Algebraic Methods Proving Sauer's Bound for Teaching Complexity

Rahim Samei, Pavel Semukhin[1], Boting Yang[2], Sandra Zilles[2]

Department of Computer Science, University of Regina, Regina, Saskatchewan, Canada

Abstract

This paper establishes an upper bound on the size of a concept class with given recursive teaching dimension (RTD, a teaching complexity parameter). The upper bound coincides with Sauer's well-known bound on classes with a fixed VC-dimension. Our result thus supports the recently emerging conjecture that the combinatorics of VC-dimension and those of teaching complexity are intrinsically interlinked.

We further introduce and study RTD-maximum classes (whose size meets the upper bound) and RTD-maximal classes (whose RTD increases if a concept is added to them), showing similarities but also differences to the corresponding notions for VC-dimension.

Another contribution is a set of new results on maximal classes of a given VC-dimension.

Methodologically, our contribution is the successful application of algebraic techniques, which we use to obtain a purely algebraic characterization of teaching sets (sample sets that uniquely identify a concept in a given concept class) and to prove our analog of Sauer's bound for RTD. Such techniques have been used before to prove results relevant to computational learning theory, e.g., by Smolensky (1997), but are not standard in the field.

Keywords: VC-dimension, teaching, Sauer's bound, maximum classes

1. Introduction

An important combinatorial result, proven by Sauer (1972) and independently by Shelah (1972), states that the size of any concept class of Vapnik-Chervonenkis dimension (VC-dimension, introduced by Vapnik and Chervonenkis (1971)) d is at most $\sum_{i=0}^{d} \binom{m}{i}$, where m is the number of instances the concept class is defined over.

In Computational Learning Theory, this bound (typically called *Sauer's bound*) has proven helpful—if not essential—for a variety of studies, most notably for the definition and analysis of *maximum classes*.

Email addresses: `samei20r@cs.uregina.ca` (Rahim Samei), `semukhip@cs.uregina.ca` (Pavel Semukhin), `boting@cs.uregina.ca` (Boting Yang), `zilles@cs.uregina.ca` (Sandra Zilles)

[1]Supported in part by a postdoctoral fellowship from the Pacific Institute for the Mathematical Sciences.
[2]Supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

A concept class of VC-dimension d over a finite instance space X is maximum, if its size meets Sauer's bound.[3] Maximum classes exhibit a number of interesting structural properties, e.g., their complements as well as their restrictions to subsets of the instance space are maximum (Rubinstein et al., 2009; Welzl, 1987). These structural properties have remarkable implications. For example, maximum classes form one of the few general cases of concept classes known to have labeled and unlabeled sample compression schemes of the size of their VC-dimension (Floyd and Warmuth, 1995; Kuzmin and Warmuth, 2007). Moreover, the *recursive teaching dimension* (RTD, a complexity parameter of the recently introduced recursive teaching model (Zilles et al., 2011)) of any maximum class equals its VC-dimension (Doliwa et al., 2010).

Recent work by Doliwa et al. (2010) indicates connections between the VC-dimension and the RTD; besides maximum classes, several other types of concept classes are shown to have an RTD upper-bounded by their VC-dimension. An open question is whether or not the RTD has an upper bound linear in the VC-dimension. Thus recursive teaching is the only model known so far that could potentially establish a close connection between the complexity of learning from a teacher and the complexity of learning from randomly chosen examples (the VC-dimension being an essential complexity parameter for the latter).

This paper establishes a further connection between RTD and VC-dimension: its main result is an analog of Sauer's bound for RTD. We prove that the size of any concept class of RTD r is at most $\sum_{i=0}^{r}\binom{m}{i}$, where m is the size of the instance space. This new evidence of a strong connection between learning from a teacher and learning from randomly chosen examples suggests that the study of the recursive teaching dimension could be of interest in a more general context than just computational teaching. Our result is proven using algebraic methods, which first provide us with a purely algebraic characterization of teaching sets. A *teaching set* for a concept c in a concept class C is a set of labeled examples that is consistent with c but with no other concept in C; thus it uniquely identifies c in C. Our algebraic characterization of teaching sets, a second contribution of this paper, is the main ingredient of our proof of Sauer's bound for RTD, but it may be of independent interest. In particular, the algebraic techniques applied here may provide new proof ideas for combinatorial studies in Computational Learning Theory, e.g., using this technique we give an alternative proof of Kuzmin and Warmuth's result that maximum classes are shortest-path-closed (Kuzmin and Warmuth, 2007). Previously, methods from algebra yielded an alternative proof of Sauer's bound for the VC-dimension (Smolensky, 1997).

Our Sauer-type bound for RTD naturally allows us to define and study the concept of *RTD-maximum classes*—classes whose size meets the upper bound. To distinguish RTD-maximum classes from maximum classes in the original sense, we refer to the latter as *VCD-maximum classes*. Although every VCD-maximum class is shown to be RTD-maximum, RTD-maximum classes turn out to exhibit slightly different properties. For example, their complements are not necessarily RTD-maximum. We further study *RTD-maximal*

[3] In this paper, we restrict ourselves to finite instance spaces.

classes—classes whose RTD increases if any new concept is added to them. Such classes are not necessarily RTD-maximum.

In studying RTD-maximum and RTD-maximal classes, we discover some new interesting properties of VCD-maximal classes. In particular, we provide bounds on the size of VCD-maximal classes, shown in the appendix.

This paper is an extension of a conference paper (Samei et al., 2012).

2. Preliminaries

Let X be a finite set, called *instance space.* Elements of X are called instances. A *concept* on X is a subset of X. Each concept c is identified with a function $c(x)$ defined as follows: $c(x) = 1$ if $x \in c$ and $c(x) = 0$ if $x \notin c$. For $\ell \in \{0, 1\}$, $\bar{\ell}$ is defined as $\bar{\ell} = 1 - \ell$.

A *concept class* C on X is a set of concepts on X, that is, $C \subseteq 2^X$. $\overline{C}$ denotes the complement of C. For $Y \subseteq X$, let $C|_Y$ denote the restriction of C to Y, that is, $C|_Y = \{c \cap Y : c \in C\}$. Similarly, $c|_Y$ means $c \cap Y$. To simplify notation, the restriction $C|_{X\setminus\{x\}}$ will also be denoted as $C - x$, and $c|_{X\setminus\{x\}}$ will be denoted as $c - x$. The *reduction* of C to Y is defined as $C^Y = \{c \subseteq X \setminus Y : \ c \cup c' \in C \text{ for all } c' \subseteq Y\}$. In other words, $c \in C^Y$ if and only if all possible extensions of the concept c from $X \setminus Y$ to X belong to C. If X_1 and X_2 are two disjoint instance spaces, $C_1 \subseteq 2^{X_1}$ and $C_2 \subseteq 2^{X_2}$, then the *direct product* of C_1 and C_2 is a concept class on $X_1 \cup X_2$ defined as $C_1 \times C_2 = \{c_1 \cup c_2 : \ c_1 \in C_1 \text{ and } c_2 \in C_2\}$. If the class C_1 contains only a single concept and $C_2 = 2^{X_2}$, then the class $C_1 \times C_2$ is called a *cube.* If $|X_2| = d$, then such a cube is called a *d-dimensional cube* (or *d-cube* for short).

A set $S \subseteq X$ is *shattered* by the class C if $C|_S = 2^S$. The *VC-dimension* of a class C is defined as $\text{VCD}(C) = \max\{|S| : \ S \text{ is shattered by } C\}$ (Vapnik and Chervonenkis, 1971). Let $\Phi_d(m) = \sum_{i=0}^{d} \binom{m}{i}$. Sauer's lemma states that if $\text{VCD}(C) = d$, then $|C| \leq \Phi_d(|X|)$ (Sauer, 1972; Shelah, 1972). Let $\text{VCD}(C) = d$; then C is called *VCD-maximum* if $|C| = \Phi_d(|X|)$, that is, if the size of C matches the upper bound from Sauer's lemma (Welzl, 1987). A class is called *maximal* with respect to VC-dimension (or *VCD-maximal*) if adding any new concept to the class increases its VC-dimension.

We will often use the formulas $\Phi_d(m) + \Phi_{m-d-1}(m) = 2^m$ and $\Phi_d(m) = \Phi_{d-1}(m-1) + \Phi_d(m-1)$.

The following proposition, proven by Rubinstein et al. (2009), follows immediately from the definition of VC-dimension.

Proposition 1. (Rubinstein et al., 2009) *Let* $C \subseteq 2^X$ *and* $|X| = m$*. Then* $\text{VCD}(C) \leq d$ *if and only if* $\overline{C}$ *contains at least one* $(m-d-1)$*-cube for each subset of* $(m-d-1)$ *instances, i.e.,* $\overline{C}^S \neq \emptyset$ *for every subset* S *of* $m-d-1$ *instances.*

The *one-inclusion graph* $G(C)$ of a concept class C is the labeled graph G with $V(G) = C$ and $E(G) = \{\{c, c'\} : |c \triangle c'| = 1\}$, where $c \triangle c'$ is the symmetric difference of c and c'. Every edge $\{c, c'\} \in E(G)$ is labeled by the instance from $c \triangle c'$. The degree of a concept $c \in C$ in $G(C)$, denoted by $\deg_C(c)$, is the number of edges incident to c.

A *labeled example* is a pair (x, ℓ), where $x \in X$ and $\ell \in \{0, 1\}$. For a set S of labeled examples, $X(S)$ denotes $X(S) = \{x \in X : (x, \ell) \in S \text{ for some } \ell\}$. A set S of labeled examples is a *teaching set* for a concept c in a class C, if c is the only concept from C which is consistent with S. The collection of all teaching sets for c in C is denoted $\mathrm{TS}(c, C)$. For simplicity, if S is a teaching set for c with respect to C, we also call $X(S)$ a teaching set for c with respect to C, since the labels of examples from S are uniquely determined by $X(S)$ and c.

The *teaching dimension* of c in C is $\mathrm{TD}(c, C) = \min\{|S| : S \in \mathrm{TS}(c, C)\}$. The teaching dimension of C is defined as $\mathrm{TD}(C) = \max_{c \in C} \mathrm{TD}(c, C)$ (Goldman and Kearns, 1995; Shinohara and Miyano, 1991). We will also refer to the *minimal* teaching dimension $\mathrm{TD}_{min}(C) = \min_{c \in C} \mathrm{TD}(c, C)$.

The following definitions are based on previous literature on recursive teaching (Doliwa et al., 2010; Zilles et al., 2011). A *teaching plan* for a concept class C is a sequence $P = ((c_1, S_1), \ldots, (c_n, S_n))$, where $C = \{c_1, \ldots, c_n\}$ and $S_i \in \mathrm{TS}(c_i, \{c_i, \ldots, c_n\})$ for all $i = 1, \ldots, n$. The *order* of the teaching plan P is $\mathrm{ord}(P) = \max_{i=1,\ldots,n} |S_i|$. The *recursive teaching dimension* of C is

$$\mathrm{RTD}(C) = \min\{\mathrm{ord}(P) : P \text{ is a teaching plan for } C\}.$$

A teaching plan of C whose order equals $\mathrm{RTD}(C)$ is called an *optimal teaching plan* for C. For an optimal teaching plan $P = ((c_1, S_1), \ldots, (c_n, S_n))$ for C, the set S_i is called a *recursive teaching set* for c_i in C with respect to the plan P, and $|S_i|$ is called the *recursive teaching dimension* of c_i in C with respect to the plan P, denoted $\mathrm{RTD}(c_i, C)$. The words "with respect to the plan P" may be omitted if there is no ambiguity.

At first glance, as a complexity notion for a model of learning from teachers, the RTD parameter seems to require that the teacher and the learner first agree on a particular teaching plan before a target concept can be taught. However, it was proven that there exists a surjective mapping from a family of sets of labeled examples to a concept class C in a way that (i) teacher and learner can communicate successfully without agreeing on a teaching plan and (ii) the size of a set of examples the teacher needs to present in order to teach a concept c in C equals $\mathrm{RTD}(c, C)$; in particular, the largest set required for teaching any concept in C has size $\mathrm{RTD}(C)$ (Zilles et al., 2011).

A teaching plan $P = ((c_1, S_1), \ldots, (c_n, S_n))$ is called *repetition-free*, if $X(S_i) \neq X(S_j)$ for all $i, j \in \{1, \ldots, n\}$ with $i \neq j$.

The RTD has the following properties (Doliwa et al., 2010; Zilles et al., 2011):

- RTD is monotonic, i.e, $\mathrm{RTD}(C') \leq \mathrm{RTD}(C)$ whenever $C' \subseteq C$.
- RTD equals the order of any *canonical teaching plan*, i.e., a teaching plan $((c_1, S_1), \ldots, (c_n, S_n))$ with $|S_i| = \mathrm{TD}_{min}(\{c_i, \ldots, c_n\})$ for all $i = 1, \ldots, n$.

- $\mathrm{RTD}(C) = \max_{C' \subseteq C} \mathrm{TD}_{min}(C')$.

In Tables 3, 4, 5, 6, and 8 the concepts are listed in the same order as they appear in a teaching plan for the given class, and the recursive teaching sets are underlined.

3. Sauer's bound with respect to the recursive teaching dimension

In this section we present our main result. A central theorem required in the proof of our main result is an algebraic characterization of the teaching sets for a concept c in a concept class C, which is interesting in its own right. We begin by introducing the algebraic setting needed for this characterization and for further proofs throughout this paper.

Let $X = \{x_1, \dots, x_m\}$ be a finite instance space, and let $C = \{c_1, \dots, c_n\}$ be a concept class on X. Consider a vector space $\mathbf{F}_2^n$ of dimension n over the field $\mathbf{F}_2$ (i.e., the field consisting of 2 elements). For each polynomial $f(x_1, \dots, x_m)$ with variables from X and coefficients from $\mathbf{F}_2$, we define a vector $f = (f_1, \dots, f_n)$ from $\mathbf{F}_2^n$ as follows

$$f_i = f(c_i(x_1), \dots, c_i(x_m)) \text{ for } i = 1, \dots, n.$$

Note that we use the same notation for a polynomial and a vector. We also associate each concept $c_i \in C$ with the ith standard basis vector $c_i = (0, \dots, 1, \dots, 0)$ of $\mathbf{F}_2^n$. Again, we are using the same notation for a concept and a vector. This should not cause confusion as the exact meaning of such notation will be clear from the context. For instance, by "the vector x_1x_2" we mean the vector in $\mathbf{F}_2^n$ that corresponds to the polynomial x_1x_2. Similarly, an equality like $c = f(x_1, x_2)$ should be interpreted as the equality between two vectors, the one corresponding to the concept c and the one corresponding to the polynomial $f(x_1, x_2)$.

To illustrate this notation, let us consider the concept class in Table 1. In our notation, $c_1 = (1, 0, 0, 0)$,

	x_1	x_2	x_3
c_1	0	1	0
c_2	1	0	1
c_3	1	1	0
c_4	0	0	1

Table 1: A concept class over three instances, consisting of four concepts.

$c_2 = (0, 1, 0, 0)$, $c_3 = (0, 0, 1, 0)$ and $c_4 = (0, 0, 0, 1)$. Moreover, $x_1 = (0, 1, 1, 0)$, $x_2 = (1, 0, 1, 0)$, $x_3 = (0, 1, 0, 1)$, $0 = (0, 0, 0, 0)$ and $1 = (1, 1, 1, 1)$. So we have $x_2 + x_3 = 1$, $x_2x_3 = 0$, $c_2 = x_1x_3$, $c_3 = x_1x_2$, $c_1 = c_3 + x_2 = x_1x_2 + x_2$, and $c_4 = c_2 + x_3 = x_1x_3 + x_3$.

The following theorem provides an algebraic description of teaching sets.

Theorem 1. *Let $C = \{c_1, \ldots, c_n\} \subseteq 2^X$. A set of instances $\{z_1, \ldots, z_k\} \subseteq X$ is a teaching set for a concept c_i if and only if $c_i = f(z_1, \ldots, z_k)$ for some polynomial f over $\mathbf{F}_2$.*

Proof. Suppose $\{z_1, \ldots, z_k\}$ is a teaching set for c_i. It is not hard to see that in this case $c_i = p_1 \cdots p_k$, where $p_t = z_t$ if $c_i(z_t) = 1$ and $p_t = z_t + 1$ if $c_i(z_t) = 0$.

To prove the other implication, consider $c_i \in C$ and assume that $c_i = f(z_1, \ldots, z_k)$ but $\{z_1, \ldots, z_k\}$ is not a teaching set for c_i. Hence there is another concept $c_j \neq c_i$ from C which coincides with c_i on $\{z_1, \ldots, z_k\}$, that is, $c_i(z_t) = c_j(z_t)$ for all $t = 1, \ldots, k$. Thus the following equalities hold

$$f_i = f(c_i(z_1), \ldots, c_i(z_k)) = f(c_j(z_1), \ldots, c_j(z_k)) = f_j.$$

So, the ith and jth coordinates of the vector $f(z_1, \ldots, z_k)$ are equal. By definition, c_i corresponds to the standard basis vector $(0, \ldots, 1, \ldots, 0)$ which has only one coordinate equal to 1, namely, the ith coordinate. Since we assumed that $c_i = f(z_1, \ldots, z_k)$ and showed that $f_i = f_j$, the vector $f(z_1, \ldots, z_k)$ must have at least two coordinates equal to 1, namely, the ith and jth coordinates. This contradicts the assumption that $c_i = f(z_1, \ldots, z_k)$. □

The next theorem is the main result of our paper. It provides a Sauer-type bound on the size of a concept class with a given RTD.

Theorem 2. *Let $C \subseteq 2^X$ and $|X| = m$. If $\mathrm{RTD}(C) = r$ then $|C| \leq \Phi_r(m)$.*

Proof. Let P_m^r be the collection of monomials over $\mathbf{F}_2$ of the form $x_{i_1} \cdots x_{i_k}$, where $0 \leq k \leq r$ and $1 \leq i_1 < \cdots < i_k \leq m$. In the case when $k = 0$ we let the corresponding monomial be equal to the constant 1. Note that $|P_m^r| = \Phi_r(m)$.

Let $c_1, c_2, \ldots, c_n$ be all the concepts from C listed in the same order as they appear in some teaching plan for C of order r. In particular, for every $s = 1, \ldots, n$, we have $\mathrm{TD}(c_s, \{c_s, \ldots, c_n\}) \leq r$.

We will show that the vector space $\mathbf{F}_2^n$ is spanned by the vectors that correspond to the monomials from P_m^r. The theorem then follows from a well-known linear algebra fact that the size of a spanning set cannot be smaller than the dimension of the vector space.

We will show by induction that each c_s lies in the span of P_m^r. Since $\mathrm{TD}(c_1, C) \leq r$, by Theorem 1, c_1 is equal to a polynomial of the form $p_{i_1} \cdots p_{i_k}$ for some $k \leq r$, where each p_t is equal to x_t or $x_t + 1$. It is not hard to see that the product $p_{i_1} \cdots p_{i_k}$ lies in the span of P_m^r, e.g., $(x_1+1)(x_2+1) = x_1x_2 + x_1 + x_2 + 1$, etc.

Now suppose that $c_1, \ldots, c_s$ are in the span of P_m^r. Let $\mathbf{F}_2^{s,0}$ be the subspace of $\mathbf{F}_2^n$ consisting of the vectors whose last $n - s$ coordinates are zeros. Similarly, let $\mathbf{F}_2^{0,n-s}$ be the subspace of $\mathbf{F}_2^n$ consisting of the vectors whose first s coordinates are zeros. Also, let $(v)_{s,0}$ and $(v)_{0,n-s}$ be the projections of a vector $v \in \mathbf{F}_2^n$ to the subspaces $\mathbf{F}_2^{s,0}$ and $\mathbf{F}_2^{0,n-s}$, respectively. In particular, we have $v = (v)_{s,0} + (v)_{0,n-s}$.

Since $\mathrm{TD}(c_{s+1}, \{c_{s+1}, \ldots, c_n\}) \leq r$, applying Theorem 1 to $\{c_{s+1}, \ldots, c_n\}$ and c_{s+1} yields that $(c_{s+1})_{0,n-s} = (p_{i_1} \cdots p_{i_k})_{0,n-s}$ for some $k \leq r$ and some $i_1, \cdots, i_k$, where each p_t is equal to x_t or $x_t + 1$. In other words,

$(c_{s+1} - p_{i_1} \cdots p_{i_k})_{0,n-s} = \mathbf{0}$, which means that $c_{s+1} - p_{i_1} \cdots p_{i_k}$ belongs to the subspace $\mathbf{F}_2^{s,0}$. As before, the product $p_{i_1} \cdots p_{i_k}$ lies in the span of P_m^r. Moreover, by the induction hypothesis, the vectors $c_1, \ldots, c_s$ are in the span of P_m^r, and hence the subspace $\mathbf{F}_2^{s,0}$ is contained in the span of P_m^r. Hence c_{s+1} lies in the span of P_m^r. □

Remark. The main ingredient of the proof of Theorem 2 is that if $\mathrm{RTD}(C) = r$, then the monomials from P_m^r span the vector space $\mathbf{F}_2^{|C|}$. This idea was used previously by Smolensky (1997) to prove Sauer's bound for VCD. Namely, he showed that if $\mathrm{VCD}(C) = d$, then the monomials from P_m^d span $\mathbf{F}_2^{|C|}$. However, the technique we used to prove that P_m^r spans $\mathbf{F}_2^{|C|}$ is different from the one used in Smolensky's argument. In particular, our technique is based on algebraic characterization of teaching sets provided in Theorem 1. To see the difference, one can compare the proof of Theorem 2 with the proof of Lemma 1 in Section 6 which is based on Smolensky's original idea.

The above result suggests that the notions of RTD and VCD are related, at least for certain types of concept classes—an observation that is in line with the recent results proven by Doliwa et al. (2010).

4. RTD-maximum classes

The Sauer-type bound in Theorem 2 is tight for any r and m, in particular, it is met by all VCD-maximum classes of VC-dimension r, as we will see in Proposition 2. This suggests the following definition.

Definition 1. Let $C \subseteq 2^X$, $|X| = m$, and $\mathrm{RTD}(C) = r$. C is called RTD-*maximum* if $|C| = \Phi_r(m)$, and C is called RTD-*maximal* if $\mathrm{RTD}(C \cup \{c\}) > r$ for any concept $c \notin C$.

In Sections 4 and 5 we will prove various properties of RTD-maximum and RTD-maximal classes and compare them with their analogs for VC-dimension. For convenience, the main results of these sections are summarized in Table 2.

4.1. RTD-*maximum classes versus* VCD-*maximum classes*

This section deals with general properties of RTD-maximum classes, in particular, in comparison to the properties that VCD-maximum classes possess. We will see that there are some similarities between them, but also many differences between their structural properties.

We begin with some simple observations relating RTD-maximum classes to VCD-maximum classes.

Proposition 2. (i) *Every* VCD-*maximum class C is also* RTD-*maximum with* $\mathrm{RTD}(C) = \mathrm{VCD}(C)$.

(ii) *There is a class C for which both C and $\overline{C}$ are* RTD-*maximum, but neither C nor $\overline{C}$ is* VCD-*maximum. In particular, there are* RTD-*maximum classes that are not* VCD-*maximum.*

Property	**K = VCD**	**K = RTD**
C is K-maximum $\Rightarrow C\|_{X'}$ is K-maximum	Yes	No (Prop. 4)
C is K-maximum $\Rightarrow \overline{C}$ is K-maximum	Yes	Yes, if $K(C) = 1$ No, in general (Prop. 5 and 6)
C is K-maximum $\Rightarrow K(C) + K(\overline{C}) = \|X\| - 1$	Yes	Yes, if $K(C) = 1$ No, in general (Prop. 6 and Table 5)
$K(C) + K(\overline{C}) = \|X\| - 1 \Rightarrow C$ is K-maximum	Yes	Yes (Prop. 7)
C is K-maximum $\Rightarrow C$ is shortest-path closed	Yes (Thm. 3)	No (Prop. 8)
C is K-maximum and $K(C) = \|X\| - 2 \Rightarrow$ $\|\{c \in C : \mathrm{TD}(c, C) = \|X\| - 2\}\| \geq \|X\| - 1$	Yes (Prop. 14)	No (Prop. 15)
C is K-maximum $\Rightarrow C$ has a sample compression scheme of size $K(C)$	Yes	No (Corollary 1)
C is K-maximal $\Rightarrow C$ shatters all subsets of size $K(C)$	No (Table 7)	Yes (Prop. 16)
$K(C) = 1$ and C is K-maximal $\Rightarrow C$ is K-maximum	Yes	Yes (Prop. 17)
There is a K-maximal class that is not K-maximum	Yes	Yes (Prop. 18)

Table 2: Summary of the main results of Sections 4 and 5, in the context of known results on VCD.

Proof. (i) For every VCD-maximum class C, $\mathrm{RTD}(C) = \mathrm{VCD}(C)$ (Doliwa et al., 2010). It follows from Theorem 2 and Definition 1 that C is RTD-maximum.

(ii) C_1 in Table 3 is RTD-maximum with $\mathrm{RTD}(C_1) = 2$, and $\overline{C_1}$ is RTD-maximum with $\mathrm{RTD}(\overline{C_1}) = 1$. As $\mathrm{VCD}(C_1) = 3$ and $\mathrm{VCD}(\overline{C_1}) = 2$, neither C_1 nor $\overline{C_1}$ is VCD-maximum. □

Since for RTD-maximum classes the VC-dimension can exceed the recursive teaching dimension, it is natural to ask how large the difference between these two parameters can be. The following proposition answers this question.

Proposition 3. *For any two integers i and d with $1 \leq i \leq d$, there is an* RTD*-maximum class C such that* $\mathrm{RTD}(C) = i$ *and* $\mathrm{VCD}(C) \geq d$*.*

Proof. Fix positive integers d and i with $1 \leq i \leq d$. Choose any integer m such that $\binom{m}{i} \geq 2^d$. Let C' be a concept class on $X = Y \cup Z$ with $|Y| = m$ and $|Z| = d$ such that C' contains all subsets of X of size at most i. So, $|C'| = \Phi_i(|X|)$ and for any concept $c' \in C'$, $|c'| = |\{x \in X : c'(x) = 1\}| \leq i$. Note that $\mathrm{RTD}(C') = i$; in particular, there is a teaching plan $((c'_1, S_1), \ldots, (c'_{|C'|}, S_{|C'|}))$ of order i for C' for which the concepts c'_1, c'_2, …, c'_{2^d} are of size i and are subsets of Y.

$c_i \in C_1$	x_1	x_2	x_3	x_4
c_1	$\underline{1}$	1	1	$\underline{1}$
c_2	0	0	$\underline{1}$	$\underline{1}$
c_3	0	$\underline{1}$	0	$\underline{1}$
c_4	0	$\underline{1}$	$\underline{1}$	0
c_5	$\underline{1}$	0	$\underline{1}$	0
c_6	$\underline{1}$	$\underline{1}$	0	0
c_7	0	0	0	$\underline{1}$
c_8	0	0	$\underline{1}$	0
c_9	0	$\underline{1}$	0	0
c_{10}	$\underline{1}$	0	0	0
c_{11}	0	0	0	0

$c_i \in \overline{C_1}$	x_1	x_2	x_3	x_4
c_1	$\underline{0}$	1	1	1
c_2	1	1	1	$\underline{0}$
c_3	1	0	$\underline{1}$	1
c_4	1	$\underline{1}$	0	1
c_5	1	0	0	1

Table 3: C_1 and $\overline{C_1}$ are RTD-maximum but neither C_1 ($\{x_1, x_2, x_3\}$ is shattered) nor $\overline{C_1}$ ($\{x_2, x_3\}$ is shattered) is VCD-maximum. Recursive teaching sets are underlined.

$c_i \in C_2$	x_1	x_2	x_3	x_4
c_1	$\underline{1}$	0	0	0
c_2	0	1	1	$\underline{1}$
c_3	0	$\underline{1}$	0	0
c_4	0	0	$\underline{1}$	0
c_5	0	0	0	0

Table 4: C_2 is RTD-maximum but $C_2 - x_4$ is not. Recursive teaching sets are underlined.

We construct an RTD-maximum concept class C with $\text{VCD}(C) = d$ and $\text{RTD}(C) = i$ in the following way. First, we fix an order over all subsets of Z. Then, we define new concepts c_k over X, for $1 \leq k \leq |C'|$, as follows.

- For $1 \leq k \leq 2^d$, let $c_k = c'_k \cup b_k$ where b_k is the kth subset of Z in the given order.
- For $k > 2^d$, let $c_k = c'_k$.

Let $C = \{c_1, \ldots, c_{|C'|}\}$. In particular, $|C| = |C'| = \Phi_i(|X|)$. On the one hand, $((c_1, S_1), \ldots, (c_{|C'|}, S_{|C'|}))$ is a teaching plan of order i for C and thus $\text{RTD}(C) \leq i$. On the other hand, since $|C| = \Phi_i(|X|)$, Theorem 2 implies $\text{RTD}(C) \geq i$. Hence, we obtain $\text{RTD}(C) = i$ and C is RTD-maximum. Furthermore, $\text{VCD}(C) \geq d$ since C shatters the set Z of size d. □

An interesting consequence of Proposition 3 is that, even for RTD-maximum classes, the recursive teaching dimension is not an upper bound on the smallest possible size of a *sample compression scheme*, as we will show below. A labeled sample compression scheme of size k for a concept class C over X (Littlestone and Warmuth, 1986; Floyd and Warmuth, 1995) consists of two mappings f and g which map labeled sets of examples to labeled sets of examples. f maps every set S that is consistent with some concept in C to a subset $S' \subseteq S$ of size at most k. Given such a set S', g returns a set $S'' \supseteq S$ containing exactly one labeled example for each instance in X; in particular, all examples originally presented to f must be contained in the set output by g. In an unlabeled compression scheme, the sets output by f and input to g are just sets of instances without any labels.

Littlestone and Warmuth (1986) showed that sample compression schemes can be interpreted as PAC-learning algorithms (Valiant, 1984) and that their size yields bounds on the sample complexity of PAC-learning for a specified concept class. These bounds would improve on the best known ones for PAC-learning if it could be shown that the smallest possible size of a sample compression scheme for a class C is linear in the VC-dimension of C. It was conjectured that such a linear relationship exists, but only in a few special cases has the conjecture been proven so far. Most notably, any VCD-maximum class C possesses a (labeled and an unlabeled) sample compression scheme of size $\mathrm{VCD}(C)$ (Floyd and Warmuth, 1995; Kuzmin and Warmuth, 2007).

Doliwa et al. (2010) revealed a strong relationship between sample compression schemes and recursive teaching sets, in particular for the case of VCD-maximum classes. In proving that RTD and VCD coincide on VCD-maximum classes, they showed a 1-1 correspondence between recursive teaching sets and the compression sets $f(S)$ used in unlabeled sample compression schemes for those classes. In particular, every VCD-maximum class C has a sample compression scheme of size $\mathrm{RTD}(C)$. This result cannot be generalized to RTD-maximum classes.

Corollary 1. *There is an* RTD*-maximum class C for which no (labeled or unlabeled) sample compression scheme of size* $\mathrm{RTD}(C)$ *exists.*

Proof. Floyd and Warmuth (1995) showed that no concept class of VC-dimension d has a sample compression scheme of size at most $d/5$. By Proposition 3, for any $d \geq 5$, there are RTD-maximum classes C with $\mathrm{VCD}(C) = d$ for which $\mathrm{RTD}(C) \leq d/5$; these cannot have sample compression schemes of size $\mathrm{RTD}(C)$. □

Due to Welzl (1987), we know that restricting a VCD-maximum class to a subset of its instance space yields another VCD-maximum class. But this property does not in general hold for RTD-maximum classes, as we show next.

Proposition 4. *There is an* RTD*-maximum class which has a restriction that is not* RTD*-maximum. Furthermore, there is an* RTD*-maximum class C that has an* RTD*-maximum restriction C' such that* $\mathrm{RTD}(C') > \mathrm{RTD}(C)$*.*

Proof. Consider C_2 in Table 4. It is easy to see that C_2 is RTD-maximum and $\mathrm{RTD}(C_2) = 1$. However, $\mathrm{RTD}(C_2 - x_4) = 2$ and $C_2 - x_4$ is not RTD-maximum. Furthermore, consider the RTD-maximum class C_1 in Table 3. Clearly, $C_1 - x_4$ is RTD-maximum and $\mathrm{RTD}(C_1) = 2 < \mathrm{RTD}(C_1 - x_4) = 3$. □

As mentioned in Section 1, the complement of any VCD-maximum class is VCD-maximum. But RTD-maximum classes do not possess this property.

Proposition 5. *There is an* RTD*-maximum class whose complement is not* RTD*-maximum.*

Proof. Consider the RTD-maximum class C with $\mathrm{RTD}(C) = 3$ in Table 5. $\overline{C}$ is not RTD-maximum because $\mathrm{RTD}(\overline{C}) = 2$ and $6 < \Phi_2(5)$. □

$c_i \in C$	x_1	x_2	x_3	x_4	x_5	$c_i \in C$	x_1	x_2	x_3	x_4	x_5
c_1	$\underline{1}$	$\underline{1}$	$\underline{1}$	1	1	c_{14}	0	$\underline{1}$	0	0	$\underline{1}$
c_2	1	1	$\underline{0}$	$\underline{1}$	$\underline{1}$	c_{15}	$\underline{1}$	0	$\underline{1}$	$\underline{1}$	0
c_3	$\underline{1}$	$\underline{1}$	0	$\underline{1}$	0	c_{16}	$\underline{1}$	0	0	$\underline{1}$	0
c_4	$\underline{1}$	$\underline{1}$	0	0	$\underline{1}$	c_{17}	0	$\underline{1}$	$\underline{1}$	0	0
c_5	0	$\underline{1}$	1	$\underline{1}$	$\underline{1}$	c_{18}	0	$\underline{1}$	0	0	0
c_6	$\underline{1}$	0	1	$\underline{1}$	$\underline{1}$	c_{19}	0	0	$\underline{1}$	$\underline{1}$	0
c_7	0	0	1	$\underline{1}$	$\underline{1}$	c_{20}	0	0	0	$\underline{1}$	0
c_8	$\underline{1}$	$\underline{1}$	0	0	0	c_{21}	$\underline{1}$	0	$\underline{1}$	0	0
c_9	$\underline{1}$	0	$\underline{1}$	0	$\underline{1}$	c_{22}	$\underline{1}$	0	0	0	0
c_{10}	$\underline{1}$	0	0	0	$\underline{1}$	c_{23}	0	0	$\underline{1}$	0	$\underline{1}$
c_{11}	0	$\underline{1}$	$\underline{1}$	$\underline{1}$	0	c_{24}	0	0	$\underline{1}$	0	0
c_{12}	0	$\underline{1}$	0	$\underline{1}$	0	c_{25}	0	0	0	0	$\underline{1}$
c_{13}	0	$\underline{1}$	$\underline{1}$	0	$\underline{1}$	c_{26}	0	0	0	0	0

$c_i \in \overline{C}$	x_1	x_2	x_3	x_4	x_5
c_1	$\underline{0}$	$\underline{0}$	0	1	1
c_2	$\underline{0}$	1	0	1	1
c_3	1	0	$\underline{0}$	1	1
c_4	1	1	1	0	$\underline{1}$
c_5	1	1	1	$\underline{0}$	0
c_6	1	1	1	1	0

Table 5: C is RTD-maximum but $\overline{C}$ is not. Recursive teaching sets are underlined.

By contrast with Proposition 5, we can show that the complement of an RTD-maximum class of RTD 1 is still RTD-maximum.

Proposition 6. *Let C be an* RTD*-maximum class over X with $|X| \geq 2$. If* $\mathrm{RTD}(C) = 1$*, then $\overline{C}$ is* RTD*-maximum and* $\mathrm{RTD}(\overline{C}) = |X| - 2$.

Proof. For $|X| = 2$, it is easy to verify that the proposition is true. Suppose it is also true for $|X| < m$. Now consider the case $|X| = m > 2$. Let $c_1 \in C$ with $\mathrm{TD}(c_1, C) = 1$. Without loss of generality, let $\{(x_1, 1)\}$ be a teaching set for c_1 in C. Then we can write C as a disjoint union of $\{c_1\}$ and $\{0\} \times C_1$, where

$C_1 = (C \setminus \{c_1\}) - x_1$ is an RTD-maximum class with $\mathrm{RTD}(C_1) = 1$ on $X \setminus \{x_1\}$. So, the complement of C is equal to the disjoint union $\overline{C} = (\{0\} \times \overline{C_1}) \cup (\{1\} \times C_2)$, where $C_2 = 2^{X \setminus \{x_1\}} \setminus \{c_1 - x_1\}$ is a class of size $2^{m-1} - 1$ on $X \setminus \{x_1\}$.

By the induction hypothesis, there is a teaching plan of order $m - 3$ for $\overline{C_1}$. Take such a plan and extend every recursive teaching set S from this plan to $S \cup \{(x_1, 0)\}$. As a result, we obtain a teaching plan for $\{0\} \times \overline{C_1}$ of order $m - 2$, which we call P_1. Note that C_2 is a VCD-maximum class with $\mathrm{VCD}(C_2) = |X \setminus \{x_1\}| - 1 = m - 2$, and hence $\mathrm{RTD}(C_2) = m - 2$. Since $\mathrm{RTD}(\{1\} \times C_2) = \mathrm{RTD}(C_2)$, there is a teaching plan of order $m - 2$ for $\{1\} \times C_2$, which we call P_2.

Every recursive teaching set from P_1 contains $(x_1, 0)$, which distinguishes the concepts in $\{0\} \times \overline{C_1}$ from those in $\{1\} \times C_2$. So, P_1 and P_2 can be merged into a teaching plan for $\overline{C}$ of order $m - 2$. Thus $\mathrm{RTD}(\overline{C}) \leq m - 2$. Furthermore, $|\overline{C}| = 2^m - |C| = 2^m - (m + 1) = \Phi_{m-2}(m)$. By Theorem 2, we have $\mathrm{RTD}(\overline{C}) \geq m - 2$, and hence $\overline{C}$ is RTD-maximum. □

The RTD-maximum class C in the proof of Proposition 5 fulfills $\mathrm{RTD}(C) + \mathrm{RTD}(\overline{C}) = |X|$. In contrast to this, note that a class C is VCD-maximum if and only if $\mathrm{VCD}(C) + \mathrm{VCD}(\overline{C}) = |X| - 1$. Necessity of the condition was proven by Rubinstein et al. (2009). Sufficiency is easy to see, as was pointed out by an anonymous reviewer of a preliminary version of this paper (Samei et al., 2012): Suppose C with $\mathrm{VCD}(C) = d$ is not VCD-maximum. Then $|C| < \Phi_d(|X|)$ and thus $|\overline{C}| > 2^{|X|} - \Phi_d(|X|) = \Phi_{|X|-d-1}(|X|)$, which implies $\mathrm{VCD}(\overline{C}) > |X| - d - 1$. The same reasoning implies that the condition is sufficient as well when VCD is replaced by RTD.

Proposition 7. *For any $C \subseteq 2^X$, if $\mathrm{RTD}(C) + \mathrm{RTD}(\overline{C}) = |X| - 1$, then C is RTD-maximum.*

Remark. The converse of Proposition 7 is false: Table 5 contains an RTD-maximum class C on 5 instances for which $\mathrm{RTD}(C) + \mathrm{RTD}(\overline{C}) = 3 + 2 > 5 - 1$. However, a weaker statement is still true: if C and $\overline{C}$ are *both* RTD maximum, then $\mathrm{RTD}(C) + \mathrm{RTD}(\overline{C}) = |X| - 1$. Indeed, let $\mathrm{RTD}(C) = r$. Since C is RTD-maximum, $|C| = \Phi_r(|X|)$. Hence $|\overline{C}| = 2^{|X|} - \Phi_r(|X|) = \Phi_{|X|-r-1}(|X|)$. By Theorem 2, $\mathrm{RTD}(\overline{C}) \geq |X| - r - 1$. If $\mathrm{RTD}(\overline{C}) > |X| - r - 1$, then again Theorem 2 implies that $\overline{C}$ is not RTD-maximum. So, $\mathrm{RTD}(\overline{C}) = |X| - r - 1$ and we obtain that $\mathrm{RTD}(C) + \mathrm{RTD}(\overline{C}) = |X| - 1$.

Another difference between VCD-maximum classes and RTD-maximum classes is that the former are always shortest-path closed (Kuzmin and Warmuth, 2007), while the latter are not. A class C is *shortest-path closed* if any two concepts $c, c' \in C$ are Hamming-connected, i.e., there are pairwise distinct instances $x_1, \ldots, x_k$ and $c_1, \ldots, c_{k-1} \in C$ such that, with $c_0 = c$ and $c_k = c'$, the concepts c_{i-1} and c_i differ only in x_i, for $1 \leq i \leq k$. In other words, there is a path in $G(C)$ between c and c', which is labeled by instances of $c \triangle c'$ and has a length of $|c \triangle c'|$.

Proposition 8. *There is an RTD-maximum class that is not shortest-path closed.*

Proof. The RTD-maximum class C_1 in Table 3 is not shortest-path closed because the concept c_{11} has Hamming distance at least 2 from any other concept in the class. □

Note that shortest-path closedness is not the distinguishing property between RTD-maximum and VCD-maximum classes.

Proposition 9. *There is an* RTD*-maximum class that is shortest-path closed but not* VCD*-maximum.*

Proof. The RTD-maximum class C in Table 6 is of RTD 2 and also shortest-path closed. However, the VC-dimension of C is 3 which means that it is not VCD-maximum. This class was found by exhaustive enumeration of all concept classes over instance spaces of size at most 5. □

$c_i \in C$	x_1	x_2	x_3	x_4
c_1	$\underline{0}$	0	0	$\underline{1}$
c_2	$\underline{0}$	0	$\underline{1}$	0
c_3	$\underline{0}$	$\underline{0}$	0	0
c_4	$\underline{0}$	1	0	0
c_5	1	$\underline{0}$	$\underline{0}$	1
c_6	1	$\underline{0}$	1	$\underline{0}$
c_7	1	$\underline{0}$	1	1
c_8	1	1	$\underline{0}$	$\underline{0}$
c_9	1	1	$\underline{0}$	1
c_{10}	1	1	1	$\underline{0}$
c_{11}	1	1	1	1

Table 6: C is RTD-maximum and shortest-path closed (found by computer experiments) but not VCD-maximum ($\{x_2, x_3, x_4\}$ is shattered). Recursive teaching sets are underlined.

On the other hand, in the case when RTD is equal to 1 we have the following result.

Proposition 10. *Let C be an* RTD*-maximum class with* $\mathrm{RTD}(C) = 1$*. If C is shortest-path closed, then C is* VCD*-maximum with* $\mathrm{VCD}(C) = 1$*.*

Proof. Let $C \subseteq 2^X$ be a shortest-path closed RTD-maximum class with $\mathrm{RTD}(C) = 1$ and $X = \{x_1, \ldots, x_m\}$. It follows from the proof of Theorem 2 that the monomials $\{1, x_1, \ldots, x_m\}$ form a spanning set for $\mathbf{F}_2^{|C|}$. Since $|\{1, x_1, \ldots, x_m\}| = m + 1 = |C|$, these monomials form a basis. In particular, we have $x_i \neq 0$ and $x_i \neq 1$ for all $i \in \{1, \ldots, m\}$.

So for any $i \in \{1, \ldots, m\}$, there is a pair of concepts $c_1, c_2 \in C$ such that $c_1(x_i) = 1$ and $c_2(x_i) = 0$. Since C is shortest-path closed, c_1 and c_2 must be Hamming-connected. In particular, the one-inclusion graph of C must have an edge with label x_i. Hence we have that $C^{\{x_i\}} \neq \emptyset$ for any $i \in \{1, \ldots, m\}$. It follows from Proposition 1 that $\mathrm{VCD}(\overline{C}) \leq m - 2$. But $|\overline{C}| = 2^m - |C| = 2^m - \Phi_1(m) = \Phi_{m-2}(m)$. So by

Sauer's lemma $\mathrm{VCD}(\overline{C}) = m - 2$ and $\overline{C}$ is VCD-maximum. Therefore, C itself is a VCD-maximum class with $\mathrm{VCD}(C) = 1$. □

4.2. Teaching plans of RTD*-maximum classes*

In this section, we analyze the structure of teaching plans of RTD-maximum classes, which is motivated by the rich structure of teaching plans of VCD-maximum classes. For example, due to Doliwa et al. (2010), we know that each VCD-maximum class C possesses a repetition-free teaching plan of order $\mathrm{RTD}(C) = \mathrm{VCD}(C)$, and the recursive teaching sets used in such repetition-free teaching plans immediately correspond to sample compression schemes whose size equals the VC-dimension. Further motivation is that understanding the structure of teaching plans of RTD-maximum classes can help us to understand the structure of these classes themselves.

While it remains open whether or not RTD-maximum classes always possess repetition-free teaching plans, the results presented in this section provide useful insights into the structural properties of RTD-maximum classes.

We begin by showing that, for any teaching plan of an RTD-maximum class C, all instance sets of size $\mathrm{RTD}(C)$ are used as recursive teaching sets. This result is a consequence of the proof of Theorem 2.

Proposition 11. *Let $C \subseteq 2^X$ be* RTD*-maximum and* $\mathrm{RTD}(C) = r$*. Let $X' \subseteq X$ be any subset of size r. Then for any teaching plan P for C of order r, there is a concept $c \in C$ and a recursive teaching set S for c with respect to P, such that $X(S) = X'$.*

Proof. Let $X' = \{x_{i_1}, \ldots, x_{i_r}\}$, and P be a teaching plan for C of order r such that $c_1, c_2, \ldots, c_n$ are all concepts from C listed in the same order as they appear in P. Assume that X' does not appear as a recursive teaching set in the plan P. Then, in the proof of Theorem 2 we can always represent the concept c_{s+1} inside the class $\{c_{s+1}, \ldots, c_n\}$ as a polynomial $f(z_1, \ldots, z_r)$ over $\mathbf{F}_2$ such that $\{z_1, \ldots, z_r\} \neq \{x_{i_1}, \ldots, x_{i_r}\}$. (This follows from Theorem 1 and the fact that X' is not used as a recursive teaching set.) As a consequence, we can span $\mathbf{F}_2^n$ without using the monomial $x_{i_1} \cdots x_{i_r}$, which implies that $|C| = \dim(\mathbf{F}_2^n) \leq \Phi_r(|X|) - 1$. Hence C is not RTD-maximum. This is a contradiction. □

Another consequence of Theorem 2 is that, for an RTD-maximum class, teaching sets of size 1 cannot be used too early in any teaching plan.

Proposition 12. *Let $C \subseteq 2^X$ be* RTD*-maximum, $|X| = m$, and* $\mathrm{RTD}(C) = r$*. For an arbitrary teaching plan for C, let $(c_1, c_2, \ldots, c_n)$ be the sequence of all concepts of C listed in the plan. Then for any positive integer $i < \Phi_{r-1}(m-1)$, we have* $\mathrm{TD}(c_i, \{c_i, \ldots, c_n\}) > 1$*.*

Proof. Assume that $\mathrm{TD}(c_i, \{c_i, \ldots, c_n\}) = 1$ for some $i < \Phi_{r-1}(m-1)$. Let $(x, \ell) \in \mathrm{TS}(c_i, \{c_i, \ldots, c_n\})$ for some $x \in X$ and $\ell \in \{0, 1\}$. Then we have $c(x) = \bar{\ell}$ for any $c \in \{c_{i+1}, \ldots, c_n\}$. So, $|\{c_{i+1}, \ldots, c_n\}| =$

$|\{c_{i+1},\ldots,c_n\}-x|$. Consequently,

$$\begin{aligned} |C| &= |\{c_1,\ldots,c_i\}|+|\{c_{i+1},\ldots,c_n\}| = i+|\{c_{i+1},\ldots,c_n\}| \\ &= i+|\{c_{i+1},\ldots,c_n\}-x| \ \le \ i+\Phi_r(m-1) \ \ \text{(by Theorem 2)} \\ &< \Phi_{r-1}(m-1)+\Phi_r(m-1)=\Phi_r(m). \end{aligned}$$

Thus C is not RTD-maximum. This is a contradiction. □

Note that Proposition 12 is tight in the sense that there is an RTD-maximum class of RTD r that possesses an optimal teaching plan in which the concept at position $i=\Phi_{r-1}(m-1)$ has a recursive teaching set of size 1. In particular, the VCD-maximum class that contains all concepts of size at most r fulfills this property. The witnessing optimal teaching plan begins with all concepts of size r that contain an arbitrary but fixed instance x, followed by those of size $r-1$ that contain x and so on. The concept c containing only x occurs at position $\binom{m-1}{r-1}+\cdots+\binom{m-1}{1}+1=\Phi_{r-1}(m-1)$ in this plan. After that point, no concept in the plan contains x, and hence $\{(x,1)\}$ is a recursive teaching set of size one for c in the chosen plan.

As a generalization of Proposition 12, we observe that, in any teaching plan for any RTD-maximum class C, recursive teaching sets S of size less than RTD(C) can only be used after all instance sets $X'\supset X(S)$ of size RTD(C) have been used as recursive teaching sets.

Proposition 13. *Let $C\subseteq 2^X$ be* RTD*-maximum with* $\mathrm{RTD}(C)=r$ *and let* $P=((c_1,S_1),\ldots,(c_n,S_n))$ *be a teaching plan for C. Suppose there is some $i\in\{1,\ldots,n\}$ such that $|X(S_i)|\le r-1$. Then for all sets X' with $X(S_i)\subset X'\subseteq X$ and $|X'|=r$, there is an index $j\in\{1,\ldots,i-1\}$ such that $X(S_j)=X'$.*

Proof. Let $X=\{x_1,\ldots,x_m\}$. Without loss of generality, assume that there is an index i such that $X(S_i)=\{x_1,\ldots,x_t\}$, where $t\le r-1$, but for every $j<i$, $X(S_j)\neq\{x_1,\ldots,x_t,x_{t+1},\ldots,x_r\}$. We will use the same notation as in the proof of Theorem 2. In particular, P_m^r is the collection of monomials over $\mathbf{F}_2$ of the form $x_{i_1}\cdots x_{i_k}$, where $0\le k\le r$ and $1\le i_1<\cdots<i_k\le m$.

Note that P_m^r is a basis for $\mathbf{F}_2^{|C|}$ because P_m^r spans $\mathbf{F}_2^{|C|}$ and $|C|=\Phi_r(m)=|P_m^r|$. Thus every concept $c\in C$ can be expressed as a linear combination of monomials from P_m^r in a unique way. We now express c_i as a linear combination of monomials from P_m^r in two different ways: one will contain the monomial $x_1x_2\cdots x_r$ and the other will not.

Since $X(S_i)=\{x_1,\ldots,x_t\}$, by Theorem 1,

$$(c_i)_{0,n-i}=x_1x_2\cdots x_t+L(x_1,\ldots,x_t),$$

where $L(x_1,\ldots,x_t)$ is a linear combination of monomials of degree less than t with the variables among $x_1,\ldots,x_t$. In particular, the linear combination for $(c_i)_{0,n-i}$ does not contain the monomial $x_1x_2\cdots x_r$. Note that c_i is equal to $(c_i)_{0,n-i}$ plus a linear combination of $c_1,\ldots,c_{i-1}$. By assumption, none of the

$c_1, \ldots, c_{i-1}$ contains the monomial $x_1x_2\cdots x_r$ in its expression as a linear combination of P_m^r. Therefore, c_i can be expressed as a linear combination of monomials from P_m^r in which the monomial $x_1x_2\cdots x_r$ does not occur.

Notice that any superset of $X(S_i)$, and in particular the set $\{x_1, \ldots, x_r\}$, is also a recursive teaching set for c_i according to plan P. Again, by Theorem 1, we can write

$$(c_i)_{0,n-i} = x_1x_2\cdots x_r + L(x_1, \ldots, x_r),$$

where $L(x_1, \ldots, x_r)$ is a linear combination of monomials of degree less than r with the variables among $x_1, \ldots, x_r$. Hence the linear combination for $(c_i)_{0,n-i}$ contains the monomial $x_1x_2\cdots x_r$. As before, c_i is equal to $(c_i)_{0,n-i}$ plus a linear combination of $c_1, \ldots, c_{i-1}$ and, by assumption, none of the $c_1, \ldots, c_{i-1}$ contains the monomial $x_1x_2\cdots x_r$ in its expression as a linear combination of P_m^r. So, $c_1, \ldots, c_{i-1}$ cannot cancel out $x_1x_2\cdots x_r$ from $(c_i)_{0,n-i}$. Therefore, c_i can be expressed as a linear combination of monomials from P_m^r which contains $x_1x_2\cdots x_r$.

Thus we have expressed c_i as a linear combination of monomials from P_m^r in two different ways. This contradicts the fact that P_m^r is a basis. □

An interesting question is how many concepts in an RTD-maximum class C have the smallest teaching dimension with respect to C. All of these concepts would be safe choices to start the construction of a canonical teaching plan, which is always an optimal plan. Kuzmin and Warmuth (2007) conjectured that VCD-maximum classes of VC-dimension d always have at least $d+1$ concepts of teaching dimension d. The latter is provably the smallest teaching dimension a concept in such a class can have.[4] We cannot offer a proof of Kuzmin and Warmuth's conjecture in the general case, but it turns out to be true for the case of $d = |X| - 2$. Since VCD-maximum classes are also RTD-maximum, this result is relevant to this section.

Proposition 14. *Let $C \subseteq 2^X$ with $|X| = m$. Suppose C is* VCD*-maximum with* $\mathrm{VCD}(C) = m - 2$. *Then there are at least $m-1$ concepts $c \in C$ with* $\mathrm{TD}(c, C) = m-2$ *(and no concepts $c \in C$ with* $\mathrm{TD}(c, C) < m-2$*).*

Proof. Let $C \subseteq 2^X$ be a VCD-maximum class with $\mathrm{VCD}(C) = m - 2$ and let $\overline{C}$ be the complement of C. Then $\overline{C}$ is a VCD-maximum class with $\mathrm{VCD}(\overline{C}) = 1$. According to Dudley (1985), the latter implies that $G(\overline{C})$ is a tree.

We now show that for any pair of adjacent edges in $G(\overline{C})$, there is a concept $c \in C$ with $\mathrm{TD}(c, C) = m-2$. Let $\overline{c}_{z_1,z_2}$ be a concept in $G(\overline{C})$ which has two adjacent edges labeled with z_1 and z_2. Let B_{z_1,z_2} be the unique 2-dimensional cube that contains $\overline{c}_{z_1,z_2}$ and the edges labeled with z_1 and z_2. By the choice of $\overline{c}_{z_1,z_2}$, the cube B_{z_1,z_2} contains three concepts from $\overline{C}$ and one concept from C, which we denote c_{z_1,z_2}. Note that

[4]This follows, for example, from Doliwa et al.'s proof showing that RTD and VCD coincide for VCD-maximum classes (Doliwa et al., 2010).

$X - \{z_1, z_2\}$ is a teaching set for c_{z_1,z_2} in C since the other three concepts that are consistent with c_{z_1,z_2} on $X - \{z_1, z_2\}$ belong to the cube B_{z_1,z_2} and thus to $\overline{C}$. Hence $\mathrm{TD}(c, C) \leq m - 2$. On the other hand, $\mathrm{TD}(c, C)$ cannot be strictly smaller than $m-2$ because C is a VCD-maximum class. So, $\mathrm{TD}(c, C) = m-2$.

Note that if $\{z_1, z_2\}$ and $\{y_1, y_2\}$ are two different pairs of adjacent edges in $G(\overline{C})$, then $c_{z_1,z_2} \neq c_{y_1,y_2}$. To see this, suppose $c_{z_1,z_2} = c_{y_1,y_2}$. We consider two cases: $\{z_1, z_2\} \cap \{y_1, y_2\} = \emptyset$ and $|\{z_1, z_2\} \cap \{y_1, y_2\}| = 1$. One can see that in the first case the concepts $\overline{c}_{z_1,z_2}$ and $\overline{c}_{y_1,y_2}$ from $\overline{C}$ are not Hamming-connected, a contradiction to the fact that $\overline{C}$ is shortest-path closed. In the second case, say for $z_2 = y_1$, it is not hard to see that there are two edges in $G(\overline{C})$ labeled with $z_2 = y_1$, which is also impossible.

To conclude the proof, we note that since $G(\overline{C})$ is a tree with $m + 1$ nodes, it has at least $m - 1$ pairs of adjacent edges.[5] □

Proposition 14 cannot be generalized to all RTD-maximum classes.

Proposition 15. *There is an* RTD*-maximum class* $C \subseteq 2^X$ *with* $|X| = m$ *and* $\mathrm{RTD}(C) = m - 2$ *for which the number of concepts* $c \in C$ *with* $\mathrm{TD}(c, C) = m - 2$ *is less than* $m - 1$.

Proof. Consider again the class C_1 in Table 3. It is RTD-maximum, where $m = 4$ and $\mathrm{RTD}(C_1) = 2 = m-2$. However, there is only one concept in this class whose teaching dimension is $m-2$, namely, the concept c_1. □

The latter result also implies that removing all concepts of teaching dimension r from an RTD-maximum class C with $\mathrm{RTD}(C) = r$ does not necessarily yield another RTD-maximum class. Consequently, there is no straightforward nested structure of RTD-maximum classes, as we find for example for VCD-maximum classes with respect to the restriction of the instance set.

5. RTD-maximal classes

In this section we present some properties of RTD-maximal classes. We first show that an RTD-maximal class shatters each subset of the instance space whose size is equal to RTD.

Proposition 16. *Let* $C \subseteq 2^X$ *be* RTD*-maximal with* $\mathrm{RTD}(C) = r$*. Then, for any subset* $X' \subseteq X$ *with* $|X'| = r$*,* C *shatters* X'*.*

Proof. Assume that X' is not shattered by C. Then $|C|_{X'}| < 2^{|X'|}$ and we can add a new concept c_{new} to C such that $c_{new}|_{X'} \notin C|_{X'}$. Thus, $\mathrm{TD}(c_{new}, C \cup \{c_{new}\}) \leq r$. Since $\mathrm{RTD}(C) = r$, C has a teaching plan of order r. So, $C \cup \{c_{new}\}$ also has a teaching plan of order r, which starts with c_{new} and then continues with any teaching plan for C of order r. Therefore, $\mathrm{RTD}(C \cup \{c_{new}\}) \leq r$ and C is not RTD-maximal. □

[5] In fact, the number of pairs of adjacent edges in $G(\overline{C})$ is exactly $m - 1$ if and only if $G(\overline{C})$ is a path.

Note that Proposition 16 is not true in general for VCD-maximal classes. Table 7 contains an example of a VCD-maximal class, found by computer experiments, that does not shatter all subsets of the instance space whose size is equal to the VC-dimension.

$c_i \in C$	x_1	x_2	x_3	x_4	x_5
c_1	0	1	0	1	1
c_2	0	1	1	0	1
c_3	0	1	1	1	0
c_4	1	0	0	0	0
c_5	1	0	0	0	1
c_6	1	0	0	1	0
c_7	1	0	1	0	0
c_8	1	1	0	0	0
c_9	1	1	0	0	1
c_{10}	1	1	0	1	0
c_{11}	1	1	0	1	1
c_{12}	1	1	1	0	0
c_{13}	1	1	1	0	1
c_{14}	1	1	1	1	0

Table 7: VCD-maximal class of VCD 2 that does not shatter the subset $\{x_1, x_2\}$ (found by computer experiments).

As a corollary of Proposition 16 we can show that, for any RTD-maximal class, the minimal teaching dimension and the recursive teaching dimension coincide.

Corollary 2. *For any* RTD*-maximal class* $C \subseteq 2^X$, $\mathrm{TD}_{min}(C) = \mathrm{RTD}(C)$.

Proof. First, note that $\mathrm{TD}_{min}(C) \leq \mathrm{RTD}(C)$. Now assume $\mathrm{TD}_{min}(C) < \mathrm{RTD}(C)$. In this case, there is a concept $c \in C$ for which $\{x_{i_1}, \ldots, x_{i_k}\}$ is a teaching set, for some $k < \mathrm{RTD}(C)$. Consider any subset $X' \subseteq X$ such that $|X'| = \mathrm{RTD}(C)$ and $\{x_{i_1}, \ldots, x_{i_k}\} \subset X'$. Then C does not shatter X', since otherwise there would exist at least one more concept $c' \in C$ with $c'|_{\{x_{i_1},\ldots,x_{i_k}\}} = c|_{\{x_{i_1},\ldots,x_{i_k}\}}$. This is impossible because $\{x_{i_1}, \ldots, x_{i_k}\}$ is a teaching set for c in C. Hence, by Proposition 16, C cannot be RTD-maximal. This is a contradiction. □

It is not hard to see that VCD-maximal classes of VC-dimension 1 are VCD-maximum (Welzl and Woeginger, 1987). We now show that the same holds for RTD-maximal classes.

Proposition 17. *Let* $C \subseteq 2^X$ *be* RTD*-maximal. If* $\mathrm{RTD}(C) = 1$, *then* C *is* RTD*-maximum.*

Proof. For $|X| = 1$ there is only one RTD-maximal class with two concepts which is clearly RTD-maximum. Suppose that the proposition holds when $|X| = m$. Now we consider the case that $|X| = m + 1$ and C is an RTD-maximal class on X with $\mathrm{RTD}(C) = 1$. Since $\mathrm{RTD}(C) = 1$, there is a concept $c \in C$ such that

$\mathrm{TD}(c, C) = 1$. Let (x, ℓ) be a teaching set for c. Then, for any $c' \in C \setminus \{c\}$, $(x, \ell) \notin c'$ or equivalently, $(x, \bar{\ell}) \in c'$, which implies that $|C \setminus \{c\}| = |(C \setminus \{c\}) - x|$. Clearly, $(C \setminus \{c\}) - x$ is RTD-maximal, otherwise C would not be RTD-maximal. So, by the induction hypothesis, $|(C \setminus \{c\}) - x| = \Phi_1(m)$. Therefore, $|C| = \Phi_1(m) + 1 = \Phi_1(m+1)$ and C is RTD-maximum. □

Surprisingly, not all RTD-maximal classes are RTD-maximum.

Proposition 18. *There is an* RTD*-maximal class that is not* RTD*-maximum.*

Proof. Consider the RTD-maximal class C in Table 8. Since $\mathrm{RTD}(C) = 2$ and $|C| = 13 < \Phi_2(5)$, C is not RTD-maximum. This class was found by exhaustive enumeration of all concept classes over instance spaces of size at most 5. □

$c_i \in C$	x_1	x_2	x_3	x_4	x_5
c_1	$\underline{0}$	$\underline{0}$	1	1	1
c_2	$\underline{1}$	$\underline{1}$	1	1	1
c_3	0	1	0	$\underline{1}$	$\underline{1}$
c_4	$\underline{0}$	1	$\underline{0}$	1	0
c_5	0	$\underline{1}$	1	0	$\underline{1}$
c_6	$\underline{0}$	1	1	$\underline{0}$	0
c_7	$\underline{0}$	1	1	1	0
c_8	1	0	$\underline{0}$	0	$\underline{1}$
c_9	1	0	1	0	$\underline{1}$
c_{10}	1	0	$\underline{0}$	$\underline{0}$	0
c_{11}	1	0	$\underline{0}$	1	0
c_{12}	1	0	1	$\underline{0}$	0
c_{13}	1	0	1	1	0

Table 8: C is RTD-maximal but not RTD-maximum (the class was found by computer experiments). Recursive teaching sets are underlined.

6. Shortest-path closedness of VCD-maximum classes and of special VCD-maximal classes

In this section, we give an example to show how the algebraic technique that is applied to obtain our main result can also yield more elegant and insightful proofs for already known results. Kuzmin and Warmuth (2007) showed that VCD-maximum classes are shortest-path closed, but algebraic methods provide an interesting alternative proof.

We will further prove a new result, showing that even some of the smallest possible VCD-maximal classes can be shortest-path closed.

For $Z \subseteq X = \{x_1, \ldots, x_m\}$ and $t \leq m$, let $P_m^t(Z)$ be the collection of monomials over $\mathbf{F}_2$ of the form $x_{i_1} \cdots x_{i_k}$ such that $0 \leq k \leq t$, $1 \leq i_1 < \cdots < i_k \leq m$ and $\{x_{i_1}, \ldots, x_{i_k}\} \subseteq Z$.

Lemma 1. *Let $|X| = m$, $C \subseteq 2^X$, and $\mathrm{VCD}(C) = d$. A set of instances $Z \subseteq X$ is a teaching set for $c \in C$ if and only if c is in the span of $P_m^d(Z)$.*

Proof. The main idea of this proof is from Smolensky (1997). Suppose that $Z \subseteq X$ is a teaching set for $c \in C$. Then, by Theorem 1, $c = f$ for some polynomial f over $\mathbf{F}_2$ whose variables are in the set Z. Each such polynomial is equal to a linear combination of monomials from $P_m^t(Z)$, where $t = |Z|$. For instance, $(x_1 + 1)(x_2 + 1)x_3 = x_1x_2x_3 + x_1x_3 + x_2x_3 + x_3$, etc.

We show that, for every $t \leq m$ and $Z \subseteq X$, the monomials from $P_m^t(Z)$ are in the span of $P_m^d(Z)$. This in turn implies that f is in the span of $P_m^d(Z)$.

Like Smolensky (1997), we use induction on t. If $t \leq d$, there is nothing to prove. Suppose that $t > d$ and every monomial from $P_m^{t-1}(Z)$ is in the span of $P_m^d(Z)$. Consider a monomial $x_{i_1} \cdots x_{i_t}$ from $P_m^t(Z)$. Since $t > d$, the set $\{x_{i_1}, \ldots, x_{i_t}\}$ is not shattered by C. Let $(a_1, \ldots, a_t)$ be a concept that is not in $C|_{\{x_{i_1},\ldots,x_{i_t}\}}$ and consider a polynomial $p(x_{i_1}, \ldots, x_{i_t}) = (x_{i_1} + a_1 + 1)(x_{i_2} + a_2 + 1) \cdots (x_{i_t} + a_t + 1)$.

As a vector in $\mathbf{F}_2^{|C|}$, p has zero coordinates because $p(c(x_{i_1}), \ldots, c(x_{i_t})) = 0$ for all $c \in C$ as at least one of the factors of p will be zero. Hence $p = \mathbf{0}$ and $x_{i_1} \cdots x_{i_t}$ can be expressed as a linear combination of monomials of smaller degree with coefficients from $\{x_{i_1}, \ldots, x_{i_t}\} \subseteq Z$, that is, the ones from $P_m^{t-1}(Z)$. To see this, consider, e.g., $(x_1 + 1)(x_2 + 1)x_3 = \mathbf{0}$; then we have $x_1x_2x_3 = x_1x_3 + x_2x_3 + x_3$. By the induction hypothesis, $P_m^{t-1}(Z)$ is in the span of $P_m^d(Z)$, and hence $x_{i_1} \cdots x_{i_t}$ is in the span of $P_m^d(Z)$. So $P_m^t(Z)$ is in the span of $P_m^d(Z)$.

The implication in the other direction follows from Theorem 1. □

With the help of this lemma, we now provide an alternative proof for the shortest-path closedness of VCD-maximum classes.

Theorem 3. (Kuzmin and Warmuth, 2007) *If C is a* VCD*-maximum class, then C is shortest-path closed.*

Proof. Let $C \subseteq 2^X$ be a VCD-maximum class with $|X| = m$ and $\mathrm{VCD}(C) = d$, and let $I(c)$ denote the set $\{x \in X \mid \text{there exists a } c' \in C \text{ such that } c \triangle c' = \{x\}\}$. We first show that, for every $c \in C$, $I(c)$ is a teaching set for c. By Theorem 1, the monomials from $P_m^d(X)$ span the vector space $\mathbf{F}_2^{|C|}$. Since $|P_m^d(X)| = \Phi_d(m) = |C|$, the set $P_m^d(X)$ is a basis for $\mathbf{F}_2^{|C|}$.

Let $c \in C$ and let $S \subseteq X$ be a minimal teaching set for c in the sense that no proper subset of S is a teaching set for c. Suppose $I(c) \neq S$ and let $x \in S \setminus I(c)$. By Lemma 1, there is a linear combination f_1 of monomials from $P_m^d(S)$ such that $c = f_1$. Note that $X \setminus \{x\}$ is also a teaching set for c, since otherwise $x \in I(c)$. Thus, there is a linear combination f_2 of monomials from $P_m^d(X \setminus \{x\})$ with $c = f_2$. Since $P_m^d(X)$ is a basis for $\mathbf{F}_2^{|C|}$, we have $f_1 = f_2$. As f_2 does not depend on x, f_1 does not depend on x either. Thus f_1

depends only on variables from $S \setminus \{x\}$. By Lemma 1, $S \setminus \{x\}$ is a teaching set for c, which contradicts the minimality of S. Therefore $S = I(c)$, and thus $I(c)$ is a teaching set for c.

Finally, we prove that any two concepts c_1 and c_2 in C are Hamming-connected, by induction on $|c_1 \triangle c_2|$. For $|c_1 \triangle c_2| = 1$ the proof is obvious. Suppose $|c_1 \triangle c_2| = n$ and any two concepts c, c' with $|c \triangle c'| < n$ are Hamming-connected. Since $I(c_1)$ is a teaching set for c_1, it cannot be disjoint from $c_1 \triangle c_2$. Hence there is an $x \in I(c_1) \cap (c_1 \triangle c_2)$. Let c' be the concept from C such that $c_1 \triangle c' = \{x\}$. Then $|c' \triangle c_2| = n - 1$ and by the induction hypothesis c' and c_2 are Hamming-connected. Therefore, c_1 and c_2 are Hamming-connected. □

Theorem 3 states that VCD-maximal classes of the *largest* possible size are shortest-path closed. Note that, in general, VCD-maximal classes need not be shortest-path closed. In the remainder of this section, we prove that some of the VCD-maximal classes of the *smallest* possible size also can be shortest-path closed.

First, we establish a non-trivial lower bound on the size of VCD-maximal classes.

Lemma 2. *Let $C \subseteq 2^X$ be a* VCD*-maximal class over a set X with $|X| = m$. If* $\mathrm{VCD}(C) = d$*, then*

$$|C| \geq 2^m - 2^{m-d-1}\binom{m}{d+1}.$$

Equivalently, if $\mathrm{VCD}(C) = m - d - 1$*, then*

$$|C| \geq 2^m - 2^d\binom{m}{d}.$$

Proof. We prove the second inequality. Suppose $\mathrm{VCD}(C) = m - d - 1$ and $|C| < 2^m - 2^d\binom{m}{d}$. In this case, we have that $|\overline{C}| > 2^d\binom{m}{d}$. By Proposition 1, $\overline{C}$ must contain at least one d-cube for each subset of d instances. Consider a union of d-cubes from $\overline{C}$ taking exactly one cube for each subset of instances of size d. Then the size of this union will be at most $2^d\binom{m}{d}$. Therefore, $\overline{C}$ must contain at least one concept c that does not belong to the above union of d-cubes. Hence, due to Proposition 1, we can add this concept c to the class C without increasing its VC-dimension, which contradicts the fact that C is VCD-maximal. □

We will use the next lemma to show that the lower bound in Lemma 2 can be met by a shortest-path closed class, when *both* VCD and $|X|$ are large.

Lemma 3. *For every d and ℓ, there is some m and a concept class $C \subseteq 2^X$ with $|X| = m$ that satisfies the following properties:*

(i) *$|C^S| = 1$ for every subset $S \subseteq X$ of size d, i.e., C contains exactly one d-cube for any subset of d instances.*

(ii) *Each concept from C belongs to exactly one d-cube.*

(iii) *For any $c, c' \in C$, if c and c' belong to different d-cubes, then $|c \triangle c'| \geq \ell$.*

(iv) *The complement of C is a* VCD*-maximal class of dimension $m - d - 1$.*

Proof. Let $k = \max\{\ell, 2\}$ and $X = \{1, \ldots, (2d+k)t\}$, where t will be chosen later. Split the instance space X into disjoint blocks of size $2d+k$ and let $\{c_1, \ldots, c_N\}$ be the concepts that are equal to all possible unions of such blocks. Note that $N = 2^t$ and $|c_i \triangle c_j| \geq 2d + k$ for $i \neq j$. To each subset $S \subseteq X$ of size d, we assign a concept c_S from the above list such that $c_S \neq c_{S'}$ for $S \neq S'$. This can be done when t is chosen large enough so that $N = 2^t$ is greater than $\binom{(2d+k)t}{d}$, the number of subsets of size d of the set X.

For each $S \subseteq X$ of size d, define a d-cube $C(S)$ based on c_S, that is, $C(S) = 2^S \times \{c_S|_{X \setminus S}\}$. Let the class C be defined as

$$C = \bigcup_{S \subseteq X:\ |S|=d} C(S).$$

Note that for $S \neq S'$ and for any $c \in C(S)$ and $c' \in C(S')$, we have

$$|c \triangle c'| \geq |c_S \triangle c_{S'}| - (|S| + |S'|) \geq 2d + k - 2d = k \geq \ell. \qquad (*)$$

This proves part (iii).

To show (i), suppose $|C^S| > 1$ and let C' be another d-cube on the instances from S. Each concept from C' must belong to some $C(S')$ for $S' \neq S$. Since C' is a cube, there are two concepts $c_1, c_2 \in C'$ such that $|c_1 \triangle c_2| = 1$ and $c_1 \in C(S_1)$, $c_2 \in C(S_2)$ for some $S_1 \neq S_2$. But we showed in (*) that $|c_1 \triangle c_2| \geq k \geq 2$. This contradiction proves part (i).

So, C contains only the cubes of the form $C(S)$ for $S \subseteq X$. Again, it follows from (*) that these cubes are disjoint, which proves part (ii).

Consider $\overline{C}$, the complement of C. By Proposition 1, we have $\mathrm{VCD}(\overline{C}) \leq m - d - 1$. On the other hand, if $\mathrm{VCD}(\overline{C}) < m - d - 1$ then, again by Proposition 1, C must contain a $(d+1)$-cube, and hence two d-cubes, which contradicts part (i). Thus $\mathrm{VCD}(\overline{C}) = m - d - 1$. To show that $\overline{C}$ is VCD-maximal, consider any concept $c \notin \overline{C}$. We have $c \in C$, and removing c from C will destroy the d-cube to which c belongs to. Hence by Proposition 1, $\mathrm{VCD}(\overline{C} \cup \{c\}) > m - d - 1$. This completes the proof of part (iv). □

Theorem 4. *There are* VCD*-maximal classes of the smallest possible size that are shortest-path closed. Namely, for any d there is some $m > d$ and a shortest-path closed concept class $C \subseteq 2^X$ such that*

- $|X| = m$ *and* $\mathrm{VCD}(C) = m - d - 1$,
- C *is* VCD*-maximal, and*
- *there is no* VCD*-maximal concept class $C' \subseteq 2^X$ with* $\mathrm{VCD}(C') = m - d - 1$ *and* $|C'| < |C|$.

The following lemma, proven by Doliwa et al. (2013), presents a sufficient condition for shortest-path closedness of a concept class, formulated in terms of its one-inclusion graph.

Lemma 4. (Doliwa et al., 2013) *Let* $C \subseteq 2^X$ *be a concept class such that for every* $c \in C$*,* $\deg_C(c) \geq |X|-1$*. Then* C *is shortest-path closed.*

Proof of Theorem 4. Take any d and let C_0 be the class constructed in Lemma 3 for this choice of d and $\ell = 3$. Let C be the complement of C_0. It follows from Lemma 3 part (iv) that C is a VCD-maximal class with $\mathrm{VCD}(C) = m - d - 1$. Since C_0 consists of disjoint d-cubes, and there are 2^d many of them, we have $|C_0| = 2^d\binom{m}{d}$. Therefore, $|C| = 2^m - 2^d\binom{m}{d}$. By Lemma 2, there is no VCD-maximal concept class $C' \subseteq 2^X$ with $\mathrm{VCD}(C') = m - d - 1$ and $|C'| < |C|$.

We now prove that C is shortest-path closed. By Lemma 4, it suffices to show that $\deg_C(c) \geq |X| - 1$ for any $c \in C$. Suppose there is some $c \in C$ such that $\deg_C(c) \leq |X| - 2$. Hence there are c_1, c_2 with $c_1 \neq c_2$ such that $|c \triangle c_1| = |c \triangle c_2| = 1$ and $c_1, c_2 \notin C$, that is, $c_1, c_2 \in C_0$. Since $|c_1 \triangle c_2| = 2 < \ell$, Lemma 3 part (iii) implies that c_1 and c_2 cannot belong to two different d-cubes in C_0. So, they are in the same d-cube. But then c must be in the same d-cube as well, which is impossible because $c \in C = \overline{C_0}$. □

7. Conclusions

Our analog of Sauer's bound for RTD establishes a new connection between teaching complexity and VC-dimension. Another main contribution besides obtaining this result is the successful application of algebraic proof techniques. Such techniques have occasionally been used to obtain results relevant to computational learning theory (see, e.g., the article by Smolensky (1997)), but our results suggest that the research community might benefit from further exploring their applicability. For example, we presented an elegant alternative proof of the fact that VCD-maximum classes are shortest-path closed. The algebraic characterization of teaching sets provided in Theorem 1 is of potential use for future studies not only in the context of the combinatorial problems we deal with in this paper.

Our results on RTD-maximum and RTD-maximal classes provide deep insights into structural properties that affect the complexity of teaching a concept class. In particular, we investigated the characteristics of teaching plans for RTD-maximum classes, which help to better understand such properties. Our study of the structure of teaching plans leaves a number of open questions. For example, Doliwa et al. (2010) proved that every VCD-maximum class possesses a repetition-free teaching plan, but it remains open whether the same is true for RTD-maximum classes in general. It would also be interesting to find out whether every class with a repetition-free teaching plan can be extended to an RTD-maximum class without increasing the RTD. If the answer is 'yes', this would imply that any RTD-maximal class with a repetition-free teaching plan must already be RTD-maximum. If the answer is 'no', this would immediately lead to the question under which conditions on the repetition-free teaching plan of a class C one could conclude that C is contained in an RTD-maximum class of the same RTD.

The comparative study of RTD-maximum classes and VCD-maximum classes points out many essential differences between these two notions. An open question remaining from this study concerns the gap between Proposition 5 and Proposition 6. By Proposition 6, the complement of any RTD-maximum class C with $\text{RTD}(C) = 1$ is still maximum, but by Proposition 5 this cannot be generalized to arbitrary values of $\text{RTD}(C)$. The counterexample we gave in the proof of Proposition 5 is a class C with $\text{RTD}(C) = 3$, but we could not find a counterexample of a class whose RTD equals 2. Interestingly, an exhaustive enumeration of all concept classes over instance spaces of size at most 5 showed that the class given in the proof of Proposition 5 is the *only* RTD-maximum class over at most 5 instances whose complement is not RTD-maximum.

So far, we do not know much about "proper" RTD-maximal classes, i.e., classes that are RTD-maximal but not RTD-maximum. Our exhaustive enumeration of all concept classes over at most 5 instances suggests that they occur even less frequently than proper VCD-maximal classes. For $|X| = 4$, there exist 402 distinct concept classes,[6] 2 of them being proper VCD-maximal classes and none of them being proper RTD-maximal classes. For $|X| = 5$, there are 1,228,158 distinct concept classes, 55 of them being proper VCD-maximal, and only 17 being proper RTD-maximal. We could not find any proper RTD-maximal classes that are shortest-path closed.

By contrast, RTD-maximum classes occur much more frequently. For $|X| = 4$, we found 25 such classes; 9 of them are VCD-maximum. For $|X| = 5$, the contrast in the numbers is large: 14,204 classes are RTD-maximum, but only 56 of them are VCD-maximum. Most RTD-maximum classes are not shortest-path closed: over 4 instances, there is only 1 shortest-path closed RTD-maximum class that is not VCD-maximum (see Table 6); over 5 instances, the corresponding number is 19. From a theoretical point of view, it would be interesting to find structural properties that characterize those RTD-maximum classes that are not VCD-maximum.

As a byproduct of our theoretical studies, we proved several new results on VCD-maximal classes, presented in the Appendix. Altogether, our results might be helpful in solving the long-standing sample compression conjecture (Floyd and Warmuth, 1995) and in establishing further connections between learning from a teacher and learning from randomly chosen examples. In particular, we hope that methods from algebra will turn out to be of further use in these contexts.

Acknowledgements

We would like to thank the anonymous reviewers for their thoughtful comments, including many suggestions that improved the presentation of this paper. In particular, we would like to express our gratitude

[6] In the tables in this paper, concept classes are represented as matrices in which the columns correspond to instances and the rows correspond to concepts. We consider two such representations equivalent if one can be obtained from the other by permuting columns, permuting rows, and swapping zeros and ones within one or more columns. Two concept classes are distinct if they don't have equivalent matrix representations.

towards one reviewer for providing us with Table 2.

Appendix: VCD-maximal classes

This appendix contains some new interesting properties of VCD-maximal classes. For instance, the next theorem provides a way of constructing an infinite series of equal-sized VCD-maximal classes starting from a given VCD-maximal class.

Theorem 5. *Let C be a class of VC-dimension d on a set of m instances $X = \{x_1, \ldots, x_m\}$.*

(1) If C is a VCD*-maximal class and for some instance $x \in X$ we have $|C - x| = |C|$, then $C + x$ is also* VCD*-maximal, where*

$$C + x = \{c \in 2^{X \cup \{x_{m+1}\}} : c \cap X \in C \text{ and } c(x_{m+1}) = c(x)\}.$$

This process can be continued to obtain a series of VCD*-maximal classes $C + x$, $(C + x) + x$, $((C + x) + x) + x$, etc.*

(2) On the other hand, if $|C - x| < |C|$, then $C + x$ is not a VCD*-maximal class.*

Proof. (1) Note that $\text{VCD}(C) = \text{VCD}((C + x) - x)$ and $\text{VCD}(C) = \text{VCD}(C + x)$. These equalities follow from the fact that C is equivalent to $(C + x) - x$, and that if $C + x$ shatters a set S, then S cannot contain both x and x_{m+1}.

Suppose C is VCD-maximal and $|C - x| = |C|$ for some $x \in X$. Consider any $c \in 2^{X \cup \{x_{m+1}\}}$ such that $c \notin C + x$ and let $c - x_{m+1} = c \cap X$. We need to show that $\text{VCD}(C + x \cup \{c\}) > \text{VCD}(C + x)$. First, suppose $c - x_{m+1} \notin C$. Then, since C is VCD-maximal, $\text{VCD}(C + x \cup \{c\}) \geq \text{VCD}(C \cup \{c - x_{m+1}\}) > \text{VCD}(C) = \text{VCD}(C + x)$.

Now suppose $c - x_{m+1} \in C$. In this case $c(x) \neq c(x_{m+1})$ since otherwise $c \in C + x$. Also note that the concept $c - x = c \cap (X \cup \{x_{m+1}\} - x)$ does not belong to $(C + x) - x$. Indeed, suppose $c - x \in (C + x) - x$ and let $c' \in C$ be the image of $c - x$ under the equivalence transformation from $(C + x) - x$ to C. We then have that C contains two concepts, namely $c - x_{m+1}$ and c', that differ only on x since $(c - x_{m+1})(x) = c(x) \neq c(x_{m+1}) = (c - x)(x_{m+1}) = c'(x)$. This contradicts the assumption that $|C - x| = |C|$. Therefore, $c - x \notin (C + x) - x$ and we have that $\text{VCD}(C + x \cup \{c\}) \geq \text{VCD}((C + x) - x \cup \{c - x\}) > \text{VCD}((C + x) - x) = \text{VCD}(C) = \text{VCD}(C + x)$. Hence $C + x$ is a VCD-maximal class.

(2) If $|C - x| < |C|$ then there are two concepts c_1 and c_2 in C that differ only in x. Consider a concept $c \notin C + x$ defined as $c = c_1 \cup \{(x_{m+1}, \ell)\}$ where ℓ is chosen so that $c(x) \neq c(x_{m+1})$. Since c coincides with c_1 on X, we have $(C + x \cup \{c\}) - x_{m+1} = C$. Furthermore, c coincides with the extension of c_2 in $C + x$ on the instances from $(X \cup \{x_{m+1}\}) - x$. Hence $(C + x \cup \{c\}) - x = (C + x) - x$, which is, of course, equivalent to C.

Let $\mathrm{VCD}(C+x) = d$ and suppose that $C + x \cup \{c\}$ shatters a set S of size $d+1$. Note that S cannot contain both x and x_{m+1} since the restriction of $C + x \cup \{c\}$ to these two instances can contain only one of the two concepts $(0,1)$ and $(1,0)$. If S does not contain x_{m+1}, then we have $\mathrm{VCD}(C+x) = \mathrm{VCD}(C) = \mathrm{VCD}((C + x \cup \{c\}) - x_{m+1}) \geq d+1$. On the other hand, if S does not contain x, we have $\mathrm{VCD}(C+x) = \mathrm{VCD}((C+x)-x) = \mathrm{VCD}((C + x \cup \{c\}) - x) \geq d+1$. These contradictions show that in fact $\mathrm{VCD}(C + x \cup \{c\}) = \mathrm{VCD}(C+x)$, and hence $C+x$ is not a VCD-maximal class. □

As a corollary of Lemma 2 from Section 6, we obtain that for a VCD-maximal class C with $\mathrm{VCD}(C) = |X| - O(1)$, the sum $\mathrm{VCD}(C) + \mathrm{VCD}(\overline{C})$ is bounded by $|X| + O(\log_2 |X|)$.

Theorem 6. *Let $|X| = m$. If $C \subseteq 2^X$ is a* VCD*-maximal class and* $\mathrm{VCD}(C) = m - d$*, then*

$$\mathrm{VCD}(C) + \mathrm{VCD}(\overline{C}) \leq m - 1 + (d-1)\log_2 m.$$

Proof. Since C is VCD-maximal, we have, by Lemma 2, that $|C| \geq 2^m - 2^{d-1}\binom{m}{d-1}$. Therefore, $|\overline{C}| \leq 2^{d-1}\binom{m}{d-1}$ and hence $\mathrm{VCD}(\overline{C}) \leq \log_2 |\overline{C}| \leq d - 1 + \log_2 \binom{m}{d-1}$. Taking into account that $\binom{m}{d-1} \leq m^{d-1}$, we obtain $\mathrm{VCD}(\overline{C}) \leq d - 1 + (d-1)\log_2 m$. Since $\mathrm{VCD}(C) = m - d$, it follows that $\mathrm{VCD}(C) + \mathrm{VCD}(\overline{C}) \leq m - 1 + (d-1)\log_2 m$. □

Another property of VCD-maximal classes is that they are indecomposable in the sense that they cannot be formed by a direct product of non-trivial smaller classes.

Theorem 7. *Let $C_0 \subseteq 2^{X_0}$ and $C_1 \subseteq 2^{X_1}$ be nonempty concept classes with*

(a) $\mathrm{VCD}(C_0) > 0$ *or* $\mathrm{VCD}(C_1) > 0$ *and*

(b) $C_0 \times C_1 \neq 2^{X_0 \cup X_1}$.

Then $C_0 \times C_1$ is not a VCD*-maximal class.*

We will need to prove the following lemma first.

Lemma 5. *Let $C_0 \subseteq 2^{X_0}$ and $C_1 \subseteq 2^{X_1}$ be nonempty concept classes and let $c_0 \in 2^{X_0}$ and $c_1 \in 2^{X_1}$ be any two concepts with the property that for each $i \in \{0,1\}$, if* $\mathrm{VCD}(C_i) = 0$ *then* $\mathrm{VCD}(C_{1-i} \cup \{c_{1-i}\}) = \mathrm{VCD}(C_{1-i})$*. Then* $\mathrm{VCD}((C_0 \times C_1) \cup \{c_0c_1\}) = \mathrm{VCD}(C_0 \times C_1) = \mathrm{VCD}(C_0) + \mathrm{VCD}(C_1)$*.*

Proof. Let $d_i = \mathrm{VCD}(C_i)$, for $i \in \{0,1\}$, and suppose that $(C_0 \times C_1) \cup \{c_0c_1\}$ shatters a set $S \subseteq X_0 \cup X_1$ of size $d_0 + d_1 + 1$. Let $S_i = S \cap X_i$ and assume, without loss of generality, that $|S_0| = d_0 + 1$ and $|S_1| = d_1$. Therefore, $\mathrm{VCD}(C_0 \cup \{c_0\}) = d_0 + 1 > \mathrm{VCD}(C_0)$, and by the assumption we have that $d_1 > 0$. So, on the one hand, we have that $(C_0 \times C_1) \cup \{c_0c_1\}$ must contain at least $2^{d_1} > 1$ concepts that extend $c_0|_{S_0}$. But, on the other hand, $(C_0 \times C_1) \cup \{c_0c_1\}$ contains only one such concept, namely c_0c_1, since $c_0|_{S_0} \notin C_0|_{S_0}$. This contradiction proves the lemma. □

Proof of Theorem 7. If $\mathrm{VCD}(C_0) > 0$ and $\mathrm{VCD}(C_1) > 0$, then by Lemma 5 for any concept $c \notin C_0 \times C_1$ (which exists by our assumption), we have that $\mathrm{VCD}((C_0 \times C_1) \cup \{c\}) = \mathrm{VCD}(C_0 \times C_1)$. Hence $C_0 \times C_1$ is not VCD-maximal.

Consider the case $\mathrm{VCD}(C_0) = 0$ and $\mathrm{VCD}(C_1) > 0$ (the other case is similar). Let $c_0 \notin C_0$ and $c_1 \in 2^{X_1}$ be such that $\mathrm{VCD}(C_1 \cup \{c_1\}) = \mathrm{VCD}(C_1)$ (e.g., any $c_1 \in C_1$). By Lemma 5, we have that $\mathrm{VCD}((C_0 \times C_1) \cup \{c_0 c_1\}) = \mathrm{VCD}(C_0 \times C_1)$. Since $c_0 c_1 \notin C_0 \times C_1$, this proves that the class $C_0 \times C_1$ is not VCD-maximal. □

Doliwa, T., Fan, G., Simon, H., Zilles, S., 2013. Recursive teaching dimension, VC-dimension, and sample compression. Unpublished manuscript.

Doliwa, T., Simon, H.-U., Zilles, S., 2010. Recursive teaching dimension, learning complexity, and maximum classes. In: Proceedings of the 21st International Conference on Algorithmic Learning Theory (ALT). Vol. 6331 of Lecture Notes in Artificial Intelligence. Springer, pp. 209–223.

Dudley, R., 1985. The structure of some Vapnik-Chervonenkis classes. In: Proceedings of the Berkeley Conference in Honor of J. Neyman and J. Kiefer. Wadsworth, pp. 495–508.

Floyd, S., Warmuth, M. K., 1995. Sample compression, learnability, and the Vapnik-Chervonenkis dimension. Machine Learning 21 (3), 269–304.

Goldman, S. A., Kearns, M. J., 1995. On the complexity of teaching. Journal of Computer and System Sciences 50, 20–31.

Kuzmin, D., Warmuth, M. K., 2007. Unlabeled compression schemes for maximum classes. Journal of Machine Learning Research 8, 2047–2081.

Littlestone, N., Warmuth, M. K., June 1986. Relating data compression and learnability. Tech. rep., Department of Computer and Information Sciences, Santa Cruz, CA.

Rubinstein, B. I. P., Bartlett, P. L., Rubinstein, J. H., 2009. Shifting: One-inclusion mistake bounds and sample compression. Journal of Computer and System Sciences 75 (1), 37–59.

Samei, R., Semukhin, P., Yang, B., Zilles, S., 2012. Sauer's bound for a notion of teaching complexity. In: Proceedings of the 23rd International Conference on Algorithmic Learning Theory (ALT). Vol. 7568 of Lecture Notes in Computer Science. Springer, pp. 96–110.

Sauer, N., 1972. On the density of families of sets. Journal of Combinatorial Theory, Series A 13 (1), 145–147.

Shelah, S., 1972. A combinatorial problem: Stability and order for models and theories in infinitary languages. Pacific Journal of Mathematics 4, 247–261.

Shinohara, A., Miyano, S., 1991. Teachability in computational learning. New Generation Computing 8 (4), 337–347.

Smolensky, R., 1997. Well-known bound for the VC-dimension made easy. Computational Complexity 6 (4), 299–300.

Valiant, L. G., 1984. A theory of the learnable. Communications of the ACM 27 (11), 1134–1142.

Vapnik, V. N., Chervonenkis, A. Y., 1971. On the uniform convergence of relative frequencies of events to their probabilities. Theory of Probability and its Applications 16, 264–280.

Welzl, E., 1987. Complete range spaces. Unpublished notes.

Welzl, E., Woeginger, G., 1987. On Vapnik-Chervonenkis dimension one. Unpublished manuscript.

Zilles, S., Lange, S., Holte, R., Zinkevich, M., 2011. Models of cooperative teaching and learning. Journal of Machine Learning Research 12, 349–384.