathcal{L}_k^B denotes the class of unions of k or fewer balls in \mathbb{R}^d , and $\mathcal{L}_k^B[\text{disj}]$ denotes the corresponding class for disjoint unions. Later, we will often use the subscript “ $= k$ ” instead of the subscript k when referring to concepts that can be written as the union of k , but no fewer than k , objects, e.g., $\mathcal{L}_{=k}^C = \mathcal{L}_k^C \setminus \mathcal{L}_{k-1}^C$ and $\mathcal{L}_{=k}^C[\text{disj}] = \mathcal{L}_k^C[\text{disj}] \setminus \mathcal{L}_{k-1}^C[\text{disj}]$.

In the sequel, we set $\text{BALLS}_k^d := \mathcal{L}_k^B$. Similarly CUBES_k^d denotes the class of unions of k or fewer d -dimensional axis-aligned cubes, and BOXES_k^d denotes the class of unions of k or fewer d -dimensional boxes. Note, however, that BOXES_k^d cannot be obtained from a single convex body by scaling and translation unless we employ different scaling factors along different dimensions.

Recall that the union of k mutually disjoint objects taken from the *same domain* is not be confused with the union of objects taken from *mutually disjoint domains* (as is done in the definition of the direct sum of concept classes).

As usual, the diameter of a set $M \subseteq \mathbb{R}^d$ is given by

$$\varnothing(M) = \sup_{x,y \in M} \|x - y\| .$$

For obvious reasons, the PBD of \mathcal{L}_k^C and $\mathcal{L}_k^C[\text{disj}]$ will not change when we replace all concepts L of the class by $\{Ax : x \in L\}$ where A is a fixed orthogonal matrix. Under such a transformation, we may assume that the diameter of C is realized by a pair of poles that differ only in their d -th coordinate. Furthermore, for the purposes of defining \mathcal{L}^C , we may, without loss of generality, make the following technical assumption:

(A) C has north pole $(0, \dots, 0, 0)$, south pole $(0, \dots, 0, -1)$, and $\varnothing(C) = 1$.

Note that if C satisfies assumption (A) then any object $s \cdot C + u \in \mathcal{L}^C$, has north pole u , south pole $u - (0, \dots, 0, s)$, and diameter s . It follows that, if $C' = s' \cdot C + u'$ and $C'' = s'' \cdot C + u''$ (with $s' \leq s''$) have the same north pole, then $C' \subseteq C''$.

We will now state the main result of this section.

Theorem 7 *Let C be a convex body in \mathbb{R}^d . Then $\text{PBD}(\mathcal{L}_k^C[\text{disj}]) \leq 3k$.*

Proof As noted above, we may assume that C satisfies the technical assumption (A). Let $K = (C^1, \dots, C^n)$ be a collection of $n \leq k$ disjoint objects from \mathcal{L}^C . We identify the collection K with the concept $U(K) = \cup_{l=1}^n C^l \in \mathcal{L}_k^C[\text{disj}]$. For $l = 1, \dots, n$, let x^l denote the north pole of C^l . Similarly y^l denotes the south pole of C^l .

For distinct points $p, q \in \mathbb{R}^d$, we write $p \triangleright q$ iff there exists $t \leq d$ such that (i) $p_s = q_s$, for $t < s \leq d$, and (ii) $p_t > q_t$. After renumbering, if necessary, we may assume that $x^i \triangleright x^j$, for all $i < j$. The sequence $P(K) = (x^1, \dots, x^n)$ is called the *profile* of K . Consider two collections K and K' with profiles $P(K) = (x^1, \dots, x^n)$ and $P(K') = (u^1, \dots, u^m)$. We say that profile $P(K)$ is *dominated* by profile $P(K')$ if either (i) $P(K')$ is an extension of $P(K)$, or (ii) for the first position j such that $u^j \neq x^j$, we have that $u^j \triangleright x^j$. Note that if $P(K) \neq P(K')$ then either $P(K)$ is dominated by $P(K')$ or $P(K')$ is dominated by $P(K)$.

We define a preference relation on collections of disjoint objects from \mathcal{L}^C by the following rules:

Rule 1: If $P(K) \neq P(K')$ then K is preferred over K' if $P(K)$ is dominated by $P(K')$.

Rule 2: If $P(K) = P(K')$ then K is preferred over K' if, for every $l \in [n]$, the scaling factor of D^l is at least as large as the scaling factor of C^l and, for at least one $l \in [n]$, the scaling factor of D^l is strictly larger than the scaling factor of C^l ; equivalently $U(K) \subset U(K')$.

Suppose now the collection $K = (C^1, \dots, C^m)$ with profile $P(K) = (x^1, \dots, x^n)$ is to be taught. The teaching set for K includes the following points:

- the north poles x^1, \dots, x^n and the south poles y^1, \dots, y^n ,
- for every $l \in [n]$ a point z^l that satisfies the following conditions:
 - On the first $d - 1$ coordinates, z^l coincides with y^l .
 - The coordinate z_d^l is smaller than y_d^l , but only slightly smaller so that the object from \mathcal{L}^C with north pole x^l and south pole z^l (which properly includes the object C^l) does not intersect any of the objects C^j with $j \in [n] \setminus \{l\}$. (By the disjointness of the objects in K , such a point z^l must exist.)

Obviously the points x^l and y^l are labeled “1” whereas the points z^l are labeled “0”.

Suppose that $K' = (D^1, \dots, D^m)$ with profile $P(K') = (u^1, \dots, u^m)$ is a different collection that is consistent with the above examples. It will suffice to show that K is preferred over K' , according to the rules above.

If $P(K') = P(K)$, then because of the negative examples z^1, \dots, z^n , the south pole y^l can be covered only by D^l , the object in K' with north pole x^l . It follows that the scaling factor of D^l is at least as large as the scaling factor of C^i and, because $K' \neq K$, for at least one $l \in [n]$, the scaling factor of D^l is strictly larger than the scaling factor of C^i . Hence, by Rule 2, K is preferred over K' .

On the other hand, if $P(K') \neq P(K)$ we can assume that the longest common prefix of $P(K)$ and $P(K')$ has length $j - 1 < n$. (Otherwise, $P(K')$ dominates $P(K)$ and so K is preferred over K' , by Rule 1.) We make the following observations:

- The objects D^1, \dots, D^{j-1} have the same north poles as C^1, \dots, C^{j-1} , respectively.
- The negative examples z^1, \dots, z^{j-1} ensure that none of the objects D^1, \dots, D^{j-1} can cover x^j .
- One of the objects $D^t \in \{D^j, \dots, D^m\}$ covers x^j , so either (i) its north pole u^t coincides with x^j or (ii) u_d^t , the d -th coordinate of its north pole, must be greater than x_d^j .
- Consequently either (i) $u^j \triangleright u^t$ or (ii) $u_d^j \geq u_d^t > x_d^j$. In either case, $u^j \triangleright x^j$ so that $P(K')$ dominates $P(K)$. Thus, by Rule 1, K is preferred over K' .

■

Remark 8 *It should be noted that, in the proof above, we only used the fact that there exists a point $p_S \in C \setminus \{p_N\}$ such that the segment $[p_N, p_S]$ is a maximal segment in C . Thus Theorem 7 still holds, for example, if C is star-shaped in any sufficiently small neighborhood of p_N , and the objects in K are pairwise linearly separable.*

We remind the reader of the following standard definitions. A set $C \in \mathbb{R}^d$ is said to be *strictly convex* if, for every choice of distinct points $x, y \in C$, the following holds: any point on the line segment connecting x and y , other than the endpoints x and y themselves, belongs to the interior of C . A *strictly convex body* in \mathbb{R}^d is a convex body in \mathbb{R}^d that is strictly convex. Note that objects like balls or, more generally, ellipsoids are strictly convex bodies whereas convex polytopes are convex bodies but are not strictly convex.

Suppose that $C \subset \mathbb{R}^d$ is a strictly convex body. We say that a collection K of objects from \mathcal{L}^C is *locally disjoint (around the north poles)* if no object in K contains a north pole of any other object in K . In other words: the north pole of every object in K can be assigned a neighborhood such that the collection of these neighborhoods is disjoint. Let $\mathcal{L}_k^C[\text{loc-disj}]$ be the class consisting of unions of k or fewer locally disjoint objects from \mathcal{L}^C . The following result is a close relative of Theorem 7:

Theorem 9 *Let C be a strictly convex body. Then $\text{PBTd}(\mathcal{L}_k^C[\text{loc-disj}]) \leq 3k$.*

The proof, which is quite similar to the proof of Theorem 7, will be given in Appendix A.

The upper bounds from Theorems 7 and 9 can be improved when one imposes the restriction that all objects participating in a union are of the same (unknown) size, i.e., they have the same scaling factor. This matches the intuition that any teaching strategy might save some examples when the size of only one of the objects in the union is to be taught.

More formally, let us consider the class $\mathcal{L}_k^C[\text{transl}]$ of unions of k or fewer objects from \mathcal{L}^C all of which are just translations of one another, i.e., for any two objects $C_1, C_2 \in \mathcal{L}^C$ participating in such a union, there is a unique vector $u \in \mathbb{R}^d$ such that $C_2 = C_1 + u$.¹ The notations $\mathcal{L}_{=k}^C[\text{transl}]$, $\mathcal{L}_k^C[\text{disj, transl}]$ and $\mathcal{L}_{=k}^C[\text{disj, transl}]$ are understood analogously.

Theorem 10 *Let C be a compact body in \mathbb{R}^d . Then $\text{PBTd}^+(\mathcal{L}_{=k}^C[\text{disj, transl}]) \leq k + 1$.*

The proof of this theorem will be given in Appendix B.

Since Theorem 7 does not apply to the class of boxes (recall that BOXES_k^d cannot be obtained from a single convex body by scaling and translation unless we employ different scaling factors along different dimensions), an upper bound for BOXES_k^d has to be found by a separate analysis:

Theorem 11 $\text{PBTd}(\text{BOXES}_k^d[\text{disj}]) \leq k(d + 2) - d$.

The proof is given in Appendix C.

6. Lower Bounds on the PBTd of Disjoint Unions of Basic Objects

We argued in Theorem 7 that $3k$ examples suffice to teach any disjoint union of up to k scaled and translated versions of a convex body. This subsection will address lower bounds for the same teaching problem and variants thereof. Its main result establishes that

$$\text{PBTd}(\mathcal{L}_k^C[\text{disj}]) = \Theta(k),$$

for any convex body C . To this end, we first make the following rather trivial observation.

1. Uniqueness easily follows from the compactness of C_1 and C_2 .

Lemma 12

1. Let \mathcal{X} be a domain with at least k elements. Suppose that \mathcal{L}_k is the family of all subsets of \mathcal{X} with at most k elements. Then $\text{TD}_{\min}(\mathcal{L}_k) = k$.
2. Let \mathcal{X} be a domain with at least $2k$ elements. Suppose that $\mathcal{L}_{=k}$ is the family of all subsets of \mathcal{X} with exactly k elements. Then $\text{TD}_{\min}(\mathcal{L}_{=k}) = k$.

Since the PBTD of a class is lower bounded by the TD_{\min} of each of its finite subclasses, we obtain the following result:

Corollary 13 *Let \mathcal{X} be a domain with at least k (resp. at least $2k$) elements. For any concept class \mathcal{L} over \mathcal{X} that contains the subclass \mathcal{L}_k (resp. $\mathcal{L}_{=k}$), we have that $\text{PBTD}(\mathcal{L}) \geq k$.*

Proof Fix a subset \mathcal{X}' of \mathcal{X} such that $|\mathcal{X}'| = k$ (resp. $|\mathcal{X}'| = 2k$). Let \mathcal{L}' be the subclass of \mathcal{L} that consists of all subsets of \mathcal{X}' with at most k (resp. with exactly k) elements. \mathcal{L}' is a finite subclass of \mathcal{L} to which Lemma 12 can be applied. ■

Corollary 14 *Let C be a convex body in \mathbb{R}^d . Then $\text{PBTD}(\mathcal{L}_{=k}^C[\text{disj}, \text{transl}]) \geq k$.*

Proof Since \mathcal{L}^C contains all singletons over \mathbb{R}^d , the class $\mathcal{L}_{=k}^C[\text{disj}, \text{transl}]$ contains all subsets of \mathbb{R}^d with at exactly k elements. Now Corollary 13 applies. ■

If \mathcal{L} is any concept class and $L \in \mathcal{L}$ any concept, a set $S \subseteq L$ is called a weak spanning set for L w.r.t. \mathcal{L} , if there is no proper subset $L' \subset L$ in \mathcal{L} that contains S . It was shown by Gao et al. (2017) that

$$\text{PBTD}^+(\mathcal{L}) \geq \sup_{L \in \mathcal{L}} \inf\{|S| : S \text{ is a weak spanning set for } L \text{ w.r.t. } \mathcal{L}\}. \quad (5)$$

We refer to the right hand-side of (5) as the *weak spanning dimension* of \mathcal{L} . Hence inequality (5) establishes the weak spanning dimension as lower bound on PBTD^+ .

Lemma 15 *Let C be a convex body in \mathbb{R}^d . Then the weak spanning dimension of the class $\mathcal{L}_{=k}^C[\text{disj}, \text{transl}]$ equals $k + 1$.*

Proof Consider a concept L in $\mathcal{L}_{=k}^C[\text{disj}, \text{transl}]$ that consists of k disjoint duplicates of C . Since the interior of a convex body is non-empty, C is neither empty nor a singleton. Any choice of k points x^1, \dots, x^k can be consistently explained by the hypothesis $\cup_{l=1}^k \{x_l\} \in \mathcal{L}_{=k}^C[\text{disj}, \text{transl}]$, which is a proper subset of L . Thus the weak spanning dimension of $\mathcal{L}_{=k}^C[\text{disj}, \text{transl}]$ is at least $k + 1$. On the other hand, the weak spanning dimension of $\mathcal{L}_{=k}^C[\text{disj}, \text{transl}]$ is at most $k + 1$ because it is upper-bounded by $\text{PBTD}^+(\mathcal{L}_{=k}^C[\text{disj}, \text{transl}])$ and the latter does not exceed $k + 1$ according to Theorem 10. ■

Corollary 16 *Let C be a convex body in \mathbb{R}^d . Then $\text{PBTD}^+(\mathcal{L}_{=k}^C[\text{disj}, \text{transl}]) = k + 1$.*

The lower bounds that have been shown in this section to hold for the class $\mathcal{L}_{=k}^C[\text{disj}, \text{transl}]$ apply also to classes containing $\mathcal{L}_{=k}^C[\text{disj}, \text{transl}]$ as a subclass (like, for instance, $\mathcal{L}_k^C[\text{disj}]$ and $\mathcal{L}_{=k}^C[\text{disj}]$). Together with the upper bound of Theorem 7, this implies

$$k \leq \text{PBDT}(\mathcal{L}_{=k}^C[\text{disj}]) \leq \text{PBDT}(\mathcal{L}_k^C[\text{disj}]) \leq 3k \quad (6)$$

for any convex body C in \mathbb{R}^d . The following simple result occasionally helps to narrow the gap between k and $3k$:

Lemma 17 *Let C be a convex body in \mathbb{R}^d . Then, for every $\mathcal{L}' \subseteq_f \mathcal{L}^C$, the following holds:*

1. $\text{PBDT}(\mathcal{L}_{=k}^C) \geq k \cdot \text{TD}_{\min}(\mathcal{L}')$.
2. *Suppose that none of the sets in \mathcal{L}' can be taught by a single negative example. Suppose furthermore that $U := \cup_{L' \in \mathcal{L}'} L' \in \mathcal{L}^C$ so that $\mathcal{L}'' := \mathcal{L}' \cup \{U\} \subseteq_f \mathcal{L}^C$. Then $\text{TD}_{\min}(\mathcal{L}'') \geq 2$ and $\text{PBDT}(\mathcal{L}_{=k}^C) \geq 2k$.*

Proof

1. Suppose that all concepts from \mathcal{L}' fit within the strip $S = \{x \in \mathbb{R}^d : a \leq x_1 \leq b\}$. Let S^1, S^2, \dots, S^k be disjoint copies of S , under translation by v^1, v^2, \dots, v^k , respectively. Consider the finite subclass \mathcal{L}'_k of $\text{PBDT}(\mathcal{L}_{=k}^C)$ which contains those unions of k objects from \mathcal{L}' which have precisely one object from \mathcal{L}' , translated by v^i , in the interior of S^i . Thus \mathcal{L}'_k is a direct sum of k versions of \mathcal{L}' (one version per strip). An application of Equation (1) yields $\text{PBDT}(\mathcal{L}_{=k}^C) \geq \text{TD}_{\min}(\mathcal{L}'_k) \geq k \cdot \text{TD}_{\min}(\mathcal{L}')$.
2. By assumption, none of the concepts in \mathcal{L}' can be taught by a single negative example. Clearly U cannot be taught by a single negative example either since any negative example for U is also a negative example for any concept in \mathcal{L}' . Moreover, none of the concepts in \mathcal{L}'' can be taught by a single positive example because each example in U is positive for at least one of the concepts in \mathcal{L}' , and vice versa. Hence $\text{TD}_{\min}(\mathcal{L}'') \geq 2$. From the first assertion of the lemma, it follows that $\text{PBDT}(\mathcal{L}_{=k}^C) \geq 2k$. ■

An inspection of Example 1 reveals that we may apply Lemma 17 to the class CUBES_k^d so that (in combination with Theorem 7) the following holds:

Corollary 18 *For all $d, k \geq 1$, we have*

$$2k \leq \text{PBDT}(\text{CUBES}_{=k}^d) \leq \text{PBDT}(\text{CUBES}_k^d) \leq 3k$$

and (because CUBES is a subclass of BOXES)

$$2k \leq \text{PBDT}(\text{BOXES}_{=k}^d) \leq \text{PBDT}(\text{BOXES}_k^d) .$$

	PBTD	PBTD ⁺
BALLS _k ^d [disj] BALLS _{=k} ^d [disj]	$k \leq \text{PBTD} \leq 3k$ (Equation (6))	$k + 1 \leq \text{PBTD}^+$ (Corollary 16)
BALLS _k ^d [loc-disj] BALLS _{=k} ^d [loc-disj]	$k \leq \text{PBTD} \leq 3k$ (Equation (6) and Theorem 9)	$k + 1 \leq \text{PBTD}^+$ (Corollary 16)
BALLS _{=k} ^d [disj, transl]	$k \leq \text{PBTD} \leq k + 1$ (Corollaries 14 and 16)	$\text{PBTD}^+ = k + 1$ (Corollary 16)
BALLS ₂ ^d BALLS ₌₂ ^d	$2 \leq \text{PBTD} \leq 4$ (Equation (6) and Theorem 20)	$\text{PBTD}^+ = 4$ (Theorem 20)
BALLS ₂ ^d [transl] BALLS ₌₂ ^d [transl]	$2 \leq \text{PBTD} \leq 3$ (Corollaries 14 and 21)	$\text{PBTD}^+ = 3$ (Corollary 21)
CUBES _k ^d [disj] CUBES _{=k} ^d [disj]	$2k \leq \text{PBTD} \leq 3k$ (Corollary 18)	$2k \leq \text{PBTD}^+$ (Corollary 18)
CUBES _{=k} ^d [disj, transl]	$k \leq \text{PBTD} \leq k + 1$ (Corollaries 14 and 16)	$\text{PBTD}^+ = k + 1$ (Corollary 16)
CUBES ₂ ^d CUBES ₌₂ ^d	$\text{PBTD} = 4$ (Theorem 22)	$\text{PBTD}^+ = 4$ (Theorem 22)
CUBES ₂ ^d [transl] CUBES ₌₂ ^d [transl]	$\text{PBTD} = 3$ (Theorem 23)	$\text{PBTD}^+ = 3$ (Theorem 23)
BOXES _k ^d [disj] BOXES _{=k} ^d [disj]	$2k \leq \text{PBTD} \leq (d + 2)k - d$ (Corollary 18 and Theorem 11)	$2k \leq \text{PBTD}^+$ (Corollary 18)

Table 1: Summary of our results on teaching specific types of unions of geometric concepts.

7. Summary

Table 1 summarizes our main results on teaching unions of balls, cubes, or boxes. All results concerning (possibly non-disjoint) unions of two objects (balls, cubes) are given in the appendix. They are mainly meant to illustrate the perfdies of the object when it comes to analyzing the PBTB of unions of mutually overlapping objects.

Acknowledgments

David Kirkpatrick and Sandra Zilles were supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), in the Discovery Grant program. Sandra Zilles was also supported through the NSERC Canada Research Chairs program.

References

- Anselm Blumer, Andrzej Ehrenfeucht, David Haussler, and Manfred K. Warmuth. Learnability and the Vapnik-Chervonenkis dimension. *JACM*, 36(4):929–965, 1989.
- Thorsten Doliwa, Gaojian Fan, Hans Ulrich Simon, and Sandra Zilles. Recursive teaching dimension, VC-dimension and sample compression. *J. Mach. Learn. Res.*, 15:3107–3131, 2014.
- Michael Frazier, Sally A. Goldman, Nina Mishra, and Leonard Pitt. Learning from a consistently ignorant teacher. In *Proceedings of the Seventh Annual ACM Conference on Computational Learning Theory, COLT*, pages 328–339, 1994.
- Ziyuan Gao, Christoph Ries, Hans Ulrich Simon, and Sandra Zilles. Preference-based teaching. *J. Mach. Learn. Res.*, 18(31):1–32, 2017.
- Paul W. Goldberg, Sally A. Goldman, and H. David Mathias. Learning unions of boxes with membership and equivalence queries. In *Proceedings of the Seventh Annual ACM Conference on Computational Learning Theory, COLT*, pages 198–207, 1994.
- Sally A. Goldman and Michael J. Kearns. On the complexity of teaching. *Journal of Computer and System Sciences*, 50(1):20–31, 1995.
- Sanjay Jain, Yen Kaow Ng, and Tiong Seng Tay. Learning languages in a union. *J. Comput. Syst. Sci.*, 73(1):89–108, 2007.
- Philip M. Long and Manfred K. Warmuth. Composite geometric concepts and polynomial predictability. *Inf. Comput.*, 113(2):230–252, 1994.
- Wolfgang Maass and György Turán. Algorithms and lower bound methods for on-line learning of geometrical concepts. *Machine Learning*, 14:251–269, 1994.
- Yiannis N. Moschovakis. *Notes on Set Theory*. Springer Texts in Electrical Engineering. Springer-Verlag, 1994.
- Yen Kaow Ng and Takeshi Shinohara. Inferring unions of the pattern languages by the most fitting covers. In *Proceedings of the 16th International Conference on Algorithmic Learning Theory, ALT*, pages 269–282, 2005.

Seishi Ouchi, Tomohiko Okayama, Keisuke Otaki, Ryo Yoshinaka, and Akihiro Yamamoto. Learning concepts and their unions from positive data with refinement operators. *Ann. Math. Artif. Intell.*, 79(1-3):181–203, 2017.

Ernest Schimmerling. *A Course on Set Theory*. Cambridge University Press, 2011.

Ayumi Shinohara and Satoru Miyano. Teachability in computational learning. *New Gen. Comput.*, 8:337–348, 1991.

Keith Wright. Identification of unions of languages drawn from an identifiable class. In *Proceedings of the Second Annual Workshop on Computational Learning Theory, COLT*, pages 328–333, 1989.

Sandra Zilles, Steffen Lange, Robert Holte, and Martin Zinkevich. Models of cooperative teaching and learning. *J. Mach. Learn. Res.*, 12:349–384, 2011.

Appendix A. Proof of Theorem 9

The proof is quite similar to the proof of Theorem 7. The same preference relation is used. Suppose that $K = (C^1, \dots, C^n)$ is a collection of objects that is to be taught. For every $l \in [n]$, we include the north pole x^l into the teaching set K (as we did in the proof of Theorem 7). The crucial difference to the proof of Theorem 7 is the following:

- Instead of the south pole y^l of C^l , we include a point y_{new}^l that lies on the boundary of C^l and is sufficiently close to x^l .
- Instead of a point z^l that is located southern of and sufficiently close to y^l , we choose a point z_{new}^l that is sufficiently close to y_{new}^l and does not belong to C^l .

The main observations are as follows. First, all examples that we have put in the teaching set are located in a neighborhood of the corresponding north pole (and this neighborhood can be chosen as small as we want). Second, these examples fit the same purpose as the original examples did in the proof of Theorem 7. Specifically, the following can be achieved:

1. The negative examples z_{new}^l make sure that every object with north pole x^l that excludes z_{new}^l does still not intersect any of the objects C^j with $j \in [n] \setminus \{l\}$.
2. The positive examples y_{new}^l make sure that every collection K' of objects with the same profile as the target collection K uses objects of the “right size” so that $K' = K$.

Given the proof of Theorem 7, the missing details are easy to fill in.

Appendix B. Proof of Theorem 10

We begin with the following:

Observation Suppose that C is any subset of \mathbb{R}^d , and let $C^1 = C + u^1$ and $C^2 = C + u^2$ be arbitrary objects in \mathcal{L}^C . Let x be an arbitrary point in C . If $C' = C + v \in \mathcal{L}^C$ contains both $x + u^1$ and $x + u^2$, then $x + u^1 - v$ and $x + u^2 - v$ both belong to C , and hence $x + u^1 + u^2 - v$ belongs to $C^1 \cap C^2$.

As in the proof of Theorem 7, we may assume that if $K = (C^1, \dots, C^k)$ a member of the class $\mathcal{L}_{=k}^C[\text{disj, transl}]$ it has a profile (x_d^1, \dots, x_d^k) , where, for every $l \in [k]$, x^l is the north pole of C^l , and, for every $i < j$, $x^i \triangleright x^j$. We define the following hierarchical preference relation:

Rule 1: Collections of objects with a smaller common scaling factor are preferred over collections of larger objects.

Rule 2: Between two collections of objects of the same scaling factor, the collection with lexicographically smaller profile is preferred.

Suppose now the disjoint collection $K = (C^1, \dots, C^k)$ with k objects of the same scaling factor, say s , is to be taught. The teaching set for K consists of the north poles x^1, \dots, x^k and the single south pole y^k of C^k . Clearly, each of these $k + 1$ examples is labeled “1”.

According to the pigeon-hole principle, a consistent hypothesis from $\mathcal{L}_{=k}^C[\text{disj, transl}]$ must cover two of the positive examples by a single object. Because of Rule 1, a smallest possible hypothesis will be chosen for this purpose. The target collection K witnesses that a scaling factor of s would be sufficient.

Since C satisfies condition (A), it follows that $\|x^k - y^k\| = \varphi(C^k) = s$. Furthermore, since $x^l \triangleright x^k$, $\|x^l - y^k\| > s$, for all $l < k$. Hence, no object from \mathcal{L}^C with scaling factor less than or equal to s can cover x^l and y^k , unless the scaling factor is s and the object is C_k .

We may conclude from our preliminary Observation that it is impossible to cover two distinct north poles by a single object from \mathcal{L}^C with scaling factor s (since, otherwise, there would be $l \neq l' \in [k]$ such that $C^l \cap C^{l'} \neq \emptyset$).

In summary, the preference relation forces us to cover x^k and y^k by C_k , and all other positive examples x^l by their own object. Given the constraint that all k objects in the collection must have scaling factor s , Rule 2 forces us to make x^1, \dots, x^k the north poles of the chosen collection. Thus, given Rules 1 and 2, the $k + 1$ chosen examples uniquely specify K as the most preferred hypothesis (among the consistent ones).

Appendix C. Proof of Theorem 11

In one dimension BOXES are just intervals. Any collection of k disjoint intervals can be taught by providing (i) positive examples at all interval endpoints, and (ii) one negative example between adjacent intervals, for a total of at most $3k - 1$ examples. The student simply prefers a hypothesis that minimizes the total length of the intervals.

This strategy generalizes in a straightforward way to higher dimensions. Let $K = (C^1, \dots, C^n)$ be any instance of $\text{BOXES}_k^d[\text{disj}]$ that is to be taught, and let δ denote the minimum distance (in any dimension) between distinct boxes in K . We consider two natural orders on points based on their coordinate vectors: first the usual *lexicographic* total order and second the *domination* partial order in which one vector dominates another if it is at least as large in every coordinate position. We say that point p is *larger than* point q if the coordinate vector of p is lexicographically larger than that of q .

For $l = 1, \dots, n$, let p^l (resp. q^l) denote the maximal (resp. minimal) corner of C^l . After renumbering, if necessary, we may assume that the sequence (p^1, p^2, \dots, p^n) is in decreasing order by size. The sequence $P(K) = (p^1, \dots, p^n)$ is called the *profile* of K . One such profile is said to be *larger than* a second if either (i) in the first position they differ the first has a larger point than the

second, or (ii) the first is an extension of the second. We identify the collection K with its coverage $U(K) = \cup_{l=1}^n C^l$.

Suppose that $K = (C^1, \dots, C^n)$ be an instance of $\text{BOXES}_k^d[\text{disj}]$ that is to be taught. The teaching set for K includes the following points associated with each object $C^l \in K$:

- the corner points p^l and q^l , and
- For $l = 1, \dots, n$ and for each dimension $i \in \{1, \dots, d\}$ a point r_i^l at distance $\delta/2$ below the other box corner whose coordinate vector differs from that of p^l in the i -th coordinate only (or $\delta/2$ below p^l itself, if box C^l is degenerate in dimension i).

Obviously the points p^l and q^l are labeled “1” whereas the points r_i^l are all labeled “0”.

We define the following hierarchical preference relation:

Rule 1: Collections with a smaller profile are preferred over collections with a larger profile.

Rule 2: Among collections K and K' with the same profile, we prefer K over K' if $U(K) \subset U(K')$.

Suppose that a student chooses an hypothesis $K' = (D^1, \dots, D^m)$ with profile $P(K')$. Note that any point that belongs to $P(K)$ must either belong to $P(K')$ or be dominated by a point (namely the maximal corner of its containing box) that belongs to $P(K')$. The latter cannot itself belong to the profile of K , since, by disjointness, no profile point belongs to the box associated with another profile point. Hence, the hypothesis K' must have a profile $P(K')$ that is no smaller than $P(K)$ (since either $P(K')$ is an extension of $P(K)$, or the largest (first) element p^j in $P(K)$ that does not belong to $P(K')$ is smaller than the largest (first) element in $P(K')$ that does not belong to $P(K)$).

Thus, guided by preference Rule 1, the student must choose an hypothesis K' with $P(K') = P(K)$. But the placement of negative examples ensures that in any such hypothesis, the point q^l can only be contained in the box with maximum corner p^l , for all l . Thus any hypothesis that satisfies in addition preference Rule 2 must consist of minimum volume boxes that contain both p^l and q^l , which is exactly the intended hypothesis K .

Note that the negative example in dimension i associated with any box C^l whose corner q^l is minimal in dimension i is unnecessary, since extending C^l in this dimension cannot possibly cover any additional corner points. Thus, a total of $k(d+2) - d$ examples suffice.

Appendix D. Teaching Unions of Two Balls

Let $\text{BALLS}_k^d[\text{loc-disj}]$ denote the subclass of BALLS_k^d consisting of all unions of at most k balls whose pairwise intersections do not contain the north pole of either ball. An application of Theorem 9 yields the following bounds:

$$\text{PBDT}(\text{BALLS}_k^d[\text{disj}]) \leq \text{PBDT}(\text{BALLS}_k^d[\text{loc-disj}]) \leq 3k . \quad (7)$$

When the balls in a union may overlap in arbitrary ways, the analysis of the teaching complexity becomes more difficult. The remainder of this section is devoted to the special case of unions of at most two balls, i.e., the case $k = 2$.

The following simple observation is illustrated in Fig. 2. Suppose that B and B' are two distinct balls in \mathbb{R}^d whose surfaces touch each other in one point. Then the two balls intersect just in this

one point (Case 1), as depicted in part a) of the figure, or the ball with the smaller radius is strictly contained in the ball of larger radius (Case 2), as depicted in part b) of the figure. For example, if the north pole (resp. south pole) of B coincides with the north pole (resp. south pole) of B' , then we are in Case 2.

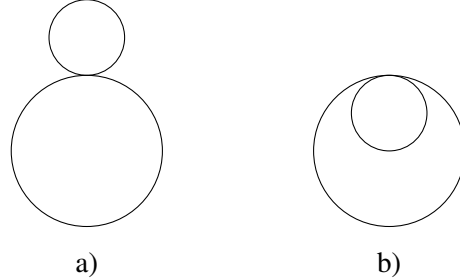


Figure 2: Two balls whose surfaces touch each other in one point.

Throughout this section, $B(z, r)$ with $z \in \mathbb{R}^d$ and $r \geq 0$ denotes the ball of radius r around z .

Lemma 19 *Consider a hypothesis H from BALLS_2^d , say $H = B_0 \cup B_1$ where $B_b = B(z_b, r_b)$ for $b = 0, 1$. Suppose that neither $B_0 \subseteq B_1$ nor $B_1 \subseteq B_0$ (which implies that $z_0 \neq z_1$). Let L be the line through z_0 and z_1 . Assume w.l.o.g. that (after performing an appropriate rotation of \mathbb{R}^d), z_0 is northern of z_1 so that L is a vertical line. Let x_0 be the north pole of B_0 and let x_1 be the south pole of B_1 . With these assumptions, the following holds:*

1. *For any pair $(x'_0, x'_1) \in B_0 \times B_1 \setminus \{(x_0, x_1)\}$, we have $\|x'_1 - x'_0\| < \|x_1 - x_0\|$ (so that $\|x_1 - x_0\|$ coincides with the diameter of H).*
2. *The point x_0 is northern of the north pole of B_1 and, symmetrically, the point x_1 is southern of the south pole of B_0 .*

Proof A straightforward proof, illustrated in Fig. 3, goes as follows. Let z be the point that resides in the middle of the line segment between x_0 and x_1 . Obviously, z is located somewhere in between z_0 and z_1 . It follows that the ball of radius $\|z - x_0\| = \|z - x_1\|$ around z touches x_0 and x_1 and (because of having a larger radius than the balls B_0 and B_1 , respectively) has the remaining points of B_0 and B_1 in its interior. Thus, any pair $(x'_0, x'_1) \in B_0 \times B_1 \setminus \{(x_0, x_1)\}$ must satisfy $\|x'_1 - x'_0\| < \|x_1 - x_0\|$. Finally note that x_0 not being northern of the north pole of B_1 would imply that $B_0 \subseteq B_1$ and, symmetrically, x_1 not being southern of the south pole of B_0 would imply $B_1 \subseteq B_0$. ■

Given two balls B_0 and B_1 such that neither one is contained in the other, we say that the pair $(x_0, x_1) \in B_0 \times B_1$ represents the diameter of $B_0 \cup B_1$ if $\|x_1 - x_0\|$ coincides with the diameter of $B_0 \cup B_1$. It follows from Lemma 19 that such a pair is unique and, after an appropriate rotation of \mathbb{R}^d , we have precisely the situation described in the assumption and conclusions of Lemma 19.

Theorem 20 $\text{PBD}^+(\text{BALLS}_2^d) = \text{PBD}^+(\text{BALLS}_{=2}^d) = 4$.

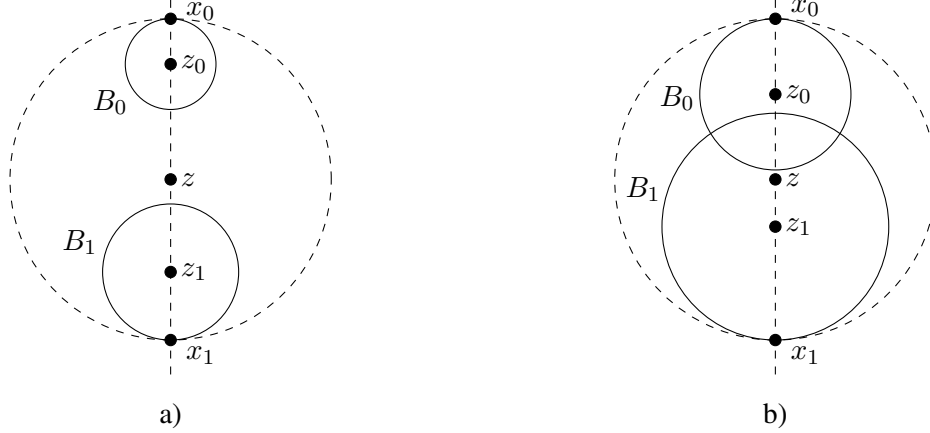


Figure 3: The diameter of $B_0 \cup B_1$ uniquely given as the line segment between x_0 and x_1 .

Proof We first show that $\text{PBD}^+(\text{BALLS}_{=2}^d) \geq 4$. To this end, it suffices to show that a union of two balls $B_0 = B(z_0, r_0)$ and $B_1 = B(z_1, r_1)$ such that $0 < r_0 \leq r_1$ and $B_0 \not\subseteq B_1$ has no weak spanning set of size 3 or less. Consider any collection of three distinct points $P_1, P_2, P_3 \in B_0 \cup B_1$. At least two of them, say P_1 and P_2 , must belong to the same ball B_b with $b \in \{0, 1\}$. But then $B_b \cup \{P_3\} = B_b \cup B(P_3, 0)$ contains the three points too and is a proper subset $B_0 \cup B_1$. Thus we need at least four points for weakly spanning every member of the class BALLS_2^d .

We now show that $\text{PBD}^+(\text{BALLS}_2^d) \leq 4$. For this purpose, we employ the following hierarchical preference relation \prec on BALLS_2^d :

1. With highest priority, pick a hypothesis with a smallest possible diameter.
2. With second highest priority, pick a hypothesis such that the maximum of the two radii is as small as possible.
3. With third highest priority, pick a hypothesis such that the minimum of the two radii is as small as possible.

Consider the union H of two balls $B_0 = B(z_0, r_0)$ and $B_1 = B(z_1, r_1)$. We have to design a positive teaching set for $B_0 \cup B_1$ w.r.t. $(\text{BALLS}_2^d, \prec)$ of size at most 4. We proceed by case analysis:

Case 1: $H = B_0$ or $H = B_1$. Say $H = B_1$.

The following discussion is illustrated in part a) of Fig. 4. Let x_0 be the north pole and let x_1 be the south pole of B_1 . Let x_2 be any point from the surface of B_1 but different from x_0 and x_1 . We claim that $T = \{x_0, x_1, x_2\}$ is a positive teaching set for B_1 . Consider any hypothesis $H' \in \text{BALLS}_2^d$ that contains x_0 and x_1 and is not inferior to B_1 w.r.t. \prec . It suffices to show that either $H' = B_1$ or H' does not contain x_2 . If H' consists of a single ball, say B' , then it easily follows that $B' = B_1$. Let us therefore assume that B' is of the form $B' = B'_0 \cup B'_1$ where, for $b = 0, 1$, $B'_b = B(z'_b, r'_b)$ and neither $B'_0 \subseteq B'_1$ nor $B'_1 \subseteq B'_0$. We first observe that the diameter of H' coincides with $\|x_1 - x_0\|$:

- It is at least $\|x_1 - x_0\|$ because $x_0, x_1 \in H'$.
- It is at most $\|x_1 - x_0\|$ because, otherwise, we would prefer B_1 over H' .

Now it follows that (x_0, x_1) represents the diameter of H' . Since neither $B'_0 \subseteq B'_1$ nor $B'_1 \subseteq B'_0$, it follows that $\max\{r'_0, r'_1\} < r_1$. But then the point x_2 (located at the surface of B_1) cannot belong to $H' = B'_0 \cup B'_1$.

Case 2: Neither $B_0 \subseteq B_1$ nor $B_1 \subseteq B_0$.

The following discussion is illustrated in part b) of Fig. 4. After an appropriate rotation of \mathbb{R}^d , we have precisely the situation that is described in Lemma 19. We may assume that $r_0 \leq r_1$. (If not, then perform another rotation that makes everything upside-down.) We claim that the following four points x_0, x_1, x_2, x_3 form a positive teaching set for H w.r.t. $(\text{BALLS}_2^d, \prec)$:

- x_0 is the north pole of B_0 and x_1 is the south pole of B_1 (which is consistent with the notation used in Lemma 19).
- x_2 is a point on the surface of B_0 that is sufficiently close to x_0 so that it does not belong to B_1 .
- x_3 is a point on the surface of B_1 that is sufficiently close to x_1 so that the distance between x_0 and x_3 is larger than the diameter of B_1 .

Consider any hypothesis $H' \in \text{BALLS}_2^d$ that contains x_0 and x_1 and is not inferior to H w.r.t. \prec . H' cannot consist of a single ball B because B (containing x_0 and x_1) would have a diameter of at least $\|x_1 - x_0\|$ (which is the diameter of H) and B would have a radius greater than $\max\{r_0, r_1\}$ so that H would be preferred over B . Thus, H' is of the form $B'_0 \cup B'_1$ where, for $b = 0, 1$, $B'_b = B(z'_b, r'_b)$ and neither $B'_0 \subseteq B'_1$ nor $B'_1 \subseteq B'_0$. As in the analysis of Case 1, it follows that (x_0, x_1) represents the diameter of H' . It follows that x_0 is the north pole of one of the balls in H' and x_1 is the south pole of the other ball in H' . After renumbering (if necessary), we may assume that x_0 is the north pole of B'_0 and x_1 is the south pole of B'_1 . The chosen location of x_2 makes sure that $x_2 \notin B'_1$ because, otherwise, the diameter of B'_1 would exceed the diameter of B_1 (in which case H would be preferred over H'). It follows that $x_2 \in B'_0$. The chosen location of x_3 similarly implies that $x_3 \notin B'_0$. It follows that $x_3 \in B'_1$. Recall that x_1 is the south pole of B_1 and also of B'_1 . Point x_3 (located at the surface of B_1) can belong to B'_1 only if $r'_1 \geq r_1$. On the other hand $r'_1 \leq r_1$ because, otherwise, H would be preferred over H' . We may conclude that $B'_1 = B_1$. A similar argument (exploiting that x_0 is the north pole of B_0 and also of B'_0 and that $x_2 \in B'_0$) shows that $B'_0 = B_0$. Thus, as claimed, $\{x_1, x_2, x_3, x_4\}$ is a positive teaching set for H w.r.t. $(\text{BALLS}_2^d, \prec)$.

This case analysis concludes the proof of Theorem 20. ■

Let $\text{BALLS}_k^d[\text{transl}]$ be the sub-class of BALLS_k^d consisting of concepts for which all balls in the defining union must have the same radius. $\text{BALLS}_{=k}^d[\text{transl}]$ is defined analogously.

Corollary 21 $\text{PBD}^+(\text{BALLS}_2^d[\text{transl}]) = \text{PBD}^+(\text{BALLS}_{=2}^d[\text{transl}]) = 3$.

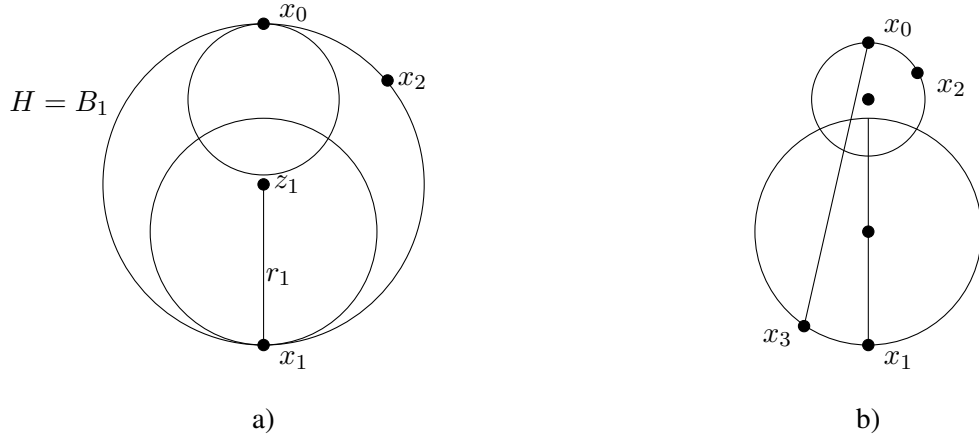


Figure 4: Positive teaching sets for the union of two (not necessarily distinct) balls in the PBT model.

Proof The lower bound is immediate from Corollary 16. For the upper bound, we may proceed as in the proof of Theorem 20 but the point x_2 occurring in the analysis of Case 2 becomes superfluous when we know that $r_0 = r_1$ and $r'_0 = r'_1$. ■

Appendix E. Teaching Unions of Two Cubes

Recall from Corollary 18 that $\text{PBTd}(\text{CUBES}_k^d[\text{disj}])$ is bounded from below by $2k$ and from above by $3k$. We will now demonstrate that 4 positive examples suffice for teaching any union of two (potentially overlapping) cubes:

Theorem 22 For all $d \geq 2$:

$$\text{PBTd}(\text{CUBES}_2^d) = \text{PBTd}^+(\text{CUBES}_2^d) = \text{PBTd}(\text{CUBES}_{=2}^d) = \text{PBTd}^+(\text{CUBES}_{=2}^d) = 4 .$$

Proof $\text{PBTd}(\text{CUBES}_{=2}^d) \geq 4$ is an immediate consequence of Corollary 18. It remains to show that $\text{PBTd}^+(\text{CUBES}_2^d) \leq 4$. For this purpose, choose points to be labeled “1” in a teaching set for a concept C as follows. Here, a concept consisting of a single cube is considered the union of two copies of the same cube. (i) If both cubes are non-degenerate, select two $(d - 1)$ -dimensional faces in C that are parallel and as far apart as possible²; from each pick two diagonally opposite corner points. See Figure 5(a) for illustration. (ii) If both cubes in C are degenerate (i.e., singletons), choose the set of points in C as positive examples. (iii) If exactly one of the two cubes in C is degenerate, choose its single point as the first positive example, and pick two more positive examples as two diagonally opposite corner points of a $(d - 1)$ -dimensional face of the non-degenerate cube; this face is to be chosen so that it is as far away as possible from the degenerate cube. The following hierarchical preference relation is chosen.

² which implies that these faces are not taken from the same cube unless C consists of one cube only

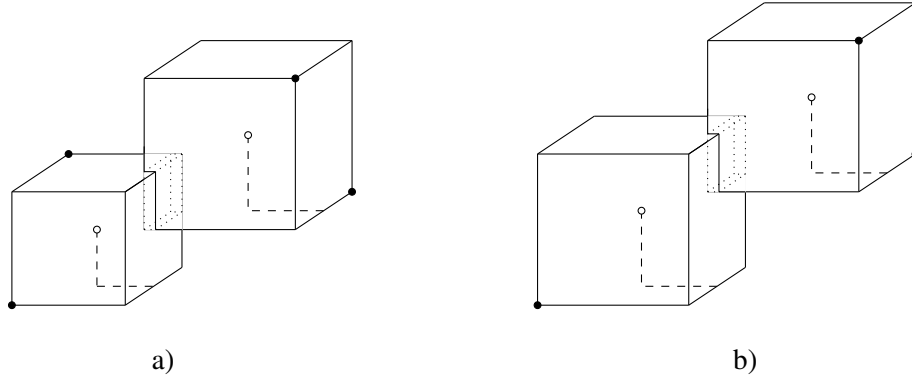


Figure 5: Positive teaching sets for the union of two axis-aligned cubes with arbitrary area and for two axis-aligned cubes with the same area in the PBT model.

Rule 1: With highest priority, prefer concepts whose largest cube is as small as possible.

Rule 2: With second highest priority, prefer concepts whose smallest cube is as small as possible.

Rule 3: With least priority, prefer concepts in which the center points of the two participating cubes are as close to one another as possible.

The first two rules obviously settle the case when the target concept C consists of only one or two points. If C consists of two non-degenerate cubes, Rules 1 and 2 will ensure that the four points are matched up in two pairs correctly, and that each pair will form a diagonal of a face of one cube (note that both cubes must be axis-aligned). Due to Rule 3, the position of each cube relative to the already discovered face is chosen correctly. If exactly one of the two cubes is degenerate, again Rules 1 and 2, together with the constraint that both cubes are axis-aligned, will assign the positive examples correctly to the two cubes, and Rule 3 will determine the position of the non-degenerate cube. ■

Similar to the case of teaching unions of balls all of which have the same (unknown) volume, requiring all cubes in the union to have the same (unknown) volume makes teaching easier.

Theorem 23 For all $d \geq 2$:

$$\begin{aligned} \text{PBTd}(\text{CUBES}_2^d[\text{transl}]) &= \text{PBTd}^+(\text{CUBES}_2^d[\text{transl}]) \\ &= \text{PBTd}(\text{CUBES}_{=2}^d[\text{transl}]) = \text{PBTd}^+(\text{CUBES}_{=2}^d[\text{transl}]) = 3 . \end{aligned}$$

Proof First, we show $\text{PBTd}(\text{CUBES}_2^d[\text{transl}]) \leq 3$. This upper bound is obtained via the same teaching strategy as described in the proof of Theorem 22; here one of the four points labeled “1” can be omitted, as the size of the second cube is determined by the size of the first cube in the union. The cube represented by only one positive example will be positioned by the learner in a way that the given positive example is a corner of the cube that is as far away as possible from the center of the first cube. See Figure 5(b) for illustration.

Second, we prove $\text{PBTd}(\text{CUBES}_{=2}^d[\text{transl}]) \geq 3$. To this end, we once again deploy the lower bound in Equation (4), i.e., we claim that there is a finite subclass \mathcal{L} of $\text{CUBES}_{=2}^d[\text{transl}]$ whose

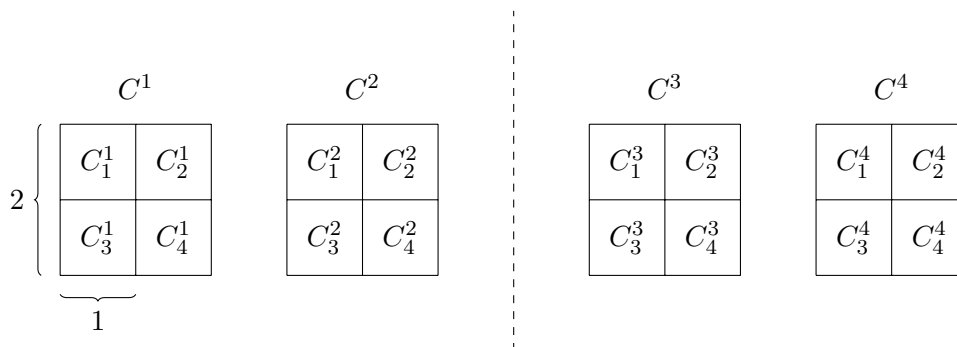


Figure 6: A finite subclass of $\text{CUBES}_{=2}^d[\text{transl}]$ whose TD_{\min} is at least 3: To form a concept consisting of exactly two equal-sized cubes, the first cube in the union is chosen from one of the ten cubes depicted to the left of the dashed line; the second cube is then chosen from the right-hand side of the dashed line, as one of two (one of eight) options if the first cube has side length 2 (side length 1, respectively).

TD_{\min} is at least 3, which will complete the proof. Let C^1, C^2, C^3, C^4 be mutually disjoint “big” cubes of side length 2 each of which decomposes into 2^d “small” cubes of side length 1. See Figure 6 for an illustration in two dimensions. We say that C^1 and C^2 (resp. C^3 and C^4) and all the corresponding subcubes belong to the first (resp. to the second) group of cubes. Thus there are two big and $2 \cdot 2^d$ small cubes in each of the two groups. The concepts in the constructed finite subclass \mathcal{L} of $\text{CUBES}_{=2}^d[\text{transl}]$ are built by choosing one cube from the first group and another one of the same size from the second group. It is now easy to see that none of the concepts in \mathcal{L} can be uniquely specified by a teaching set of size 2. Hence $\text{TD}_{\min}(\mathcal{L}) \geq 3$, which completes the proof. ■

As for the class BOXES_2^d , we have some preliminary results but, as we think that they can be improved further, we do not include them here.