THE PROPER REPAIR DIMENSION ON A FIXED DOMAIN: A SHARP BOUND, ITS EQUALITY CASES, AND A DOMAIN-MINIMAL SEPARATION FROM THE LITTLESTONE DIMENSION

KOREA SUPERINTELLIGENCE LABS

ABSTRACT. For a finite binary concept class $\mathcal{H} \subseteq \{0,1\}^X$ on a finite domain X, Li [1] defines the *proper repair dimension* $\mathrm{PRD}(\mathcal{H})$ by a recursion on version spaces — a live hypothesis is eliminated by a label it does not predict, and PRD is the worst case, over the adversary's choices, of the best case over the learner's — and proves $\mathrm{Ldim}(\mathcal{H}) \le \mathrm{PRD}(\mathcal{H}) \le |\mathcal{H}| - 1$, both bounds tight. Every bound in that paper is in terms of $|\mathcal{H}|$; nothing there constrains PRD in terms of the domain. We prove that $\mathrm{PRD}(V) \le |\operatorname{free}(V)|$, where $\operatorname{free}(V)$ is the set of coordinates on which V is not constant, so that $\mathrm{PRD}(\mathcal{H}) \le n$ for every class on n points, against the $2^n - 1$ that $|\mathcal{H}| - 1$ gives at $\mathcal{H} = \{0,1\}^{[n]}$. We then determine the equality cases exactly: $\mathrm{PRD}(V) = |\operatorname{free}(V)|$ holds if and only if V is a subcube, so $\max\{\mathrm{PRD}(\mathcal{H}) : \emptyset \ne \mathcal{H} \subseteq \{0,1\}^{[n]}\} = n$ with the full class as the *unique* maximiser. It follows that a class of size $m \ge 3$ attaining Li's bound $\mathrm{PRD}(\mathcal{H}) = |\mathcal{H}| - 1$ needs a domain of at least m points. That bound is attained: the one-hot class $\{e_1, \ldots, e_n\} \subseteq \{0,1\}^{[n]}$ has $|\mathcal{H}| = n$, $\mathrm{Ldim}(\mathcal{H}) = 1$ and $\mathrm{PRD}(\mathcal{H}) = n - 1 = |\mathcal{H}| - 1$. This class is Li's own separating example U_n with its domain cut from $2^{[n]}$ down to the n singletons, and it improves that example on two axes at once — domain size $2^n \to n$ and Littlestone dimension $\lfloor \log_2 n \rfloor \to 1$ — at the same $|\mathcal{H}|$ and the same PRD; by the previous sentence the domain cannot be cut further. We also record the complete joint distribution of $(\mathrm{Ldim}, \mathrm{PRD})$ over all nonempty classes on at most three points, and, as computations outside the formal development, over all 65 535 classes on four points and, as a count of orbits under the automorphism group of the cube, over the 4 294 967 295 on five. All the general statements are proved uniformly in n and machine-checked in Lean 4.

1. Introduction

The dimension. Fix a finite domain X and a finite binary concept class $\mathcal{H} \subseteq \{0,1\}^X$. For a *version space* $V \subseteq \mathcal{H}$, a point $x \in X$ and a bit $b \in \{0,1\}$ write

$$V_{x,b} = \{g \in V : g(x) = b\}.$$

Following Li [1, Definition 2.2], for a nonempty V and a *live hypothesis* $h \in V$ the *legal eliminations* of the pair (V,h) are

$$E(V,h) = \big\{\, V_{x,b} \;:\; V_{x,b} \ne \emptyset \;\text{ and }\; h(x) \ne b \,\big\},$$

and

$$D(V,h) = \begin{cases} 0 & \text{if } E(V,h) = \emptyset, \\ \max_{W \in E(V,h)} \big(1 + D(W)\big) & \text{otherwise,} \end{cases} \qquad D(V) = \min_{h \in V} D(V,h),$$

with $\mathrm{PRD}(\mathcal{H}) = D(\mathcal{H})$ the *proper repair dimension* of $\mathcal{H}$. The recursion terminates because each $W \in E(V,h)$ omits h and is therefore strictly smaller than V. Read operationally, $D(V,h)$ is the number of further changes of the live hypothesis that an adversary can force on a learner currently committed to h with version space V, and the minimum over h is the learner's choice of what to commit to.

Li compares PRD with the Littlestone dimension Ldim, taken in the matching recursive form [1, Definition 3.1]: $\mathrm{Ldim}(V) = 0$ if for every point x at most one of $V_{x,0}$, $V_{x,1}$ is nonempty, and otherwise

$$\mathrm{Ldim}(V) = 1 + \max_{x \,:\, V_{x,0} \ne \emptyset,\; V_{x,1} \ne \emptyset} \min\big\{\mathrm{Ldim}(V_{x,0}), \mathrm{Ldim}(V_{x,1})\big\}.$$

The two recursions are guarded by the same condition. Since b is a bit, $h(x) \neq b$ forces $b = 1-h(x)$, so $E(V,h)$ is the family of nonempty sets $V_{x,1-h(x)}$ with x ranging over X; and for $h \in V$ the set $V_{x,h(x)}$ already contains h, so $V_{x,1-h(x)}$ is nonempty exactly when *both* $V_{x,0}$ and $V_{x,1}$ are, that is, exactly when x *splits* V in the sense of the Ldim recursion. We use this throughout.

What is known. Li proves [1, Theorem 3.2] that every finite binary concept class satisfies

$$\mathrm{Ldim}(\mathcal{H}) \;\leq\; \mathrm{PRD}(\mathcal{H}) \;\leq\; |\mathcal{H}| - 1, \tag{1}$$

and that both bounds are tight. For the left one, the full class $C_d = \{0,1\}^{[d]}$ has $\mathrm{Ldim}(C_d) = \mathrm{PRD}(C_d) = d$ [1, Theorem 3.3]. For the right one, and for the separation between the two dimensions, the paper exhibits the *universal coordinate class*

$$U_n = \{h_1, \dots, h_n\} \subseteq \{0,1\}^{2^{[n]}}, \qquad h_i(S) = 1 \iff i \in S,$$

a class of n concepts on the domain $2^{[n]}$ of all 2^n subsets of $[n]$, and proves $\mathrm{PRD}(U_n) = n-1$ and $\mathrm{Ldim}(U_n) = \lfloor \log_2 n \rfloor$ [1, Theorem 3.4]. Taking $n = 2^d$ gives that paper's headline gap: $\mathrm{Ldim}(U_{2^d}) = d$ while $\mathrm{PRD}(U_{2^d}) = 2^d - 1$.

What that leaves open. Both bounds in (1) are in terms of $|\mathcal{H}|$, and [1] states no bound of any kind in terms of the domain. At the full class the two sides of the comparison are far apart: $\mathrm{PRD}(C_n) = n$, while the bound $|\mathcal{H}| - 1$ available for it is $2^n - 1$. Nor does [1] characterise equality in either of its bounds — it proves both tight by exhibiting one witness each — or ask how small a domain can carry a class attaining $\mathrm{PRD}(\mathcal{H}) = |\mathcal{H}| - 1$; the separating witness U_n realises $|\mathcal{H}| = n$ and $\mathrm{PRD} = n-1$ on a domain of 2^n points, and no lower bound is offered against which to measure that. Finally, the paper tabulates no small values: beyond C_d and U_n nothing is recorded about how $(\mathrm{Ldim}, \mathrm{PRD})$ is distributed over the classes on a small domain.

What is settled here. Write $\mathrm{free}(V) \subseteq X$ for the set of coordinates on which V is not constant — that is, the coordinates that split V — and $\mathrm{fix}(V)$ for its complement. Our results, all for classes on the domain $X = [n]$ and all proved uniformly in n, are the following.

- **A domain bound** (Theorem 3.1): $\mathrm{PRD}(V) \leq |\,\mathrm{free}(V)|$ for every version space V, and hence $\mathrm{PRD}(\mathcal{H}) \leq n$ for every class on n points. This is a bound of a kind absent from [1], and at the full class it is exact where $|\mathcal{H}| - 1$ is exponentially off.
- **The equality cases** (Theorem 4.1): for nonempty V, $\mathrm{PRD}(V) = |\,\mathrm{free}(V)|$ if and only if V is a subcube of $\{0,1\}^{[n]}$. There are exactly 3^n nonempty subcubes.
- **The maximum on a fixed domain** (Theorem 4.3): $\max\{\mathrm{PRD}(\mathcal{H}) : \emptyset \neq \mathcal{H} \subseteq \{0,1\}^{[n]}\} = n$, and the full class is the *unique* maximiser. Consequently (Corollary 4.4) a class of size $m \geq 3$ with $\mathrm{PRD}(\mathcal{H}) = |\mathcal{H}| - 1$ cannot live on fewer than m points.
- **A domain-minimal separating class** (Theorem 5.1): for $n \geq 2$ the one-hot class $\{e_1, \dots, e_n\} \subseteq \{0,1\}^{[n]}$, where $e_j(x) = 1 \iff x = j$, has $|\mathcal{H}| = n$, $\mathrm{Ldim}(\mathcal{H}) = 1$ and $\mathrm{PRD}(\mathcal{H}) = n - 1 = |\mathcal{H}| - 1$. It is U_n with its domain cut from $2^{[n]}$ to the n singletons $\{x\}$, since $h_i(\{x\}) = 1 \iff i = x$; so this is an improvement to Li's own object, not a different construction, and it improves it on two axes simultaneously, the domain from 2^n points to n and the Littlestone dimension from $\lfloor \log_2 n \rfloor$ to 1, at unchanged $|\mathcal{H}|$ and unchanged PRD. By Corollary 4.4 the domain cannot be cut further.
- **Small domains** (Section 6): the complete joint distribution of $(\mathrm{Ldim}, \mathrm{PRD})$ over all $2^{2^n} - 1$ nonempty classes for $n \leq 3$, machine-checked, together with the same distribution for $n = 4$ and an orbit count for $n = 5$, both computations outside the formal development.

Theorem 5.1 sharpens [1]'s separating example at every $n \geq 2$: it realises the same $|\mathcal{H}|$ and the same PRD on n points instead of 2^n, and with $\mathrm{Ldim} = 1$ in place of $\lfloor \log_2 n \rfloor$ — a strict improvement in the Littlestone dimension as soon as $n \geq 4$, and in the domain size throughout. At $n = 2^d$ the one-hot class realises $\mathrm{PRD} = 2^d - 1$ with $\mathrm{Ldim} = 1$ on a domain of 2^d points, where U_{2^d} realises the same PRD with $\mathrm{Ldim} = d$ on 2^{2^d} points. Neither parameter can be pushed further: $\mathrm{Ldim} = 0$ forces the class to be a singleton (Remark 2.4), and the domain is minimal by Corollary 4.4.

Scope. This paper is about the finite combinatorics of the two dimensions and nothing else. In particular we do not touch the second half of [1], which converts $\mathrm{PRD}(\mathcal{H})$ into an exact competitive bound for an online-buying problem; no statement here concerns learning algorithms, sample complexity, or the semantics of a "repair". We take PRD and Ldim strictly as the two recursions displayed above.

All general statements below are formalised and machine-checked in Lean 4 [3]; Section 7 says exactly which parts of the paper that covers and which finite computations the remaining parts rest on.

2. Preliminaries

Throughout, $n \geq 0$, the domain is $[n] = \{1, \dots, n\}$, a *concept* is a map $g : [n] \to \{0, 1\}$, and a *class* or *version space* is a set V of concepts. We write $C_n = \{0, 1\}^{[n]}$ for the full class.

Definition 2.1. A point $x \in [n]$ *splits* V when $V_{x,0} \neq \emptyset$ and $V_{x,1} \neq \emptyset$. We put $\mathrm{free}(V) = \{x \in [n] : x \text{ splits } V\}$ and $\mathrm{fix}(V) = [n] \setminus \mathrm{free}(V)$.

For nonempty V the point x splits V exactly when V is non-constant at x, and for $x \in \mathrm{fix}(V)$ all members of V share a common value at x, which we denote $\sigma_V(x)$.

Definition 2.2. A *subcube* of $\{0, 1\}^{[n]}$ is a set of the form $\{g : g(x) = \tau(x) \text{ for all } x \in F\}$ for some $F \subseteq [n]$ and some $\tau : F \to \{0, 1\}$. There are 3^n of them — one for each map $[n] \to \{0, 1, *\}$ — and every one is nonempty.

We call a concept g *achievable* for a nonempty V when $g(x) = \sigma_V(x)$ for every $x \in \mathrm{fix}(V)$, that is, when g lies in the smallest subcube containing V; and we say V *is* a subcube when every achievable concept belongs to V. The reverse inclusion is automatic, so this says exactly that V coincides with the smallest subcube containing it, which is the subcube on $F = \mathrm{fix}(V)$ with $\tau = \sigma_V$; in particular V is then a subcube in the sense of Definition 2.2, with free coordinate set $\mathrm{free}(V)$ and $|V| = 2^{|\,\mathrm{free}(V)|}$.

Two facts from [1] are used below and are re-proved from the definitions in our formal development, so that nothing in this paper depends on an unchecked import.

Proposition 2.3 ([1, Theorems 3.2 and 3.3])**.** *For every nonempty version space $V \subseteq \{0, 1\}^{[n]}$,*

$$\mathrm{Ldim}(V) \;\leq\; \mathrm{PRD}(V) \;\leq\; |V| - 1,$$

and $\mathrm{PRD}(C_n) = n$ for the full class $C_n = \{0, 1\}^{[n]}$.

Both are Li's, with Li's proofs: the upper bound and the comparison are inductions on $|V|$, and the value at the full class is recovered here as the case $V = C_n$ of Theorem 4.1 below. We restate them because every later statement is measured against them.

Remark 2.4. If $\mathrm{Ldim}(V) = 0$ then no point splits V, so V is constant at every coordinate and $|V| \leq 1$; hence $\mathrm{Ldim}(V) \geq 1$ for any class with at least two members, and $\mathrm{Ldim}(V) = 1$ is the least value compatible with $\mathrm{PRD}(V) \geq 1$. This one-line observation is elementary and, unlike the numbered statements, is not part of the formal development.

3. The domain bound

Theorem 3.1. *For every version space $V \subseteq \{0, 1\}^{[n]}$,*

$$\mathrm{PRD}(V) \;\leq\; |\,\mathrm{free}(V)|,$$

and consequently $\mathrm{PRD}(\mathcal{H}) \leq n$ for every class $\mathcal{H} \subseteq \{0, 1\}^{[n]}$.

Proof sketch. Induction on $|V|$; the empty case is trivial, since $\mathrm{PRD}(\emptyset) = 0$. Let V be nonempty and pick any $h \in V$; it suffices to bound $D(V, h)$, because $\mathrm{PRD}(V) = \min_{h'} D(V, h') \leq D(V, h)$. Every legal elimination of (V, h) is of the form $W = V_{x,1-h(x)}$ for some x, and if it is nonempty then x splits V, so $x \in \mathrm{free}(V)$. The set W is constant at x, hence $x \notin \mathrm{free}(W)$, and $W \subseteq V$ gives $\mathrm{free}(W) \subseteq \mathrm{free}(V)$; therefore

$$\mathrm{free}(W) \subseteq \mathrm{free}(V) \setminus \{x\}, \qquad \text{so} \qquad |\,\mathrm{free}(W)| + 1 \leq |\,\mathrm{free}(V)|.$$

Also $|W| < |V|$, since $h \in V \setminus W$. By the inductive hypothesis $\mathrm{PRD}(W) \le |\operatorname{free}(W)|$, whence $1 + \mathrm{PRD}(W) \le |\operatorname{free}(V)|$ for every legal elimination, and taking the maximum gives $D(V,h) \le |\operatorname{free}(V)|$. The second assertion follows from $|\operatorname{free}(\mathcal{H})| \le n$. □

The content of the bound is that a legal elimination *consumes* a free coordinate and can never restore one: once the version space has been cut down to a fixed value at x, the point x is dead for the rest of the game. Theorem 3.1 is not comparable with Li's $\mathrm{PRD}(\mathcal{H}) \le |\mathcal{H}| - 1$ in general — on $n \ge 2$ points a class of two concepts differing everywhere has $|\mathcal{H}| - 1 = 1 < n = |\operatorname{free}(\mathcal{H})|$ — but it is strictly stronger whenever $|\mathcal{H}| > n + 1$, and at the full class it replaces $2^n - 1$ by the exact value n.

4. Equality: the subcube characterisation

Theorem 4.1. *Let $V \subseteq \{0,1\}^{[n]}$ be nonempty. Then*

$$\mathrm{PRD}(V) = |\operatorname{free}(V)| \iff V \text{ is a subcube.}$$

The case $V = C_n$ recovers $\mathrm{PRD}(C_n) = n$, that is, the half of [1, Theorem 3.3] concerning PRD.

Proof sketch. ($\Leftarrow$) Induction on $|V|$. If V is a subcube and $x \in \operatorname{free}(V)$, then for either bit b the restriction $V_{x,b}$ is nonempty, is again a subcube, and has $\operatorname{free}(V_{x,b}) = \operatorname{free}(V) \setminus \{x\}$: the first two because a subcube restricted at a free coordinate is a subcube, the third because a coordinate $y \ne x$ free in V stays free after fixing x, which is where being a subcube is used. Given a live $h \in V$, the elimination at x produces $W = V_{x,1-h(x)}$ with, inductively, $\mathrm{PRD}(W) = |\operatorname{free}(W)| = |\operatorname{free}(V)| - 1$, so $D(V,h) \ge 1 + \mathrm{PRD}(W) = |\operatorname{free}(V)|$. As h was arbitrary, $\mathrm{PRD}(V) \ge |\operatorname{free}(V)|$; Theorem 3.1 gives the reverse.

($\Rightarrow$) Suppose $\mathrm{PRD}(V) = f := |\operatorname{free}(V)|$. Since $D(V,h) \le f$ for every h (the argument of Theorem 3.1 bounds each pair value, not just the minimum) and the minimum over h equals f, in fact $D(V,h) = f$ for *every* $h \in V$. Assume $f \ge 1$ and fix $h \in V$. The maximum defining $D(V,h)$ is attained at some point x, which therefore splits V and has $\mathrm{PRD}(W) = f - 1$ for $W = V_{x,1-h(x)}$. Combining $|\operatorname{free}(W)| \le f - 1$ with $\mathrm{PRD}(W) \le |\operatorname{free}(W)|$ forces $|\operatorname{free}(W)| = f - 1$ and $\mathrm{PRD}(W) = |\operatorname{free}(W)|$; by induction W is a subcube, and $\operatorname{free}(W) = \operatorname{free}(V) \setminus \{x\}$ then identifies W as the set of *all* concepts achievable for V with value $1 - h(x)$ at x. We have shown:

> for every $h \in V$ there is a splitting point x such that every concept achievable for V taking the value $1 - h(x)$ at x lies in V. $(*)$

Now let g be achievable and apply $(*)$ to some $h_0 \in V$, obtaining x_0: either $g \in V$ and we are done, or $g(x_0) = h_0(x_0)$. Let $\hat{g}$ agree with g on $\operatorname{fix}(V)$ and take the opposite value at every free coordinate. Then $\hat{g}$ is achievable and $\hat{g}(x_0) = 1 - h_0(x_0)$, so $\hat{g} \in V$ by $(*)$. Applying $(*)$ to $h := \hat{g}$ gives a splitting x_1; since x_1 is free, $g(x_1) = 1 - \hat{g}(x_1)$, and $(*)$ yields $g \in V$. The case $f = 0$ is immediate: then V is constant everywhere and is the singleton subcube. □

Since a nonempty subcube is determined by its pattern in $\{0,1,*\}^{[n]}$, the classes realising equality are counted exactly:

Corollary 4.2. *The number of nonempty $V \subseteq \{0,1\}^{[n]}$ with $\mathrm{PRD}(V) = |\operatorname{free}(V)|$ is 3^n.*

Section 6 records the machine-checked confirmation of this count for $n \le 3$ and its unformalised confirmation for $n = 4, 5$.

Theorem 4.3. *Let $V \subseteq \{0,1\}^{[n]}$ be nonempty. Then $\mathrm{PRD}(V) = n$ if and only if $V = \{0,1\}^{[n]}$. In particular*

$$\max\bigl\{\mathrm{PRD}(\mathcal{H}) : \emptyset \ne \mathcal{H} \subseteq \{0,1\}^{[n]}\bigr\} = n,$$

attained only at the full class.

Proof sketch. If $\mathrm{PRD}(V) = n$, then $n \le |\operatorname{free}(V)| \le n$ by Theorem 3.1, so $\operatorname{free}(V) = [n]$ and $\mathrm{PRD}(V) = |\operatorname{free}(V)|$; by Theorem 4.1 V is a subcube, and a subcube with no fixed coordinate is the full class. Conversely $\mathrm{PRD}(C_n) = n$ by Theorem 4.1 applied to $V = C_n$, which is Proposition 2.3. □

Corollary 4.4. *Let $\mathcal{H} \subseteq \{0,1\}^{[n]}$ with $|\mathcal{H}| = m \geq 3$ and $\mathrm{PRD}(\mathcal{H}) = |\mathcal{H}| - 1$. Then $m \leq n$: a class of size $m \geq 3$ attaining Li's bound needs a domain of at least m points.*

Proof. By Theorem 3.1, $m - 1 = \mathrm{PRD}(\mathcal{H}) \leq n$. If $m > n$ then $m - 1 = n$, so $\mathrm{PRD}(\mathcal{H}) = n$ and $\mathcal{H} = \{0,1\}^{[n]}$ by Theorem 4.3, whence $m = 2^n = 2^{m-1}$. But $m < 2^{m-1}$ for every $m \geq 3$, a contradiction. Hence $m \leq n$. □

The hypothesis $m \geq 3$ cannot be dropped: on a one-point domain the full class has $m = 2$, $\mathrm{PRD} = 1 = m - 1$ and $m = 2 > 1 = n$ (Proposition 6.4).

5. A domain-minimal separating class

For $j \in [n]$ let e_j be the concept with $e_j(x) = 1$ if and only if $x = j$, and for $J \subseteq [n]$ put $S_J = \{e_j : j \in J\}$; the *one-hot class* is $S_{[n]}$.

Theorem 5.1. *For every nonempty $J \subseteq [n]$ we have $\mathrm{PRD}(S_J) = |J| - 1$, and $\mathrm{Ldim}(S_J) = 1$ as soon as $|J| \geq 2$. In particular, for $n \geq 2$ the one-hot class $\mathcal{H} = \{e_1, \ldots, e_n\} \subseteq \{0,1\}^{[n]}$ satisfies*

$$|\mathcal{H}| = n, \qquad \mathrm{Ldim}(\mathcal{H}) = 1, \qquad \mathrm{PRD}(\mathcal{H}) = n - 1 = |\mathcal{H}| - 1.$$

Proof sketch. Everything follows from the two restrictions of S_J. For every x one has $(S_J)_{x,0} = S_{J \setminus \{x\}}$, and $(S_J)_{x,1} = \{e_x\}$ if $x \in J$ and $\emptyset$ otherwise; hence x splits S_J if and only if $x \in J$ and $J \setminus \{x\} \neq \emptyset$.

For Ldim: every split of S_J has the singleton $\{e_x\}$ as one of its two branches, so the inner minimum in the recursion is 0 and the recursion returns 1 as soon as some point splits, that is, as soon as $|J| \geq 2$.

For PRD: the upper bound $\mathrm{PRD}(S_J) \leq |S_J| - 1 = |J| - 1$ is Proposition 2.3. For the lower bound, induct on $|J|$; given a live $h = e_i$ with $i \in J$ and $|J| \geq 2$, the point $x = i$ splits S_J and $e_i(i) = 1$, so the elimination at i with label 0 is legal and lands in $(S_J)_{i,0} = S_{J \setminus \{i\}}$. Therefore $D(S_J, e_i) \geq 1 + \mathrm{PRD}(S_{J \setminus \{i\}}) = 1 + (|J| - 2) = |J| - 1$ by the inductive hypothesis. As i was arbitrary, $\mathrm{PRD}(S_J) \geq |J| - 1$. □

Remark 5.2. The one-hot class is Li's U_n with its domain replaced by the n singletons of $[n]$: under $x \mapsto \{x\}$ one has $h_i(\{x\}) = 1 \iff i = x$, which is $e_i(x)$. So Theorem 5.1 improves [1, Theorem 3.4] on its own object. At the same $|\mathcal{H}| = n$ and the same $\mathrm{PRD} = n - 1$, the domain shrinks from 2^n points to n and the Littlestone dimension from $\lfloor \log_2 n \rfloor$ to 1. At $n = 2^d$, where [1] reports $\mathrm{Ldim}(U_{2^d}) = d$ against $\mathrm{PRD}(U_{2^d}) = 2^d - 1$, the same PRD is realised with $\mathrm{Ldim} = 1$ on 2^d points instead of 2^{2^d}. Both improvements are final: $\mathrm{Ldim} = 1$ is least possible for a class with $\mathrm{PRD} \geq 1$ (Remark 2.4), and by Corollary 4.4 a class of size $m \geq 3$ with $\mathrm{PRD} = |\mathcal{H}| - 1$ cannot live on fewer than m points, which is what the one-hot class uses.

The one-hot class is not the only class realising this extreme: so does every translate of it, by the invariance we record next. That invariance is also half of what the orbit-organised count of Section 6 would need.

Proposition 5.3. *For any fixed pattern $v \in \{0,1\}^{[n]}$ let $\varphi_v(g) = g \oplus v$ denote the coordinatewise exclusive-or. Then $\mathrm{Ldim}(\varphi_v(V)) = \mathrm{Ldim}(V)$ and $\mathrm{PRD}(\varphi_v(V)) = \mathrm{PRD}(V)$ for every V.*

Proof sketch. Induction on $|V|$, from the identity $(\varphi_v(V))_{x,b} = \varphi_v\big(V_{x,\, b \oplus v(x)}\big)$, which also shows that x splits $\varphi_v(V)$ if and only if x splits V. For Ldim the two branches at a split are exchanged when $v(x) = 1$, which leaves the inner minimum unchanged; for PRD one checks $D(\varphi_v(V), \varphi_v(h)) = D(V, h)$ and notes that $h \mapsto \varphi_v(h)$ is a bijection of V onto $\varphi_v(V)$, so the two minima agree. □

These 2^n maps are the translations of the cube; together with the $n!$ coordinate permutations they generate the full automorphism group of $\{0,1\}^{[n]}$, of order $2^n\, n!$. The translations are the half of that group we prove invariance for, and the half Corollary 5.4 uses.

Corollary 5.4. *Fix $n \geq 2$ and $c \in [n]$, and let*

$$\mathcal{H}_n = \{0^n\} \cup \{\, e_c + e_j \;:\; j \in [n],\ j \neq c \,\} \subseteq \{0,1\}^{[n]},$$

where $e_c + e_j$ *denotes the concept supported exactly on* $\{c, j\}$. *Then* $|\mathcal{H}_n| = n$, $\mathrm{Ldim}(\mathcal{H}_n) = 1$ *and* $\mathrm{PRD}(\mathcal{H}_n) = n - 1$.

Proof. $\mathcal{H}_n = \varphi_{e_c}(S_{[n]})$: the member e_c of the one-hot class maps to 0^n, and e_j for $j \neq c$ maps to the concept supported on $\{c, j\}$. Now apply Proposition 5.3 and Theorem 5.1. $\square$

6. Small domains

Nothing in [1] records the distribution of $(\mathrm{Ldim}, \mathrm{PRD})$ over the classes on a small domain. We computed it. For $n \leq 3$ the computation is inside the formal development: each cell below is the cardinality of an explicitly filtered finite type, evaluated over all 2^{2^n} subsets of $\{0,1\}^{[n]}$, with a separate statement asserting that the listed cells exhaust the 255 nonempty classes at $n = 3$.

Proposition 6.1. *The joint distribution of* $(\mathrm{Ldim}, \mathrm{PRD})$ *over the nonempty classes* $\mathcal{H} \subseteq \{0,1\}^{[n]}$ *is*

n	*nonempty classes*	$(\mathrm{Ldim}, \mathrm{PRD})$ *distribution*
1	3	$(0,0)\!:\!2$, $(1,1)\!:\!1$
2	15	$(0,0)\!:\!4$, $(1,1)\!:\!10$, $(2,2)\!:\!1$
3	255	$(0,0)\!:\!8$, $(1,1)\!:\!84$, $(1,2)\!:\!8$, $(2,2)\!:\!154$, $(3,3)\!:\!1$

and at $n = 3$ *these five cells account for all* 255 *classes. Moreover the number of nonempty* V *with* $\mathrm{PRD}(V) = |\,\mathrm{free}(V)|$ *is* 3, 9, 27 *at* $n = 1, 2, 3$, *as Corollary 4.2 predicts.*

The census is a check on the general theorems rather than a new argument: the (n, n) cell has a single member at each n, as Theorem 4.3 requires, and the 3^n classes with $\mathrm{PRD} = |\,\mathrm{free}\,|$ are exactly the nonempty subcubes. It also shows that the separation $\mathrm{Ldim} < \mathrm{PRD}$ is genuinely rare on small domains: it does not occur at all for $n \leq 2$, and at $n = 3$ it occurs in 8 of the 255 classes, all in the cell $(1, 2)$.

Remark 6.2 (computations outside the formal development). The same distribution was computed for $n = 4$ and $n = 5$ by a program in C, which is not part of the formal development and whose output is reported here as data. Per class the reachable version spaces are exactly the intersections of the class with the 3^n subcubes, so one class costs a dynamic program over 3^n states. For $n \leq 4$ every one of the $2^{2^n} - 1$ nonempty classes was swept directly. For $n = 5$ that is 2^{32} classes; there the program keeps only the minimum of each orbit under the full automorphism group of the cube, of order $2^5 \cdot 5! = 3840$, and weights each representative by its orbit size. The $n = 5$ line below is therefore an *orbit count*: for each cell it gives the total size of the orbits whose minimal representative falls in that cell, so every one of the $2^{32} - 1$ nonempty classes is counted exactly once — the weights do sum to $4\,294\,967\,295 = 2^{32} - 1$ — but each is counted at the value the pair takes on the representative of its orbit rather than on the class itself. The two readings coincide exactly when both dimensions are constant on orbits. Half of that is proved here: invariance under the 2^n translations is Proposition 5.3. The other half, invariance under the $n!$ coordinate permutations, is not proved anywhere in this paper and is not part of the formal development, so we state the $n = 5$ line as the orbit count it is and claim nothing more for it. (The two recursions do treat the points of the domain symmetrically, so the permutation invariance is expected; the evidence offered here is only that at $n = 3$ and $n = 4$ the orbit path reproduces the direct sweep cell for cell, which is a test of exactly this.) The $n = 4$ line, by contrast, is a direct count of all 65 535 classes and needs no invariance of any kind.

n	nonempty classes	$(\mathrm{Ldim}, \mathrm{PRD})$ distribution ($n = 5$: orbit count)
4	65 535	$(0,0)\!:\!16$, $(1,1)\!:\!696$, $(1,2)\!:\!160$, $(1,3)\!:\!16$, $(2,2)\!:\!23\,052$, $(2,3)\!:\!8\,088$, $(3,3)\!:\!33\,506$, $(4,4)\!:\!1$
5	4 294 967 295	$(0,0)\!:\!32$, $(1,1)\!:\!5\,968$, $(1,2)\!:\!2\,080$, $(1,3)\!:\!480$, $(1,4)\!:\!32$, $(2,2)\!:\!3\,823\,000$, $(2,3)\!:\!6\,236\,384$, $(2,4)\!:\!442\,672$, $(3,3)\!:\!1\,315\,856\,404$, $(3,4)\!:\!1\,205\,986\,440$, $(4,4)\!:\!1\,762\,613\,802$, $(5,5)\!:\!1$

Two independent implementations — the C program above and a dynamic program in Python over all 2^{2^n} subsets — agree cell for cell at $n \le 4$, and the orbit path of the C program reproduces its own direct sweep exactly at $n = 3$ and $n = 4$. These runs confirm both general theorems independently of the formal development: the count of $\mathrm{PRD} = |\,\mathrm{free}\,|$ is 81 at $n = 4$ and 243 at $n = 5$, that is 3^n as in Corollary 4.2, and exactly one class attains $\mathrm{PRD} = n$, as in Theorem 4.3. The count of $\mathrm{Ldim} < \mathrm{PRD}$ is $0, 0, 8, 8\,264, 1\,212\,668\,088$ for $n = 1, \ldots, 5$, and the count of $\mathrm{Ldim} = 1$ is $1, 10, 92, 872, 8\,560$. In each of these the $n = 5$ entry is again an orbit count in the sense above, the others are counts of classes. None of these sequences, and none of the census rows, matches an entry of the OEIS; that database contains no entry mentioning "Littlestone" at all.

Remark 6.3 (an open question, from the same computation). Among the classes counted above, those with $\mathrm{Ldim} = 1$ and $\mathrm{PRD} = |\mathcal{H}| - 1 \ge 2$ number $0, 0, 8, 176, 2592$ for $n = 1, \ldots, 5$ (the last again an orbit count), and those of them with $|\mathcal{H}| = n$ number 8 and 16 at $n = 3$ and $n = 4$. Explicit orbit enumeration at those two values shows that they form a single orbit under the automorphism group of the cube, namely that of the one-hot class, and at $n = 5$ the cell $(\mathrm{Ldim}, \mathrm{PRD}) = (1, 4)$ carries the weight $32 = 2^5$, which is the size of that orbit. Whether the one-hot class is, for every $n \ge 3$, the unique such orbit is open; the evidence for it is the computation just described and nothing more.

Finally we record the sharpness statements that keep the results above from being satisfiable by a degenerate reading; each was verified in the formal development, five of the six as consequences of the general theorems above and the sixth, item (4), by direct evaluation at $n = 3$.

Proposition 6.4. (1) *The bound* $\mathrm{PRD}(\mathcal{H}) \le n$ *of Theorem 3.1 cannot be improved to* $n - 1$*: the full class attains* n*.*
(2) *Li's bound* $\mathrm{PRD}(\mathcal{H}) \le |\mathcal{H}| - 1$ *cannot be improved to* $|\mathcal{H}| - 2$*: the one-hot class attains* $|\mathcal{H}| - 1$ *at every size.*
(3) *The inequality* $\mathrm{Ldim} \le \mathrm{PRD}$ *is strict somewhere: the one-hot class on three points has* $\mathrm{Ldim} = 1 < 2 = \mathrm{PRD}$.
(4) *The subcube hypothesis in Theorem 4.1 is not vacuous:* $\{0^3, 1^3\}$ *has three free coordinates and* $\mathrm{PRD} = 1$.
(5) *The hypothesis* $|J| \ge 2$ *in Theorem 5.1 is needed for the value of* Ldim*: a one-hot class of size* 1 *is a singleton, with* $\mathrm{Ldim} = 0$.
(6) *The hypothesis* $m \ge 3$ *in Corollary 4.4 is needed: on a one-point domain the full class has* $m = 2$*,* $\mathrm{PRD} = 1 = m - 1$ *and* $m > n$*.*

7. Verification

The whole development is written in Lean 4 [3] against Mathlib [4], in four modules — `Defs`, `Extremal`, `Subcube`, `Census` — totalling 1 233 lines and 115 declarations, of which 101 are theorems or lemmas and 14 are definitions; every one of those 101 was audited for its axiom dependencies. Formalised and machine-checked are: Proposition 2.3 (both inequalities and the value at the full class, re-proved from the definitions, not imported); Theorems 3.1, 4.1, 4.3 and 5.1; Corollaries 4.4 and 5.4; Proposition 5.3; Proposition 6.1; and Proposition 6.4. All of these except the finite evaluations named next are ordinary kernel proofs, depending on no axioms beyond `propext`, `Classical.choice` and `Quot.sound`; in particular every general theorem of Sections 3–5 is proved uniformly in n and uses no computation. The exceptions, which do rest on exhaustive computation, are Proposition 6.1, evaluated by Lean's compiled evaluator (`native_decide`) over all 2^{2^n} subsets of $\{0,1\}^{[n]}$ for $n \le 3$ — that is 4, 16 and 256 classes — and item (4) of Proposition 6.4, two such evaluations at a single class on three points. Each of those statements carries one evaluation axiom of its own, and that evaluator is the only trusted step in them; the other five items of Proposition 6.4 are kernel proofs, read off the general theorems. The audit, re-run on 3 September 2026 against the compiled modules, gives the following totals over the 101 theorems and lemmas: none depends on `sorryAx`, and there is no `sorry` anywhere in the sources; exactly 19 carry a native evaluation axiom, and they are precisely the `native_decide` statements just described (the thirteen census cells and totals, the exhaustiveness statement at $n = 3$, the three subcube counts, and the two evaluations behind item (4)); the remaining 82 depend on nothing beyond `propext`, `Classical.choice` and `Quot.sound`, 66 on all three of them and 16 on a strict subset. Not formalised, and labelled as such where they occur, are: Corollary 4.2, an elementary count of subcubes; Remark 2.4; and the computations of Remarks 6.2 and 6.3, which are the runs of the C program at $n = 4, 5$ and the orbit enumerations, and on which nothing else in the paper depends. Mathlib supplies no prior art for either dimension: it carries a VC-dimension API (`Finset.Shatters`, `Finset.shatterer`, `Finset.vcDim` and both forms of the Sauer–Shelah lemma, in `Mathlib/Combinatorics/SetFamily/Shatter.lean`) but no Littlestone dimension, no mistake bound and nothing about version spaces, and neither dimension used here is defined through shattering, so that API does not apply. (These are statements about the pinned snapshot named in [4], not about Mathlib in general.) The repository is private; repository reference to be added at submission.

References

[1] Y. Li, *Learnable Predictions Need Not Be Actionable: Proper Repair Dimension for Online Buying*, arXiv:2607.22608v1, submitted 14 June 2026; cs.DS is the sole listed category. All quotations and statements above are from the v1 e-print, which was still the live version when the abstract page was re-fetched on 3 September 2026; the arXiv metadata records no journal reference. **The numbering below is that of v1**, obtained by compiling the e-print's own LaTeX source, and would have to be rechecked against any later version: Definition 2.2 (legal eliminations, $D(V, h)$, $D(V)$ and $\mathrm{PRD}(\mathcal{H}) = D(\mathcal{H})$), Definition 3.1 (the recursive Littlestone dimension), Theorem 3.2 ("Every finite binary concept class $\mathcal{H}$ satisfies $\mathrm{Ldim}(\mathcal{H}) \leq \mathrm{PRD}(\mathcal{H}) \leq |\mathcal{H}| - 1$"), Theorem 3.3 ("Let $C_d = \{0, 1\}^{[d]}$. Then $\mathrm{Ldim}(C_d) = \mathrm{PRD}(C_d) = d$"), and Theorem 3.4 (the class U_n, with $\mathrm{PRD}(U_n) = n - 1$ and $\mathrm{Ldim}(U_n) = \lfloor \log_2 n \rfloor$).

[2] N. Littlestone, *Learning quickly when irrelevant attributes abound: a new linear-threshold algorithm*, Machine Learning **2** (1988), no. 4, 285–318; `doi:10.1007/BF00116827`. Journal, volume, issue, pages and date were checked on 3 September 2026 against the Crossref and OpenAlex records for that DOI, which agree; an extended abstract had appeared in the *28th Annual Symposium on Foundations of Computer Science* (1987), 68–77, `doi:10.1109/SFCS.1987.37`. This is the paper [1] cites for the Littlestone dimension. We use Ldim only in the recursive form [1, Definition 3.1] states, which is the form formalised here; its equivalence with the mistake-bound notion of the present reference is not verified in this paper and is used nowhere above.

[3] L. de Moura and S. Ullrich, *The Lean 4 theorem prover and programming language*, in: A. Platzer and G. Sutcliffe (eds.), Automated Deduction — CADE 28, Lecture Notes in Computer Science **12699**, Springer, Cham, 2021, pp. 625–635; `doi:10.1007/978-3-030-79876-5_37`.

[4] The mathlib Community, *The Lean mathematical library*, in: Proceedings of the 9th ACM SIGPLAN International Conference on Certified Programs and Proofs (CPP 2020), ACM, New York, 2020, pp. 367–381; `doi:10.1145/3372885.3373824`. Mathlib is a moving target, so the statements above about what it does and does not contain are statements about one snapshot: the revision this development is pinned to in `lake-manifest.json`, namely commit `81a5d257c8e410db227a6665ed08f64fea08e997` (13 July 2026) under toolchain `leanprover/lean4:v4.32.0`. They were rechecked against that snapshot on 3 September 2026.

[5] OEIS Foundation Inc., *The On-Line Encyclopedia of Integer Sequences*, `https://oeis.org`. The searches reported in Remark 6.2 were run on 2 September 2026.