Electronic Colloquium on Computational Complexity, Report No. 65 (2016)

A Note on Teaching for VC Classes

Xi Chen* Yu Cheng† Bo Tang‡§

April 8, 2016

Given a concept class $\mathcal{C} \subseteq \{0,1\}^n$ (a set of binary strings of length n), $X \subseteq [n]$ is a *teaching set* for a concept $c \in \mathcal{C}$ (a binary string in $\mathcal{C}$) if X satisfies

$$c|_X \neq c'|_X, \quad \text{for all other concepts } c' \in \mathcal{C},$$

where we use $c|_X$ to denote the projection of c on X. The *teaching dimension* of $\mathcal{C}$ is the smallest number t such that *every* $c \in \mathcal{C}$ has a teaching set of size no more than t [GK95]. However, teaching dimension does not always capture the cooperation in teaching and learning, and the notion of *recursive teaching dimension* has been introduced and studied extensively in the literature [Kuh99, DSZ10, ZLHZ11, WY12, DFSZ14, SSYZ14, MSWY15]. The recursive teaching dimension of a class $\mathcal{C} \subseteq \{0,1\}^n$, denoted by $\mathrm{RTD}(\mathcal{C})$, is the smallest number t where one can order all the concepts of $\mathcal{C}$ as a sequence $c_1, \ldots, c_{|\mathcal{C}|}$ such that every concept c_i, $i < |\mathcal{C}|$, has a teaching set of size no more than t in $\{c_i, \ldots, c_{|\mathcal{C}|}\}$. Hence, $\mathrm{RTD}(\mathcal{C})$ measures the worst-case number of labelled examples needed to learn any target concept in $\mathcal{C}$, if the teacher and the learner agree *a priori* on a specific order of the concepts of the class $\mathcal{C}$.

In this note, we study the recursive teaching dimension of concept classes of low VC-dimension. Recall that the VC-dimension [VC71] of $\mathcal{C} \subseteq \{0,1\}^n$, denoted by $\mathrm{VCD}(\mathcal{C})$, is the maximum size of a *shattered* subset of $[n]$, where $Y \subseteq [n]$ is shattered if for every binary string $\boldsymbol{b}$ of length $|Y|$, there is a concept $c \in \mathcal{C}$ such that $c|_Y = \boldsymbol{b}$.

Our main result is the following upper bound for $\mathrm{RTD}(\mathcal{C})$.

Theorem 1. *Let $\mathcal{C}$ be a concept class with $\mathrm{VCD}(\mathcal{C}) = d$. Then $\mathrm{RTD}(\mathcal{C}) \leq 2^{d+1}(d-2) + d + 4$.*

This is the first upper bound for $\mathrm{RTD}(\mathcal{C})$ that depends only on $\mathrm{VCD}(\mathcal{C})$, but not $|\mathcal{C}|$, the size of the concept class. Previously, Moran et al. [MSWY15] showed an upper bound of $O(d2^d \log\log|\mathcal{C}|)$ for $\mathrm{RTD}(\mathcal{C})$; our result removes the $\log\log|\mathcal{C}|$ factor, and answers positively an open problem posed in [MSWY15]. Theorem 1 is also a step towards answering the following question:

$$\textit{Is } \mathrm{RTD}(\mathcal{C}) = O(\mathrm{VCD}(\mathcal{C}))\,?$$

posed by Simon and Zilles [SZ15]. Given that the current best lower bound for $\mathrm{RTD}(\mathcal{C})$, in terms of $d = \mathrm{VCD}(\mathcal{C})$, is only $3d/2$ for $d \geq 2$ [DFSZ14], an exponential gap remains. The simplest case that is still open is when $d = 2$ ([Kuh99] showed that $\mathrm{RTD}(\mathcal{C}) = 1$ when $d = 1$): [DFSZ14] presented a concept class $\mathcal{C}$ (Warmuth's class) with $\mathrm{RTD}(\mathcal{C}) = 3$; Theorem 1 shows that $\mathrm{RTD}(\mathcal{C}) \leq 6$ when $d = 2$.

*Columbia University. Email: xichen@cs.columbia.edu.
†University of Southern California. Email: yu.cheng.1@usc.edu.
‡Oxford University. Email: tangbonk1@gmail.com.
§This work was done in part while the authors were visiting the Simons Institute for the Theory of Computing.

ISSN 1433-8092

Proof of Theorem 1

Theorem 1 follows directly from the following lemma and the observation that the VC-dimension of a concept class cannot go up after a concept is removed.

Lemma 2. *Let $\mathcal{C} \subseteq \{0,1\}^n$ be a concept class with VC-dimension d. Then there exists a concept $c \in \mathcal{C}$ with a teaching set of size at most $2^{d+1}(d-2)+d+4$.*

Proof. We prove by induction on d. Let

$$f(d) = \max_{\mathcal{C}:\,\mathrm{VCD}(\mathcal{C})\le d} \mathrm{RTD}(\mathcal{C}),$$

and our goal is to prove the following upper bound for $f(d)$:

$$f(d) \le 2^{d+1}(d-2)+d+4 \tag{1}$$

for all $d \ge 1$. The base case of $d=1$ follows from [Kuh99].

For the induction step, we show that condition (1) holds for some $d>1$, assuming that it holds for $d-1$. Take any concept class $\mathcal{C} \subseteq \{0,1\}^n$ with $\mathrm{VCD}(\mathcal{C}) \le d$. Let $k = 2^d(d-1)+1$. If $n \le k$ then we are already done; assume in the rest of the proof that $n > k$. Any set of k coordinates $Y \subset [n]$ partitions $\mathcal{C}$ into 2^k (possibly empty) subsets, denoted by

$$\mathcal{C}_{\boldsymbol{b}}^{Y} = \{c \in \mathcal{C} : c\,|_Y = \boldsymbol{b}\}, \qquad \text{for each } \boldsymbol{b} \in \{0,1\}^k.$$

We follow the idea of [MSWY15] to choose a set of k coordinates $Y^* \subset [n]$ and a vector $\boldsymbol{b}^* \in \{0,1\}^k$ such that $\mathcal{C}_{\boldsymbol{b}^*}^{Y^*}$ is *nonempty* and has the *smallest* size among all nonempty $\mathcal{C}_{\boldsymbol{b}}^{Y}$ over all choices of Y and $\boldsymbol{b}$. Without loss of generality, we assume below that $Y^* = [k]$ and $\boldsymbol{b}^*$ is the all-zero vector. Also for notational convenience, we write $\mathcal{C}_{\boldsymbol{b}}$ to denote $\mathcal{C}_{\boldsymbol{b}}^{Y^*}$ for $\boldsymbol{b} \in \{0,1\}^k$.

Notice that if $\mathcal{C}_{\boldsymbol{b}^*} = \mathcal{C}_{\boldsymbol{b}^*}^{Y^*}$ has VC-dimension at most $d-1$, then we have

$$\mathrm{VCD}(\mathcal{C}) \le k + f(d-1) \le 2^{d+1}(d-2)+d+4,$$

using the inductive hypothesis. This is because according to the definition of f one of the concepts $c \in \mathcal{C}_{\boldsymbol{b}^*}$ has a teaching set $T \subseteq [n]\setminus Y^*$ of size at most $f(d-1)$ to distinguish it from other concepts of $\mathcal{C}_{\boldsymbol{b}^*}$. Thus, $[k] \cup T$ is a teaching set of c in the original class $\mathcal{C}$, of size at most $k + f(d-1)$.

Finally we prove by contradiction that $\mathcal{C}_{\boldsymbol{b}^*}$ has VC-dimension at most $d-1$. Assume that $\mathcal{C}_{\boldsymbol{b}^*}$ has VC-dimension d. Then by definition, there exist a set of d coordinates $Z \subseteq [n]\setminus Y^*$ that is shattered by $\mathcal{C}_{\boldsymbol{b}^*}$ (i.e., all the 2^d possible vectors appear in $\mathcal{C}_{\boldsymbol{b}^*}$ on Z). Observe that for each $i \in Y^*$, the union of all $\mathcal{C}_{\boldsymbol{b}}$ with $b_i = 1$ (recall that $\boldsymbol{b}^*$ is all-zero) must miss at least one vector on Z, which we denote by $\boldsymbol{p}_i$ (choose one arbitrarily if more than one are missing); otherwise, $\mathcal{C}$ has a shattered set of size $d+1$, i.e., $Z \cup \{i\}$, contradicting with the assumption that $\mathrm{VCD}(\mathcal{C}) \le d$. However, given that there are only 2^d possibilities for each $\boldsymbol{p}_i$ (and $|Y^*| = k = 2^d(d-1)+1$), it follows from the pigeonhole principle that there exists a subset $K \subset Y^*$ of size d such that $\boldsymbol{p}_i = \boldsymbol{p}$ for every $i \in K$, for some $\boldsymbol{p} \in \{0,1\}^d$. Let $Y' = (Y^* \setminus K) \cup Z$ be a new set of k coordinates and let $\boldsymbol{b}' = \boldsymbol{0}_{k-d} \circ \boldsymbol{p}$. Then $\mathcal{C}_{\boldsymbol{b}'}^{Y'}$ is indeed a nonempty and proper subset of $\mathcal{C}_{\boldsymbol{b}^*}^{Y^*}$, a contradiction with our choice of Y^* and $\boldsymbol{b}^*$. ☐

References

[DFSZ14] T. Doliwa, G. Fan, H.-U. Simon, and S. Zilles. Recursive teaching dimension, VC-dimension and sample compression. *Journal of Machine Learning Research*, 15(1):3107–3131, 2014.

[DSZ10] T. Doliwa, H.-U. Simon, and S. Zilles. Recursive teaching dimension, learning complexity, and maximum classes. In *Proceedings of the 21st International Conference on Algorithmic Learning Theory*, pages 209–223, 2010.

[GK95] S. A. Goldman and M. J. Kearns. On the complexity of teaching. *Journal of Computer and System Sciences*, 50(1):20–31, 1995.

[Kuh99] C. Kuhlmann. On teaching and learning intersection-closed concept classes. In *Proceedings of the 4th European Conference on Computational Learning Theory*, pages 168–182, 1999.

[MSWY15] S. Moran, A. Shpilka, A. Wigderson, and A. Yehudayoff. Compressing and teaching for low VC-dimension. In *Proceedings of the 56th IEEE Annual Symposium on Foundations of Computer Science*, pages 40–51, 2015.

[SSYZ14] R. Samei, P. Semukhin, B. Yang, and S. Zilles. Algebraic methods proving Sauer's bound for teaching complexity. *Theoretical Computer Science*, 558:35–50, 2014.

[SZ15] H.-U. Simon and S. Zilles. Open problem: Recursive teaching dimension versus VC dimension. In *Proceedings of the 28th Conference on Learning Theory*, pages 1770–1772, 2015.

[VC71] V. N. Vapnik and A. Ya. Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and Its Applications*, 16:264–280, 1971.

[WY12] A. Wigderson and A. Yehudayoff. Population recovery and partial identification. In *Proceedings of the 53rd IEEE Annual Symposium on Foundations of Computer Science*, pages 390–399, 2012.

[ZLHZ11] S. Zilles, S. Lange, R. Holte, and M. Zinkevich. Models of cooperative teaching and learning. *Journal of Machine Learning Research*, 12:349–384, 2011.

ECCC ISSN 1433-8092

http://eccc.hpi-web.de