# Sauer's Bound for a Notion of Teaching Complexity

Rahim Samei, Pavel Semukhin, Boting Yang, and Sandra Zilles

Department of Computer Science, University of Regina, Canada
{samei20r,semukhip,boting,zilles}@cs.uregina.ca

**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.

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

## 1 Introduction

An important combinatorial result, proven by Sauer [7] and independently by Shelah [8], states that the size of any concept class of Vapnik-Chervonenkis dimension (VC-dimension, [11]) $d$ 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*. A concept class of VC-dimension $d$ over a finite instance space $X$ is maximum, if its size meets Sauer's bound.[1] 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 [6, 12]. 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

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

schemes of the size of their VC-dimension [3, 5]. Moreover, the *recursive teaching dimension* (RTD, a complexity parameter of the recently introduced recursive teaching model [13]) of any maximum class equals its VC-dimension [2].

Recent work [2] 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 deserves more attention. 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 highlight 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., we give an example for an alternative proof to Kuzmin and Warmuth's result that maximum classes are shortest-path-closed [5]. Previously, methods from algebra yielded an alternative proof of Sauer's bound for the VC-dimension [10].

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 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.

## 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 be also denoted as $C - x$, and $c|_{X\setminus\{x\}}$ will be denoted as $c - x$. The *reduction* of $C$ to $Y$ is defined as $C^Y = \{c \subseteq Y : c \cup c' \in C \text{ for all } c' \subseteq X\setminus Y\}$. In other words, $c \in C^Y$ if and only if all possible extensions of the concept $c$ from $Y$ 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 $\mathrm{VCD}(C) = \max\{|S| : S \text{ is shattered by } C\}$ [11]. Let $\Phi_d(m) = \sum_{i=0}^{d} \binom{m}{i}$. Sauer's lemma states that if $\mathrm{VCD}(C) = d$, then $|C| \le \Phi_d(|X|)$ [7,8]. Let $\mathrm{VCD}(C) = d$; 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 (cf. [12]). 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.

A *labeled example* is a pair $(x,\ell)$, 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 $S$. For simplicity, we then also call $X(S)$ a teaching set since the labels of examples from $S$ are uniquely determined by $X(S)$ and $c$. The collection of all teaching sets for $c$ in $C$ is denoted $\mathrm{TS}(c,C)$.

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 $\mathrm{TD}(C) = \max_{c\in C} \mathrm{TD}(c,C)$ [4,9]. 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 [2,13]. 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

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

For a teaching plan $P = ((c_1,S_1),\ldots,(c_n,S_n))$ of $C$ whose order is equal to $\mathrm{RTD}(C)$, 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. We will also use the notation $\mathrm{RTD}^*(C) = \max_{X'\subseteq X} \mathrm{RTD}(C|_{X'})$.

The RTD has the following properties [2,13]:

- RTD is monotonic, i.e, $\mathrm{RTD}(C') \le \mathrm{RTD}(C)$ 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 $i = 1, \ldots, n$.
– $\mathrm{RTD}(C) = \max_{C' \subseteq C} \mathrm{TD}_{min}(C')$.

## 3 Algebraic characterization of teaching sets

In this section we give an algebraic characterization of the teaching sets for a concept $c$ in a concept class $C$. Let $X = \{x_1, \ldots, x_m\}$ be a finite instance space, and let $C = \{c_1, \ldots, c_n\}$ 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 $f(x_1, \ldots, x_m)$ with variables from $X$ and coefficients from $\mathbf{F}_2$, we define a vector $f = (f_1, \ldots, f_n)$ from $\mathbf{F}_2^n$ as follows

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

Note that we use the same notation for a polynomial and a vector. We also associate each concept $c_i \in C$ with the $i$th standard basis vector $c_i = (0, \ldots, 1, \ldots, 0)$ 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 these notations, let us consider the following concept class:

| | $x_1$ | $x_2$ | $x_3$ |
|---|---|---|---|
| $c_1$ | 0 | 1 | 0 |
| $c_2$ | 1 | 0 | 1 |
| $c_3$ | 0 | 1 | 1 |

In this class, $x_1 = (0, 1, 0)$, $x_2 = (1, 0, 1)$, $x_3 = (0, 1, 1)$, $0 = (0, 0, 0)$ and $1 = (1, 1, 1)$. In our notations, $c_1 = (1, 0, 0)$, $c_2 = (0, 1, 0)$ and $c_3 = (0, 0, 1)$. So we have $x_1 + x_2 = 1$, $x_1x_2 = 0$, $c_1 = x_3 + 1$, $x_2x_3 = (0, 0, 1)$ and hence $c_3 = x_2x_3$.

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 $i$th and $j$th 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 $i$th 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 $i$th and $j$th coordinates. This contradicts the assumption that $c_i = f(z_1, \ldots, z_k)$. □

## 4 RTD-maximum classes

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 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., $(x_1 + 1)(x_2 + 1) = x_1 x_2 + x_1 + x_2 + 1$, 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 the 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 the 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 $P_m^r$. □

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 $r$. 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 $c \notin C$.

RTD-maximum classes have the following properties.

**Proposition 1.** (i) *Every* VCD-*maximum class* $C$ *is also* RTD-*maximum with* $\mathrm{RTD}(C) = \mathrm{VCD}(C)$.

(ii) *There are* RTD-*maximum classes that are not* VCD-*maximum.*

(iii) *There is a class* $C$ *for which both* $C$ *and* $\overline{C}$ *are* RTD-*maximum, but neither* $C$ *nor* $\overline{C}$ *is* VCD-*maximum.*

(iv) *There are* RTD-*maximum classes whose restrictions are 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.* (i) For every VCD-maximum class $C$, $\mathrm{RTD}(C) = \mathrm{VCD}(C)$ [2]. It follows from Theorem 2 and Definition 1 that $C$ is RTD-maximum.

(ii) If an RTD-maximum class $C$ is not VCD-maximum, then $\mathrm{RTD}(C) < \mathrm{VCD}(C)$. Table 1 shows an RTD-maximum class $C_1$ with $\mathrm{RTD}(C_1) = 2$ and $\mathrm{VCD}(C_1) = 3$.

(iii) $C_1$ in Table 1 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.

(iv) $C_2$ in Table 1 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 1. Clearly, $C_1 - x_4$ is RTD-maximum and $\mathrm{RTD}(C_1) = 2 < \mathrm{RTD}(C_1 - x_4) = 3$. □

A consequence of the proof of Theorem 2 is that, for RTD-maximum classes, all instance sets of size $\mathrm{RTD}(C)$ are used as recursive teaching sets.

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

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

| $c_i \in C_2$ | $x_1$ | $x_2$ | $x_3$ | $x_4$ |
|---|---|---|---|---|
| $c_1$ | 0 | 0 | 0 | 0 |
| $c_2$ | 1 | 0 | 0 | 0 |
| $c_3$ | 0 | 1 | 0 | 0 |
| $c_4$ | 0 | 0 | 1 | 0 |
| $c_5$ | 0 | 1 | 1 | 1 |

**Table 1.** $C_1$ and $\overline{C_1}$ are RTD-maximum but neither $C_1$ nor $\overline{C_1}$ is VCD-maximum. $C_2$ is RTD-maximum but $C_2 - x_4$ is not.

**Corollary 1.** *Let $C \subseteq 2^X$ be* RTD-*maximum, $|X| = m$, 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 $|C| = \dim(\mathbf{F}_2^n) \leq \Phi_r(m) - 1$. Hence $C$ is not RTD-maximum. This is a contradiction. □

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

**Corollary 2.** *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 there is a teaching plan for $C$ such 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 for any $c \in \{c_{i+1}, \ldots, c_n\}$, $c(x) = \bar{\ell}$. So, $|\{c_{i+1}, \ldots, c_n\}| = |\{c_{i+1}, \ldots, c_n\}|_{X \setminus \{x\}}|$. Consequently,

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

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

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

**Proposition 2.** *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 2. $\overline{C}$ is not RTD-maximum because $\mathrm{RTD}(\overline{C}) = 2$ and $6 < \Phi_2(5)$. □

Still, the complement of an RTD-maximum class of RTD 1 is RTD-maximum.

**Proposition 3.** *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* $\mathrm{RTD}(\overline{C}) = |X| - 2$.

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

| $c_i \in C$ | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ |
|---|---|---|---|---|---|
| $c_{14}$ | 0 | $\underline{1}$ | 0 | 0 | $\underline{1}$ |
| $c_{15}$ | $\underline{1}$ | 0 | $\underline{1}$ | $\underline{1}$ | 0 |
| $c_{16}$ | $\underline{1}$ | 0 | 0 | $\underline{1}$ | 0 |
| $c_{17}$ | 0 | $\underline{1}$ | $\underline{1}$ | 0 | 0 |
| $c_{18}$ | 0 | $\underline{1}$ | 0 | 0 | 0 |
| $c_{19}$ | 0 | 0 | $\underline{1}$ | $\underline{1}$ | 0 |
| $c_{20}$ | 0 | 0 | 0 | $\underline{1}$ | 0 |
| $c_{21}$ | $\underline{1}$ | 0 | $\underline{1}$ | 0 | 0 |
| $c_{22}$ | $\underline{1}$ | 0 | 0 | 0 | 0 |
| $c_{23}$ | 0 | 0 | $\underline{1}$ | 0 | $\underline{1}$ |
| $c_{24}$ | 0 | 0 | $\underline{1}$ | 0 | 0 |
| $c_{25}$ | 0 | 0 | 0 | 0 | $\underline{1}$ |
| $c_{26}$ | 0 | 0 | 0 | 0 | 0 |

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

**Table 2.** $C$ is RTD-maximum (recursive teaching sets are underlined), but $\overline{C}$ is not.

*Proof.* By induction on $|X|$. For $|X| = 2$ the proof is trivial. Suppose that for $|X| < m$ the statement of the theorem is true. Now consider the case $|X| = m > 2$. Let $c_1 \in C$ with $\mathrm{TD}(c_1, C) = 1$, and w.l.o.g., 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 a maximum class of $\mathrm{RTD}(C_1) = 1$ 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 to a teaching plan for $\overline{C}$ of order $m - 2$. Thus $\mathrm{RTD}(\overline{C}) \le m - 2$. Further, $|\overline{C}| = 2^m - |C| = 2^m - (m + 1) = \Phi_{m-2}(m)$. Hence, by Theorem 2, $\mathrm{RTD}(\overline{C}) = m - 2$, and $\overline{C}$ is RTD-maximum. □

The RTD-maximum class $C$ in the proof of Proposition 2 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. [6]. Sufficiency is easy to see, as was pointed out by an anonymous reviewer of this paper: 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 throughout.

**Proposition 4.** *Let $C \subseteq 2^X$ and $|X| = m$. If $\mathrm{RTD}(C) + \mathrm{RTD}(\overline{C}) = m - 1$, then $C$ is RTD-maximum.*

Recall that $\mathrm{RTD}^*(C) = \max_{X' \subseteq X} \mathrm{RTD}(C|_{X'})$. We obtain the following property.

**Proposition 5.** *Let $C \subseteq 2^X$ and $|X| = m$. If $\mathrm{RTD}^*(C) \leq r$, then $|C| \leq \Phi_r(m)$. The inverse statement is not true in general.*

*Proof.* Since $\mathrm{RTD}^*(C) \leq r$, $\mathrm{RTD}(C) \leq r$ and by Theorem 2, $|C| \leq \Phi_r(m)$. An example[2] for a class $C$ with $|C| \leq \Phi_r(m)$ and $\mathrm{RTD}^*(C) > \mathrm{RTD}(C) = r$ is the class $C = \{\emptyset,\ \{x_2, x_3\},\ \{x_1, x_3\},\ \{x_1, x_2, x_3\}\}$, for which $|C| = 4$, $\mathrm{RTD}(C) = 1$ and $\mathrm{RTD}^*(C) = 2$. □

## 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 6.** *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. □

As a corollary we obtain that for an RTD-maximal class, the minimal and the recursive teaching dimensions coincide.

**Corollary 3.** *For any* RTD*-maximal class $C \subseteq 2^X$,* $\mathrm{TD}_{min}(C) = \mathrm{RTD}(C)$*.*

*Proof.* $\mathrm{TD}_{min}(C) \leq \mathrm{RTD}(C)$ is easy to see. Assume $\mathrm{TD}_{min}(C) < \mathrm{RTD}(C)$. Then, 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 6, $C$ 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. We now show that the same holds for RTD-maximal classes.

**Proposition 7.** *Let $C \subseteq 2^X$ be* RTD*-maximal. If* $\mathrm{RTD}(C) = 1$*, then $C$ is* RTD*-maximum.*

[2] This example also provides a simpler proof of the second part of Proposition 1(iv). The latter in turn implies that the inverse of Proposition 5 is not true in general.

*Proof.* By induction on the size of $X$. For $|X| = 1$ there is only one RTD-maximal class with two concepts which is clearly RTD-maximum. Suppose that the theorem 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 $(x, \ell)$ 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 8.** (Doliwa [1]) *There is an* RTD*-maximal class that is not* RTD*-maximum.*

*Proof.* Consider the RTD-maximal class $C$ in Table 3. Since $\mathrm{RTD}(C) = 3$ and $|C| = 40 < \Phi_3(6)$, $C$ is not RTD-maximum. □

| $c_i$ | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $c_i$ | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $c_i$ | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ | $c_i$ | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ | $x_6$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $c_1$ | 0 | 1 | 0 | 1 | 1 | 0 | $c_{11}$ | 1 | 0 | 1 | 1 | 1 | 1 | $c_{21}$ | 0 | 1 | 0 | 0 | 1 | 0 | $c_{31}$ | 0 | 0 | 0 | 0 | 0 | 0 |
| $c_2$ | 0 | 1 | 1 | 1 | 0 | 1 | $c_{12}$ | 0 | 0 | 1 | 0 | 0 | 0 | $c_{22}$ | 1 | 1 | 0 | 1 | 1 | 0 | $c_{32}$ | 1 | 1 | 0 | 1 | 0 | 1 |
| $c_3$ | 1 | 0 | 0 | 0 | 0 | 0 | $c_{13}$ | 1 | 1 | 1 | 0 | 0 | 1 | $c_{23}$ | 1 | 0 | 0 | 0 | 1 | 0 | $c_{33}$ | 0 | 0 | 0 | 1 | 0 | 0 |
| $c_4$ | 1 | 0 | 0 | 1 | 1 | 1 | $c_{14}$ | 0 | 1 | 1 | 0 | 1 | 0 | $c_{24}$ | 1 | 1 | 0 | 1 | 1 | 1 | $c_{34}$ | 0 | 0 | 0 | 0 | 1 | 0 |
| $c_5$ | 0 | 0 | 1 | 1 | 0 | 0 | $c_{15}$ | 1 | 0 | 1 | 0 | 1 | 1 | $c_{25}$ | 1 | 1 | 0 | 0 | 1 | 1 | $c_{35}$ | 1 | 1 | 0 | 0 | 0 | 0 |
| $c_6$ | 1 | 0 | 0 | 1 | 1 | 0 | $c_{16}$ | 0 | 0 | 1 | 1 | 0 | 1 | $c_{26}$ | 0 | 1 | 0 | 0 | 0 | 0 | $c_{36}$ | 1 | 0 | 1 | 0 | 1 | 0 |
| $c_7$ | 0 | 0 | 1 | 0 | 1 | 1 | $c_{17}$ | 1 | 1 | 1 | 1 | 0 | 0 | $c_{27}$ | 1 | 0 | 0 | 0 | 0 | 1 | $c_{37}$ | 0 | 1 | 0 | 0 | 0 | 1 |
| $c_8$ | 1 | 1 | 1 | 0 | 1 | 0 | $c_{18}$ | 1 | 1 | 1 | 0 | 1 | 1 | $c_{28}$ | 0 | 1 | 0 | 1 | 0 | 1 | $c_{38}$ | 1 | 1 | 1 | 1 | 1 | 0 |
| $c_9$ | 0 | 1 | 1 | 0 | 0 | 1 | $c_{19}$ | 0 | 0 | 1 | 1 | 1 | 0 | $c_{29}$ | 0 | 1 | 1 | 1 | 1 | 0 | $c_{39}$ | 1 | 1 | 1 | 1 | 0 | 1 |
| $c_{10}$ | 1 | 0 | 1 | 0 | 0 | 0 | $c_{20}$ | 1 | 1 | 1 | 1 | 1 | 1 | $c_{30}$ | 1 | 1 | 0 | 0 | 1 | 0 | $c_{40}$ | 0 | 1 | 1 | 0 | 0 | 0 |

**Table 3.** RTD-maximal class that is not RTD-maximum.

## 6 Algebraic Proof of Shortest-Path-Closedness of VCD-Maximum Classes

In this section, we give an example of how the algebraic techniques applied to obtain our main result can also yield more elegant and insightful proofs for already known results. Our example is the proof showing that VCD-maximum classes are shortest-path-closed.

A shortest-path-closed class is a class $C$ in which any two concepts $c, c'$ 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 $1 \le i \le k$. It is known that VCD-maximum classes are shortest-path-closed [5], but algebraic methods provide an elegant alternative proof.

For $Z \subseteq X = \{x_1, \ldots, x_m\}$ and $t \le m$, 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 $0 \le k \le t$, $1 \le i_1 < \cdots < i_k \le m$ 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.* Suppose $Z \subseteq X$ 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 $P_m^d(Z)$.

As in [10], we use induction on $t$: If $t \leq d$, there is nothing to prove. Suppose $t > d$ 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 inductive 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. □

**Theorem 3.** *If $C$ is a* VCD*-maximum class, then $C$ is shortest-path-closed.*

*Proof.* In this proof, we use the symbol $\triangle$ to denote symmetric difference.

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 inductive hypothesis $c'$ and $c_2$ are Hamming-connected. Therefore, $c_1$ and $c_2$ are Hamming-connected. □

## 7 Conclusions

Our analog of Sauer's bound for RTD establishes a new connection between teaching complexity and VC-dimension. A main contribution besides obtaining this result is the successful application of algebraic proof techniques. The characterization of teaching sets obtained this way is of potential use for future studies not only in the context of the combinatorial questions we asked 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. As a byproduct of our studies, we proved several new results on VCD-maximal classes. Altogether, our results might be helpful in solving the long-standing sample compression conjecture [3] 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.** This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC). Pavel Semukhin was also supported in part by a postdoctoral fellowship from the Pacific Institute for the Mathematical Sciences.

## References

1. Doliwa, T.: Personal communication (2011)
2. Doliwa, T., Simon, H.U., Zilles, S.: Recursive teaching dimension, learning complexity, and maximum classes. In: ALT. (2010) 209–223
3. Floyd, S., Warmuth, M.K.: Sample compression, learnability, and the Vapnik-Chervonenkis dimension. Machine Learning **21**(3) (1995) 269–304
4. Goldman, S.A., Kearns, M.J.: On the complexity of teaching. Journal of Computer and System Sciences **50** (1995) 20–31
5. Kuzmin, D., Warmuth, M.K.: Unlabeled compression schemes for maximum classes. J. Mach. Learn. Res. **8** (2007) 2047–2081
6. Rubinstein, B.I.P., Bartlett, P.L., Rubinstein, J.H.: Shifting: One-inclusion mistake bounds and sample compression. J. Comput. Syst. Sci. **75**(1) (2009) 37–59
7. Sauer, N.: On the density of families of sets. J. Comb. Theory, Ser. A **13**(1) (1972) 145–147
8. Shelah, S.: A combinatorial problem: Stability and order for models and theories in infinitary languages. Pac. J. Math. **4** (1972) 247–261
9. Shinohara, A., Miyano, S.: Teachability in computational learning. New Generation Comput. **8**(4) (1991) 337–347
10. Smolensky, R.: Well-known bound for the VC-dimension made easy. Computational Complexity **6**(4) (1997) 299–300
11. Vapnik, V.N., Chervonenkis, A.Y.: On the uniform convergence of relative frequencies of events to their probabilities. Theory Probab. Appl. **16** (1971) 264–280
12. Welzl, E.: Complete range spaces. Unpublished notes (1987)
13. Zilles, S., Lange, S., Holte, R., Zinkevich, M.: Models of cooperative teaching and learning. Journal of Machine Learning Research **12** (2011) 349–384

## 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 maximal classes starting from a given maximal class.

**Theorem 4.** *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 maximal class and for some instance $x \in X$ we have $|C - x| = |C|$, then $C + x$ is also 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 maximal classes $C + x$, $(C + x) + x$, $((C + x) + x) + x$, etc.*

*(2) If $|C - x| < |C|$, then $C + x$ is not a maximal class.*

*Proof.* (1) Note that $\mathrm{VCD}(C) = \mathrm{VCD}((C+x)-x)$ and $\mathrm{VCD}(C) = \mathrm{VCD}(C+x)$. 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 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 $\mathrm{VCD}(C{+}x{\cup}\{c\}) > \mathrm{VCD}(C{+}x)$. First, suppose $c{-}x_{m+1} \notin C$. Then, since $C$ is maximal, $\mathrm{VCD}(C{+}x{\cup}\{c\}) \geq \mathrm{VCD}(C{\cup}\{c{-}x_{m+1}\}) > \mathrm{VCD}(C) = \mathrm{VCD}(C{+}x)$.

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 $\mathrm{VCD}(C{+}x{\cup}\{c\}) \geq \mathrm{VCD}((C{+}x){-}x{\cup}\{c{-}x\}) > \mathrm{VCD}((C{+}x){-}x) = \mathrm{VCD}(C) = \mathrm{VCD}(C + x)$. Hence $C + x$ is a 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 $\ell$ 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 maximal class. □

The following proposition by Rubinstein et al. [6] follows immediately from the definition of VC-dimension.

**Proposition 9.** $\mathrm{VCD}(C) \le d$ *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.*

We now establish a non-trivial lower bound for the size of VCD-maximal classes and show that this bound can be met when *both* VCD and $|X|$ are large.

**Theorem 5.** *Let* $C \subseteq 2^X$ *be a* VCD*-maximal class over a set* $X$ *with* $|X| = m$*. If* $\mathrm{VCD}(C) = d$*, then*

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

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

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

*This lower bound can be met when* $m \gg d$*, that is, when* $|X| - \mathrm{VCD}(C)$ *is small compared to* $|X|$*.*

*Proof.* We prove the second inequality. Suppose $\mathrm{VCD}(C) = m-d$ and $|C| < 2^m - 2^{d-1}\binom{m}{d-1}$. In this case, we have that $|\overline{C}| > 2^{d-1}\binom{m}{d-1}$. By Proposition 9, $\overline{C}$ must contain at least one $(d-1)$-cube for each subset of $d-1$ instances. Consider a union of $(d-1)$-cubes from $\overline{C}$ taking exactly one cube for each subset of instances of size $d-1$. Then the size of this union will be at most $2^{d-1}\binom{m}{d-1}$. Therefore, $\overline{C}$ must contain at least one concept $c$ that does not belong to the above union of $(d-1)$-cubes. Hence, due to Proposition 9, we can add this concept $c$ to the class $C$ without increasing its VC-dimension, which contradicts the fact that $C$ is maximal.

To show that the lower bound is exact for large $m$, we need to construct a disjoint union of $(d-1)$-cubes which consists of exactly one cube for each choice of $d-1$ instances; then the complement of such union will be a maximal class $C$ with $\mathrm{VCD}(C) = m-d$ and $|C| = 2^m - 2^{d-1}\binom{m}{d-1}$. To do this, let us split the instance space $X$ into disjoint blocks of size $2d$ and let $\{c_1, \ldots, c_N\}$ be the concepts that are equal to unions of such blocks. Note that $N = 2^{\lfloor m/2d \rfloor}$ and $|c_i \triangle c_j| \ge 2d$ for $i \neq j$. Now to each subset $S \subseteq X$ of size $d-1$, 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 since for $m \gg d$, $N = 2^{\lfloor m/2d \rfloor}$ is greater than $\binom{m}{d-1}$, the number of all subsets of size $d-1$.

For each $S \subseteq X$ of size $d-1$, define a $(d-1)$-cube $C(S)$ based on $c_S$, that is, $C(S) = 2^S \times \{c_S|_{X\setminus S}\}$. Note that for $S \neq S'$, the cubes $C(S)$ and $C(S')$ are disjoint because, by construction, $|c_S \triangle c_{S'}| \ge 2d$. Therefore, the class $C$, defined as

$$C = 2^X \setminus \bigcup_{S \subseteq X:\ |S|=d-1} C(S),$$

is a maximal class of VC-dimension $m-d$ and size $2^m - 2^{d-1}\binom{m}{d-1}$. □

As a corollary we obtain that for a 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 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 maximal, we have, by Theorem 5, 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 maximal class.*

We will need to prove the following lemma first.

**Lemma 2.** *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 w.l.o.g. 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 2 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 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 2, we have that $\mathrm{VCD}((C_0 \times C_1) \cup \{c_0c_1\}) = \mathrm{VCD}(C_0 \times C_1)$. Since $c_0c_1 \notin C_0 \times C_1$, this proves that the class $C_0 \times C_1$ is not maximal. □