BEST-CASE TEACHING DIMENSION AGAINST VC DIMENSION ON AT MOST FIVE POINTS: THE JOINT CENSUS, THE LEAST DOMAIN FOR A GAP AT VC DIMENSION THREE, AND WHERE THE GREEDY TEACHING-SET ALGORITHM FIRST FAILS

KOREA SUPERINTELLIGENCE LABS

Abstract. For a finite concept class $\mathcal{C}$, the best-case teaching dimension $\mathrm{TS}_{\min}(\mathcal{C})$ is the least number of domain points on which some concept of $\mathcal{C}$ differs from every other concept of $\mathcal{C}$; whether $\mathrm{TS}_{\min}(\mathcal{C}) = O(\mathrm{VCD}(\mathcal{C}))$ is open, and everything known about the two parameters together is asymptotic. We report the exact joint distribution of $(\mathrm{VCD}, \mathrm{TS}_{\min})$ over every concept class on at most five points — all 4 294 967 295 nonempty classes on five — and hence the table $M(n,d) = \max\{\mathrm{TS}_{\min}(\mathcal{C}) : \mathrm{VCD}(\mathcal{C}) = d\}$ for $n \leq 5$, whose last row reads $0, 1, 3, 4, 4, 5$. Two cells lie above the diagonal. One is Warmuth's class ($\mathrm{VCD} = 2$, $\mathrm{TS}_{\min} = 3$), which on five points is unique up to relabelling and complementation of points. The other consists of classes with $\mathrm{VCD} = 3$ and $\mathrm{TS}_{\min} = 4$ on *five* points — three orbits, of 20, 21 and 22 concepts — whereas the standard construction, the disjoint union of Warmuth's class with a two-concept class, needs six; by a theorem of Doliwa, Fan, Simon and Zilles five is the least possible domain size, and each of the three contains a copy of Warmuth's class. Second, the greedy teaching-set algorithm whose limitations Compton, Pabbaraju and Zhivotovskiy study is already suboptimal on four points and returns twice the optimum on five, under every tie-breaking; the largest gap on $n \leq 5$ points is $0, 0, 0, 1, 2$. The witness classes, their parameters and the one-step correctness of the algorithm are verified in Lean 4 against *Mathlib* by kernel reduction; the censuses are external computations and are labelled as such. The four extremal orbits also arise in a companion paper, under an equivalent condition; what is new here is their $\mathrm{TS}_{\min}$ values, the joint census, the greedy analysis and the machine-checked witnesses.

1. Introduction

1.1. **Two parameters.** Let X be a finite set, the *domain*, and $\mathcal{C}$ a nonempty family of subsets of X, a *concept class*; a concept is identified with the set of points it labels 1. Two parameters of $\mathcal{C}$ meet in this paper. The first is the Vapnik–Chervonenkis dimension $\mathrm{VCD}(\mathcal{C})$, the largest size of a set $Y \subseteq X$ every subset of which is $c \cap Y$ for some $c \in \mathcal{C}$. The second comes from machine teaching. A *teaching set* for $c \in \mathcal{C}$ is a set $S \subseteq X$ on which c differs from every other concept of $\mathcal{C}$ — no $c' \in \mathcal{C} \setminus \{c\}$ has $c' \cap S = c \cap S$ — and, writing $\mathrm{TS}(c;\mathcal{C})$ for the least size of one,

$$\mathrm{TS}_{\min}(\mathcal{C}) \;=\; \min_{c\in\mathcal{C}} \mathrm{TS}(c;\mathcal{C})$$

is the *best-case teaching dimension* of $\mathcal{C}$, in the name and notation of Doliwa, Fan, Simon and Zilles [2, p. 3112] ("In the sequel, $\mathrm{TS}_{\min}(\mathcal{C})$ is called the best-case teaching dimension of $\mathcal{C}$"); Simon and Zilles [4] write $\mathrm{TD}_{\min}$ for the same quantity.

1.2. **The question.** Compton, Pabbaraju and Zhivotovskiy [1] open with the sentence

> "A fundamental open problem in learning theory is to characterize the best-case teaching dimension $\mathrm{TS}_{\min}$ of a concept class $\mathcal{C}$ with finite VC dimension d."

The target is $\mathrm{TS}_{\min} = O(d)$, which by an observation of [2] recorded in [4] is equivalent to the COLT 2015 open problem of Simon and Zilles, $\mathrm{RTD} = O(d)$ for the recursive teaching dimension. Everything that [1] records as known about the two parameters together is asymptotic: "a line of work exploiting no properties of the concept class beyond its finite VC dimension was initiated by Kuhlmann (1999), who proved that $\mathrm{TS}_{\min} = 1$ for classes with $d = 1$"; then $O(\log|\mathcal{C}|)$ by halving, $O(d\,2^d \log\log|\mathcal{C}|)$ by Moran, Shpilka, Wigderson and Yehudayoff [7], $O(d\,2^d)$ by Chen, Cheng and Tang [3], and $O(d^2)$ by Hu, Wu, Li and Wang [5]. All of these upper bounds come from one greedy procedure — Algorithm 1 of [1], recalled in Section 2 — which repeatedly adds to the teaching set the k labelled points that shrink the class most. The contribution of [1] is a lower bound on that procedure: its Theorem 1 builds classes of axis-aligned rectangles on which the $k = 1$ version returns $\Omega(\log|\mathcal{C}|)$ points, while (its Appendix C) the classes have $\mathrm{TS}_{\min} \le 2$.

What is known at small scale is little and is exact. Doliwa et al. [2] print a ten-concept class $\mathcal{C}_W$ on five points, communicated to them by Manfred Warmuth, with "$\mathrm{RTD}(\mathcal{C}_W) = \mathrm{TS}_{\min}(\mathcal{C}_W) = 3$ and $\mathrm{VCD}(\mathcal{C}_W) = 2$" (their p. 3119, by brute-force enumeration), and prove (their Theorem 15) that any class with RTD > VCD has at least ten concepts and at least five points. Chen, Cheng and Tang [3] encode the existence of a class with prescribed $(\mathrm{VCD}, \mathrm{RTD}, \text{domain size})$ as a satisfiability problem and print, as their Figure 3, seven decided cells; none of them asks about VC dimension 3 on five points. No table of the largest $\mathrm{TS}_{\min}$ at a given VC dimension on a given number of points is printed anywhere we could find, and nothing is printed about the greedy procedure on small domains.

1.3. **What is settled here.** Both parameters are finite and elementary, so small domains are decidable, and we decide them. Points are numbered $0, 1, \ldots, n-1$ throughout, to match the machine-checked statements of Appendix A.

Five points, VC dimension 3*, best-case teaching dimension* 4*.* Theorem 4.1 exhibits three concept classes on five points — of 20, 21 and 22 concepts, listed in full — with $\mathrm{VCD} = 3$ and $\mathrm{TS}_{\min} = 4$, each containing a copy of Warmuth's class up to relabelling and complementation of points. The natural way to reach the pair $(3, 4)$ is the additivity of both parameters under disjoint union ([2], Lemma 16), which glues $\mathcal{C}_W$ to a two-concept class on a fresh point and lands on *six* points with 20 concepts; the classes here have the same 20 concepts on five. Five is least possible, by Theorem 15 of [2] (Corollary 4.3). The cell $(\mathrm{VCD}, \mathrm{RTD}, n) = (3, 4, 5)$ is absent from the table of [3], whose search is confined to ratios $\mathrm{RTD} / \mathrm{VCD} > 3/2$.

The greedy algorithm fails on four points. Theorem 6.2 gives a nine-concept class on four points with $\mathrm{TS}_{\min} = 2$ on which every run of the greedy procedure at $k = 1$ returns three points, and a seventeen-concept class on five points with $\mathrm{TS}_{\min} = 2$ on which every run returns four. Lemma 6.1 is the one-step correctness of the procedure, for every domain size: one restriction costs at most one point of teaching set, so its output is never below $\mathrm{TS}_{\min}$ and the gap is well defined.

The censuses. Section 5 reports the joint distribution of $(\mathrm{VCD}, \mathrm{TS}_{\min})$ over all $2^{32} - 1$ nonempty classes on five points and over the smaller domains, hence the table $M(n, d)$ for $n \le 5$; the two cells above the diagonal are the orbit of $\mathcal{C}_W$ and the three orbits of Theorem 4.1. Section 6 reports the exact largest gap between the greedy output and $\mathrm{TS}_{\min}$ over all classes on n points, $0, 0, 0, 1, 2$ for $n = 1, \ldots, 5$, with its distribution, and reproduces the rectangle construction of [1] at its three smallest sizes. These computations are external to the formal development, support no theorem, and are labelled as such wherever they appear.

A calibration and an elementary layer. Proposition 3.1 records the values $\mathrm{VCD}(\mathcal{C}_W) = 2$ and $\mathrm{TS}_{\min}(\mathcal{C}_W) = 3$ of [2] as verified, not as new; Lemma 2.5 is the textbook layer over the definitions, formalised because every statement about a specific class here is a finite evaluation against those same definitions.

1.4. **Overlap with a companion paper.** The four orbits above the diagonal are not new to the authors' own work. A companion paper [9] on non-clashing teaching enumerates, in its Remark 5.5, the classes on five points that have no *fragment of size* VCD *and frequency* 1 — no pattern on a $\mathrm{VCD}(\mathcal{C})$-element set realised by exactly one concept — and finds exactly 960 of them, in four orbits of sizes $192, 192, 384, 192$ with $10, 20, 21, 22$ concepts. That condition is equivalent to $\mathrm{TS}_{\min}(\mathcal{C}) > \mathrm{VCD}(\mathcal{C})$ (Lemma 2.6), so those are the same four orbits. What [9] does not contain is any value of $\mathrm{TS}_{\min}$: it reports the non-clashing teaching dimension of the four representatives, $2, 2, 2, 3$. What is new here is the $\mathrm{TS}_{\min}$ values of the four orbits, the joint $(\mathrm{VCD}, \mathrm{TS}_{\min})$ census the table $M(n, d)$ is read from, the statement about the least domain size at VC dimension 3, everything about the greedy algorithm, and the machine-checked witnesses. We say so again where Theorem 4.1 is proved.

1.5. **What is not claimed.** Nothing is claimed about six or more points: not the value of $M(6, d)$ for any d, not whether $\mathrm{TS}_{\min} - \mathrm{VCD}$ can reach 2, not whether 20 is the least number of concepts of a class with VCD = 3 and $\mathrm{TS}_{\min} = 4$ over all domains, not the greedy gap beyond five points. Theorem 1 of [1] is not re-proved; Remark 6.4 reproduces its construction at $N \le 3$ only. Kuhlmann's theorem $\mathrm{TS}_{\min} = 1$ at $d = 1$ is not re-proved and his paper [6] was not obtained (see the reference list); the fact is used only as the literature states it and is reproduced by the census on $n \le 5$. The "No" rows of the table of [3] at seven and eight points rest on satisfiability solvers and are not re-derived here beyond their shadow on at most five points. Nothing is claimed about RTD beyond the consequence $\mathrm{TS}_{\min}(\mathcal{C}) \le \mathrm{RTD}(\mathcal{C})$ of Lemma 4 of [2].

1.6. **Where this was looked for.** The search was made before anything was claimed, on 4 September 2026. The table of [3] (Figure 3, read from both the NeurIPS and the ECCC versions, which agree) has exactly seven rows, $(\mathrm{VCD}, \mathrm{RTD}, n) = (2, 3, 5)$ satisfiable by $\mathcal{C}_W$, $(2, 4, 7)$, $(3, 5, 7)$, $(3, 6, 8)$, $(4, 6, 7)$, $(4, 7, 8)$ unsatisfiable, and $(3, 5, 12)$ satisfiable; its authors write that "brute-force search quickly becomes infeasible as the domain size n gets larger" and that "for $n > 8$, even these SAT solvers are no longer feasible". Doliwa et al. [2] run no exhaustive search; their two uses of brute force are verifications of named classes. Hu et al. [5] define (their Definition 5) a table $f(x, y) = \sup_{\mathcal{C}} \mathrm{TD}_{\min}(\mathcal{C})$ over classes with at most y patterns on every x-element set, with the domain size unbounded — a different table from $M(n, d)$ — and record "Kuhlmann (1999) proved $f(2, 3) = 1$". The source [1] has no recorded citation on OpenAlex or Semantic Scholar, and still none when this was rechecked in September 2026, after its appearance in the COLT 2025 proceedings; all eight recorded citers of [3] and all 56 of [2] were listed and read by title, and none is a census or a table. A full-text corpus of arXiv through August 2026 contains the phrase "best-case teaching dimension" in exactly two papers, the source and a commented-out to-do list. The oldest paper on witness sets, Kushilevitz, Linial, Rabinovich and Saks [8], bounds the *average* witness size and prints no table. The OEIS returns nothing for the orbit sizes $192, 768, 960$, for the census counts, or for $3408, 432$. Read these negatives as "not found", not as "new": every computation here is a few processor-hours for anyone who asks the question.

2. Preliminaries

Throughout $n \in \mathbb{N}$, $X_n = \{0, 1, \ldots, n-1\}$, and a concept $c \subseteq X_n$ is written as the bit string $c_0 c_1 \cdots c_{n-1}$ with $c_i = 1$ iff $i \in c$. A concept class is a nonempty $\mathcal{C} \subseteq 2^{X_n}$; 2^{X_n} itself is the class of all 2^n concepts.

Definition 2.1 (shattering, VCD)**.** $\mathcal{C}$ shatters $Y \subseteq X_n$ if for every $p \subseteq Y$ there is $c \in \mathcal{C}$ with $c \cap Y = p$. $\mathrm{VCD}(\mathcal{C})$ is the largest $|Y|$ over shattered Y.

Definition 2.2 (teaching sets, $\mathrm{TS}_{\min}$)**.** $S \subseteq X_n$ is a teaching set for $c \in \mathcal{C}$ if every $c' \in \mathcal{C}$ with $c' \cap S = c \cap S$ equals c. $\mathrm{TS}_{\min}(\mathcal{C})$ is the least $|S|$ over all $c \in \mathcal{C}$ and all teaching sets S for c. We write $\mathrm{TS}_{\min}(\mathcal{C}) \le k$ to mean that some concept of $\mathcal{C}$ has a teaching set of at most k points, and $\mathrm{VCD}(\mathcal{C}) \ge d$ to mean that some d-element set is shattered; upper bounds on $\mathrm{TS}_{\min}$ and lower bounds on VCD are proved in that form, and the opposite bounds by refuting it.

In [2] a teaching set is a set of labelled examples consistent with c and with no other concept; since the labels are forced by c, that is the same object. A superset of a teaching set is a teaching set, and X_n is one for every concept, so $\mathrm{TS}_{\min}(\mathcal{C}) \le n$.

The *recursive teaching dimension* RTD of Zilles et al., as used in [2], enters only through Lemma 4 of [2], $\mathrm{RTD}(\mathcal{C}) = \max_{\mathcal{C}' \subseteq \mathcal{C}} \mathrm{TS}_{\min}(\mathcal{C}')$, whose case $\mathcal{C}' = \mathcal{C}$ gives

$$\mathrm{TS}_{\min}(\mathcal{C}) \;\le\; \mathrm{RTD}(\mathcal{C}) \qquad \text{for every concept class } \mathcal{C}. \tag{1}$$

Definition 2.3 (the table)**.** For $0 \le d \le n$ let

$$M(n, d) \;=\; \max\{\mathrm{TS}_{\min}(\mathcal{C}) : \mathcal{C} \text{ a concept class on } X_n \text{ with } \mathrm{VCD}(\mathcal{C}) = d\}.$$

The full power set of a d-element $Y \subseteq X_n$ has VCD $= d$ and $\mathrm{TS}_{\min} = d$ (a teaching set for any of its concepts must contain all of Y), so $M(n, d) \ge d$ always; the question is where $M(n, d) > d$.

Symmetry. The hyperoctahedral group $G_n = \mathbb{Z}_2^n \rtimes S_n$, of order $2^n n!$, acts on concepts by $c \mapsto \pi(c) \triangle v$ for $\pi \in S_n$ and $v \subseteq X_n$ (relabel the points, then complement the labels at the points of v), and hence on classes.

Lemma 2.4 (invariance)**.** VCD *and* $\mathrm{TS}_{\min}$ *are invariant under* G_n*.*

Proof. Let $g = (\pi, v)$ and $\mathcal{C}' = g\mathcal{C}$. For $Y \subseteq X_n$ the map $p \mapsto \pi(p)\triangle(v \cap \pi(Y))$ is a bijection from the subsets of Y to the subsets of $\pi(Y)$ carrying the traces $\{c \cap Y : c \in \mathcal{C}\}$ onto $\{c' \cap \pi(Y) : c' \in \mathcal{C}'\}$, so $\mathcal{C}$ shatters Y iff $\mathcal{C}'$ shatters $\pi(Y)$. Likewise $c' \cap S = c \cap S$ iff $gc' \cap \pi(S) = gc \cap \pi(S)$, so S teaches c in $\mathcal{C}$ iff $\pi(S)$ teaches gc in $\mathcal{C}'$. □

A class on X_n is encoded, when a compact name is wanted, by its *mask*: the 2^n-bit integer whose bit number $\sum_{i\in c} 2^i$ is set for each $c \in \mathcal{C}$. The *canonical form* of a class is the numerically least mask in its G_n-orbit. Every class named below is given by its list of bit strings; the mask is quoted beside it only so that the census output of Section 5 can be compared literally.

The greedy procedure. For $x \in X_n$ and $b \in \{0,1\}$ let $\mathcal{C}|_{x,b} = \{c \in \mathcal{C} : [x \in c] = b\}$ be the *restriction* of $\mathcal{C}$ to the labelled point (x, b). Algorithm 1 of [1] with greediness parameter $k = 1$ is:

> $S \leftarrow \emptyset$; while $|\mathcal{C}| > 1$: choose (x, b) with $\mathcal{C}|_{x,b} \neq \emptyset$ and $|\mathcal{C}|_{x,b}|$ least; $S \leftarrow S \cup \{x\}$; $\mathcal{C} \leftarrow \mathcal{C}|_{x,b}$. Return S (and the surviving concept c, $\mathcal{C} = \{c\}$).

(For general k the minimum runs over sets T of at most k points with a pattern $b \in \{0,1\}^{|T|}$; only $k = 1$ is studied here.) The footnote of [1] to Algorithm 1, "ties are broken in favor of a T that has smaller size", decides nothing at $k = 1$, where every candidate T is a single point, so the procedure is genuinely nondeterministic. Write $g_{\min}(\mathcal{C})$ and $g_{\max}(\mathcal{C})$ for the least and the greatest $|S|$ returned over all tie-breakings. Two elementary facts frame the definitions. First, once x has been chosen every concept of the current class agrees at x, so the restriction to (x, b) is either empty or the whole class, and the whole class is never a least nonempty restriction while $|\mathcal{C}| > 1$ (two concepts differ at some y, and both restrictions at y are then nonempty and smaller); hence no run chooses a point twice, every run stops within n iterations, and $|S|$ is the number of iterations. Second, $g_{\min}(\mathcal{C}) > \mathrm{TS}_{\min}(\mathcal{C})$ means that *no* tie-breaking is optimal. The formal statements of Section 6 carry the bound n on the number of iterations explicitly, as a fuel parameter; by the first fact that bound is never active. That first fact — that n iterations always suffice, so that the fuel-bounded functions of the development compute $g_{\min}$ and $g_{\max}$ and not a truncation of them — is the elementary argument just given and is *not* formalised; the development checks only, for one witness, that raising the fuel by one changes nothing. Every statement below about $g_{\min}$ or $g_{\max}$ should be read as a statement about the fuel-bounded functions, which is what is machine-checked.

The elementary layer. The following facts are textbook. They are recorded because every statement about a particular class below is a finite evaluation against the definitions above, and those definitions therefore need an independent check that they mean what the prose says.

Lemma 2.5 (elementary layer)**.** *Let* $\mathcal{C}$ *be a concept class on* X_n*.*

(1) $\mathrm{TS}_{\min}(\mathcal{C}) \le n$.
(2) $\mathrm{TS}_{\min}(\mathcal{C}) = 0$ *if and only if* $|\mathcal{C}| = 1$.
(3) $\mathrm{VCD}(\mathcal{C}) \ge 1$ *if and only if* $|\mathcal{C}| \ge 2$.
(4) *If* $\mathrm{VCD}(\mathcal{C}) = 0$ *then* $\mathrm{TS}_{\min}(\mathcal{C}) = 0$*; in particular* $\mathrm{TS}_{\min}(\mathcal{C}) \le \mathrm{VCD}(\mathcal{C})$ *whenever* $\mathrm{VCD}(\mathcal{C}) = 0$.

Proof. (1) X_n teaches every concept. (2) $\emptyset$ teaches c iff every concept equals c. (3) Two distinct concepts differ at some x, and then $\{x\}$ is shattered; conversely a shattered singleton needs two concepts. (4) By (3), $\mathrm{VCD}(\mathcal{C}) = 0$ forces $|\mathcal{C}| = 1$, and (2) applies. Part (4) is the case $d = 0$ of the question of Section 1 and the reason the tables below start at $d = 1$. □

Lemma 2.6 (fragments)**.** *Let* $d = \mathrm{VCD}(\mathcal{C})$*. Then* $\mathrm{TS}_{\min}(\mathcal{C}) \le \mathrm{VCD}(\mathcal{C})$ *if and only if some* d*-element* $Y \subseteq X_n$ *and some* $p \subseteq Y$ *have exactly one* $c \in \mathcal{C}$ *with* $c \cap Y = p$*.*

Proof. If S teaches c with $|S| \leq d$, enlarge S to a d-element Y (possible since $d \leq n$); Y still teaches c, i.e. c is the only concept with trace $c \cap Y$ on Y. Conversely such a Y is a d-element teaching set for that unique c. □

The right-hand side of Lemma 2.6 is the negation of the condition "no fragment of size VCD and frequency 1" of [9]; this is the equivalence used in Section 1.4.

3. Warmuth's class, as a calibration

The class $\mathcal{C}_W$ printed by Doliwa et al. as their Figure 1(a) has, in canonical form, the ten concepts

$$\mathcal{C}_W : \quad \texttt{11000 10100 01010 10110 01110 11001 00101 01101 00011 10011}$$

(mask `0x03586428`). The rows of the printed table (on their points $x_1, \ldots, x_5$: 11000, 01100, 00110, 00011, 10001, 01011, 01101, 10110, 10101, 11010) canonicalise to this list under G_5; both parameters are unchanged by Lemma 2.4, so the proposition below is about their class.

Proposition 3.1 ([2], p. 3119; not new)**.** $\mathrm{VCD}(\mathcal{C}_W) = 2$ *and* $\mathrm{TS}_{\min}(\mathcal{C}_W) = 3$*.*

Proof. A finite check against the definitions: some pair is shattered and no triple is (VCD = 2); some concept has a three-point teaching set and none has a two-point one ($\mathrm{TS}_{\min} = 3$). The values are those printed in [2] ("Brute-force enumeration shows that $\mathrm{RTD}(\mathcal{C}_W) = \mathrm{TS}_{\min}(\mathcal{C}_W) = 3$ and $\mathrm{VCD}(\mathcal{C}_W) = 2$"); here they are recomputed, from the list above, by three independent implementations and by the kernel-checked development of Section 8. □

4. Five points, VC dimension three, best-case teaching dimension four

Theorem 4.1. *The following three concept classes on* $X_5 = \{0, 1, 2, 3, 4\}$ *have* $\mathrm{VCD} = 3$ *and* $\mathrm{TS}_{\min} = 4$*:*

$$\begin{aligned}
\mathcal{C}_{20} : \;& \texttt{00000 10000 01000 00100 11100 10010 01010 10110 01110 11110} \\
& \texttt{10001 11001 00101 01101 11101 00011 01011 11011 00111 10111} \qquad (\texttt{0x3ddae697}), \\
\mathcal{C}_{21} : \;& \mathcal{C}_{20} \cup \{\texttt{01100}\} \qquad (\texttt{0x3ddae6d7}), \\
\mathcal{C}_{22} : \;& \texttt{00000 10000 01000 10100 01100 10010 11010 00110 10110 01110 11110} \\
& \texttt{00001 01001 11001 10101 01101 11101 00011 01011 11011 00111 10111} \qquad (\texttt{0x3dedfa67}).
\end{aligned}$$

Consequently there is a concept class on a five-point domain with $\mathrm{VCD} = 3$ *and* $\mathrm{TS}_{\min} = 4$*. Moreover the ten-concept class*

$$\mathcal{C}_W^a : \quad \texttt{00000 10000 01010 10110 11110 11001 00101 11101 01011 00111},$$

which has $\mathrm{VCD} = 2$ *and* $\mathrm{TS}_{\min} = 3$*, is contained in* $\mathcal{C}_{20}$ *and in* $\mathcal{C}_{21}$*, and the ten-concept class*

$$\mathcal{C}_W^b : \quad \texttt{00000 01100 10010 10110 01110 00001 10101 11101 01011 11011},$$

also with $\mathrm{VCD} = 2$ *and* $\mathrm{TS}_{\min} = 3$*, is contained in* $\mathcal{C}_{22}$*.*

Proof. Every clause is a finite check against the definitions of Section 2: for each of the three classes, some triple is shattered and no four-element set is; some concept has a four-point teaching set and no concept has a three-point one (for $\mathcal{C}_{20}$ the empty concept is taught by $\{0, 1, 2, 3\}$, and there are 50 pairs (concept, four-point teaching set) in all). The two ten-concept subclasses are checked the same way, and the three containments are checked element by element. The checks are carried out by kernel reduction in the development of Section 8; nothing in the theorem depends on the census that found the classes. □

Remark 4.2 (the subclasses are copies of $\mathcal{C}_W$; computation)*.* $\mathcal{C}_W^a$ and $\mathcal{C}_W^b$ are not merely classes with the parameters of Warmuth's class: each is a G_5-image of $\mathcal{C}_W$, since both have canonical form `0x03586428`, which is the canonical form of $\mathcal{C}_W$. That identification is a finite computation carried out outside the development and is not part of Theorem 4.1, whose machine-checked content is the containments together with $\mathrm{VCD} = 2$ and $\mathrm{TS}_{\min} = 3$. It also follows from the census of Section 5, where the 192 five-point classes with $\mathrm{VCD} = 2$ and $\mathrm{TS}_{\min} = 3$ turn out to be a single G_5-orbit — that of $\mathcal{C}_W$ — so that on five points those two parameter values already force the class to be a copy of Warmuth's.

Corollary 4.3. *Five is the least size of a domain carrying a concept class with* $\mathrm{VCD} = 3$ *and* $\mathrm{TS}_{\min} = 4$*; more generally, no class on at most four points has* $\mathrm{TS}_{\min} > \mathrm{VCD}$.

Proof. The upper bound is Theorem 4.1. For the lower bound, if $\mathrm{TS}_{\min}(\mathcal{C}) > \mathrm{VCD}(\mathcal{C})$ then $\mathrm{RTD}(\mathcal{C}) > \mathrm{VCD}(\mathcal{C})$ by (1), and Theorem 15 of [2] — "Let $\mathcal{C}$ be a concept class over domain X such that $\mathrm{RTD}(\mathcal{C}) > \mathrm{VCD}(\mathcal{C})$. Then $|\mathcal{C}| \geq 10$ and $|X| \geq 5$" — gives $|X| \geq 5$. This half is theirs and is not re-proved here; it is also the one statement in this paper with no machine-checked content of its own, and the corollary is correspondingly not part of the formal development. □

Three comments. First, on the construction the theorem replaces. Lemma 16 of [2] states that VCD, $\mathrm{TS}_{\min}$ and RTD are additive under the disjoint union $\mathcal{C}_1 \uplus \mathcal{C}_2 = \{A \cup B : A \in \mathcal{C}_1, B \in \mathcal{C}_2\}$ of classes on disjoint domains. Applied to $\mathcal{C}_W$ and the two-concept class $\{\emptyset, \{p\}\}$ on a fresh point p it yields a class with $\mathrm{VCD} = 3$, $\mathrm{TS}_{\min} = 4$ and 20 concepts — on six points. The class $\mathcal{C}_{20}$ has the same number of concepts on five. Second, on the table of [3]. Its Lemma 5 (the same additivity, for Cartesian products) is what lets its authors "focus on finding small concept classes with $\mathrm{RTD}(\mathcal{C}) > (3/2) \cdot \mathrm{VCD}(\mathcal{C})$", and $4/3 < 3/2$; the cell $(\mathrm{VCD}, \mathrm{RTD}, n) = (3, 4, 5)$ is therefore absent from their Figure 3 because it was not asked, not because it was out of reach — their encoding would decide it in seconds. Third, on the companion paper, as promised in Section 1.4: the three orbits of $\mathcal{C}_{20}, \mathcal{C}_{21}, \mathcal{C}_{22}$ under G_5, of sizes 192, 384 and 192, are the three orbits of 20-, 21- and 22-concept classes without a frequency-1 fragment listed in Remark 5.5 of [9], by Lemma 2.6; the value $\mathrm{TS}_{\min} = 4$ is not in that paper, and neither is the containment of $\mathcal{C}_W$.

Remark 4.4 (non-vacuity). $\mathrm{VCD} = 3$ does not force $\mathrm{TS}_{\min} = 4$ on five points: the power set of $\{0, 1, 2\}$, viewed on X_5, has $\mathrm{VCD} = 3$ and $\mathrm{TS}_{\min} = 3$; and the full power set 2^{X_5} has $\mathrm{VCD} = \mathrm{TS}_{\min} = 5$. Both are checked in the development as negative controls, together with $\mathrm{TS}_{\min}(\mathcal{C}_W) \neq 4$, $\mathrm{VCD}(\mathcal{C}_W) \neq 3$, $\mathrm{TS}_{\min}(\mathcal{C}_{20}) \neq 3$ and $\mathrm{TS}_{\min}(\mathcal{C}_W) \neq 2$.

5. The census

This section reports computations that are not part of the formal development. They support no theorem above; the theorems above do not depend on them; and they are what found the classes of Theorem 4.1.

A program in C scans every nonempty class on X_n for $n \leq 5$ — at $n = 5$ all $2^{32} - 1 = 4\,294\,967\,295$ masks — and computes VCD and $\mathrm{TS}_{\min}$ from one table of "coset masks" $\{c : c \cap S = v\}$ for $S \subseteq X_n$, $v \subseteq S$: $\mathrm{TS}_{\min}(\mathcal{C}) \leq t$ iff for some $|S| = t$ and some $v \subseteq S$ exactly one concept of $\mathcal{C}$ has trace v on S, and S is shattered iff every $v \subseteq S$ is the trace of some concept. At $n = 5$ this took 483 processor-seconds with 2.1 MB of memory. The joint distribution of $(\mathrm{VCD}, \mathrm{TS}_{\min})$ over the classes on five points is

	$\mathrm{TS}_{\min}$ 0	1	2	3	4	5
VCD 0	32					
1		44 880				
2		6 545 920	159 484 456	**192**		
3		3 279 040	1 903 792 960	2 051 732 972	**768**	
4		320	7 073 440	159 316 040	3 696 274	
5						1

and the tables for $n = 3$ and $n = 4$ are $(\mathrm{VCD}, \mathrm{TS}_{\min}) \mapsto$ count: $(0,0)\,8$, $(1,1)\,116$, $(2,1)\,48$, $(2,2)\,82$, $(3,3)\,1$ for $n = 3$, and $(0,0)\,16$, $(1,1)\,1928$, $(2,1)\,8960$, $(2,2)\,38\,636$, $(3,1)\,128$, $(3,2)\,12\,528$, $(3,3)\,3338$, $(4,4)\,1$ for $n = 4$. Each table sums to $2^{2^n} - 1$, which is the arithmetic check on the scan; the $n \leq 4$ tables were reproduced cell for cell by an independent implementation in Python. Reading off the largest $\mathrm{TS}_{\min}$ in each row:
Five readings.

Only two cells lie above the diagonal, and both by exactly one. On at most five points $\mathrm{TS}_{\min} - \mathrm{VCD} \in \{0, 1\}$, and the value 1 occurs only at $n = 5$, only at $d \in \{2, 3\}$.

n	$M(n,0)$	$M(n,1)$	$M(n,2)$	$M(n,3)$	$M(n,4)$	$M(n,5)$
1	0	1				
2	0	1	2			
3	0	1	2	3		
4	0	1	2	3	4	
5	0	1	**3**	**4**	4	5

The $(2,3)$ *cell is Warmuth's class, and nothing else.* A second program, an independent implementation of both predicates, canonicalises every class of the two exceptional cells under G_5: the 192 classes with $(\mathrm{VCD}, \mathrm{TS}_{\min}) = (2,3)$ form a single orbit, that of $\mathcal{C}_W$, all of ten concepts. So on five points a class with $\mathrm{VCD} = 2$ and $\mathrm{TS}_{\min} = 3$ is, up to relabelling and complementation, Warmuth's class — sharpening, on five points, the minimality statement of [2]. The 768 classes with $(3,4)$ form the three orbits of $\mathcal{C}_{20}$, $\mathcal{C}_{21}$, $\mathcal{C}_{22}$, of sizes 192, 384, 192; and $192 + 192 + 384 + 192 = 960$, the two cell counts again. A further check, in Python over the full 192-element orbit of $\mathcal{C}_W$ — not part of the C scan — reports that every one of the 768 contains a G_5-image of $\mathcal{C}_W$. For the three representatives, Theorem 4.1 kernel-checks the containment of a concrete ten-concept subclass with $\mathrm{VCD} = 2$ and $\mathrm{TS}_{\min} = 3$, and Remark 4.2 identifies that subclass as an image of $\mathcal{C}_W$ by its canonical form.

What the literature already implies. The rows $n \leq 4$ follow from Theorem 15 of [2] through (1) (no class on at most four points has $\mathrm{TS}_{\min} > \mathrm{VCD}$) together with $M(n,d) \geq d$; the entries $M(n,0) = 0$ and $M(n,1) = 1$ are Lemma 2.5 and Kuhlmann's theorem as restated in Corollary 9 of [2] and Definition 5 of [5]; and $M(5,2) = 3$ follows from $\mathcal{C}_W$ together with the unsatisfiable row $(2,4,7)$ of [3] and (1), if one trusts the solver. The entry $M(5,3) = 4$ is Theorem 4.1 with the matching upper bound from the scan, and the entries $M(5,4) = 4$, $M(5,5) = 5$ are the scan's. Conversely the scan re-derives, without a solver, the shadow on at most five points of the two rows $(2,4,7)$ and $(3,5,7)$ of [3]: no class on X_5 with $\mathrm{VCD} \leq 2$ has $\mathrm{TS}_{\min} \geq 4$, and none with $\mathrm{VCD} \leq 3$ has $\mathrm{TS}_{\min} \geq 5$, so by Lemma 4 of [2] none has $\mathrm{RTD} \geq 4$, respectively ≥ 5.

Least sizes. The least number of concepts in the $(2,3)$ cell is 10 and in the $(3,4)$ cell is 20. The first is the bound $|\mathcal{C}| \geq 10$ of Theorem 15 of [2], valid over every domain; whether 20 is least over every domain is open (Section 7).

Comparison with the companion census. The four orbits are the four fragment-free orbits of Remark 5.5 of [9], where their representatives have non-clashing teaching dimension $2,2,2,3$; here they have $\mathrm{TS}_{\min} = 3,4,4,4$.

6. The greedy algorithm on four and five points

6.1. One step is sound.

Lemma 6.1 (one restriction costs at most one point)**.** *Let $\mathcal{C}$ be a concept class on X_n, $x \in X_n$, $b \in \{0,1\}$ and $k \in \mathbb{N}$. If some concept of $\mathcal{C}|_{x,b}$ has a teaching set of at most k points inside $\mathcal{C}|_{x,b}$, then some concept of $\mathcal{C}$ has a teaching set of at most $k+1$ points inside $\mathcal{C}$.*

Proof. Let S teach c in $\mathcal{C}|_{x,b}$ with $|S| \leq k$. Then $S \cup \{x\}$ teaches c in $\mathcal{C}$: if $c' \in \mathcal{C}$ agrees with c on $S \cup \{x\}$, then in particular $[x \in c'] = [x \in c] = b$, so $c' \in \mathcal{C}|_{x,b}$, and agreement on S forces $c' = c$. □

Iterating the lemma along a run of the procedure is why the returned set S teaches the surviving concept: the final class is a singleton, taught by $\emptyset$, and each earlier step adds one point. Hence $g_{\min}(\mathcal{C}) \geq \mathrm{TS}_{\min}(\mathcal{C})$, the gap $g_{\min} - \mathrm{TS}_{\min}$ is well defined and non-negative, and so is $g_{\max} - \mathrm{TS}_{\min}$. The lemma holds for every n and is proved, not evaluated, in the development. The *iteration* is not: that the single step composes along a run, and hence that $g_{\min} \geq \mathrm{TS}_{\min}$ in general, is an argument in this prose only, and so is the fuel bound of Section 2. What is machine-checked is the one step, for every n, and the individual values of the next subsection.

6.2. The first failures.

Theorem 6.2. (1) *Let $\mathcal{C}_9$ be the class on X_4 consisting of the empty concept and the eight concepts that contain exactly one of the points $2, 3$:*

$$\texttt{0000 0010 1010 0110 1110 0001 1001 0101 1101} \qquad (\texttt{0x00000ff1}).$$

Then $\mathrm{TS}_{\min}(\mathcal{C}_9) = 2$, $\mathrm{VCD}(\mathcal{C}_9) = 3$, and $g_{\min}(\mathcal{C}_9) = g_{\max}(\mathcal{C}_9) = 3$: every run of the greedy procedure at $k = 1$ returns three points.

(2) *Let $\mathcal{C}_{17}$ be the class on X_5 consisting of the empty concept and the sixteen concepts that contain exactly one of the points $3, 4$:*

$$\texttt{00000 00010 10010 01010 11010 00110 10110 01110 11110}$$
$$\texttt{00001 10001 01001 11001 00101 10101 01101 11101}$$

(mask `0x00ffff01`*). Then $\mathrm{TS}_{\min}(\mathcal{C}_{17}) = 2$, $\mathrm{VCD}(\mathcal{C}_{17}) = 4$, and $g_{\min}(\mathcal{C}_{17}) = g_{\max}(\mathcal{C}_{17}) = 4$: every run returns four points.*

Proof. Both parts are finite checks against the definitions, carried out by kernel reduction in the development of Section 8, with the procedure implemented there as the pair of functions returning the least and the greatest output size over all tie-breakings and with the iteration bound of Section 2 set to n. The reason is transparent and worth recording. In $\mathcal{C}_{17}$ every restriction $\mathcal{C}|_{x,1}$ has eight concepts and every $\mathcal{C}|_{x,0}$ has nine, so the first step of every run takes some $(x, 1)$. For $x \in \{0, 1, 2\}$ the result is the set of concepts containing x and exactly one of $3, 4$, in which every restriction at every remaining point has four concepts; for $x \in \{3, 4\}$ it is a full power set on $\{0, 1, 2\}$. In either case each further step halves the class, and three more steps are needed: four in all. The empty concept, however, is taught by $\{3, 4\}$ — it is the only concept containing neither — and no concept has a one-point teaching set, since every restriction has at least eight concepts; so $\mathrm{TS}_{\min} = 2$. The optimal teaching set is reached only through the *larger* restriction $\mathcal{C}|_{3,0}$, which the procedure never takes. The same account holds for $\mathcal{C}_9$ with $\{2, 3\}$, restrictions of sizes four and five, and one fewer step. □

The two classes have the same shape, on four and on five points. Their gaps are the largest possible on their domains and both are unavoidable, in the sense that no tie-breaking helps (Remark 6.3); $\mathcal{C}_9$ is the numerically least mask on X_4 on which *some* run of the procedure is suboptimal, and also the least on which *every* run is; on X_3 there is no such mask at all. Neither is a counterexample to $\mathrm{TS}_{\min} = O(d)$ — their VC dimensions exceed their $\mathrm{TS}_{\min}$ — and neither is meant to be: as [1] writes of its own constructions, "Our constructions are not counterexamples to the general $\mathrm{TS}_{\min} = O(d)$ conjecture as they have small $\mathrm{TS}_{\min}$." The theorem is the exact small-domain shadow of Theorem 1 of [1], which makes the gap grow like $\Omega(\log|\mathcal{C}|)$. Two negative controls sit beside it in the development: on the full power set 2^{X_4} the procedure is optimal under every tie-breaking ($g_{\min} = g_{\max} = \mathrm{TS}_{\min} = 4$), so the two functions do not simply overcount; and adding one more unit to the iteration bound changes nothing on $\mathcal{C}_9$.

Remark 6.3 (the greedy gap census; computation). A third program computes $g_{\min}$ and $g_{\max}$ for every class on X_n, $n \le 5$, by a forward dynamic program over masks — every restriction of a class is a sub-mask, hence numerically smaller — together with $\mathrm{TS}_{\min}$ and VCD; at $n = 5$ it took 1 367 processor-seconds of user time and 4.2 GB of resident memory, as recorded, and the $n \le 4$ results were reproduced in Python. The largest value of $g_{\max} - \mathrm{TS}_{\min}$ over all classes on n points is

$$0,\ 0,\ 0,\ 1,\ 2 \qquad \text{for } n = 1, 2, 3, 4, 5,$$

and the largest value of $g_{\min} - \mathrm{TS}_{\min}$ is the same; the numerically least witnesses are $\mathcal{C}_9$ and $\mathcal{C}_{17}$, each its own canonical form. On four points 3 408 classes have $g_{\max} > \mathrm{TS}_{\min}$ and 432 have $g_{\min} > \mathrm{TS}_{\min}$, all of them with $\mathrm{VCD} = 3$ and $\mathrm{TS}_{\min} = 2$, the smallest with eight concepts (nine among those with $g_{\min} > \mathrm{TS}_{\min}$). On five points, counting by $g_{\max} - \mathrm{TS}_{\min}$ and excluding the 32 one-concept classes, 411 929 440 classes have gap 1 at $\mathrm{TS}_{\min} = 2$, 95 213 440 have gap 1 at $\mathrm{TS}_{\min} = 3$, and 275 040 have gap 2, all at $\mathrm{TS}_{\min} = 2$; the largest gap by VC dimension $d = 1, \dots, 5$ is $0, 1, 2, 2, 0$. The $\mathrm{TS}_{\min}$ marginals of this table agree with those of the census of Section 5 (9 870 160, 2 070 350 856, 2 211 049 204, 3 697 042, 1 at $\mathrm{TS}_{\min} = 1, \dots, 5$), which is the cross-check between the two programs.

Remark 6.4 (the rectangle construction at $N \leq 3$; computation). Theorem 1 of [1] builds, for every N, a class $\mathcal{F}_N$ of axis-aligned rectangles on a domain $V_1 \cup \cdots \cup V_N$, where V_i has a centre z_i and w_i points above and below it, with four families of prefix rectangles per level, the "level 2" ones enlarged to contain all lower levels; it takes $w_i = 2^{10i}$ and uses that choice only through its Claim 2, that the lower levels together have fewer than half as many concepts as level i. The construction was re-implemented from that paragraph for arbitrary weights w and run at $N \leq 3$, with the procedure, $\mathrm{TS}_{\min}$ and VCD computed exactly ($|\mathcal{F}|$ counts distinct concepts; at level 1 the enlarged and unenlarged rectangles coincide):

w	$\lvert X\rvert$	$\lvert\mathcal{F}\rvert$	Claim 2 holds	$(g_{\min}, g_{\max})$	$\mathrm{TS}_{\min}$	VCD
(1)	3	3	—	$(1,1)$	1	1
$(1,2)$	8	13	yes	$(2,2)$	2	3
$(1,1,1)$	9	15	no, at both levels	$(3,3)$	2	3
$(1,2,3)$	15	27	no, at level 3	$(3,3)$	2	3
$(1,2,7)$	23	43	yes, at every level	$(3,3)$	2	3

So at $N \leq 3$ the conclusion of Theorem 1 (output at least N) and the statement of Appendix C of [1] ("The $\mathrm{TS}_{\min}$ for this class is at most 2") are reproduced, the latter with equality for $N \geq 2$; and the weights are far from necessary at this size: $w = (1,1,1)$ on nine points already forces every run to three points, although the sufficient condition of Claim 2 fails at both of its levels, while the paper's own $w_i = 2^{10i}$ would place $N = 3$ on $2(2^{10} + 2^{20} + 2^{30}) + 3 \approx 2.1 \cdot 10^9$ points. The least w satisfying Claim 2 at every level is $(1,2,7)$, on 23 points. This is a small-N observation and nothing is claimed for general N; $N = 4$ was not completed (stopped after ten minutes of wall-clock time on the tie-breaking tree). Cost about 0.17 processor-hours.

7. What remains

Six points. The census is a scan over 2^{64} masks and is out of reach by the same route; the orbit-representative trick of [9] does not help, because enumerating representatives of 2^{64} subsets under a group of order 46 080 is the same size of task. The tractable sub-question is whether $M(6,3) = 5$, i.e. whether $\mathrm{TS}_{\min} - \mathrm{VCD}$ reaches 2 on six points; a satisfiability encoding in the style of [3] would settle it. The unsatisfiable rows of [3] at seven and eight points rest on solver runs with no proof certificate; a certified refutation, or a structural argument, would put them on the same footing as Corollary 4.3.

Least size at $(3,4)$. The census gives 20 as the least number of concepts of a five-point class with $\mathrm{VCD} = 3$ and $\mathrm{TS}_{\min} = 4$. Is 20 least over all domains? The argument of Lemma 14 of [2], which gives $|\mathcal{C}[x,\ell]| \geq 5$ for every labelled point when $\mathrm{VCD} = 2$ and $\mathrm{TS}_{\min} = 3$, may generalise to $|\mathcal{C}[x,\ell]| \geq 10$ at $\mathrm{VCD} = 3$, $\mathrm{TS}_{\min} = 4$, which would give exactly 20; we have not carried this out.

The greedy gap. The witnesses of Theorem 6.2 are the cases $n = 4, 5$ of one shape — the empty concept together with every concept containing exactly one of two fixed points — and the census gives the largest gap $0,0,0,1,2$ for $n \leq 5$. Is the largest gap on n points equal to $n - 3$ for every $n \geq 3$, and is this shape always extremal? The census says nothing beyond $n = 5$. The procedure at $k = 2$ is not studied here at all; on five points it is the same dynamic program with $2n + 4\binom{n}{2} = 50$ candidate restrictions per class instead of $2n = 10$, priced at about 1.4 processor-hours.

$\mathrm{TS}_{\min}$ *in a proof library. Mathlib* has shattering, the VC dimension of a finite set family and the Sauer–Shelah lemma, but nothing about teaching sets; the development below defines its own shattering predicate so that every statement is one evaluation away from the definitions, and a bridge to the library's VC dimension would make these statements composable with Sauer–Shelah.

8. Verification

Lemmas 2.5 and 6.1, Proposition 3.1 and Theorems 4.1 and 6.2 have been formally verified in Lean 4 against *Mathlib* [10, 11]; their statements are reproduced verbatim in Appendix A. The development is three modules of about 510 lines in all: `Defs` (the bit encoding of concepts and classes, the predicates "shatters", "VCD $\leq d$", "VCD $\geq d$", "teaches" and "$\mathrm{TS}_{\min} \leq k$" with their decidability instances, and Lemma 2.5), `Witnesses` ($\mathcal{C}_W$, $\mathcal{C}_{20}$, $\mathcal{C}_{21}$, $\mathcal{C}_{22}$, the two copies of $\mathcal{C}_W$, the two non-vacuity classes;

Proposition 3.1, Theorem 4.1 and three negative controls), and `Greedy` (the restriction, the set of least nonempty restrictions, the two output-size functions, Lemma 6.1, $\mathcal{C}_9$, $\mathcal{C}_{17}$, Theorem 6.2 and three controls). They hold 28 theorem declarations in all — 6, 14 and 8 respectively — of which fourteen carry the statements of this paper. Every statement about a specific class is discharged by `decide`, that is, by reduction inside the kernel; no compiled evaluation (`native_decide`) is used anywhere, there is no `sorry`, and no axiom is declared by hand. A print of the axiom set of each of the 28 returns `propext`, `Classical.choice` and `Quot.sound` for 27 of them and `propext`, `Quot.sound` for the monotonicity lemma $\mathrm{TS}_{\min} \le k \Rightarrow \mathrm{TS}_{\min} \le k'$, and nothing else. The non-vacuity control, which evaluates the full 32-concept power set, needed a raised recursion limit and nothing else. Checking the three files one at a time took 45 s, 60 s wall / 126 processor-seconds with 9.3 GB peak, and 26 s wall / 30 processor-seconds with 7.9 GB peak respectively, as recorded. The negative controls are those named in Sections 4 and 6: a too-large claim refuted ($\mathrm{TS}_{\min}(\mathcal{C}_W) \ne 4$, $\mathrm{VCD}(\mathcal{C}_W) \ne 3$), a too-small claim refuted ($\mathrm{TS}_{\min}(\mathcal{C}_{20}) \ne 3$, $\mathrm{TS}_{\min}(\mathcal{C}_W) \ne 2$), non-vacuity ($\mathrm{VCD} = 3$ with $\mathrm{TS}_{\min} = 3$; $\mathrm{VCD} = \mathrm{TS}_{\min} = 5$), a zero-gap control for the greedy functions, the value $g_{\min}(\mathcal{C}_9) \ne 2$, and the iteration bound.

Outside the formal development, and labelled as such where they occur, are the census of Section 5 (two C programs, the second an independent implementation of both predicates used for the orbit analysis; 483 processor-seconds at $n = 5$), the greedy gap census of Remark 6.3 (a third C program; 1 367 processor-seconds of user time and 4.2 GB), the rectangle reproduction of Remark 6.4 (Python; 0.17 processor-hours), and the Python re-implementations that reproduced every $n \le 4$ table. Their arithmetic checks are stated where they are used: the row sums $2^{2^n} - 1$, the orbit weights $960 = 192 + 768$, and the agreement of the $\mathrm{TS}_{\min}$ marginals between the two censuses. Total cost of everything reported, including the formal checks, about 1.1 processor-hours. The source of the development and of the programs is currently held in a private repository: repository reference to be added at submission.

On the reference list. Statement numbers attributed to [1] (Algorithm 1 and its tie-breaking footnote 3, Theorem 1, Claim 2, Theorem 4, Appendix C) are those printed in the COLT 2025 proceedings PDF. They were compared against the arXiv version 1 PDF, which the ledger of this project cites: Algorithm 1, footnote 3, Theorem 1, Claim 2 and Theorem 4 are identically numbered and worded there, and the appendix bounding $\mathrm{TS}_{\min}$ for the constructions, Appendix C of the proceedings, is Appendix A of the arXiv version. A reader following the arXiv version should read “Appendix A” for “Appendix C” throughout. The e-print source was also read in full but does not compile with a current LaTeX distribution (a package clash in its class file). Statement and page numbers attributed to [2] are those of the journal PDF, and those attributed to [3] agree between the NeurIPS and the ECCC versions. The numbering of [9] is that of its version 1.

Page ranges for the three papers published by *Proceedings of Machine Learning Research* — [1], [4] and [5] — are the volume-wide ranges printed by the PMLR volume indexes and recorded by DBLP, which agree in all three cases. The camera-ready PDFs themselves carry a different, per-paper numbering in their running heads (“vol 291:1–16”, “vol 40:1–3”, “vol 65:1–10”), and never print the volume-wide numbers; a page-level reference to one of those PDFs must be offset accordingly.

Appendix A. The formal statements

The Lean statements corresponding to the results above, verbatim (statements only; proofs, docstrings and attributes omitted), grouped by module. The domain is `Fin n` $= X_n$; `conc n m` is the concept with bit mask `m` (bit i set iff point i belongs to it) and `cls n l` the class with the masks in the list `l`; `Shatters`, `VCDLe C d`, `VCDGe C d`, `Teaches C c S` and `TSMinLe C k` are “shatters”, $\mathrm{VCD}(\mathcal{C}) \le d$, $\mathrm{VCD}(\mathcal{C}) \ge d$, “$S$ teaches c in $\mathcal{C}$” and $\mathrm{TS}_{\min}(\mathcal{C}) \le k$; `restrictC C x b` is $\mathcal{C}|_{x,b}$; `greedyMin f C` and `greedyMax f C` are $g_{\min}(\mathcal{C})$ and $g_{\max}(\mathcal{C})$ with the iteration bound `f`; `CW`, `E20`, `E21`, `E22`, `CWa`, `CWb`, `Sub9`, `Gap17` are $\mathcal{C}_W$, $\mathcal{C}_{20}$, $\mathcal{C}_{21}$, $\mathcal{C}_{22}$, the two copies of $\mathcal{C}_W$ in Theorem 4.1, $\mathcal{C}_9$ and $\mathcal{C}_{17}$.

Defs.lean

```
def conc (n m : ℕ) : Finset (Fin n) := Finset.univ.filter fun i => m / 2 ^ (i : ℕ) % 2 = 1
def cls (n : ℕ) (l : List ℕ) : Finset (Finset (Fin n)) := (l.map (conc n)).toFinset
def Shatters (C : Finset (Finset (Fin n))) (X : Finset (Fin n)) : Prop :=
  ∀ p, p ⊆ X → ∃ c ∈ C, c ∩ X = p
def VCDLe (C : Finset (Finset (Fin n))) (d : ℕ) : Prop :=
```

```
  ∀ X : Finset (Fin n), Shatters C X → X.card ≤ d
def VCDGe (C : Finset (Finset (Fin n))) (d : ℕ) : Prop :=
  ∃ X : Finset (Fin n), Shatters C X ∧ X.card = d
def Teaches (C : Finset (Finset (Fin n))) (c S : Finset (Fin n)) : Prop :=
  ∀ c' ∈ C, c' ∩ S = c ∩ S → c' = c
def TSMinLe (C : Finset (Finset (Fin n))) (k : ℕ) : Prop :=
  ∃ c ∈ C, ∃ S : Finset (Fin n), S.card ≤ k ∧ Teaches C c S
theorem tsMinLe_of_nonempty {C : Finset (Finset (Fin n))} (h : C.Nonempty) : TSMinLe C n
theorem tsMinLe_zero_iff {C : Finset (Finset (Fin n))} : TSMinLe C 0 ↔ C.card = 1
theorem vcdGe_one_iff {C : Finset (Finset (Fin n))} : VCDGe C 1 ↔ 2 ≤ C.card
theorem tsMinLe_zero_of_vcdLe_zero {C : Finset (Finset (Fin n))} (hne : C.Nonempty)
    (h : VCDLe C 0) : TSMinLe C 0
```

Witnesses.lean

```
def CW : Finset (Finset (Fin 5)) := cls 5 [3, 5, 10, 13, 14, 19, 20, 22, 24, 25]
theorem CW_vcd : VCDLe CW 2 ∧ VCDGe CW 2
theorem CW_tsmin : TSMinLe CW 3 ∧ ¬ TSMinLe CW 2
def E20 : Finset (Finset (Fin 5)) :=
  cls 5 [0, 1, 2, 4, 7, 9, 10, 13, 14, 15, 17, 19, 20, 22, 23, 24, 26, 27, 28, 29]
def E21 : Finset (Finset (Fin 5)) :=
  cls 5 [0, 1, 2, 4, 6, 7, 9, 10, 13, 14, 15, 17, 19, 20, 22, 23, 24, 26, 27, 28, 29]
def E22 : Finset (Finset (Fin 5)) :=
  cls 5 [0, 1, 2, 5, 6, 9, 11, 12, 13, 14, 15, 16, 18, 19, 21, 22, 23, 24, 26, 27, 28, 29]
theorem E20_spec : (VCDLe E20 3 ∧ VCDGe E20 3) ∧ TSMinLe E20 4 ∧ ¬ TSMinLe E20 3
theorem E21_spec : (VCDLe E21 3 ∧ VCDGe E21 3) ∧ TSMinLe E21 4 ∧ ¬ TSMinLe E21 3
theorem E22_spec : (VCDLe E22 3 ∧ VCDGe E22 3) ∧ TSMinLe E22 4 ∧ ¬ TSMinLe E22 3
theorem exists_vcd_three_tsMin_four :
    ∃ C : Finset (Finset (Fin 5)), (VCDLe C 3 ∧ VCDGe C 3) ∧ TSMinLe C 4 ∧ ¬ TSMinLe C 3
def CWa : Finset (Finset (Fin 5)) := cls 5 [0, 1, 10, 13, 15, 19, 20, 23, 26, 28]
def CWb : Finset (Finset (Fin 5)) := cls 5 [0, 6, 9, 13, 14, 16, 21, 23, 26, 27]
theorem CW_copies :
    ((VCDLe CWa 2 ∧ VCDGe CWa 2) ∧ TSMinLe CWa 3 ∧ ¬ TSMinLe CWa 2) ∧
    ((VCDLe CWb 2 ∧ VCDGe CWb 2) ∧ TSMinLe CWb 3 ∧ ¬ TSMinLe CWb 2) ∧
    CWa ⊆ E20 ∧ CWa ⊆ E21 ∧ CWb ⊆ E22
theorem control_too_large : TSMinLe CW 3 ∧ ¬ VCDGe CW 3
theorem control_too_small : ¬ TSMinLe E20 3 ∧ ¬ TSMinLe CW 2
def Cube3 : Finset (Finset (Fin 5)) := cls 5 [0, 1, 2, 3, 4, 5, 6, 7]
def Full5 : Finset (Finset (Fin 5)) := cls 5 (List.range 32)
theorem control_nonvacuous :
    ((VCDLe Cube3 3 ∧ VCDGe Cube3 3) ∧ TSMinLe Cube3 3 ∧ ¬ TSMinLe Cube3 2) ∧
    ((VCDLe Full5 5 ∧ VCDGe Full5 5) ∧ TSMinLe Full5 5 ∧ ¬ TSMinLe Full5 4)
```

Greedy.lean

```
def restrictC (C : Finset (Finset (Fin n))) (x : Fin n) (b : Bool) :
    Finset (Finset (Fin n)) := C.filter fun c => decide (x ∈ c) = b
def pairs (n : ℕ) : List (Fin n × Bool) :=
  (List.finRange n).map (fun x => (x, true)) ++ (List.finRange n).map (fun x => (x, false))
def minBlock (C : Finset (Finset (Fin n))) : ℕ :=
  (pairs n).foldr (fun p m => if 0 < (restrictC C p.1 p.2).card
    then min (restrictC C p.1 p.2).card m else m) C.card
def choices (C : Finset (Finset (Fin n))) : List (Fin n × Bool) :=
  (pairs n).filter fun p =>
    decide (0 < (restrictC C p.1 p.2).card ∧ (restrictC C p.1 p.2).card = minBlock C)
def greedyMax : ℕ → Finset (Finset (Fin n)) → ℕ
  | 0, _ => 0
  | f + 1, C =>
      if C.card ≤ 1 then 0
      else 1 + ((choices C).map fun p => greedyMax f (restrictC C p.1 p.2)).foldr max 0
def greedyMin : ℕ → Finset (Finset (Fin n)) → ℕ
  | 0, _ => 0
  | f + 1, C =>
      if C.card ≤ 1 then 0
      else 1 + ((choices C).map fun p => greedyMin f (restrictC C p.1 p.2)).foldr min n
theorem tsMinLe_restrict_succ {C : Finset (Finset (Fin n))} {x : Fin n} {b : Bool} {k : ℕ}
    (h : TSMinLe (restrictC C x b) k) : TSMinLe C (k + 1)
def Sub9 : Finset (Finset (Fin 4)) := cls 4 [0, 4, 5, 6, 7, 8, 9, 10, 11]
theorem Sub9_greedy_gap :
    (TSMinLe Sub9 2 ∧ ¬ TSMinLe Sub9 1) ∧ greedyMin 4 Sub9 = 3 ∧ greedyMax 4 Sub9 = 3 ∧
      (VCDLe Sub9 3 ∧ VCDGe Sub9 3)
def Gap17 : Finset (Finset (Fin 5)) :=
  cls 5 [0, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23]
theorem Gap17_greedy_gap :
    (TSMinLe Gap17 2 ∧ ¬ TSMinLe Gap17 1) ∧ greedyMin 5 Gap17 = 4 ∧ greedyMax 5 Gap17 = 4 ∧
      (VCDLe Gap17 4 ∧ VCDGe Gap17 4)
def Full4 : Finset (Finset (Fin 4)) := cls 4 (List.range 16)
theorem control_no_gap :
```

```
    (TSMinLe Full4 4 ∧ ¬ TSMinLe Full4 3) ∧ greedyMin 4 Full4 = 4 ∧ greedyMax 4 Full4 = 4
theorem control_greedy_not_two : greedyMin 4 Sub9 ≠ 2 ∧ greedyMax 4 Sub9 ≠ 4
theorem control_fuel : greedyMin 5 Sub9 = 3 ∧ greedyMax 5 Sub9 = 3
```

References

[1] S. Compton, C. Pabbaraju and N. Zhivotovskiy, *Lower bounds for greedy teaching set constructions*, in: Proceedings of the 38th Conference on Learning Theory (COLT 2025), Proceedings of Machine Learning Research **291** (2025), 1314–1329; preprint arXiv:2505.03223v1 [stat.ML], 6 May 2025 (cross-listed cs.DS, cs.LG, math.CO). The source of the question, of Algorithm 1 and its tie-breaking footnote 3, of Theorem 1 ("Rectangles Lower Bound for $k = 1$": the classes $\mathcal{F}_N$ of axis-aligned rectangles on which the $k = 1$ procedure returns a teaching set of size at least N), of Claim 2 ("A level dominates all lower levels"), of Theorem 4 ("Lower Bound for $k \geq 2$"), and of Appendix C, "Bounding $\mathrm{TS}_{\min}$ for our constructions" ("The $\mathrm{TS}_{\min}$ for this class is at most 2"). Both versions were read in full from their PDFs, and the arXiv e-print source as well. Algorithm 1, footnote 3, Theorem 1, Claim 2 and Theorem 4 carry the same numbers and the same wording in both; the only shift among the items cited here is the appendix, which is Appendix A of the arXiv version and Appendix C of the proceedings version (the latter adds two earlier appendices). Numbers quoted in this paper are the proceedings version's.

[2] T. Doliwa, G. Fan, H. U. Simon and S. Zilles, *Recursive teaching dimension, VC-dimension and sample compression*, Journal of Machine Learning Research **15** (2014), 3107–3131. Read in full from the journal's PDF; page numbers are the PDF's own. Source of the name and notation $\mathrm{TS}_{\min}$ (p. 3112), of Lemma 4 ($\mathrm{RTD}(\mathcal{C}) = \max_{\mathcal{C}' \subseteq \mathcal{C}} \mathrm{TS}_{\min}(\mathcal{C}')$, p. 3113), of Corollary 9(1) ("If $\mathrm{VCD}(\mathcal{C}) = 1$, then $\mathrm{RTD}(\mathcal{C}) = \mathrm{SDC}(\mathcal{C}) = 1$", p. 3116), of Warmuth's class as Figure 1(a) with the sentence "Brute-force enumeration shows that $\mathrm{RTD}(\mathcal{C}_W) = \mathrm{TS}_{\min}(\mathcal{C}_W) = 3$ and $\mathrm{VCD}(\mathcal{C}_W) = 2$" (p. 3119), of Lemma 14 (p. 3120), of Theorem 15 (p. 3121), and of Lemma 16, additivity of VCD, $\mathrm{TS}_{\min}$ and RTD under disjoint union (p. 3122). Its p. 3119 records that the smallest class of Kuhlmann's family has 24 concepts on 16 points, and its p. 3128 records a flaw in the proof of Theorem 4 of [6] — a statement about the *average*-case teaching dimension, not about $\mathrm{TS}_{\min}$, whose claim it then reproves as its own Theorem 32.

[3] X. Chen, Y. Cheng and B. Tang, *On the recursive teaching dimension of VC classes*, Advances in Neural Information Processing Systems 29 (NIPS 2016), 2164–2171; also Electronic Colloquium on Computational Complexity, Report No. 65 (2016), Revision 1. Both PDFs were read in full and agree on everything cited: the bound $O(d\,2^d)$, Lemma 5 (additivity under Cartesian product, attributed there to Lemma 16 of [2]), the sentence on brute force, Figure 3 with its seven rows, the solvers Lingeling and Glucose, and the sentence "Unfortunately for $n > 8$, even these SAT solvers are no longer feasible". The page range is the one recorded by the proceedings' own metadata and by DBLP, which agree; the reference list of [1] gives "151–159" for this paper, which is not a page range of that volume. Revision 1 is the one to cite: the original report under the same number is titled *A note on teaching for VC classes* and proves a weaker bound.

[4] H. U. Simon and S. Zilles, *Open problem: recursive teaching dimension versus VC dimension*, Proceedings of the 28th Conference on Learning Theory (COLT 2015), JMLR Workshop and Conference Proceedings **40** (2015), 1770–1772. Read from the PDF: the definition "By $\mathrm{TD}_{\min}(\mathcal{C}) = \min_{C \in \mathcal{C}} \mathrm{TD}(C;\mathcal{C})$ we refer to the best-case teaching dimension of $\mathcal{C}$", the question $\mathrm{RTD}(\mathcal{C}) \in O(\mathrm{VCD}(\mathcal{C}))$, and the sentence attributing to [2] its equivalence with $\mathrm{TD}_{\min}(\mathcal{C}) \in O(\mathrm{VCD}(\mathcal{C}))$. It prints no small-domain values.

[5] L. Hu, R. Wu, T. Li and L. Wang, *Quadratic upper bound for recursive teaching dimension of finite VC classes*, Proceedings of the 30th Conference on Learning Theory (COLT 2017), Proceedings of Machine Learning Research **65** (2017), 1147–1156. Read from the PMLR PDF: the bound $O(d^2)$, and Definition 5 with the sentence following it, "Kuhlmann (1999) proved $f(2, 3) = 1$, and Moran et al. (2015) proved $f(3, 6) \leq 3$".

[6] C. Kuhlmann, *On teaching and learning intersection-closed concept classes*, in: Computational Learning Theory, 4th European Conference (EuroCOLT '99), Lecture Notes in Computer Science **1572**, Springer, 1999, 168–182, doi:10.1007/3-540-49097-3_14. **Not obtained**: the publisher's PDF is behind an institutional login, no open-access copy is recorded by Unpaywall or Semantic Scholar, and the author's own link is dead since 2006. The bibliographic data above were confirmed against DBLP and Crossref, not taken from a citing paper. Cited here only for $\mathrm{TS}_{\min} = 1$ at VC dimension 1, and only as restated in Definition 5 of [5] and Corollary 9 of [2]: the statement is used here *secondhand* and is not re-proved. The flaw that [2] records in its Theorem 4 concerns a different quantity, the average-case teaching dimension, and does not bear on the fact used here.

[7] S. Moran, A. Shpilka, A. Wigderson and A. Yehudayoff, *Compressing and teaching for low VC-dimension*, in: Proceedings of the 56th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2015), IEEE, 40–51, doi:10.1109/focs.2015.12; the preprint arXiv:1502.06187 and the later book chapter carry the transposed title *Teaching and compressing for low VC-dimension*, which is the form [1] uses for the conference version. Cited only for the bound $O(d\,2^d \log\log|\mathcal{C}|)$ as [1] states it; not read for this work beyond its title page.

[8] E. Kushilevitz, N. Linial, Y. Rabinovich and M. Saks, *Witness sets for families of binary vectors*, Journal of Combinatorial Theory, Series A **73** (1996), no. 2, 376–380, doi:10.1006/jcta.1996.0031. Read in full: it bounds the average witness size ($O(n^{1/2})$, tight) and states no exact value at fixed VC dimension; [2] (p. 3128) cite it for the fact that the average-case teaching dimension is not bounded by the VC dimension. Cited here only as part of the search record of Section 1.6.

[9] Korea Superintelligence Labs, *Non-clashing teaching on small domains: a complete five-point census, a planar closed-neighbourhood census, and the class where a withdrawn proof fails*, companion manuscript, version 1, September 2026. Cited for its Section 4 (the five-point census of the non-clashing teaching dimension), its Theorem 5.2 ($\mathcal{C}_W$ has no fragment of size 2 and frequency 1) and its Remark 5.5 (the 960 fragment-free classes on five points, in four orbits of sizes $192, 192, 384, 192$ with $10, 20, 21, 22$ concepts and non-clashing teaching dimensions $2, 2, 2, 3$). Statement numbers are those of its version 1, obtained by compiling it.

[10] L. de Moura and S. Ullrich, *The Lean 4 theorem prover and programming language*, in: Automated Deduction — CADE 28, Lecture Notes in Computer Science, Springer, 2021, 625–635, doi:10.1007/978-3-030-79876-5_37.

[11] The mathlib Community, *The Lean mathematical library*, in: Proceedings of the 9th ACM SIGPLAN International Conference on Certified Programs and Proofs (CPP 2020), ACM, 2020, 367–381, doi:10.1145/3372885.3373824.