Journal of Machine Learning Research x (2011) xx-xx Submitted 6/10; Revised 12/10; Published x/11

Models of Cooperative Teaching and Learning

Sandra Zilles ZILLES@CS.UREGINA.CA
Department of Computer Science
University of Regina, Regina, SK, Canada

Steffen Lange S.LANGE@FBI.H-DA.DE
Department of Computer Science
Darmstadt University of Applied Sciences, Darmstadt, Germany

Robert Holte HOLTE@CS.UALBERTA.CA
Department of Computing Science
University of Alberta, Edmonton, AB, Canada

Martin Zinkevich MAZ@YAHOO-INC.COM
Yahoo! Research, Mission College, CA, USA

Editor: Nicolò Cesa-Bianchi

Abstract

While most supervised machine learning models assume that training examples are sampled at random or adversarially, this article is concerned with models of learning from a cooperative teacher that selects "helpful" training examples. The number of training examples a learner needs for identifying a concept in a given class C of possible target concepts (sample complexity of C) is lower in models assuming such teachers, *i.e.*, "helpful" examples can speed up the learning process.

The problem of how a teacher and a learner can cooperate in order to reduce the sample complexity, yet without using "coding tricks", has been widely addressed. Nevertheless, the resulting teaching and learning protocols do not seem to make the teacher select intuitively "helpful" examples. The two models introduced in this paper are built on what we call *subset teaching sets* and *recursive teaching sets*. They extend previous models of teaching by letting both the teacher and the learner exploit *knowing* that the partner is cooperative. For this purpose, we introduce a new notion of "coding trick"/"collusion".

We show how both resulting sample complexity measures (the *subset teaching dimension* and the *recursive teaching dimension*) can be arbitrarily lower than the classic teaching dimension and known variants thereof, without using coding tricks. For instance, monomials can be taught with only two examples independent of the number of variables.

The subset teaching dimension turns out to be nonmonotonic with respect to subclasses of concept classes. We discuss why this nonmonotonicity might be inherent in many interesting cooperative teaching and learning scenarios.

Keywords: Teaching dimension, Learning Boolean functions, Interactive learning, Collusion

1. Introduction

A central problem in machine learning is that learning algorithms often require large quantities of data. Data may be available only in limited quantity, putting successful deployment of standard machine learning techniques beyond reach. This problem is addressed by models

of machine learning that are enhanced by interaction between a learning algorithm (learner, for short) and its environment, whose main purpose is to reduce the amount of data needed for learning. Interaction here means that at least one party actively controls which information is exchanged about the target object to be learned. Most classic machine learning models address the "average case" of data presentation to a learner (labeled examples are drawn independently at random from some distribution) or even the "worst case" (examples are drawn in an adversarial fashion). This results in the design of learners requiring more data than would be necessary under more optimistic (and often realistic) assumptions. As opposed to that, interactive learning refers to a "good case" in which representative examples are selected, whereby the number of examples needed for successful learning may shrink significantly.

Interactive machine learning is of high relevance for a variety of applications, *e.g.*, those in which a human interacts with and is observed by a learning system. A systematic and formally founded study of interactive learning is expected to result in algorithms that can reduce the cost of acquiring training data in real-world applications.

This paper focusses on particular formal models of interactive concept learning. Considering a finite instance space and a class of (thus finite) concepts over that space, it is obvious that each concept can be uniquely determined if enough examples are known. Much less obvious is how to minimize the number of examples required to identify a concept, and with this aim in mind models of *cooperative learning* and learning from *good examples* were designed and analyzed. The selection of good examples to be presented to a learner is often modeled using a teaching device (teacher) that is assumed to be benevolent by selecting examples expediting the learning process (see for instance Angluin and Kriķis (1997); Jackson and Tomkins (1992); Goldman and Mathias (1996); Mathias (1997)).

Throughout this paper we assume that teaching/learning proceeds in a simple protocol; the teacher presents a batch of labeled examples (that is, a set of instances, each paired with a label 1 or 0, according to whether or not the instance belongs to the target concept) to the learner and the learner returns a concept it believes to be the target concept. If the learner's conjecture is correct, the exchange is considered successful. The sample size, *i.e.*, the number of examples the teacher presents to the learner, is the object of optimization; in particular we are concerned with the worst case sample size measured over all concepts in the underlying class C of all possible target concepts. Other than that, computational complexity issues are not the focus of this paper.

A typical question is *How can a teacher and a learner cooperatively minimize the worst case sample size without using coding tricks?*—a coding trick being, *e.g.*, any *a priori* agreement on encoding concepts in examples, depending on the concept class C. For instance, if teacher and learner agreed on a specific order for the concept representations and the instances and agreed to use the j^{th} instance in this ordering to teach the j^{th} concept, that would be a coding trick. In practice, the teacher and the learner might not be able to agree on such an order, for instance, if the teacher is a human who does not have the same representation of a concept as the machine has. There is so far no generally accepted definition of the term "coding trick" (sometimes also called "collusion"); the reader is referred to Angluin and Kriķis (1997); Ott and Stephan (2002); Goldman and Mathias (1996). It is often more convenient to define what constitutes a valid pair of teacher and learner.

The most popular teaching model is the one introduced by Goldman and Mathias (1996). Here a team of teacher and learner is considered valid if, for every concept c in the underlying class C the following properties hold.

- The teacher selects a set S of labeled examples consistent with c.
- On input of *any superset* of S of examples that are labeled consistently with c, the learner will return a hypothesis representing c.

The idea behind this definition is that the absence of examples in the sample S cannot be used for encoding knowledge about the target concept. This is completely in line with notions of inductive inference from good examples, see Freivalds et al. (1993); Lange et al. (1998).

One way for a teacher and a learner to form a valid team under these constraints is for the teacher to select, for every concept $c \in C$, a sample S that is consistent with c but inconsistent with every other concept in C. The size of the minimum such sample is called the *teaching dimension* of c in C. The teaching dimension of the class C is the maximum teaching dimension over all concepts in C. For more information, the reader is referred to the original literature on teaching dimension and variants thereof (Shinohara and Miyano (1991); Goldman and Kearns (1995); Anthony et al. (1992)).

The teaching dimension however does not always seem to capture the intuitive idea of cooperation in teaching and learning. Consider the following simple example. Let C_0 consist of the empty concept and all singleton concepts over a given instance space $X = \{x_1, \ldots, x_n\}$. Each singleton concept $\{x_i\}$ has a teaching dimension of 1, since the single positive example $(x_i, +)$ is sufficient for determining $\{x_i\}$. This matches our intuition that concepts in this class are easy to teach. In contrast to that, the empty concept has a teaching dimension of n—every example has to be presented. However, if the learner assumed the teacher was cooperative—and would therefore present a positive example if the target concept was non-empty—the learner could confidently conjecture the empty concept upon seeing just one negative example.

Let us extend this reasoning to a slightly more complex example, the class of all boolean functions that can be represented as a monomial over m variables ($m = 4$ in this example). Imagine yourself in the role of a learner knowing your teacher will present helpful examples. If the teacher sent you the examples

$$(0100, +), (0111, +)\,,$$

what would be your conjecture? Presumably most people would conjecture the monomial $M \equiv \overline{v_1} \wedge v_2$, as does for instance the algorithm proposed by Valiant (1984). Note that this choice is not uniquely determined by the data: the empty (always true) monomial and the monomials $\overline{v_1}$ and v_2 are also consistent with these examples. And yet M seems the best choice, because we'd think the teacher would not have kept any bit in the two examples constant if it was not in the position of a relevant variable. In this example, the natural conjecture is the most specific concept consistent with the sample, but that does not, in general, capture the intuitive idea of cooperative learning. In particular, if, instead of the class of all monomials, the class of all complements of these concepts over the same instance space is chosen, then a cooperative teacher and learner would need only two negatively

labeled example for teaching the complement of the concept associated with $\overline{v_1} \wedge v_2$, which is now the least specific concept in the class. Going further, one could swap + for − and vice versa only for some of the instances. In effect, only the labels in the examples chosen by the teacher would change, but not the instances as such. The concepts guessed by the learner would then be neither the most specific nor the least specific concepts.

Could the learner's reasoning about the teacher's behavior in these examples be implemented without a coding trick? We will argue below that, for a very intuitive, yet mathematically rigorous definition of coding tricks, no coding trick is necessary to achieve exactly this behavior of teacher and learner; there are general strategies that teachers and learners can independently implement to cooperatively learn any finite concept class. When applied to the class of monomials this principle enables any monomial to be learned from just two examples, regardless of the number m of variables.

Our approach is to define a new model of cooperation in learning, based on the idea that each partner in the cooperation tries to reduce the sample size by exploiting the assumption that the other partner does so. If this idea is iteratively propagated by both partners, one can refine teaching sets iteratively ending up with a framework for highly efficient teaching and learning without any coding tricks. It is important to note that teacher and learner do not agree on any order of the concept class or any order of the instances. All they know about each others' strategies is a general assumption about how cooperation should work independent of the concept class or its representation.

We show that the resulting variant of the teaching dimension—called the *subset teaching dimension (STD)*—is not only a uniform lower bound of the teaching dimension but can be constant where the original teaching dimension is exponential, even in cases where only one iteration is needed. For example, as illustrated above, the STD of the class of monomials over $m \geq 2$ variables is 2, in contrast to its original teaching dimension of 2^m.

Some examples however will reveal a nonmonotonicity of the subset teaching dimension: some classes possess subclasses with a higher subset teaching dimension, which is at first glance not very intuitive. We will explain below why in a cooperative model such a nonmonotonicity does not have to contradict intuition; additionally we introduce a second model of cooperative teaching and learning, that results in a monotonic dimension, called the *recursive teaching dimension (RTD)*. Recursive teaching is based on the idea to let the teacher and the learner exploit a hierarchical structure that is intrinsic in the concept class. The canonical hierarchy associated with a concept class C is a nesting of C, starting with the class of all concepts in C that are easiest to teach (*i.e.,* have the lowest teaching dimension) and then applying the nesting process recursively to the remaining set of concepts. At every stage, the recursive teaching sets for the concepts that are easiest to teach are the teaching sets for these concepts with respect to the class of remaining concepts. The recursive teaching dimension is the size of the largest recursive teaching set constructed this way.

The RTD-model is not as intuitive a model of cooperative teaching and learning as the STD-model is, but it displays a surprising set of properties. Besides its monotonicity, the RTD corresponds to teacher-learner protocols that do not violate Goldman and Mathias's definition of teaching and learning without coding tricks. Nevertheless, substantial improvements over the classical teaching dimension are obtained. A recent study furthermore shows that the recursive teaching dimension is a combinatorial parameter of importance

when analyzing the complexity of learning problems from the perspective of active learning, teaching, learning from random examples, and sample compression, see Doliwa et al. (2010).

Both our teaching protocols significantly improve sample efficiency compared to previously studied variants of the teaching dimension.

This paper is a correction and extension of an earlier publication (Zilles et al. (2008)). In this earlier publication, both Proposition 5(1) and the conjecture in Lemma 23 were wrong.

2. Related work

The problem of defining what are "good" or "helpful" examples in learning has been studied in several fields of learning theory.

Various learning models, which each involve one particular type of teacher, were proposed by Goldman and Kearns (1995); Goldman and Mathias (1996); Mathias (1997); Jackson and Tomkins (1992); Shinohara and Miyano (1991); Angluin and Kriķis (1997, 2003); Balbach (2008); Kobayashi and Shinohara (2009); these studies mostly focus on learning boolean functions. See also Balbach and Zeugmann (2009) for a recent survey. The teaching dimension model, independently introduced by Goldman and Kearns (1991; 1995) and by Shinohara and Miyano (1991), is concerned with the sample complexity of teaching arbitrary consistent learners. Samples that will allow any consistent learner to identify the target concept are called *teaching sets*; the maximum size of minimal teaching sets of all concepts in the underlying concept class C is called the *teaching dimension* of C. The problem of avoiding unfair "coding tricks" between teachers and learners is addressed in particular by Goldman and Mathias (1996) with the introduction of a formal notion of collusion-free learning. It is known that computing (the size of) minimal teaching sets is in general intractable, see Servedio (2001), which is one reason why the polynomial-time models introduced by Jackson and Tomkins (1992) are of interest. Jackson and Tomkins no longer require that teachers choose samples that make any consistent learner successful; they rather focus on specific teacher/learner pairs. Loosening the requirement of learners being consistent, Kobayashi and Shinohara (2009) analyze how restrictions on the number of examples given by the teacher influence the worst-case error of the hypothesis returned by a learner.

The teaching dimension was analyzed in the context of online learning, cf. Ben-David and Eiron (1998); Rivest and Yin (1995), and in the model of learning from queries, *e.g.*, by Hegedűs (1995) and by Hanneke (2007), with a focus on active learning in the PAC framework. In contrast to these models, in inductive inference the learning process is not necessarily considered to be finite. Approaches to defining learning infinite concepts from good examples (Freivalds et al. (1993); Lange et al. (1998)) do not focus on the size of a finite sample of good examples, but rather on characterizing the cases in which learners can identify concepts from only finitely many examples.

One of the two approaches we present in this paper is mainly based on an idea by Balbach (2008). He defined and analyzed a model in which, under the premise that the teacher uses a minimal teaching set (as defined by Goldman and Kearns (1991; 1995)) as a sample, a learner can reduce the size of a required sample by eliminating concepts which possess a

teaching set smaller than the number of examples provided by the teacher so far. Iterating this idea, the size of the teaching sets might be gradually reduced significantly. Though our approach is syntactically quite similar to Balbach's, the underlying idea is a different one (we do not consider elimination by the sample size but elimination by the sample content as compared to all possible teaching sets). The resulting variant of the teaching dimension in general yields different performance results in terms of sample size than Balbach's model does.

3. The teaching dimension and the Balbach teaching dimension

Let $\mathbb{N}$ denote the set of all non-negative integers, $\emptyset$ denote the empty set, and $|M|$ denote the cardinality of a finite set M. For any $k \in \mathbb{N}$, the power set of $\{1, \ldots, k\}$ will be denoted by $2^{[k]}$.

In the models of teaching and learning to be defined below, we will always assume that the goal in an interaction between a teacher and a learner is to make the learner identify a (finite) concept over a (finite) instance space X.

Most of the recent work on teaching (cf. Balbach (2008); Zilles et al. (2008); Balbach and Zeugmann (2009); Kobayashi and Shinohara (2009)) defines a concept simply as a subset of X and a concept class as a set of subsets of X. In effect, this is exactly the definition we would need for introducing the teaching models we define below. However, the definition and discussion of the notion of collusion (*i.e.*, the conceptualization of what constitutes a coding trick), cf. Section 4, motivates a more general definition of concepts and concept classes. This more general definition considers the instance space X as an ordered set and every concept class C as an ordered set of subsets of X.

To formalize this, let $X = \{1, \ldots, n\}$. Concepts and concept classes are defined as follows.

Definition 1 *Let $z \in \mathbb{N}$.*

A concept class of cardinality z is defined by an injective mapping $C : \{1, \ldots, z\} \to 2^{[n]}$. Every $i \in \{1, \ldots, z\}$ and thus every concept $C(i)$ is associated with a membership function on $X = \{1, \ldots, n\}$, given by $C(i)(j) = +$ if $j \in C(i)$, and $C(i)(j) = -$ if $j \notin C(i)$, where $1 \leq j \leq n$. Thus a concept class C of cardinality $z \in \mathbb{N}$ is represented as a matrix $(C(i)(j))_{1 \leq i \leq z, 1 \leq j \leq n}$ over $\{+, -\}$.

$\mathcal{C}_z$ denotes the collection of all concept classes of cardinality z. $\mathcal{C} = \bigcup_{z \in \mathbb{N}} \mathcal{C}_z$ denotes the collection of all concept classes (of any cardinality).

Consequently, concepts and concept classes considered below will always be finite.

Definition 2 *Let $z \in \mathbb{N}$ and $C \in \mathcal{C}_z$.*

A sample is a set $S = \{(j_1, l_1), \ldots, (j_r, l_r)\} \subseteq X \times \{+, -\}$, where every element (j, l) of S is called a (labeled) example.

Let $i \in \{1, \ldots, z\}$. $C(i)$ is consistent with S (and S is consistent with $C(i)$) if $C(i)(j_t) = l_t$ for all $t \in \{1, \ldots, r\}$. Denote

$$Cons(S, C) = \{i \in \{1, \ldots, z\} \mid C(i) \text{ is consistent with } S\}.$$

The power set of $\{1, \ldots, n\} \times \{+, -\}$, i.e.*, the set of all samples, is denoted by $\mathcal{S}$.*

3.1 Protocols for teaching and learning in general

In what follows, we assume that a teacher selects a sample for a given target concept and that a learner, given any sample S, always returns an index of a concept from the underlying concept class C. Formally, if $z \in \mathbb{N}$ and $(C(i)(j))_{1\leq i\leq z, 1\leq j\leq n}$ is a concept class in $\mathcal{C}_z$, a *teacher* for C is a function $\tau : \{1,\ldots,z\} \to \mathcal{S}$; a *learner* for C is a function $\lambda : \mathcal{S} \to \{1,\ldots,z\}$.

In order to constrain the definition of validity of a teacher/learner pair to a desired form of interaction in a learning process, the notion of adversaries will be useful. Adversaries will be considered third parties with the option to modify a sample generated by a teacher before this sample is given to a learner. Formally, an *adversary* is a relation $Ad \subseteq \mathcal{S}^3$. Intuitively, if $(\tau(i), C(i), S) \in Ad$ for some $i \in \{1,\ldots,z\}$ and some teacher τ for $C = (C(i)(j))_{1\leq i\leq z, 1\leq j\leq n}$, then the adversary has the option to modify $\tau(i)$ to S and the learner communicating with τ will get S rather than $\tau(i)$ as input. A special adversary is the so-called *trivial adversary* Ad^*, which satisfies $(S_1, S_2, S) \in Ad^*$ if and only if $S_1 = S$. This adversary does not modify the samples generated by the teacher at all.

All teaching and learning models introduced below will involve a very simple *protocol* between a teacher and a learner (and an adversary).

Definition 3 *Let P be a mapping that maps every concept class $C \in \mathcal{C}$ to a pair $P(C) = (\tau, \lambda)$ where τ is a teacher for C and λ is a learner for C. P is called a protocol; given $C \in \mathcal{C}$, the pair $P(C)$ is called a protocol for C.*

1. *Let $z \in \mathbb{N}$ and let $C \in \mathcal{C}_z$ be a concept class. Let Ad_C be an adversary. $P(C) = (\tau, \lambda)$ is called successful for C with respect to Ad_C if $\lambda(S) = i$ for all pairs (i, S) where $i \in \{1,\ldots,z\}$, $S \in \mathcal{S}$, and $(\tau(i), C(i), S) \in Ad_C$.*

2. *Let $\mathcal{A} = (Ad_C)_{C\in\mathcal{C}}$ be a family of adversaries. P is called successful with respect to $\mathcal{A}$ if, for all $C \in \mathcal{C}$, $P(C)$ is successful for C with respect to Ad_C.*

Protocols differ in the strategies according to which the teacher and the learner operate, *i.e.*, according to which the teacher selects a sample and according to which the learner selects a concept.

In all protocols considered below, teachers always select consistent samples for every given target concept and learners, given any sample S, always return a concept consistent with S if such a concept exists in the underlying class C. Formally, all teachers τ for a concept class $C \in \mathcal{C}_z$ will fulfill $i \in \mathit{Cons}(\tau(i), C)$ for all $i \in \{1,\ldots,z\}$; all learners λ for a class C will fulfill $\lambda(S) \in \mathit{Cons}(S, C)$ for all $S \in \mathcal{S}$ with $\mathit{Cons}(S, C) \neq \emptyset$. Moreover, all the adversaries Ad we present below will have the following property:

$$\text{for any three samples } S_1, S_2, S \in \mathcal{S}, \text{ if } (S_1, S_2, S) \in Ad \text{ then } S_1 \subseteq S \subseteq S_2\,.$$

However, this does not mean that we consider other forms of teachers, learners, or adversaries illegitimate. They are just beyond the scope of this paper.

The goal in sample-efficient teaching and learning is to design protocols that, for every concept class C, are successful for C while reducing the (worst-case) size of the samples the teacher presents to the learner for any target concept in C. At the same time, by introducing adversaries, one tries to avoid certain forms of collusion, an issue that we will discuss in Section 4.

3.2 Protocols using minimal teaching sets and Balbach teaching sets

The fundamental model of teaching we consider here is based on the notion of *minimal teaching sets*, which is due to Goldman and Kearns (1995) and Shinohara and Miyano (1991).

Let $z \in \mathbb{N}$ and let $C \in \mathcal{C}_z$ be a concept class. Let S be a sample. S is called a *teaching set* for i with respect to C if $Cons(S, C) = \{i\}$. A teaching set allows a learning algorithm to uniquely identify a concept in the concept class C. Teaching sets of minimal size are called *minimal teaching sets*. The *teaching dimension* of i in C is the size of such a minimal teaching set, *i.e.*, $TD(i, C) = \min\{|S| \mid Cons(S, C) = \{i\}\}$, the worst case of which defines the teaching dimension of C, *i.e.*, $TD(C) = \max\{ TD(i, C) \mid 1 \leq i \leq z\}$. To refer to the set of all minimal teaching sets of i with respect to C, we use

$$TS(i, C) = \{S \mid Cons(S, C) = \{i\} \text{ and } |S| = TD(i, C)\}\,.$$

Minimal teaching sets induce the following protocol.

Protocol 4 *Let P be a protocol. P is called a teaching set protocol (TS-protocol for short) if the following two properties hold for every $C \in \mathcal{C}$, where $P(C) = (\tau, \lambda)$.*

1. *$\tau(i) \in TS(i, C)$ for all $i \in \{1, \ldots, z\}$,*
2. *$\lambda(S) \in Cons(S, C)$ for all $S \in \mathcal{S}$ with $Cons(S, C) \neq \emptyset$.*

This protocol is obviously successful with respect to the family consisting only of the trivial adversary. The teaching dimension of a concept class C is then a measure of the worst case sample size required in this protocol with respect to Ad^* when teaching/learning any concept in C.

The reason that, for every concept class $C \in \mathcal{C}_z$, the protocol $P(C)$ is successful (with respect to Ad^*) is simply that a teaching set eliminates all but one concept due to inconsistency. However, if the learner knew $TD(i, C)$ for every $i \in \{1, \ldots, z\}$ then sometimes concepts could also be eliminated by the mere number of examples presented to the learner. For instance, assume a learner knows that all but one concept $C(i)$ have a teaching set of size one and that the teacher will teach using teaching sets. After having seen 2 examples, no matter what they are, the learner could eliminate all concepts but $C(i)$. This idea, referred to as elimination by sample size, was introduced by Balbach (2008). If a teacher knew that a learner eliminates by consistency and by sample size then the teacher could consequently reduce the size of some teaching sets (*e.g.*, here, if $TD(i, C) \geq 3$, a new "teaching set" for i could be built consisting of only 2 examples).

More than that—this idea is iterated by Balbach (2008): if the learner knew that the teacher uses such reduced "teaching sets" then the learner could adapt his assumption on the size of the samples to be expected for each concept, which could in turn result in a further reduction of the "teaching sets" by the teacher and so on. The following definition captures this idea formally.

Definition 5 (Balbach (2008))
Let $z \in \mathbb{N}$ and let $C \in \mathcal{C}_z$ be a concept class. Let $i \in \{1, \ldots, z\}$ and S a sample. Let $BTD^0(i, C) = TD(i, C)$. We define iterated dimensions for all $k \in \mathbb{N}$ as follows.

- $Cons_{size}(S, C, k) = \{i \in Cons(S, C) \mid BTD^k(i, C) \geq |S|\}$.
- $BTD^{k+1}(i, C) = \min\{|S| \mid Cons_{size}(S, C, k) = \{i\}\}$

Let κ be minimal such that $BTD^{\kappa+1}(i, C) = BTD^{\kappa}(i, C)$ for all $i \in \{1, \ldots, z\}$. The Balbach teaching dimension $BTD(i, C)$ of i in C is defined by $BTD(i, C) = BTD^{\kappa}(i, C)$ and the Balbach teaching dimension $BTD(C)$ of the class C is $BTD(C) = \max\{BTD(i, C) \mid 1 \leq i \leq z\}$.[1] *For every $i \in \{1, \ldots, z\}$ we define*

$$BTS(i, C) = \{S \mid Cons_{size}(S, C, \kappa) = \{i\} \text{ and } |S| = BTD(i, C)\}$$

and call every set in $BTS(i, C)$ a minimal Balbach teaching set of i with respect to C.

By $Cons_{size}(S, C)$ we denote the set $Cons_{size}(S, C, \kappa)$.

The Balbach teaching dimension measures the sample complexity of the following protocol with respect to the trivial adversary.

Protocol 6 *Let P be a protocol. P is called a Balbach teaching set protocol (BTS-protocol for short) if the following two properties hold for every $C \in \mathcal{C}$, where $P(C) = (\tau, \lambda)$.*

1. *$\tau(i) \in BTS(i, C)$ for all $i \in \{1, \ldots, z\}$,*
2. *$\lambda(S) \in \{i \mid$ there is some $S' \in BTS(i, C)$ such that $S' \subseteq S\}$ for all $S \in \mathcal{S}$ that contain a set $S' \in BTS(i, C)$ for some $i \in \{1, \ldots, z\}$.*

Obviously, $BTD(C) \leq TD(C)$ for every concept class $C \in \mathcal{C}$. How much the sample complexity can actually be reduced by a cooperative teacher/learner pair according to this "elimination by sample size" principle, is illustrated by the concept class C_0 which consists of the empty concept and all singleton concepts over X. The teaching dimension of this class is n, whereas the BTD is 2.

3.3 Teaching monomials

A standard example of a class of boolean functions studied in learning theory is the class $\mathcal{F}_m$ of monomials over a set $\{v_1, \ldots, v_m\}$ of m variables, for any $m \geq 2$.[2] Usually, this class is just defined by choosing $X = \{0, 1\}^m$ as the underlying instance space. Then, for any monomial M, the corresponding concept is defined as the set of those assignments in $\{0, 1\}^m$ for which M evaluates positively. Within our more general notion of concept classes, there is more than just one class of all monomials over m variables (which we will later consider as equivalent). This is due to distinguishing different possible orderings over X and over the class of monomials itself.

Definition 7 *Let $m \in \mathbb{N}$, $m \geq 2$ and assume $n = 2^m$,* i.e., *$X = \{1, \ldots, 2^m\}$.*

1. Balbach (2008) denotes this by *IOTTD*, called iterated optimal teacher teaching dimension; we deviate from this notation for the sake of convenience.
2. A monomial over $\{v_1, \ldots, v_m\}$ is a conjunction of literals over $\{v_1, \ldots, v_m\}$, also called a 1-CNF or a 1-term DNF.

Let $\mathrm{bin} : \{1,\ldots,2^m\} \to \{0,1\}^m$ *be a bijection,* i.e.*, a repetition-free enumeration of all bit strings of length* m*. Let* $\mathrm{mon} : \{1,\ldots,3^m\} \to \mathcal{F}_m$ *be a bijective enumeration of all monomial functions over* m *variables* $v_1,\ldots,v_m$*.*

A mapping $C : \{1,\ldots,3^m\} \to 2^{[2^m]}$ *is called a concept class of all monomials over* m *variables if, for all* $i \in \{1,\ldots,3^m\}$ *and all* $j \in \{1,\ldots,2^m\}$*,*

$$C(i)(j) = \begin{cases} +, & \text{if } \mathrm{mon}(i) \text{ evaluates to TRUE when assigning } \mathrm{bin}(j) \text{ to } (v_1,\ldots,v_m)\,, \\ -, & \text{if } \mathrm{mon}(i) \text{ evaluates to FALSE when assigning } \mathrm{bin}(j) \text{ to } (v_1,\ldots,v_m)\,. \end{cases}$$

It turns out that a class of all monomials contains only one concept for which the *BTD*-iteration yields an improvement.

Theorem 8 (Balbach (2008)) *Let* $m \in \mathbb{N}$*,* $m \geq 2$*. Let* $C : \{1,\ldots,3^m\} \to 2^{[2^m]}$ *be a concept class of all monomials over* m *variables. Let* $i^* \in \{1,\ldots,3^m\}$ *with* $C(i^*) = \emptyset$ *be an index for the concept representing the contradictory monomial.*

1. $BTD(i^*, C) = m + 2 < 2^m = TD(i^*, C)$*.*
2. $BTD(i, C) = TD(i, C)$ *for all* $i \in \{1,\ldots,3^m\} \setminus \{i^*\}$*.*

The intuitive reason for $BTD(i^*, C) = m + 2$ in Theorem 8 is that samples for $C(i^*)$ of size $m + 1$ or smaller are consistent also with monomials different from $C(i^*)$, namely those monomials that contain every variable exactly once (each such monomial is positive for exactly one of the 2^m instances). These other monomials hence cannot be eliminated—neither by size nor by inconsistency.

4. Avoiding coding tricks

Intuitively, the trivial adversary of course does not prevent teacher and learner from using coding tricks. One way of defining what a coding trick is—or what a valid (collusion-free) behaviour of a teacher/learner is supposed to look like—is to require success with respect to a specific non-trivial type of adversary.

Goldman and Mathias (1996) called a pair of teacher and learner valid for a concept class $C \in \mathcal{C}_z$ if, for every concept $C(i)$ in the class C, the following properties hold.

- The teacher selects a set S of labeled examples consistent with $C(i)$.
- On input of *any superset* of S of examples that are labeled consistently with $C(i)$, the learner will return a hypothesis representing $C(i)$.

In other words, they considered a teacher-learner pair (τ, λ) a valid protocol for C if and only if it is successful with respect to *any* adversary Ad_C that fulfills $\tau(i) \subseteq S \subseteq C(i)$ for all $i \in \{1,\ldots,z\}$ and all $S \in \mathcal{S}$ with $(\tau(i), C(i), S) \in Ad_C$.

Obviously, teacher/learner pairs using minimal teaching sets according to the *TS*-protocol (Protocol 4) are valid in this sense.

Theorem 9 *Let* $z \in \mathbb{N}$ *and let* $C \in \mathcal{C}_z$ *be a concept class. Let* τ *be a teacher for* C*,* λ *a learner for* C*. If* (τ, λ) *is a TS-protocol for* C *then* (τ, λ) *is successful with respect to any adversary* Ad_C *that fulfills* $\tau(i) \subseteq S \subseteq C(i)$ *for all* $i \in \{1,\ldots,z\}$*.*

Proof. Immediate from the definitions. □

Not only the protocol based on the teaching dimension (Protocol 4), but also the protocol based on the Balbach teaching dimension (Protocol 6) yields only valid teacher/learner pairs according to this definition—a consequence of Theorem 10.

Theorem 10 *Let $z \in \mathbb{N}$ and let $C \in \mathcal{C}_z$ be a concept class. Let $i \in \{1, \ldots, z\}$, $S \in BTS(i, C)$, and $T \supseteq S$ such that $i \in Cons(T, C)$. Then there is no $i' \in Cons(T, C)$ such that $i \neq i'$ and $S' \subseteq T$ for some $S' \in BTS(i', C)$.*

Proof. Assume there is some $i' \in Cons(T, C)$ such that $i \neq i'$ and some $S' \in BTS(i', C)$ such that $S' \subseteq T$. Since both $C(i)$ and $C(i')$ are consistent with T and both S and S' are subsets of T, we have $i \in Cons(S', C)$ and $i' \in Cons(S, C)$. Now let $\kappa \geq 1$ be minimal such that $BTD^{\kappa}(i^*, C) = BTD(i^*, C)$ for all $i^* \in C$. From $i' \in Cons(S, C)$ and $S \in BTS(i, C)$ we obtain

$$|S'| = BTD^{\kappa}(i', C) \leq BTD^{\kappa-1}(i', C) < |S| \,.$$

Similarly, $i \in Cons(S', C)$ and $S' \in BTS(i', C)$ yields

$$|S| = BTD^{\kappa}(i, C) \leq BTD^{\kappa-1}(i, C) < |S'| \,.$$

This is a contradiction. □

This implies that every BTS-protocol is valid in the sense of the definition given by Goldman and Mathias (1996).

Corollary 11 *Let $z \in \mathbb{N}$ and let $C \in \mathcal{C}_z$ be a concept class. Let τ be a teacher for C, λ a learner for C. If (τ, λ) is a BTS-protocol for C then (τ, λ) is successful with respect to any adversary Ad_C that fulfills $\tau(i) \subseteq S \subseteq C(i)$ for all $i \in \{1, \ldots, z\}$.*

Goldman and Mathias's definition of valid teacher/learner pairs encompasses a broad set of scenarios. It accommodates all consistent learners even those that do not make any prior assumptions about the source of information (the teacher) beyond it being noise-free. However, in many application scenarios (*e.g.*, whenever a human interacts with a computer or in robot-robot interaction) it is reasonable to assume that (almost) all the examples selected by the teacher are helpful or particularly important for the target concept in the context of the underlying concept class. Processing a sample S selected by a teacher, a learner could exploit such an assumption by excluding not only all concepts that are inconsistent with S but also all concepts for which some examples in S would not seem particularly helpful/important. This would immediately call Goldman and Mathias's definition of validity into question.

Here we propose a more relaxed definition of what a valid teacher/learner pair is (and thus, implicitly, a new definition of collusion). It is important to notice, first of all, that in parts of the existing literature, teaching sets and teaching dimension are defined via properties of *sets* rather than properties of *representations* of sets, cf. Balbach (2008); Kobayashi and Shinohara (2009). Whenever this is the case, teacher/learner pairs cannot make use of the language they use for representing instances in X or concepts in C. For example, teacher and learner cannot agree on an *order* over the instance space or over the

concept class in order to encode information in samples just by the rank of their members with respect to the agreed-upon orders.

We want to make this an explicit part of the definition of collusion-free teacher/learner pairs.

Intuitively, the complexity of teaching/learning concepts in a class should not depend on certain representational features, such as any order over X or over C itself. Moreover, negating the values of all concepts on a single instance should not affect the complexity of teaching and learning either. In other words, we want protocols to be "invariant" with respect to the following equivalence relation over $\mathcal{C}$.

Definition 12 *Let $z \in \mathbb{N}$. Let $C = (C(i)(j))_{1\leq i\leq z, 1\leq j\leq n}$ and $C' = (C'(i)(j))_{1\leq i\leq z, 1\leq j\leq n}$ be two concept classes in $\mathcal{C}_z$. C and C' are called equivalent if there is a bijection $\imath : \{1,\ldots,z\} \to \{1,\ldots,z\}$, a bijection $\jmath : \{1,\ldots,n\} \to \{1,\ldots,n\}$, and for every $j \in \{1,\ldots,n\}$ a bijection $\ell_j : \{+,-\} \to \{+,-\}$, such that*

$$C(i)(j) = \ell_j(C'(\imath(i))(\jmath(j))) \text{ for all } i \in \{1,\ldots,z\},\ j \in \{1,\ldots,n\}\,.$$

In this case, $(\imath, \jmath, (\ell_j)_{1\leq j\leq n})$ is said to witness that C and C' are equivalent.

We call a protocol collusion-free if it obeys this equivalence relation in the following sense.

Definition 13 *Let P be a protocol. P is collusion-free if, for every $z \in \mathbb{N}$ and $C, C' \in \mathcal{C}_z$, where C and C' are equivalent as witnessed by $(\imath, \jmath, (\ell_j)_{1\leq j\leq n})$, the following two properties hold for $P(C) = (\tau, \lambda)$ and $P(C') = (\tau', \lambda')$.*

1. *If $1 \leq i \leq z$ and $\tau(i) = \{(j_1, l_1), \ldots, (j_r, l_r)\}$, then*

$$\tau'(\imath(i)) = \{(\jmath(j_1), \ell_j(l_1)), \ldots, (\jmath(j_r), \ell_j(l_r))\}\,.$$

2. *If $\{(j_1, l_1), \ldots, (j_r, l_r)\} \in \mathcal{S}$ and $\lambda(\{(j_1, l_1), \ldots, (j_r, l_r)\}) = i$, then*

$$\lambda'(\{(\jmath(j_1), \ell_j(l_1)), \ldots, (\jmath(j_r), \ell_j(l_r))\}) = \imath(i)\,.$$

It is obvious that both protocols introduced above are collusion-free.

Theorem 14 *1. Every teaching set protocol is collusion-free.*

2. Every Balbach teaching set protocol is collusion-free.

Proof. Immediate from the definitions. □

The new protocols we define below are collusion-free as well. This means that all protocols studied in this article are defined independently of the order over X and C. Concept classes can hence be considered as sets of sets rather than matrices. Consequently, Definition 1 is more general than required in the rest of this paper. We therefore ease notation as follows.

$X = \{x_1, \ldots, x_n\}$ denotes the instance space. A concept c is a subset of X and a concept class C is a subset of the power set of X. We identify every concept c with its

membership function given by $c(x_i) = +$ if $x_i \in c$, and $c(x_i) = -$ if $x_i \notin c$, where $1 \leq i \leq n$. Given a sample $S = \{(y_1, l_1), \ldots, (y_r, l_r)\} \subseteq X \times \{+, -\}$, we call c consistent with S if $c(y_i) = l_i$ for all $i \in \{1, \ldots, r\}$. If C is a concept class then $\mathit{Cons}(S, C) = \{c \in C \mid c \text{ is consistent with } S\}$. S is called a teaching set for c with respect to C if $\mathit{Cons}(S, C) = \{c\}$. Then $\mathit{TD}(c, C) = \min\{|S| \mid \mathit{Cons}(S, C) = \{c\}\}$, $\mathit{TD}(C) = \max\{\mathit{TD}(c, C) \mid c \in C\}$, and $\mathit{TS}(c, C) = \{S \mid \mathit{Cons}(S, C) = \{c\} \text{ and } |S| = \mathit{TD}(c, C)\}$. The notations concerning the Balbach teaching model are adapted by analogy.

5. The subset teaching dimension

The approach studied by Balbach (2008) does not always meet the intuitive idea of teacher and learner exploiting the knowledge that either partner behaves cooperatively. Consider for instance one more time the class C_0 containing the empty concept and all singletons over $X = \{x_1, \ldots, x_n\}$. Each concept $\{x_i\}$ has the unique minimal teaching set $\{(x_i, +)\}$ in this class, whereas the empty concept only has a teaching set of size n, namely $\{(x_1, -), \ldots, (x_n, -)\}$. The idea of elimination by size allows a learner to conjecture the empty concept as soon as two examples have been provided, due to the fact that all other concepts possess a teaching set of size one. This is why the empty concept has a BTD equal to 2 in this example.

However, as we have argued in Section 1, it would also make sense to devise a learner in a way to conjecture the empty concept as soon as a first example for that concept is provided—knowing that the teacher would not use a negative example for any other concept in the class. In terms of teaching sets this means to reduce the teaching sets to their minimal subsets that are not contained in minimal teaching sets for other concepts in the given concept class.

In fact, a technicality in the definition of the Balbach teaching dimension (Definition 5) disallows the Balbach teaching dimension to be 1 unless the teaching dimension itself is already 1, as the following proposition states.

Proposition 15 *Let C be a concept class. If $\mathit{BTD}(C) = 1$ then $\mathit{TD}(C) = 1$.*

Proof. Let $\mathit{BTD}(C) = 1$. Assume $\mathit{TD}(C) > 1$.

Since $\mathit{TD}(C) > 1$, there exists a concept $\hat{c} \in C$ such that $\mathit{TD}(\hat{c}, C) > 1$. Since $\mathit{BTD}(\hat{c}, C) = 1$, there exists a minimal $\kappa \geq 1$ such that $\mathit{BTD}^{\kappa}(\hat{c}, C) = \mathit{BTD}(\hat{c}, C) = 1$. In particular, there exists a sample S such that $|S| = 1$ and

$$\{c \in \mathit{Cons}(S, C) \mid \mathit{BTD}^{\kappa-1}(c, C) \geq 1\} = \{\hat{c}\}\,.$$

Since $\mathit{BTD}^{\kappa-1}(c, C) \geq 1$ trivially holds for all $c \in C$, we obtain $\mathit{Cons}(S, C) = \{\hat{c}\}$. Consequently, as $|S| = 1$, it follows that $\mathit{TD}(\hat{c}, C) = 1$. This contradicts the choice of $\hat{c}$. Thus $\mathit{TD}(C) = 1$. □

So, if the Balbach model improves on the worst case teaching complexity, it does so only by improving the teaching dimension to a value of at least 2.

5.1 The model

We formalize the idea of cooperative teaching and learning using subsets of teaching sets as follows.

Definition 16 *Let C be a concept class, $c \in C$, and S a sample. Let $STD^0(c,C) = TD(c,C)$, $STS^0(c,C) = TS(c,C)$. We define iterated sets for all $k \in \mathbb{N}$ as follows.*

- $Cons_{sub}(S,C,k) = \{c \in C \mid S \subseteq S' \text{ for some } S' \in STS^k(c,C)\}$.
- $STD^{k+1}(c,C) = \min\{|S| \mid Cons_{sub}(S,C,k) = \{c\}\}$
- $STS^{k+1}(c,C) = \{S \mid Cons_{sub}(S,C,k) = \{c\},\ |S| = STD^{k+1}(c,C)\}$.

Let κ be minimal such that $STS^{\kappa+1}(c,C) = STS^{\kappa}(c,C)$ for all $c \in C$.[3]

A sample S such that $Cons_{sub}(S,C,\kappa) = \{c\}$ is called a subset teaching set for c in C. The subset teaching dimension $STD(c,C)$ of c in C is defined by $STD(c,C) = STD^{\kappa}(c,C)$ and we denote by $STS(c,C) = STS^{\kappa}(c,C)$ the set of all minimal subset teaching sets for c in C. The subset teaching dimension $STD(C)$ of C is defined by $STD(C) = \max\{STD(c,C) \mid c \in C\}$.

For illustration, consider again the concept class C_0, *i.e.*, $C_0 = \{c_i \mid 0 \leq i \leq n\}$, where $c_0 = \emptyset$ and $c_i = \{x_i\}$ for all $i \in \{1,\ldots,n\}$. Obviously, for $k \geq 1$,

$$STS^k(c_i) = \{\{(x_i,+)\}\} \text{ for all } i \in \{1,\ldots,n\}$$

and

$$STS^k(c_0) = \{\{(x_i,-)\} \mid 1 \leq i \leq n\}.$$

Hence $STD(C_0) = 1$ although $TD(C_0) = n$.

Note that the example of the concept class C_0 establishes that the subset teaching dimension can be 1 even if the teaching dimension is larger, in contrast to Proposition 15.

The definition of $STS(c,C)$ induces a protocol for teaching and learning: For a target concept c, a teacher presents the examples in a subset teaching set for c to the learner. The learner will also be able to pre-compute all subset teaching sets for all concepts and determine the target concept from the sample provided by the teacher.[4]

Protocol 17 *Let P be a protocol. P is called a subset teaching set protocol (STS-protocol for short) if the following two properties hold for every $C \subseteq \mathcal{C}$, where $P(C) = (\tau,\lambda)$.*

1. *$\tau(c) \in STS(c,C)$ for all $c \in C$,*
2. *$\lambda(S) \in \{c \mid \text{there is some } S' \in STS(c,C) \text{ such that } S' \subseteq S\}$ for all $S \in \mathcal{S}$ that contain a set $S' \in STS(c,C)$ for some $c \in C$.*

3. Such a κ exists because $STD^0(c,C)$ is finite and can hence be reduced only finitely often.
4. Note that we focus on sample size here, but neglect efficiency issues arising from the pre-computation of all subset teaching sets.

Note that Definition 16 does not presume any special order of the concept representations or of the instances, *i.e.*, teacher and learner do not have to agree on any such order to make use of the teaching and learning protocol. That means, given a special concept class C, the computation of its subset teaching sets does not involve any special coding trick depending on C—it just follows a general rule.

By definition, every subset teaching set protocol is collusion-free. However, teacher-learner pairs following a subset teaching set protocol are not necessarily valid in the sense of Goldman and Mathias's definition. This is easily seen for the concept class C_θ of all linear threshold functions over three instances x_1, x_2, x_3. This class has four concepts, namely $c_1 = \{x_1, x_2, x_3\}$, $c_2 = \{x_2, x_3\}$, $c_3 = \{x_3\}$, and $c_4 = \{\}$. It is easy to verify that $\{(x_1, -)\}$ is a subset teaching set for c_2 and is consistent with c_3. Similarly, $\{(x_3, +)\}$ is a subset teaching set for c_3 and is consistent with c_2. Hence $\{(x_1, -), (x_3, +)\}$ is consistent with both c_2 and c_3 and contains a subset teaching set for c_2 as well as a subset teaching set for c_3. Obviously, there exists a teacher-learner pair (τ, λ) satisfying the properties of an $STS - protocol$ for this class, such that $\tau(c_2) = \{(x_1, -)\}$, $\tau(c_3) = \{(x_3, +)\}$, and $\lambda(\{(x_1, -), (x_3, +)\}) = c_2$. However, there is no learner λ' such that (τ, λ') is a valid teacher-learner pair for C_θ. Such a learner λ' would have to hypothesize both c_2 and c_3 on input $\{(x_1, -), (x_3, +)\}$. See Table 1 for illustration of this example.

concept	x_1	x_2	x_3	STS^0	STS^1
$\{x_1, x_2, x_3\}$	+	+	+	$\{(x_1, +)\}$	$\{(x_1, +)\}$
$\{x_2, x_3\}$	−	+	+	$\{(x_1, -), (x_2, +)\}$	$\{(x_1, -)\}$, $\{(x_2, +)\}$
$\{x_3\}$	−	−	+	$\{(x_2, -), (x_3, +)\}$	$\{(x_2, -)\}$, $\{(x_3, +)\}$
$\{\}$	−	−	−	$\{(x_3, -)\}$	$\{(x_3, -)\}$

Table 1: Iterated subset teaching sets for the class C_θ.

5.2 Comparison to the Balbach teaching dimension

Obviously, when using the trivial adversary, Protocol 17 based on the subset teaching dimension never requires a sample larger than a teaching set; often a smaller sample is sufficient. However, compared to the Balbach teaching dimension, the subset teaching dimension is superior in some cases and inferior in others. The latter may seem unintuitive, but is possible because Balbach's teaching sets are not restricted to be subsets of the original teaching sets.

Theorem 18 *1. For each $u \in \mathbb{N}$ there is a concept class C such that $STD(C) = 1$ and $BTD(C) = u$.*

2. For each $u \geq 3$ there is a concept class C such that $BTD(C) = 3$ and $STD(C) = u$.

Proof. Assertion 1. Let $n = 2^u + u$ be the number of instances in X. Define a concept class $C = C^u_{\text{pair}}$ as follows. For every $s = (s_1, \ldots, s_u) \in \{+, -\}^u$, C contains the concepts $c_{s,0} = \{x_i \mid 1 \leq i \leq u \text{ and } s_i = +\}$ and $c_{s,1} = c_{s,0} \cup \{x_{u+1+int(s)}\}$. Here $int(s) \in \mathbb{N}$ is defined as the sum of all values 2^{u-i} for which $s_i = +$, $1 \leq i \leq u$. We claim that $STD(C) = 1$ and $BTD(C) = u$. See Table 2 for the case $u = 2$.

Let $s = (s_1, \ldots, s_u) \in \{+,-\}^u$. Then

$$\begin{aligned} TS(c_{s,0}, C) &= \{\{(x_i, s_i) \mid 1 \leq i \leq u\} \cup \{(x_{u+1+int(s)}, -)\}\} \\ \text{and } TS(c_{s,1}, C) &= \{\{(x_{u+1+int(s)}, +)\}\} \end{aligned}$$

Since for each $c \in C$ the minimal teaching set for c with respect to C contains an example that does not occur in the minimal teaching set for any other concept $c' \in C$, one obtains $STD(C) = 1$ in just one iteration.

In contrast to that, we obtain

$$\begin{aligned} BTD^0(c_{s,0}, C) &= u + 1\,, \\ BTD^1(c_{s,0}, C) &= u\,, \\ \text{and } BTD^0(c_{s,1}, C) &= 1 \text{ for all } s \in \{+,-\}^u\,. \end{aligned}$$

Consider any $s \in \{+,-\}^u$ and any sample $S \subseteq \{(x, c_{s,0}(x)) \mid x \in X\}$ with $|S| = u - 1$. Clearly there is some $s' \in \{+,-\}^u$ with $s' \neq s$ such that $c_{s',0} \in Cons(S, C)$. So $|Cons(S, C, +)| > 1$ and in particular $Cons(S, C, +) \neq \{c_{s,0}\}$. Hence $BTD^2(c_{s,0}, C) = BTD^1(c_{s,0}, C)$, which finally implies $BTD(C) = u$.

Assertion 2. Let $n = u + 1$ be the number of instances in X. Define a concept class $C = C^u_{1/2}$ as follows. For every $i, j \in \{1, \ldots, u+1\}$, C contains the concept $\{x_i\}$ and the concept $\{x_i, x_j\}$. See Table 3 for the case $u = 4$.

Then the only minimal teaching set for a singleton $\{x_i\}$ is the sample $S^i = \{(x, -) \mid x \neq x_i\}$ with $|S^i| = u$. The only minimal teaching set for a concept $\{x_i, x_j\}$ with $i \neq j$ is the sample $S^{i,j} = \{(x_i, +), (x_j, +)\}$.

On the one hand, every subset of every minimal teaching set for a concept $c \in C$ is contained in some minimal teaching set for some concept $c' \in C$ with $c \neq c'$. Thus $STS^k(c, C) = TS(c, C)$ for all $c \in C$ and all $k \in \mathbb{N}$. Hence $STD(C) = TD(C) = u$.

On the other hand, any sample S containing $(x_i, +)$ and two negative examples $(x_\alpha, -)$ and $(x_\beta, -)$ (where i, α, and β are pairwise distinct) is in $BTS(\{x_i\}, C)$. This holds because every other concept in C that is consistent with this sample is a concept containing two instances and thus has a teaching set of size smaller than 3 ($= |S|$). Thus $BTD(C) = 3$. □

5.3 Teaching monomials

This section provides an analysis of the *STD* for a more natural example, the monomials, showing that the very intuitive example given in the introduction is indeed what a cooperative teacher and learner in an *STS*-protocol would do. The main result is that the *STD* of the class of all monomials is 2, independent on the number m of variables, whereas its teaching dimension is exponential in m and its *BTD* is linear in m, cf. Balbach (2008).

Theorem 19 *Let $m \in \mathbb{N}$, $m \geq 2$ and C the class of all boolean functions over m variables that can be represented by a monomial. Then $STD(C) = 2$.*

Proof. Let $m \in \mathbb{N}$, $m \geq 2$ and $s = (s_1, \ldots, s_m)$, $s' = (s'_1, \ldots, s'_m)$ elements in $\{0,1\}^m$. Let $\Delta(s, s')$ denote the Hamming distance of s and s', *i.e.*, $\Delta(s, s') = \sum_{1 \leq i \leq m} |s(i) - s'(i)|$.

We distinguish the following types of monomials M over m variables.

concept	x_1	x_2	x_3	x_4	x_5	x_6	STS^0	STS^1
$\emptyset$	[−]	[−]	[−]	−	−	−	$\{(x_1,-),(x_2,-),(x_3,-)\}$	$\{(x_3,-)\}$
$\{x_3\}$	−	−	[+]	−	−	−	$\{(x_3,+)\}$	$\{(x_3,+)\}$
$\{x_2\}$	[−]	[+]	−	[−]	−	−	$\{(x_1,-),(x_2,+),(x_4,-)\}$	$\{(x_4,-)\}$
$\{x_2,x_4\}$	−	+	−	[+]	−	−	$\{(x_4,+)\}$	$\{(x_4,+)\}$
$\{x_1\}$	[+]	[−]	−	−	[−]	−	$\{(x_1,+),(x_2,-),(x_5,-)\}$	$\{(x_5,-)\}$
$\{x_1,x_5\}$	+	−	−	−	[+]	−	$\{(x_5,+)\}$	$\{(x_5,+)\}$
$\{x_1,x_2\}$	[+]	[+]	−	−	−	[−]	$\{(x_1,+),(x_2,+),(x_6,-)\}$	$\{(x_6,-)\}$
$\{x_1,x_2,x_6\}$	+	+	−	−	−	[+]	$\{(x_6,+)\}$	$\{(x_6,+)\}$

Table 2: Iterated subset teaching sets for the class C^u_{pair} with $u = 2$, where $C^u_{\text{pair}} = \{c_{--,0}, c_{--,1} \ldots, c_{++,0}, c_{++,1}\}$ with $c_{--,0} = \emptyset$, $c_{--,1} = \{x_3\}$, $c_{-+,0} = \{x_2\}$, $c_{-+,1} = \{x_2, x_4\}$, $c_{+-,0} = \{x_1\}$, $c_{+-,1} = \{x_1, x_5\}$, $c_{++,0} = \{x_1, x_2\}$, $c_{++,1} = \{x_1, x_2, x_6\}$. All labels contributing to minimal teaching sets are highlighted by square brackets.

concept	x_1	x_2	x_3	x_4	x_5	TS (equal to STS)
$\{x_1\}$	+	−	−	−	−	$\{(x_2,-),(x_3,-),(x_4,-),(x_5,-)\}$
$\{x_2\}$	−	+	−	−	−	$\{(x_1,-),(x_3,-),(x_4,-),(x_5,-)\}$
$\{x_3\}$	−	−	+	−	−	$\{(x_1,-),(x_2,-),(x_4,-),(x_5,-)\}$
$\{x_4\}$	−	−	−	+	−	$\{(x_1,-),(x_2,-),(x_3,-),(x_5,-)\}$
$\{x_5\}$	−	−	−	−	+	$\{(x_1,-),(x_2,-),(x_3,-),(x_4,-)\}$
$\{x_1,x_2\}$	+	+	−	−	−	$\{(x_1,+),(x_2,+)\}$
$\{x_1,x_3\}$	+	−	+	−	−	$\{(x_1,+),(x_3,+)\}$
$\{x_1,x_4\}$	+	−	−	+	−	$\{(x_1,+),(x_4,+)\}$
$\{x_1,x_5\}$	+	−	−	−	+	$\{(x_1,+),(x_5,+)\}$
$\{x_2,x_3\}$	−	+	+	−	−	$\{(x_2,+),(x_3,+)\}$
$\{x_2,x_4\}$	−	+	−	+	−	$\{(x_2,+),(x_4,+)\}$
$\{x_2,x_5\}$	−	+	−	−	+	$\{(x_2,+),(x_5,+)\}$
$\{x_3,x_4\}$	−	−	+	+	−	$\{(x_3,+),(x_4,+)\}$
$\{x_3,x_5\}$	−	−	+	−	+	$\{(x_3,+),(x_5,+)\}$
$\{x_4,x_5\}$	−	−	−	+	+	$\{(x_4,+),(x_5,+)\}$

Table 3: Iterated subset teaching sets for the class $C^u_{1/2}$ with $u = 4$.

Type 1: M is the empty monomial (i.e., the always true concept).

Type 2: M involves m variables, $M \not\equiv v_1 \wedge \overline{v_1}$.[5]

Type 3: M involves k variables, $1 \leq k < m$, $M \not\equiv v_1 \wedge \overline{v_1}$.

Type 4: M is contradictory, *i.e.*, $M \equiv v_1 \wedge \overline{v_1}$.

The following facts summarize some rather obvious properties of the corresponding minimal teaching sets for monomials (cf., *e.g.*, Balbach (2008), for more details).

Fact 1: Let M be of Type 1 and let $s, s' \in \{0,1\}^m$ such that $\Delta(s, s') = m$. Then $S = \{(s,+),(s',+)\}$ forms a minimal teaching set for M, *i.e.*, $S \in STS^0(M, C)$.

Fact 2: Let M be of Type 2 and let $s \in \{0,1\}^m$ be the unique assignment for which M evaluates positively. Moreover, let $s_1, \ldots, s_m \in \{0,1\}^m$ be the m unique assignments

5. The symbols $\equiv$ and $\not\equiv$ denote functional equivalence and semantic non-equivalence of boolean formulae, respectively.

with $\Delta(s, s_1) = \cdots = \Delta(s, s_m) = 1$. Then $S = \{(s, +), (s_1, -), \ldots, (s_m, -)\}$ forms the one and only minimal teaching set for M, *i.e.*, $S \in STS^0(M, C)$. (Note that any two negative examples in S have Hamming distance 2.)

Fact 3: Let M be of Type 3 and let $s \in \{0,1\}^m$ be one assignment for which M evaluates positively. Moreover, let $s' \in \{0,1\}^m$ be the unique assignment with $\Delta(s, s') = m - k$ for which M evaluates positively and let $s_1, \ldots, s_k \in \{0,1\}^m$ be the k unique assignments with $\Delta(s, s_1) = \cdots = \Delta(s, s_k) = 1$ for which M evaluates negatively. Then $S = \{(s, +), (s', +), (s_1, -), \ldots, (s_k, -)\}$ forms a minimal teaching set for M, *i.e.*, $S \in STS^0(M, C)$. (Note that any two negative examples in S have Hamming distance 2.)

Fact 4: Let M be of Type 4 and let $S = \{(s, -) \mid s \in \{0,1\}^m\}$. Then S forms the one and only minimal teaching set for M, *i.e.*, $S \in STS^0(M, C)$.

After the first iteration the following facts can be observed.

Fact 1(a): Let M be of Type 1 and let $S \in STS^0(M, C)$. Then $S \in STS^1(M, C)$.

This is due to the observation that any singleton subset $S' \subseteq S$ is a subset of a teaching set in $STS^0(M', C)$ for some M' of Type 2.

Fact 2(a): Let M be of Type 2 and let $S \in STS^0(M, C)$. Then $S \in STS^1(M, C)$.

This is due to the observation that any proper subset $S' \subset S$ is a subset of a teaching set in $STS^0(M', C)$ for some M' of Type 3, if S' contains one positive example, or for some M' of Type 4, otherwise.

Fact 3(a): Let M be of Type 3 and let $s \in \{0,1\}^m$ be one assignment for which M evaluates positively. Moreover, let $s' \in \{0,1\}^m$ be the unique assignment with $\Delta(s, s') = m - k$ for which M evaluates positively and let $S = \{(s, +), (s', +)\}$. Then $S \in STS^1(M, C)$.

This is due to the following observations: (i) S is not a subset of any teaching set S' in $STS^0(M', C)$ for some M' of Type 1, since the two positive examples in S' have Hamming distance m. (ii) S is obviously not a subset of any teaching set S' in $STS^0(M', C)$ for some $M' \not\equiv M$ of Type 3. (iii) Any sufficiently small "different" subset S' of some teaching set in $STS^0(M, C)$—*i.e.*, S' contains at most two examples, but not two positive examples—is a subset of any teaching set in $STS^0(M', C)$ for some M' of Type 2, if S' contains one positive example, or for some M' of Type 4, otherwise.

Fact 4(a): Let M be of Type 4 and let $s \in \{0,1\}^m$ be any assignment. Moreover, let $s' \in \{0,1\}^m$ be any assignment with $\Delta(s, s') \neq 2$ and let $S = \{(s, -), (s', -)\}$. Then $S \in STS^1(M, C)$.

This is due to the following observations: (i) S is not a subset of any teaching set S' in $STS^0(M', C)$ for some M' of Type 2 or of Type 3, since any two negative examples in S' have Hamming distance 2. (ii) Any sufficiently small "different" subset S' of the unique teaching set in $STS^0(M, C)$—*i.e.*, S' contains at most two negative examples, but two having Hamming distance 2—is a subset of a teaching set in $STS^0(M', C)$ for some M' of Type 2.

After the second iteration the following facts can be observed.

Fact 1(b): Let M be of Type 1 and let $S \in STS^1(M, C)$. Then $S \in STS^2(M, C)$.

This is due to the observation that any singleton subset $S' \subseteq S$ is a subset of a teaching set in $STS^1(M', C)$ for some M' of Type 2.

Fact 2(b): Let M be of Type 2 and let $s \in \{0,1\}^m$ be the unique assignment for which M evaluates positively. Moreover, let $s' \in \{0,1\}^m$ be any assignments with $\Delta(s, s') = 1$ and let $S = \{(s, +), (s', -)\}$. Then $S \in STS^2(M, C)$.

This is due to the following observations: (i) S is not a subset of any teaching set S' in $STS^1(M', C)$ for some M' of Type 1, of Type 3 or of Type 4, since none of these teaching sets contains one positive and one negative example. (ii) S is obviously not a subset of any teaching set S' in $STS^1(M', C)$ for some $M' \not\equiv M$ of Type 2. (iii) Any sufficiently small "different" subset S' of a teaching set in $STS^1(M, C)$—*i.e.*, S' contains at most two examples, but not a positive and a negative example—is a subset of a teaching set in $STS^1(M', C)$ for some M' of Type 3, if S' contains one positive example, or for some $M' \not\equiv M$ of Type 2, otherwise.

Fact 3(b): Let M be of Type 3 and let $S \in STS^1(M, C)$. Then $S \in STS^2(M, C)$.

This is due to the observation that any singleton subset $S' \subseteq S$ is a subset of a teaching set in $STS^1(M', C)$ for some M' of Type 2.

Fact 4(b): Let M be of Type 4 and let $S \in STS^1(M, C)$. Then $S \in STS^2(M, C)$.

This is due to the observation that any singleton subset $S' \subseteq S$ is a subset of a teaching set in $STS^1(M', C)$ for some monomial M' of Type 2.

Note at this point that, for any monomial M of any type, we have $STD^2(M, C) = 2$.

Finally, it is easily seen that $STD^3(M, C) = STD^2(M, C) = 2$ for all $M \in C$. □

For illustration of this proof in case $m = 2$ see Table 4.

A further simple example showing that the STD can be constant as compared to an exponential teaching dimension, this time with an STD of 1, is the following.

Let $C^m_{\vee DNF}$ contain all boolean functions over $m \geq 2$ variables that can be represented by a 2-term DNF of the form $v_1 \vee M$, where M is a monomial that contains, for each i with $2 \leq i \leq m$, either the literal v_i or the literal $\overline{v_i}$. Moreover, $C^m_{\vee DNF}$ contains the boolean function that can be represented by the monomial $M' \equiv v_1$.[6]

Theorem 20 *Let $m \in \mathbb{N}$, $m \geq 2$.*

1. $TD(C^m_{\vee DNF}) = 2^{m-1}$.

2. $STD(C^m_{\vee DNF}) = 1$.

Proof. Assertion 1. Let S be a sample that is consistent with M'. Assume that for some $s \in \{0,1\}^m$, the sample S does not contain the negative example $(s, -)$. Obviously, there is a 2-term DNF $D \equiv v_1 \vee M$ such that D is consistent with $S \cup \{(s, +)\}$ and $D \not\equiv M'$. Hence S is not a teaching set for M'. Since there are exactly 2^{m-1} 2-term DNFs that represent pairwise distinct functions in C, a teaching set for M' must contain at least 2^{m-1} examples.

Assertion 2. The proof is straightforward: Obviously, $TD(D, C) = 1$ for all $D \in C$ with $D \not\equiv M'$. In particular, $STD(D, C) = 1$ for all $D \in C$ with $D \not\equiv M'$. It remains to show that $STD(M', C) = 1$. For this it suffices to see that a minimal teaching set for M' in C must contain negative examples, while no minimal teaching set for any $D \in C$ with $D \not\equiv M'$ contains any negative examples. Hence $STD^2(M', C) = 1$ and thus $STD(M', C) = 1$. □

6. Here and in the proof of Theorem 20, as in the proof of Theorem 19, the symbol $\equiv$ denotes functional equivalence of boolean formulae.

monomial	00	01	10	11	STS^0	STS^1
v_1	−	−	+	+	{(10,+),(11,+),(00,-)} {(10,+),(11,+),(01,-)}	{(10,+),(11,+)}
$\overline{v_1}$	+	+	−	−	{(00,+),(01,+),(10,-)} {(00,+),(01,+),(11,-)}	{(00,+),(01,+)}
v_2	−	+	−	+	{(01,+),(11,+),(00,-)} {(01,+),(11,+),(10,-)}	{(01,+),(11,+)}
$\overline{v_2}$	+	−	+	−	{(00,+),(10,+),(01,-)} {(00,+),(10,+),(11,-)}	{(00,+),(10,+)}
$v_1 \wedge v_2$	−	−	−	+	{(11,+),(01,-),(10,-)}	{(11,+),(01,-),(10,-)}
$v_1 \wedge \overline{v_2}$	−	−	+	−	{(10,+),(00,-),(11,-)}	{(10,+),(00,-),(11,-)}
$\overline{v_1} \wedge v_2$	−	+	−	−	{(01,+),(00,-),(11,-)}	{(01,+),(00,-),(11,-)}
$\overline{v_1} \wedge \overline{v_2}$	+	−	−	−	{(00,+),(01,-),(10,-)}	{(00,+),(01,-),(10,-)}
$v_1 \wedge \overline{v_1}$	−	−	−	−	{(00,-),(01,-),(10,-),(11,-)}	{(00,-),(01,-)} {(00,-),(10,-)} {(01,-),(11,-)} {(10,-),(11,-)}
T	+	+	+	+	{(00,+),(11,+)} {(01,+),(10,+)}	{(00,+),(11,+)} {(01,+),(10,+)}

monomial	00	01	10	11	STS^2	STS^3
v_1	−	−	+	+	{(10,+),(11,+)}	{(10,+),(11,+)}
$\overline{v_1}$	+	+	−	−	{(00,+),(01,+)}	{(00,+),(01,+)}
v_2	−	+	−	+	{(01,+),(11,+)}	{(01,+),(11,+)}
$\overline{v_2}$	+	−	+	−	{(00,+),(10,+)}	{(00,+),(10,+)}
$v_1 \wedge v_2$	−	−	−	+	{(11,+),(01,-)} {(11,+),(10,-)}	{(11,+),(01,-)} {(11,+),(10,-)}
$v_1 \wedge \overline{v_2}$	−	−	+	−	{(10,+),(00,-)} {(10,+),(11,-)}	{(10,+),(00,-)} {(10,+),(11,-)}
$\overline{v_1} \wedge v_2$	−	+	−	−	{(01,+),(00,-)} {(01,+),(11,-)}	{(01,+),(00,-)} {(01,+),(11,-)}
$\overline{v_1} \wedge \overline{v_2}$	+	−	−	−	{(00,+),(01,-)} {(00,+),(10,-)}	{(00,+),(01,-)} {(00,+),(10,-)}
$v_1 \wedge \overline{v_1}$	−	−	−	−	{(00,-),(01,-)} {(00,-),(10,-)} {(01,-),(11,-)} {(10,-),(11,-)}	{(00,-),(01,-)} {(00,-),(10,-)} {(01,-),(11,-)} {(10,-),(11,-)}
T	+	+	+	+	{(00,+),(11,+)} {(01,+),(10,+)}	{(00,+),(11,+)} {(01,+),(10,+)}

Table 4: Iterated subset teaching sets for the class of all monomials over $m = 2$ variables. Here T denotes the empty monomial. For better readability, the instances (denoting the second through fifth columns) are written in the form of bit strings representing truth assignments to the two variables.

6. Why smaller classes can be harder to teach

Interpreting the subset teaching dimension as a measure of complexity of a concept class in terms of cooperative teaching and learning, we observe a fact that is worth discussing, namely the nonmonotonicity of this complexity notion, as stated by the following theorem.

Theorem 21 *There is a concept class C such that $STD(C') > STD(C)$ for some subclass $C' \subset C$.*

Proof. This is witnessed by the concept classes $C = C^u_{1/2} \cup \{\emptyset\}$ and its subclass $C' = C^u_{1/2}$ used in the proof of Theorem 18.2, for any $u > 2$ (see Table 3 and Table 5 for $u = 4$). $STD(C^u_{1/2} \cup \{\emptyset\}) = 2$ while $STD(C^u_{1/2}) = u$. □

concept	x_1	x_2	x_3	x_4	x_5	STS^0
$\emptyset$	−	−	−	−	−	$\{(x_1,-),(x_2,-),(x_3,-),(x_4,-),(x_5,-)\}$
$\{x_1\}$	+	−	−	−	−	$\{(x_1,+),(x_2,-),(x_3,-),(x_4,-),(x_5,-)\}$
$\{x_2\}$	−	+	−	−	−	$\{(x_1,-),(x_2,+),(x_3,-),(x_4,-),(x_5,-)\}$
$\{x_3\}$	−	−	+	−	−	$\{(x_1,-),(x_2,-),(x_3,+),(x_4,-),(x_5,-)\}$
$\{x_4\}$	−	−	−	+	−	$\{(x_1,-),(x_2,-),(x_3,-),(x_4,+),(x_5,-)\}$
$\{x_5\}$	−	−	−	−	+	$\{(x_1,-),(x_2,-),(x_3,-),(x_4,-),(x_5,+)\}$
$\{x_1,x_2\}$	+	+	−	−	−	$\{(x_1,+),(x_2,+)\}$
$\{x_1,x_3\}$	+	−	+	−	−	$\{(x_1,+),(x_3,+)\}$
$\{x_1,x_4\}$	+	−	−	+	−	$\{(x_1,+),(x_4,+)\}$
$\{x_1,x_5\}$	+	−	−	−	+	$\{(x_1,+),(x_5,+)\}$
$\{x_2,x_3\}$	−	+	+	−	−	$\{(x_2,+),(x_3,+)\}$
$\{x_2,x_4\}$	−	+	−	+	−	$\{(x_2,+),(x_4,+)\}$
$\{x_2,x_5\}$	−	+	−	−	+	$\{(x_2,+),(x_5,+)\}$
$\{x_3,x_4\}$	−	−	+	+	−	$\{(x_3,+),(x_4,+)\}$
$\{x_3,x_5\}$	−	−	+	−	+	$\{(x_3,+),(x_5,+)\}$
$\{x_4,x_5\}$	−	−	−	+	+	$\{(x_4,+),(x_5,+)\}$
concept	x_1	x_2	x_3	x_4	x_5	STS^1
$\emptyset$	−	−	−	−	−	$\{(x_1,-),(x_2,-),(x_3,-),(x_4,-),(x_5,-)\}$
$\{x_1\}$	+	−	−	−	−	$\{(x_1,+),(x_2,-)\},\ldots,\{(x_1,+),(x_5,-)\}$
$\{x_2\}$	−	+	−	−	−	$\{(x_1,-),(x_2,+)\},\ldots,\{(x_2,+),(x_5,-)\}$
$\{x_3\}$	−	−	+	−	−	$\{(x_1,-),(x_3,+)\},\ldots,\{(x_3,+),(x_5,-)\}$
$\{x_4\}$	−	−	−	+	−	$\{(x_1,-),(x_4,+)\},\ldots,\{(x_4,+),(x_5,-)\}$
$\{x_5\}$	−	−	−	−	+	$\{(x_1,-),(x_5,+)\},\ldots,\{(x_4,-),(x_5,+)\}$
$\{x_1,x_2\}$	+	+	−	−	−	$\{(x_1,+),(x_2,+)\}$
$\{x_1,x_3\}$	+	−	+	−	−	$\{(x_1,+),(x_3,+)\}$
$\{x_1,x_4\}$	+	−	−	+	−	$\{(x_1,+),(x_4,+)\}$
$\{x_1,x_5\}$	+	−	−	−	+	$\{(x_1,+),(x_5,+)\}$
$\{x_2,x_3\}$	−	+	+	−	−	$\{(x_2,+),(x_3,+)\}$
$\{x_2,x_4\}$	−	+	−	+	−	$\{(x_2,+),(x_4,+)\}$
$\{x_2,x_5\}$	−	+	−	−	+	$\{(x_2,+),(x_5,+)\}$
$\{x_3,x_4\}$	−	−	+	+	−	$\{(x_3,+),(x_4,+)\}$
$\{x_3,x_5\}$	−	−	+	−	+	$\{(x_3,+),(x_5,+)\}$
$\{x_4,x_5\}$	−	−	−	+	+	$\{(x_4,+),(x_5,+)\}$

Table 5: Iterated subset teaching sets for the class $C^u_{1/2} \cup \{\emptyset\}$ with $u = 4$; two iterations. In the third iteration, the sample for the empty concept (first row) will be reduced to all its subsets of size two, thus witnessing an STD of 2.

In contrast to that, it is not hard to show that BTD in fact is monotonic, see Theorem 22.

Theorem 22 *If C is a concept class and $C' \subseteq C$ a subclass of C, then $BTD(C') \leq BTD(C)$.*

Proof. Fix C and $C' \subseteq C$. We will prove by induction on k that

$$BTD^k(c, C') \leq BTD^k(c, C) \text{ for all } c \in C' \tag{1}$$

for all $k \in \mathbb{N}$.

$k = 0$: Property (1) holds because of $BTD^0(c, C') = TD(c, C') \leq TD(c, C) = BTD^0(c, C)$ for all $c \in C'$.

Induction hypothesis: assume (1) holds for a fixed k.

$k \rightsquigarrow k + 1$: First, observe that

$$\begin{aligned} Cons_{size}(S, C', k) &= \{c \in Cons(S, C') \mid BTD^k(c, C') \geq |S|\} \\ &\subseteq \{c \in Cons(S, C') \mid BTD^k(c, C) \geq |S|\} \text{ (ind. hyp.)} \\ &\subseteq \{c \in Cons(S, C) \mid BTD^k(c, C) \geq |S|\} \\ &= Cons_{size}(S, C, k) \end{aligned}$$

Second, for all $c \in C'$ we obtain

$$\begin{aligned} BTD^{k+1}(c, C') &= \min\{|S| \mid Cons_{size}(S, C', k) = \{c\}\ \} \\ &\leq \min\{|S| \mid Cons_{size}(S, C, k) = \{c\}\ \} \\ &\leq BTD^{k+1}(c, C) \end{aligned}$$

This completes the proof. □

6.1 Nonmonotonicity after elimination of redundant instances

Note that the nonmonotonicity of the subset teaching dimension holds with a fixed number of instances n. In fact, if n was not considered fixed then every concept class C' would have a superset C (via addition of instances) of lower subset teaching dimension. However, the same even holds for the teaching dimension itself which we yet consider monotonic since it is monotonic given fixed n. So whenever we speak of monotonicity we assume a fixed instance space X.

Of course such an instance space X might contain *redundant* instances the removal of which would not affect the subset teaching dimension and would retain a non-redundant subset of the set of all subset teaching sets. In the following subsection, where we discuss a possible intuition behind the nonmonotonicity of the STD, redundancy conditions on instances will actually play an important role and show the usefulness of the following technical discussion. However, it is not straightforward to impose a suitable redundancy condition characterizing when an instance can be removed.

We derive such a condition starting with a redundancy condition for the original variant of teaching sets. For that purpose we introduce the notion C^{-x} for the concept class resulting from C after removing the instance x from the instance space X. Here C is any concept class over X and $x \in X$ is any instance. For example, if $X = \{x_1, x_2, x_3\}$ and $C = \{\{x_1\}, \{x_1, x_2\}, \{x_2, x_3\}\}$ then

$$C^{-x_3} = \{\{x_1\}, \{x_1, x_2\}, \{x_2\}\}$$

considered over the instance space $\{x_1, x_2\}$.

To ease notation, we use a single name c for both a concept $c \in C$ and its corresponding concept in the class C^{-x} for any $x \in X$. It will always be clear from the context which concept is referred to.

Lemma 23 *Let C be a concept class over X and $x \in X$. Suppose for all $c \in C$ and for all $S \in TS(c, C)$*

$$(x, c(x)) \in S \Rightarrow \exists y \neq x \; [(S \setminus \{(x, c(x))\}) \cup \{(y, c(y))\} \in TS(c, C)] \, .$$

Then the following two assertions are true.

1. $|C^{-x}| = |C|$.
2. *For all $c \in C$ and for all samples S*

$$S \in TS(c, C^{-x}) \iff [S \in TS(c, C) \; \wedge \; (x, c(x)) \notin S] \, .$$

Proof. Assertion 1. Assume $|C^{-x}| < |C|$.

Then there must be two distinct concepts $c, c' \in C$ such that c and c' disagree only in x, *i.e.*, $c(y) = c'(y)$ for all $y \in X \setminus \{x\}$ and $c(x) \neq c'(x)$. Consequently, $(x, c(x))$ must be contained in some $S \in TS(c, C)$. By the premise of the lemma, this implies that there is some $y \in X \setminus \{x\}$ such that $(S \setminus \{(x, c(x))\}) \cup \{(y, c(y))\} \in TS(c, C)$. Hence $(S \setminus \{(x, c(x))\}) \cup \{(y, c(y))\}$ is a teaching set for c in C that does not contain $(x, c(x))$. However, $(S \setminus \{(x, c(x))\}) \cup \{(y, c(y))\}$ is consistent with c', which is a contradiction. Therefore $|C^{-x}| = |C|$.

Assertion 2. Let $c \in C$ be an arbitrary concept and let S be any sample over X.

First assume $S \in TS(c, C)$ and $(x, c(x)) \notin S$. By Assertion 1, $|C^{-x}| = |C|$ and therefore $TD(c, C^{-x}) \geq TD(c, C)$. Thus we immediately obtain $S \in TS(c, C^{-x})$.

Second assume $S \in TS(c, C^{-x})$. By definition, we have $(x, c(x)) \notin S$. Hence it remains to prove that $S \in TS(c, C)$. If $S \notin TS(c, C)$ then there exists some $T \in TS(c, C)$ such that $|T| < |S|$, because otherwise $|C^{-x}|$ would be smaller than $|C|$. We distinguish two cases.

Case 1. $(x, c(x)) \notin T$.

Then $T \in TS(c, C^{-x})$ in contradiction to the facts $S \in TS(c, C^{-x})$ and $|S| \neq |T|$.

Case 2. $(x, c(x)) \in T$.

Then by the premise of the lemma there exists a $y \neq x$ such that

$$A \stackrel{\text{def}}{=} (S \setminus \{(x, c(x))\}) \cup \{(y, c(y))\} \in TS(c, C) \, .$$

Since $(x, c(x)) \notin A$ we have $A \in TS(c, C^{-x})$ and $|A| = |T| \neq |S|$. This again contradicts $S \in TS(c, C^{-x})$.

Since both cases reveal a contradiction, we obtain $S \in TS(c, C)$. □

For illustration see Table 6. In this example the instances x_4 and x_5 meet the redundancy condition. After eliminating x_5, the instance x_4 still meets the condition and can be removed as well. The new representation of the concept class then involves only the instances x_1, x_2, x_3.

concept in C	x_1	x_2	x_3	x_4	x_5	TS
$\emptyset$	$-$	$-$	$-$	$-$	$-$	$\{(x_1,-),(x_3,-)\}$, $\{(x_1,-),(x_4,-)\}$, $\{(x_1,-),(x_5,-)\}$
$\{x_1\}$	$+$	$-$	$-$	$-$	$-$	$\{(x_1,+),(x_2,-)\}$, $\{(x_1,+),(x_5,-)\}$
$\{x_3,x_4,x_5\}$	$-$	$-$	$+$	$+$	$+$	$\{(x_2,-),(x_3,+)\}$, $\{(x_2,-),(x_4,+)\}$, $\{(x_2,-),(x_5,+)\}$
$\{x_2,x_3,x_4,x_5\}$	$-$	$+$	$+$	$+$	$+$	$\{(x_1,-),(x_2,+)\}$, $\{(x_2,+),(x_4,+)\}$
$\{x_1,x_2,x_5\}$	$+$	$+$	$-$	$-$	$+$	$\{(x_2,+),(x_3,-)\}$, $\{(x_3,-),(x_5,+)\}$
$\{x_1,x_2,x_3,x_5\}$	$+$	$+$	$+$	$-$	$+$	$\{(x_1,+),(x_3,+)\}$, $\{(x_3,+),(x_4,-)\}$

concept in $(C^{-x_5})^{-x_4}$	x_1	x_2	x_3	TS
$\emptyset$	$-$	$-$	$-$	$\{(x_1,-),(x_3,-)\}$
$\{x_1\}$	$+$	$-$	$-$	$\{(x_1,+),(x_2,-)\}$
$\{x_3\}$	$-$	$-$	$+$	$\{(x_2,-),(x_3,+)\}$
$\{x_2,x_3\}$	$-$	$+$	$+$	$\{(x_1,-),(x_2,+)\}$
$\{x_1,x_2\}$	$+$	$+$	$-$	$\{(x_2,+),(x_3,-)\}$
$\{x_1,x_2,x_3\}$	$+$	$+$	$+$	$\{(x_1,+),(x_3,+)\}$

Table 6: Teaching sets for a class C before and after elimination of two redundant instances.

Lemma 23 provides a condition on an instance x. If that instance is eliminated from the instance space then the resulting concept class C^{-x} not only has the same teaching dimension as C but, even more, for each of its concepts c the teaching sets are exactly those that are teaching sets for c in C and do not contain an example involving the eliminated instance x. Note that even though several instances might meet that condition at the same time, only one at a time may be removed. For the remaining instances it has to be checked whether the condition still holds after elimination of the first redundant instance.

In the example in Table 6, x_4 and x_5 are exactly those instances that could be eliminated without reducing the size of the concept class, *i.e.*,

$$|C| = |C^{-x_4}| = |C^{-x_5}| = |(C^{-x_4})^{-x_5}| = |(C^{-x_5})^{-x_4}|\,.$$

However, if we were to simply eliminate all instances x as long as $|C| = |C^{-x}|$, then the consequence of Lemma 23 would not necessarily be fulfilled any longer. For example, consider the concept class C in Table 7. Here $|C| = |C^{-x_1}|$, but removing x_1 from the instance space would increase the teaching dimension of c_1, namely $TD(c_1, C) = 1 < 2 = TD(c_1, C^{-x_1})$.

So one legitimate redundancy condition for instances—considering the preservation of teaching sets—is the one given in the premise of Lemma 23. This condition can be extended to a redundancy condition with respect to subset teaching sets.

Theorem 24 *Let C be a concept class over X and $x \in X$. Suppose for all $k \in \mathbb{N}$, for all $c \in C$, and for all $S \in STS^k(c, C)$*

$$(x, c(x)) \in S \Rightarrow \exists y \neq x\ [(S \setminus \{(x, c(x))\}) \cup \{(y, c(y))\} \in STS^k(c, C)]\,,$$

Then the following two assertions are true.

1. $|C^{-x}| = |C|$.

concept in C	x_1	x_2	x_3	TS
$c_1 = \{x_1, x_2, x_3\}$	$+$	$+$	$+$	$\{(x_1, +)\}$
$c_2 = \{x_2\}$	$-$	$+$	$-$	$\{(x_2, +), (x_3, -)\}$, $\{(x_2, +), (x_1, -)\}$
$c_3 = \{x_3\}$	$-$	$-$	$+$	$\{(x_3, +), (x_2, -)\}$, $\{(x_3, +), (x_1, -)\}$
$c_4 = \emptyset$	$-$	$-$	$-$	$\{(x_2, -), (x_3, -)\}$

concept in C^{-x_1}	x_2	x_3	TS
$c_1 = \{x_2, x_3\}$	$+$	$+$	$\{(x_2, +), (x_3, +)\}$
$c_2 = \{x_2\}$	$+$	$-$	$\{(x_2, +), (x_3, -)\}$
$c_3 = \{x_3\}$	$-$	$+$	$\{(x_3, +), (x_2, -)\}$
$c_4 = \emptyset$	$-$	$-$	$\{(x_2, -), (x_3, -)\}$

Table 7: Teaching sets for a class C before and after elimination of the instance x_1 not satisfying the premises of Lemma 23, despite fulfilling the property $|C| = |C^{-x_1}|$.

2. For all $k \in \mathbb{N}$, for all $c \in C$, and for all samples S

$$S \in STS^k(c, C^{-x}) \iff [S \in STS^k(c, C) \ \wedge \ (x, c(x)) \notin S] .$$

Proof. Assertion 1. This follows immediately by applying Lemma 23.1 for $k = 0$.

Assertion 2. We prove the second assertion by induction on k.

For $k = 0$ the assertion follows immediately from Lemma 23.2. So assume that the assertion is proven for some k (induction hypothesis). It remains to show that it then also holds for $k + 1$.

For that purpose note that

$$\forall c \in C \ \forall A \in STS^k(c, C) \ \exists B \in STS^k(c, C^{-x}) \ [|A| = |B| \ \wedge \ A \setminus \{(x, c(x))\} \subseteq B] \quad (*)$$

by combination of the induction hypothesis with the premise of the theorem.

Choose an arbitrary $c \in C$.

First assume $S \in STS^{k+1}(c, C)$ and $(x, c(x)) \notin S$. By the definition of subset teaching sets, there is an $S' \in STS^k(c, C)$ such that

$$S \subseteq S' . \tag{2}$$

Using $(*)$ we can assume without loss of generality that

$$S' \in STS^k(c, C^{-x}) . \tag{3}$$

Moreover, again by the definition of subset teaching sets, one obtains $S \not\subseteq S''$ for every $S'' \in STS^k(c', C)$ with $c' \neq c$. The induction hypothesis then implies

$$S \not\subseteq S'' \text{ for every } S'' \in STS^k(c', C^{-x}) \text{ with } c' \neq c . \tag{4}$$

Due to (2), (3), (4) we get either $S \in STS^{k+1}(c, C^{-x})$ or $|S| > STD^{k+1}(c, C^{-x})$. In the latter case there would be a set $T \in STS^{k+1}(c, C^{-x})$ such that $|T| < |S|$. T is a subset of some set in $STS^k(c, C^{-x})$ and thus also of some set in $STS^k(c, C)$ by induction hypothesis.

If T was contained in some $T' \in STS^k(c', C)$ for some $c' \neq c$ then we could again assume without loss of generality, using $(*)$ and $(x, c(x)) \notin T$, that T is contained in some set in $STS^k(c', C^{-x})$—in contradiction to $T \in STS^{k+1}(c, C^{-x})$. Therefore $T \in STS^{k+1}(c, C)$ and so $|T| = |S|$—a contradiction. This implies $S \in STS^{k+1}(c, C^{-x})$.

Second assume that $S \in STS^{k+1}(c, C^{-x})$. Obviously, $(x, c(x)) \notin S$, so that it remains to show $S \in STS^{k+1}(c, C)$.

Because of $S \in STS^{k+1}(c, C^{-x})$ there exists some set $S' \in STS^k(c, C^{-x})$ such that

$$S \subseteq S' . \tag{5}$$

The induction hypothesis implies

$$S' \in STS^k(c, C) . \tag{6}$$

Moreover, by the definition of subset teaching sets, one obtains $S \not\subseteq S''$ for every $S'' \in STS^k(c', C^{-x})$ with $c' \neq c$. If there was a set $S'' \in STS^k(c', C)$ such that $c' \neq c$ and $S \subseteq S''$ then $(*)$ would imply that without loss of generality $S'' \in STS^k(c', C^{-x})$. So we have

$$S \not\subseteq S'' \text{ for every } S'' \in STS^k(c', C) \text{ with } c' \neq c . \tag{7}$$

Combining (5), (6), (7) we get either $S \in STS^{k+1}(c, C)$ or $|S| > STD^{k+1}(c, C)$. In the latter case there would be a set $T \in STS^{k+1}(c, C)$ such that $|T| < |S|$. T is a subset of some set $T' \in STS^k(c, C)$. We can assume without loss of generality, using $(*)$, that $T' \in STS^k(c, C^{-x})$. If T was contained in some set in $STS^k(c', C^{-x})$ for some $c' \neq c$ then by induction hypothesis T would be contained in some set in $STS^k(c', C)$ for some $c' \neq c$. This is a contradiction to $T \in STS^{k+1}(c, C)$. So $T \in STS^{k+1}(c, C^{-x})$ and hence $|T| = |S|$—a contradiction. Thus $S \in STS^{k+1}(c, C)$. □

The example in Table 7 illustrates that eliminating instances x satisfying $|C^{-x}| = |C|$, without any additional constraints, can actually change the subset teaching dimension of a class. In the given example, the subset teaching dimension of C is 1, while the subset teaching dimension of C^{-x_1} is 2. The stronger condition on the instance x in the premise of Theorem 24 guarantees that eliminating x does not change the subset teaching dimension.

6.2 Nonmonotonicity and the role of nearest neighbours

From a general point of view, it is not obvious how to explain why a teaching dimension resulting from a cooperative model should be nonmonotonic.

First of all, this is a counter-intuitive observation when considering *STD* as a notion of complexity—intuitively any subclass of C should be at most as complex for teaching and learning as C.

However, there is in fact an intuitive explanation for the nonmonotonicity of the complexity in cooperative teaching and learning: when teaching $c \in C$, instead of providing examples that eliminate all concepts in $C \setminus \{c\}$ (as is the idea underlying minimal teaching sets) cooperative teachers would rather pick only those examples that distinguish c from its "most similar" concepts in C. Similarity here is measured by the number of instances on which two concepts agree (*i.e.*, dissimilarity is given by the Hamming distance between

the concepts, where a concept c is represented as a bit vector $(c(x_1), \ldots, c(x_n)))$. This is reflected in the subset teaching sets in all illustrative examples considered above.

Considering a class $C = C^u_{\text{pair}}$ (see the proof of Theorem 18.1), one observes that a subset teaching set for a concept $c_{s,0}$ contains only the negative example $(x_{u+1+int(s)}, -)$ distinguishing it from $c_{s,1}$ (its nearest neighbor in terms of Hamming distance). A learner will recognize this example as the one that separates only that one pair $(c_{s,0}, c_{s,1})$ of nearest neighbors. In contrast to that, if we consider only the subclass $C' = \{c_{s,0} \mid s \in \{0,1\}^u\}$, the nearest neighbors of each $c_{s,0}$ are different ones, and every single example separating one nearest neighbor pair also separates other nearest neighbor pairs. Thus no single example can be recognized by the learner as a separating example for one unique pair of concepts.

This intuitive idea of subset teaching sets being used for distinguishing a concept from its nearest neighbors has to be treated with care though. The reason is that the concept class may contain "redundant" instances, *i.e.*, instances that could be removed from the instance space according to Theorem 24.

Such redundant instances might on the other hand affect Hamming distances and nearest neighbor relations. Only after their elimination does the notion of nearest neighbors in terms of Hamming distance become useful. Consider for instance Table 6. In the concept class C over 5 instances the only nearest neighbor of $\emptyset$ is $\{x_1\}$ and an example distinguishing $\emptyset$ from $\{x_1\}$ would be $(x_1, -)$. Moreover, no other concept is distinguished from its nearest neighbors by the instance x_1. According to the intuition explained here, this would suggest $\{(x_1, -)\}$ being a subset teaching set for $\emptyset$ although the subset teaching sets here equal the teaching sets and are all of cardinality 2.

After instance elimination of x_4, x_5 there is only one subset teaching set for $\emptyset$, namely $\{(x_1, -), (x_3, -)\}$. This is still of cardinality 2 but note that now $\emptyset$ has two nearest neighbors, namely $\{x_1\}$ and $\{x_3\}$. The two examples in the subset teaching set are those that distinguish $\emptyset$ from its nearest neighbors. Note that either one of these two examples is not unique as an example used for distinguishing a concept from its nearest neighbors: $(x_1, -)$ would be used by $\{x_2, x_3\}$ for distinguishing itself from its nearest neighbor $\{x_1, x_2, x_3\}$, and $(x_3, -)$ would be used by $\{x_1, x_2\}$ for distinguishing itself from its nearest neighbor $\{x_1, x_2, x_3\}$. So the subset teaching set for $\emptyset$ has to contain both examples.

This illustrates why a subclass of a class C can have a higher complexity than C if crucial nearest neighbors of some concepts are missing in the subclass.

To summarize,

- nonmonotonicity has an intuitive reason and is not an indication of an ill-defined version of the teaching dimension,

- nonmonotonicity would in fact be a consequence of implementing the idea that the existence of specific concepts (*e.g.*, nearest neighbours) associated with a target concept is beneficial for teaching and learning.

So, the STD captures certain intuitions about teaching and learning that monotonic dimensions *cannot* capture; at the same time monotonicity might in other respects itself be an intuitive property of teaching and learning which then the STD cannot capture.

In particular there are two underlying intuitive properties that seem to not be satisfiable by a single variant of the teaching dimension.

7. The recursive teaching dimension

On the one hand, we have the teaching framework based on the subset teaching dimension which results in a nonmonotonic dimension, and on the other hand we have a monotonic dimension in the *BTD* framework, which unfortunately does not always meet our idea of a cooperative teaching and learning protocol. That raises the question whether nonmonotonicity is necessary to achieve certain positive results. In fact, the nonmonotonicity concerning the class C^u_{pair} is not counter-intuitive, but would a dimension that is monotonic also result in a worse sample complexity than the *STD* in general, such as, *e.g.*, for the monomials?

In other words, is there a teaching/learning framework

- resulting in a monotonic variant of a teaching dimension and
- achieving low teaching complexity results similar to the subset teaching dimension?

At this point of course it is difficult to define what "similar to the subset teaching dimension" means. However, we would like to have a constant dimension for the class of all monomials, as well as, *e.g.*, a teaching set of size 1 for the empty concept in our often used concept class C_0.

We will now via several steps introduce a monotonic variant of the teaching dimension and show that for most of the examples studied above, it is as low as the subset teaching dimension. General comparisons will be made in Section 8, in particular in order to show that this new framework is uniformly at least as efficient as the *BTD* framework, while sometimes being less efficient than the *STD* framework. This reflects to a certain extent that monotonicity constraints might affect sample efficiency.

7.1 The model

We will first define our new variant of teaching dimension and show its monotonicity.

The nonmonotonicity of *STD* is caused by considering every STS^k-set for every concept when computing an STS^{k+1}-set for a single concept. Hence the idea in the following approach is to impose a canonical order on the concept class, in terms of the "teaching complexity" of the concepts. This is what the teaching dimension does as well, but our design principle is a recursive one. After selecting a set of concepts each of which is "easy to teach" because of possessing a small minimal teaching set, we eliminate these concepts from our concept class and consider only the remaining concepts. Again we determine those with the lowest teaching dimension, now however measured with respect to the class of remaining concepts, and so on. The resulting notion of dimension is therefore called the *recursive teaching dimension*.

Definition 25 *Let C be a concept class. The teaching hierarchy for C is the sequence $H = ((C_1, d_1), \ldots, (C_h, d_h))$ that fulfills, for all $j \in \{1, \ldots, h\}$,*

$$C_j = \{c \in \overline{C}_j \mid d_j = TD(c, \overline{C}_j) \leq TD(c', \overline{C}_j) \text{ for all } c' \in \overline{C}_j\},$$

where $\overline{C}_1 = C$ and $\overline{C}_{i+1} = C \setminus (C_1 \cup \ldots \cup C_i)$ for all $i \in \{1, \ldots, h-1\}$.

For any $j \in \{1, \ldots, h\}$ and any $c \in C_j$, a sample $S \in TS(c, \overline{C}_j)$ is called a recursive teaching set for c in C. The recursive teaching dimension $RTD(c, C)$ of c in C is then defined as $RTD(c, C) = d_j$ and we denote by $RTS(c, C) = TS(c, \overline{C}_j)$ the set of all recursive teaching sets for c in C.

The recursive teaching dimension $RTD(C)$ of C is defined by

$$RTD(C) = \max\{d_j \mid 1 \leq j \leq h\}.$$

The desired monotonicity property, see Proposition 26, follows immediately from the definition.

Proposition 26 *If C is a concept class and $C' \subseteq C$ is a subclass of C, then $RTD(C') \leq RTD(C)$.*

The definition of teaching hierarchy induces a protocol for teaching and learning: for a target concept c, a teacher uses the teaching hierarchy $H = ((C_1, d_1), \ldots, (C_h, d_h))$ for C to determine the unique index j with $c \in C_j$. The teacher then presents the examples in a teaching set from $TS(c, \overline{C}_j)$, *i.e.*, a recursive teaching set for c in C, to the learner. The learner will use the teaching hierarchy to determine the target concept from the sample provided by the teacher.

Protocol 27 *Let P be a protocol. P is called a recursive teaching set protocol (RTS-protocol for short) if the following two properties hold for every $C \subseteq \mathcal{C}$, where $P(C) = (\tau, \lambda)$.*

1. *$\tau(c) \in RTS(c, C)$ for all $c \in C$,*
2. *$\lambda(S) \in \{c \mid$ there is some $S' \in RTS(c, C)$ such that $S' \subseteq S\}$ for all $S \in \mathcal{S}$ that contain a set $S' \in RTS(c, C)$ for some $c \in C$.*

Note again that Definition 25 does not presume any special order of the concept representations or of the instances, *i.e.*, teacher and learner do not have to agree on any such order to make use of the teaching and learning protocol. The partial order resulting from the teaching hierarchy is still well-defined.

The following definition of canonical teaching plans yields an alternative definition of the recursive teaching dimension.

Definition 28 *Let C be a concept class, $|C| = z$. A teaching plan for C is a sequence $p = ((c_1, S_1), \ldots, (c_z, S_z)) \in (C \times 2^{X \times \{0,1\}})^z$ such that*

1. *$C = \{c_1, \ldots, c_z\}$.*
2. *$S_j \in TS(c_j, \{c_j, \ldots, c_z\})$ for $1 \leq j \leq z$.*

The order of p is given by $ord(p) = \max\{|S_j| \mid 1 \leq j \leq z\}$.

p is called a canonical teaching plan for C, if for any $i, j \in \{1, \ldots, z\}$:

$$i < j \Rightarrow TD(c_i, \{c_i, \ldots, c_z\}) \leq TD(c_j, \{c_i, \ldots, c_z\}).$$

Note that every concept class has a canonical teaching plan. It turns out that a canonical teaching plan has the lowest possible order over all teaching plans; this order coincides with the recursive teaching dimension, see Theorem 29.

Theorem 29 *Let C be a concept class and p^* a canonical teaching plan for C. Then $ord(p^*) = \min\{ord(p) \mid p \text{ is a teaching plan for } C\} = RTD(C)$.*

Proof. Let C and p^* as in the theorem be given, $p^* = ((c_1, S_1), \ldots, (c_z, S_z))$. $ord(p^*) = RTD(C)$ follows by definition. It needs to be shown that

$$ord(p^*) = \min\{ord(p) \mid p \text{ is a teaching plan for } C\}\,.$$

Let $p' = ((c'_1, S'_1), \ldots, (c'_z, S'_z))$ be any teaching plan for C. It remains to prove that $ord(p^*) \leq ord(p')$.

For that purpose choose the minimal $j \in \{1, \ldots, z\}$ such that $|S_j| = ord(p^*)$. By definition of a teaching plan, $TD(c_j, \{c_j, \ldots, c_z\}) = ord(p^*)$. Let $i \in \{1, \ldots, z\}$ be minimal such that $c'_i \in \{c_j, \ldots, c_z\}$. Let $k \in \{1, \ldots, z\}$ fulfill $c_k = c'_i$. By definition of a canonical teaching plan, $TD(c_k, \{c_j, \ldots, c_z\}) \geq TD(c_j, \{c_j, \ldots, c_z\}) = ord(p^*)$. This obviously yields $ord(p') \geq TD(c'_i, \{c'_i, \ldots, c'_z\}) \geq TD(c_k, \{c_j, \ldots, c_z\}) \geq ord(p^*)$. □

To summarize briefly, the recursive teaching dimension is a monotonic complexity notion which in fact has got some of the properties we desired; *e.g.*, it is easily verified that $RTD(C_0) = 1$ (by any teaching plan in which the empty concept occurs last) and that the RTD of the class of all monomials equals 2 (see below). Thus the RTD overcomes some of the weaknesses of BTD, while at the same time preserving monotonicity.

Interestingly, unlike for subset teaching set protocols, the teacher-learner pairs based on recursive teaching set protocols are valid in the sense of Goldman and Mathias's definition Goldman and Mathias (1996). This is an immediate consequence of the following theorem.

Theorem 30 *Let C be any concept class and $c \in C$. Let S be any sample. If S is consistent with c and there is some $T \in RTS(c, C)$ such that $T \subseteq S$ then there is no concept $c' \in Cons(S, C)$ with $c' \neq c$ and $T' \subseteq S$ for some $T' \in RTS(c', C)$.*

Proof. Let C, c, S, and T as in the theorem be given. Let $H = ((C_1, d_1), \ldots, (C_h, d_h))$ be the teaching hierarchy for C and let $i \in \{1, \ldots, h\}$ be such that $c \in C_i$.

Assume there was a concept $c' \in Cons(S, C)$ with $c' \neq c$ and $T' \subseteq S$ for some $T' \in RTS(c', C)$. Let $j \in \{1, \ldots, h\}$ be such that $c' \in C_j$.

Since c is consistent with S and S contains the recursive teaching set T' for c', c is also consistent with T'. As $c \in C_i$ is consistent with a recursive teaching set for $c' \in C_j$, we obtain $j > i$.

Similarly, since c' is consistent with S and S contains the recursive teaching set T for c, we obtain $i > j$.

This is clearly a contradiction. Hence there is no concept $c' \in Cons(S, C)$ with $c' \neq c$ and $T' \subseteq S$ for some $T' \in RTS(c', C)$. □

7.2 Comparison to the Balbach teaching dimension

Unlike the subset teaching dimension, the recursive teaching dimension lower-bounds the Balbach dimension. To prove this, we first observe that the smallest teaching dimension of all concepts in a given concept class is a lower bound on the Balbach dimension. This is stated formally in the following lemma.

Lemma 31 *Let C be a concept class. Then $BTD(C) \geq \min\{ TD(c,C) \mid c \in C\}$.*

Proof. Let $u = \min\{ TD(c,C) \mid c \in C\}$. To show that $BTD(C) \geq u$, we will prove by induction on k that $u \leq BTD^k(c,C)$ for all $k \in \mathbb{N}$ for all $c \in C$.

$k = 0$: $BTD^0(c,C) = TD(c,C) \geq u$ for all $c \in C$.

Induction hypothesis: assume $u \leq BTD^k(c,C)$ for all $c \in C$ holds for a fixed k.

$k \rightsquigarrow k+1$: Suppose by way of contradiction that there is a concept $c^* \in C$ such that $u > BTD^{k+1}(c^*,C)$. In particular, there exists a sample S^* such that $|S^*| < u$ and $Cons_{size}(S^*,C,k) = \{c^*\}$.

By induction hypothesis, the set $Cons_{size}(S^*,C,k)$ defined by $\{c \in Cons(S^*,C) \mid BTD^k(c,C) \geq |S^*|\}$ is equal to $Cons(S^*,C)$. Note that $TD(c,C) \geq u$ for all $c \in C$ implies either $|Cons(S^*,C)| \geq 2$ or $Cons(S^*,C) = \emptyset$. We obtain a contradiction to $Cons_{size}(S^*,C,k) = \{c^*\}$.

This completes the proof. □

This lemma helps to prove that the recursive teaching dimension cannot exceed the Balbach dimension.

Theorem 32 *1. If C is a concept class then $RTD(C) \leq BTD(C)$.*

2. There is a concept class C such that $RTD(C) < BTD(C)$.

Proof. Assertion 1. To prove this assertion, let C be a concept class such that $RTD(C) = u$. By Theorem 29 there is a canonical teaching plan $p = ((c_1,S_1),\ldots,(c_z,S_z))$ for C such that $ord(p) = u$. Fix $j \leq \mathbb{N}$ minimal such that $|S_j| = u$ and define $C' = \{c_j \ldots, c_z\}$. Obviously, $RTD(C') = u$. Moreover, using Theorem 22, $BTD(C') \leq BTD(C)$. Thus it suffices to prove $u \leq BTD(C')$. This follows from Lemma 31, since $u = \min\{ TD(c,C') \mid c \in C'\}$.

This completes the proof of Assertion 1.

Assertion 2. The second assertion is witnessed by the concept class C_0 containing the empty concept and all singletons. Obviously, $RTD(C_0) = 1$ and $BTD(C_0) = 2$. □

7.3 Teaching monomials

In this subsection, we pick up the two examples from Subsection 5.3 again, this time in order to determine the recursive teaching dimension of the corresponding classes of concepts represented by boolean functions. As in the case of the subset teaching dimension, see Theorem 19, we obtain that the recursive teaching dimension of the class of all monomials over m ($m \geq 2$) variables is 2, independent of m.

Theorem 33 *Let $m \in \mathbb{N}$, $m \geq 2$, and C the class of all boolean functions over m variables that can be represented by a monomial. Then $RTD(C) = 2$.*

Proof. Fix m and C. For all $i \in \{0, \ldots, m\}$ let C^i be the subclass of all $c \in C$ that can be represented by a non-contradictory monomial M that involves i variables. There is exactly one concept in C not belonging to any subclass C^i of C, namely the concept c^* representable by a contradictory monomial.

The proof is based on the following observation.

Observation. For any $i \in \{0, \ldots, m\}$ and any $c \in C^i$: $TD(c, C' \cup \{c^*\}) \leq 2$, where $C' = \bigcup_{i \leq j \leq m} C^j$.

Now it is easily seen that $ord(p) \leq 2$ for every teaching plan $p = ((c_1, S_1), \ldots, (c_z, S_z))$ for C that meets the following requirements:

(a) $c_1 \in C^0$ and $c_z = c^*$.

(b) For any $k, k' \in \{1, \ldots, z-1\}$: If $k < k'$, then $c_k \in C^i$ and $c_{k'} \in C^j$ for some $i, j \in \{0, \ldots, m\}$ with $i \leq j$.

Therefore $RTD(C) \leq 2$.
Since obviously $TD(c, C) \geq 2$ for all $c \in C$, we obtain $RTD(C) = 2$.
(For illustration of the case $m = 2$ see Table 8.) □

monomial	subclass	00	01	10	11	*RTS*
T	C^0	+	+	+	+	{(00,+),(11,+)}
v_1	C^1	−	−	+	+	{(10,+),(11,+)}
$\overline{v_1}$	C^1	+	+	−	−	{(00,+),(01,+)}
v_2	C^1	−	+	−	+	{(01,+),(11,+)}
$\overline{v_2}$	C^1	+	−	+	−	{(00,+),(10,+)}
$v_1 \wedge v_2$	C^2	−	−	−	+	{(11,+)}
$v_1 \wedge \overline{v_2}$	C^2	−	−	+	−	{(10,+)}
$\overline{v_1} \wedge v_2$	C^2	−	+	−	−	{(01,+)}
$\overline{v_1} \wedge \overline{v_2}$	C^2	+	−	−	−	{(00,+)}
$v_1 \wedge \overline{v_1}$		−	−	−	−	{}

Table 8: Recursive teaching sets in the teaching hierarchy (corresponding to teaching plans of order 2) for the class of all monomials over $m = 2$ variables. T denotes the empty monomial. For better readability, the instances (denoting the third through sixth columns) are written in the form of bit strings representing truth assignments to the two variables.

For the sake of completeness, note that $RTD(C^m_{\vee DNF}) = 1$ where $C^m_{\vee DNF}$ is the class of boolean functions over m variables as defined in Subsection 5.3.

Theorem 34 *$RTD(C^m_{\vee DNF}) = 1$ for all $m \in \mathbb{N}$, $m \geq 2$.*

Proof. This follows straightforwardly from the fact that $TD(c, C^m_{\vee DNF}) = 1$ for every concept c corresponding to a 2-term DNF of form $v_1 \vee M$.

(For illustration see Table 4.) □

8. Subset teaching dimension versus recursive teaching dimension

Comparing the *STD* to the *RTD* turns out to be a bit more complex. We can show that the recursive teaching dimension can be arbitrarily larger than the subset teaching dimension; it can even be larger than the maximal *STD* computed over all subsets of the concept class.

Theorem 35 *1. For each $u \in \mathbb{N}$ there is a concept class C such that $STD(C) = 1$ and $RTD(C) = u$.*

2. There is a concept class C such that $\max\{STD(C') \mid C' \subseteq C\} < RTD(C)$.

Proof. Assertion 1. This is witnessed by the classes C^u_{pair} defined in the proof of Theorem 18.1.

Assertion 2. To verify Assertion 2, consider the concept class $C = \{c_1, \ldots, c_6\}$ given by $c_1 = \emptyset$, $c_2 = \{x_1\}$, $c_3 = \{x_1, x_2\}$, $c_4 = \{x_2, x_3\}$, $c_5 = \{x_2, x_4\}$, $c_6 = \{x_2, x_3, x_4\}$. It is not hard to verify that $TD(c, C) = 2$ for all $c \in C$ and thus $ord(p) = 2$ for every teaching plan p for C. Therefore $RTD(C) = 2$. Moreover $STD(C') = 1$ for all $C' \subseteq C$ (the computation of $STD(C)$ is shown in Table 9; further details are omitted). □

concept	x_1	x_2	x_3	x_4	STS^0	STS^1	STS^2
$\emptyset$	$-$	$-$	$-$	$-$	$\{(x_1,-),(x_2,-)\}$	$\{(x_1,-)\}$	$\{(x_1,-)\}$
$\{x_1\}$	$+$	$-$	$-$	$-$	$\{(x_1,+),(x_2,-)\}$	$\{(x_1,+),(x_2,-)\}$	$\{(x_1,+)\}$ $\{(x_2,-)\}$
$\{x_1,x_2\}$	$+$	$+$	$-$	$-$	$\{(x_1,+),(x_2,+)\}$	$\{(x_2,+)\}$	$\{(x_2,+)\}$
$\{x_2,x_3\}$	$-$	$+$	$+$	$-$	$\{(x_3,+),(x_4,-)\}$	$\{(x_4,-)\}$	$\{(x_4,-)\}$
$\{x_2,x_4\}$	$-$	$+$	$-$	$+$	$\{(x_3,-),(x_4,+)\}$	$\{(x_3,-)\}$	$\{(x_3,-)\}$
$\{x_2,x_3,x_4\}$	$-$	$+$	$+$	$+$	$\{(x_3,+),(x_4,+)\}$	$\{(x_3,+),(x_4,+)\}$	$\{(x_3,+)\}$ $\{(x_4,+)\}$

Table 9: Iterated subset teaching sets for the class $C = \{c_1, \ldots, c_6\}$ given by $c_1 = \emptyset$, $c_2 = \{x_1\}$, $c_3 = \{x_1, x_2\}$, $c_4 = \{x_2, x_3\}$, $c_5 = \{x_2, x_4\}$, $c_6 = \{x_2, x_3, x_4\}$.

Similarly, we can prove that the subset teaching dimension can be arbitrarily larger than the recursive teaching dimension.

Theorem 36 *For each $u \geq 2$ there is a concept class C such that $RTD(C) = 2$ and $STD(C) = u$.*

Proof. This is witnessed by the class $C = C^u_{1/2}$ used in the proof of Theorem 18.2. □

Due to the incomparability of *STD* and *RTD* it seems worth analyzing conditions under which they become comparable. To this end, we define a property that is sufficient for a concept class to have a recursive teaching dimension equal to its subset teaching dimension.

Definition 37 (Subset Teaching Property) *Let C be a concept class. C fulfills the Subset Teaching Property if for every teaching plan $p = ((c_1, S_1), \ldots, (c_z, S_z))$ for C with $ord(p) = RTD(C)$ and every j with $ord(p) = |S_j|$ and $STD(c_j, C) \geq ord(p)$ there exists a teaching plan*

$$p' = ((c_1, S'_1), \ldots, (c_z, S'_z))$$

for C and a sample $S \in STS(c_j, C)$ such that $S'_j \subseteq S$ and $|S'_i| = |S_i|$ for all i.

Theorem 38 *Let C be a concept class with the Subset Teaching Property. Then $STD(C) = RTD(C)$.*

Proof. $STD(C) \geq RTD(C)$ follows trivially: if $p = ((c_1, S_1), \ldots, (c_z, S_z))$ is a teaching plan for C with $ord(p) = RTD(C)$ and if j fulfils $RTD(C) = ord(p) = |S_j|$, then there is some S'_j with $|S'_j| = |S_j|$ and some $S \in STS(c_j, C)$ such that $S'_j \subseteq S$; hence $STD(C) \geq |S| \geq |S'_j| = |S_j| = ord(p)$.

In order to show that $STD(C) \leq RTD(C)$, we prove property (P_j) for all $j \in \{1, \ldots, |C|\}$. The proof is done by induction on j.

> (P_j):
> If $p = ((c_1, S_1), \ldots, (c_z, S_z))$ is a teaching plan for C with $ord(p) = RTD(C)$ then $STD(c_j, C) \leq ord(p)$.

For $j = 1$ this is obvious, because

$$STD(c_1, C) \leq TD(c_1, C) \leq ord(p)$$

for any teaching plan $p = ((c_1, S_1), \ldots, (c_z, S_z))$ for C.

The induction hypothesis is that (P_i) holds for all $i \leq j$, j fixed.

To prove (P_{j+1}), let $p = ((c_1, S_1), \ldots, (c_z, S_z))$ be any teaching plan for C with $ord(p) = RTD(C)$. Consider the $(j+1)^{st}$ concept c_{j+1} in p.

Case 1. $|S_{j+1}| < ord(p)$.

In this case we swap c_j and c_{j+1} and get a new teaching plan

$$\begin{aligned} p' \quad = \quad & ((c_1, S_1), \ldots, (c_{j-1}, S_{j-1}), \\ & (c_{j+1}, T), (c_j, T'), \ldots, (c_z, S_z)) \end{aligned}$$

for C.

Note that $|T'| \leq |S_j|$. Moreover, $|T| \leq |S_{j+1}| + 1 \leq ord(p) = RTD(C)$. Hence $ord(p') = RTD(C)$.

Now c_{j+1} is in j^{th} position in the teaching plan p' whose order is equal to $RTD(C)$. By induction hypothesis we get $STD(c_{j+1}, C) \leq ord(p') = ord(p)$.

Case 2. $|S_{j+1}| = ord(p)$.

In this case we use the Subset Teaching Property. Assume that $STD(c_{j+1}, C) > ord(p)$. By the Subset Teaching Property, there is a teaching plan

$$p' = ((c_1, S'_1), \ldots, (c_z, S'_z))$$

for C and a sample $S \in STS(c_{j+1}, C)$ such that $S'_{j+1} \subseteq S$ and $|S'_i| = |S_i|$ for all i.

First, note that S'_{j+1} is not contained in any subset teaching set for any $c \in C \setminus \{c_{j+1}\}$: The concepts c_{j+2}, $\ldots$, c_z are not consistent with the sample S'_{j+1}, because $S'_{j+1} \in TS(c_{j+1}, \{c_{j+1}, \ldots, c_z\})$. The concepts c_1, $\ldots$, c_j have, by induction hypothesis, a subset teaching dimension upper-bounded by $ord(p) = |S_{j+1}| = |S'_{j+1}|$. If S'_{j+1} was contained in a subset teaching set for some concept among c_1, $\ldots$, c_j, this would imply that S'_{j+1}

equaled some subset teaching set for some concept among $c_1, \ldots, c_j$, and thus S'_{j+1} could not be contained in the subset teaching set S for c_{j+1}.

Second, since S'_{j+1} is contained in the subset teaching set S for c_{j+1} and not contained in any subset teaching set for any $c \in C \setminus \{c_{j+1}\}$, S'_{j+1} equals S and is itself a subset teaching set for c_{j+1}. Consequently,

$$|S'_{j+1}| = STD(c_{j+1}, C) > ord(p) = ord(p') \geq |S'_{j+1}| .$$

This is obviously a contradiction in itself.

Hence $STD(c_{j+1}, C) \leq ord(p)$.

This concludes the induction step. □

For example, the class of all linear threshold functions, cf. C_θ in Table 1, has the subset teaching property. Every teaching plan $p = ((c_1, S_1), \ldots, (c_z, S_z))$ for C_θ with $ord(p) = RTD(C_\theta) = 1$ starts either with the concept $c_1 = X$ or with the concept $c_1 = \emptyset$. In either case, S_1 actually is a subset teaching set for c_1. A recursive argument for the subsequent concepts in the teaching plan shows that C_θ has the subset teaching property.

A similar argument proves that the class of all monomials over m instances, for any $m \geq 2$, has the subset teaching property.

9. Conclusions

We introduced two new models of teaching and learning of finite concept classes, based on the idea that learners can learn from a smaller number of labeled examples if they assume that the teacher chooses helpful examples. These models contrast with the classic teaching dimension model in which no more assumptions on the learner are made than it being consistent with the information presented. As a consequence, the information-theoretic complexity resulting from our new models is in general much lower than the teaching dimension.

Such results have to be interpreted with care since one constraint in modeling teaching is that coding tricks have to be avoided. However, one of our two models, the one based on *recursive teaching sets*, complies with Goldman and Mathias's original definition of valid teaching without coding tricks, cf. Goldman and Mathias (1996). The model based on *subset teaching sets* does not comply with the same definition of valid teaching. As we argued though, Goldman and Mathias's definition may be too restrictive when modeling cooperation in teaching and learning. Intuitively, their definition requires a learner to hypothesize a concept c as soon as any teaching set for c is contained in the given sample. This artificially precludes the possibility of a model in which a learner assumes that all the examples selected by the teacher are representative. Hence we introduced a less restrictive definition of coding trick. Each of the protocols presented in this paper can be regarded as meta-algorithms generating teacher/learner pairs that do not involve coding tricks.

The subset teaching protocol questions not only classic definitions of coding trick but also the intuitive idea that information-theoretic complexity measures should be monotonic with respect to the inclusion of concept classes. We discussed why non-monotonicity in this context may be a natural phenomenon in cooperative teaching and learning.

For many "natural" concept classes, the subset teaching dimension and the recursive teaching dimension turn out to be equal, but in general the two measures are not comparable. This immediately implies that neither one of the two corresponding teaching models is optimal among protocols that yield collusion-free teacher-learner pairs. One could easily design a protocol that, for every concept class C, would follow the subset teaching protocol if $STD(C) \leq RTD(C)$ and would follow the recursive teaching protocol if $RTD(C) < STD(C)$. Such a protocol would comply with our definition of collusion-freeness and would strictly dominate both the subset teaching protocol and the recursive teaching protocol. In this paper, we did not address the question of optimality of teaching protocols; we focused on intuitiveness of the protocols and the resulting teaching sets.

Acknowledgments

Many thanks are due to David Kirkpatrick, Shai Ben-David, Will Evans, and Wael Emara for pointing out flaws in an earlier version of this paper and for inspiring and helpful discussions and comments.

We furthermore gratefully acknowledge the support of Laura Zilles and Michael Geilke who developed and provided a software tool for computing subset teaching sets and recursive teaching sets.

This work was partly funded by the Alberta Ingenuity Fund (AIF) and by the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

D. Angluin and M. Kriķis. Teachers, learners and black boxes. In *Proc. of the 10th Annual Conference on Computational Learning Theory (COLT'97)*, pages 285–297. ACM, New York, 1997.

D. Angluin and M. Kriķis. Learning from different teachers. *Mach. Learn.*, 51:137–163, 2003.

M. Anthony, G. Brightwell, D.A. Cohen, and J. Shawe-Taylor. On exact specification by examples. In *Proc. of 5th Annual Workshop on Computational Learning Theory (COLT'92)*, pages 311–318. ACM, New York, 1992.

F. Balbach. Measuring teachability using variants of the teaching dimension. *Theoret. Comput. Sci.*, 397(1-3):94–113, 2008.

F. Balbach and T. Zeugmann. Recent developments in algorithmic teaching. In *Proc. 3rd Intl. Conf. on Language and Automata Theory and Applications*, pages 1–18, 2009.

S. Ben-David and N. Eiron. Self-directed learning and its relation to the VC-dimension and to teacher-directed learning. *Machine Learning*, 33(1):87–104, 1998.

T. Doliwa, H.U. Simon, and S. Zilles. Recursive teaching dimension, learning complexity, and maximum classes. In *Proc. of the 21st International Conference on Algorithmic Learning Theory (ALT'10)*, volume 6331 of *Lecture Notes in Artificial Intelligence*, pages 209–223. Springer, 2010.

R. Freivalds, E.B. Kinber, and R. Wiehagen. On the power of inductive inference from good examples. *Theoret. Comput. Sci.*, 110(1):131–144, 1993.

S. Goldman and M. Kearns. On the complexity of teaching. In *Proc. of the 4th Annual Workshop on Computational Learning Theory, COLT'91*, pages 303–314. Morgan Kaufmann, 1991.

S.A. Goldman and M.J. Kearns. On the complexity of teaching. *J. Comput. Syst. Sci.*, 50 (1):20–31, 1995.

S.A. Goldman and H.D. Mathias. Teaching a smarter learner. *J. Comput. Syst. Sci.*, 52(2): 255–267, 1996.

S. Hanneke. Teaching dimension and the complexity of active learning. In *Proc. of the 20th Annual Conference on Learning Theory (COLT 2007)*, pages 66–81. LNCS 4539, Springer, Berlin, 2007.

T. Hegedűs. Generalized teaching dimensions and the query complexity of learning. In *Proc. of the 8th Annual Conference on Computational Learning Theory (COLT'95)*, pages 108–117. ACM, New York, 1995.

J. Jackson and A. Tomkins. A computational model of teaching. In *Proc. of 5th Annual Workshop on Computational Learning Theory (COLT'92)*, pages 319–326. ACM, New York, 1992.

H. Kobayashi and A. Shinohara. Complexity of teaching by a restricted number of examples. In *Proc. of the 22nd Annual Conference on Learning Theory, COLT'09*, pages 293–302, 2009.

S. Lange, J. Nessel, and R. Wiehagen. Learning recursive languages from good examples. *Ann. Math. Artif. Intell.*, 23(1-2):27–52, 1998.

H.D. Mathias. A model of interactive teaching. *J. Comput. Syst. Sci.*, 54(3):487–501, 1997.

M. Ott and F. Stephan. Avoiding coding tricks by hyperrobust learning. *Theoret. Comput. Sci.*, 284(1):161–180, 2002.

R.L. Rivest and Y.L. Yin. Being taught can be faster than asking questions. In *Proc. of the 8th Annual Conference on Computational Learning Theory (COLT'95)*, pages 144–151. ACM, New York, 1995.

R.A. Servedio. On the limits of efficient teachability. *Inf. Process. Lett.*, 79(6):267–272, 2001.

A. Shinohara and S. Miyano. Teachability in computational learning. *New Generation Comput.*, 8(4):337–348, 1991.

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

S. Zilles, S. Lange, R. Holte, and M. Zinkevich. Teaching dimensions based on cooperative learning. In *Proc. of the 21st Annual Conference on Learning Theory (COLT'08)*, pages 135–146. Omnipress, 2008.