PACKING SET SYSTEMS OF FINITE VC DIMENSION

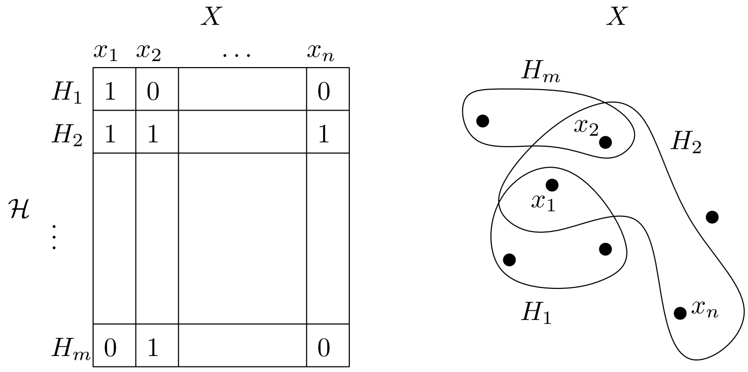
FLORESTAN BRUNCK AND LUCY DAVIDSON

Abstract. Given a set of points X together with a metric d, an ϵ-set $N \subseteq X$ is a set of point such that for every point $x \in X$, there exists a point $n(x) \in N$ such that $d(x, n(x)) \leqslant \epsilon$. It is a classical problem of discrete and computational geometry to find ϵ-nets of small size. In the late 1980s, Haussler and Welzl [4] showed bounds on the size of such sets in the realm of much more abstract and less geometric sets of points, using only the finiteness of their VC-dimension. Their result shows that sets of VC-dimension at most d behave much like Euclidean balls of dimension d.

1. Introduction

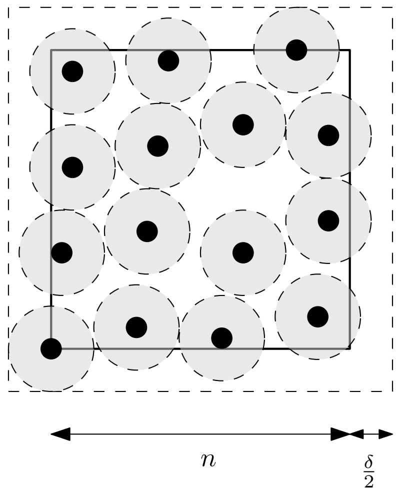
In everything that follows, X denotes a set of n elements and $\mathcal{H} = \{H_1, H_2, \ldots, H_m\}$ a set system on X.

VC-dimension of Set Systems. We say that $\mathcal{H}$ *shatters* a set $A \subseteq X$ if every subset B of A can be obtained as the intersection of A with an element of $\mathcal{H}$, i.e. for all $B \subseteq A$, there exists $H \in \mathcal{H}$ such that $H \cap A = B$. The *VC-dimension* of $\mathcal{H}$ is the maximal size of a set A that is shattered by $\mathcal{H}$. Pictorially, it will prove useful to look at set systems as Boolean matrices with columns indexed by elements of X and rows indexed by elements of $\mathcal{H}$, so that the corresponding cell in the matrix is a 1 if the set contains the element, and 0 otherwise.



Packing of Euclidean Balls. Let $\mathcal{C}_n^d = [0, n]^d$ denote the Euclidean cube of side-length n and dimension d. In this setting, finding a small ϵ-net for the cube is equivalent to finding a small set of Euclidean balls that cover it. Equivalently, we may look at it from the perspective of packings, since finding minimal covers is dual to finding a maximal packing. Thus our aim is instead to find a set P of ϵ-separated points in $\mathcal{C}_n^d$, i.e. points pairwise at a Euclidean distance further than ϵ away from each other. Note that a maximal ϵ-separated set is a ϵ-net. The problem at hand is then to pack Euclidean balls of radius $\frac{\epsilon}{2}$ inside the slightly larger cube $[-\frac{\epsilon}{2}, n + \frac{\epsilon}{2}]^d$. A volume argument gives us an upper bound on the maximal cardinality of P: each such ball has a volume greater than $(\frac{\epsilon}{2})^d$ and the total volume of the cube

$[-\frac{\epsilon}{2}, n+\frac{\epsilon}{2}]^d$ is $(n+\epsilon)^d$. So it must be that $|P|(\frac{\epsilon}{2})^d \leqslant (n+\epsilon)^d$ which gives the upper bound of $|P| = O((\frac{n}{\epsilon})^d)$.



Packing Abstract Set Systems. The following theorem, known as Haussler's Lemma [3][6][7][Chapter 5.2][5][Chapter 6.15], aims to extend this geometric bound to abstract set systems that do not necessarily arise from geometrical settings. It reassuringly confirms the intuition that the VC-dimension of a set system matches the natural Euclidean notion of dimension and shows that sets of VC-dimension d behave like d-dimensional balls. The first step to make sense of packings for more abstract objects is to metrise these objects and introduce a notion of distance for set systems. A natural choice is the cardinality of the *symmetric difference*. We thus say that two sets A and B are a distance ϵ apart if $|\Delta(A,B)| = |(A-B) \cup (B-A)| = \epsilon$.

Haussler's Packing Lemma. *Let X be a set of n elements and $\mathcal{H} = \{H_1, H_2, \ldots, H_m\}$ be a set system of VC dimension d on X. If $\mathcal{H}$ is ϵ-separated then:* $|\mathcal{H}| = O\left(\left(\frac{n}{\epsilon}\right)^d\right)$.

2. Primal Shatter Lemma

A key lemma in the proof of the theorem is known as the Primal Shatter Lemma or Sauer-(Perles-)Shelah Lemma. It gives a tight upper bound on how big sets of VC dimension d can be. The essence of the result is to confirm the intuition that the worst complexity case is really to include all subsets of size at most d.

Primal Shatter Lemma. *Let X be a set of n elements and $\mathcal{H} = \{H_1, H_2, \ldots, H_m\}$ be a set system of VC dimension d on X. For all $A \subseteq X$, we have*

$$|\mathcal{H}_{|A}| \leqslant \sum_{i=0}^{d} \binom{|A|}{i}$$

When $|A| \geqslant d$, the previous sum can be upper bounded by $(\frac{e|A|}{d})^d = O(|A|^d)$.

The main concept in the proof of this lemma, which we shall reuse later on, is that of *shifting*. We also mention the existence of an elegant linear algebra proof due to Pach and Frankl [1][2].

Shifting Sets. Given the set system $\mathcal{H} = \{H_1, H_2, \dots, H_m\}$ on X and a fixed element $a \in X$, the *shift* of $\mathcal{H}$ by a is defined as the set system $\mathcal{H}_a$ obtained by removing the element a from each set of $\mathcal{H}$ as long as it is not a duplicate of a set already existing in $\mathcal{H}$. Formally, for all $i \in [m]$ we define:

$$H'_i = \begin{cases} H_i, & \text{if } H_i - a \in \mathcal{H} \\ H_i - a, & \text{otherwise} \end{cases} \tag{1}$$

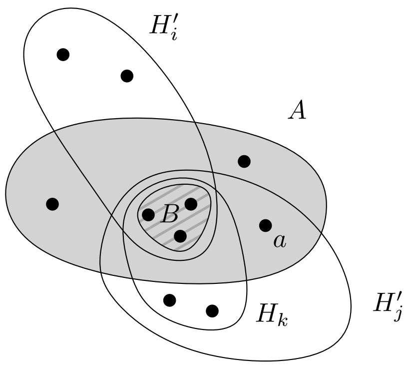
and say that the set H'_i is H_i *shifted* by a.

The core idea behind shifting is that the size of the sets is of very little interest to the VC dimension. Rather, it is the way that the sets intersect and overlap which is captured by this measure of complexity. The idea behind shifting is to get rid of the thrills of a set system and boil its sets down to a condensed form which still captures the same complexity. It is an operation on a set system which maintains the cardinality of the set system, but lowers the complexity of each set, all the while making sure to not increase the VC-dimension of the resulting set system. This is captured by the following two claims, the first of which is obvious by definition.

Claim 2.1. Shifting does not affect the cardinality of a set system: $|\mathcal{H}_a| = |\mathcal{H}|$.

Claim 2.2. Shifting does not increase the VC dimension of a set system: $\text{VC-dim}(\mathcal{H}_a) \leqslant \text{VC-dim}(\mathcal{H})$.

Proof. We want to show that any set shattered by $\mathcal{H}_a$ is also shattered by $\mathcal{H}$. In other words, we want to show that every intersection of A with a member of $\mathcal{H}_a$ we can obtain as an intersection with a set of $\mathcal{H}$. Suppose then that the set A is shattered by $\mathcal{H}_a$. If $a \notin A$, the intersection from members of $\mathcal{H}_a$ with A are the same as the intersections with their non-shifted counterparts in $\mathcal{H}$. We can thus assume that $a \in A$. By assumption, $\mathcal{H}_a$ shatters A so there exists a set $H'_i \in \mathcal{H}_a$ such that $H'_i \cap A = B$.
Case 1: $a \in B$**.** This means that H'_i was not shifted, so then $H_i = H'_i \in \mathcal{H}$ gives us the same intersection.
Case 2: $a \notin B$**.** A is shattered by $\mathcal{H}_a$ so there exists H'_j such that $H'_j \cap A = B \cup \{a\}$. This implies that $H'_j - \{a\}$ must already have been in $\mathcal{H}$ and thus corresponds to some $H_k \in \mathcal{H}$, otherwise H_j would have been shifted to not contain a. ☐



The natural idea after introducing shifting is to look at the *shifting closure* of a set. This is the set system $\overline{\mathcal{H}}$ obtained after repeatedly shifting $\mathcal{H}$ by every element of X until no more shifts may be applied. Note that this set system must exist since shifting strictly decreases the sum of the cardinalities of the sets H_i. The key observation about $\overline{\mathcal{H}}$ is the following claim.

Claim 2.3. The shifting closure $\overline{\mathcal{H}}$ is downwards closed.

Proof. By contradiction, if it were not, we could find at least one element to shift it with. □

For convenience we group here the relevant properties of $\overline{\mathcal{H}}$:

Shifting Closure Properties. *If $\mathcal{H}$ has VC dimension at most d, then the following hold:*

(1) $|\overline{\mathcal{H}}| = |\mathcal{H}|$.
(2) $\text{VC-dim}(\overline{\mathcal{H}}) \leqslant \text{VC-dim}(\mathcal{H})$.
(3) $\overline{\mathcal{H}}$ *is downwards closed.*
(4) $E(\overline{\mathcal{H}}) \geqslant E(\mathcal{H})$. *(the definition and proof of this statement is delayed until the next section).*

Proof of the Primal Shatter Lemma . $\mathcal{H}$ has VC-dimension at most d. Thus by Shifting Closure Properties (2) and Shifting Closure Properties (3) we know that the largest set in $\overline{\mathcal{H}}$ has at most d elements. Summing up the number of sets of $\mathcal{H}$ by their cardinalities we immediately get:

$$|\mathcal{H}_{|A}| = |\overline{\mathcal{H}_{|A}}| \leqslant \sum_{i}^{d} \binom{|A|}{i}$$

□

3. The Unit Distance Graph

Given the set system $\mathcal{H}$ on X, let $G_U(\mathcal{H})$ be the graph with vertex set $\mathcal{H}$ such that $\{H_i, H_j\}$ is an edge if and only if $d(H_i, H_j) = |\Delta(H_i, H_j)| = 1$.

Lemma A. *If $\mathcal{H}$ has VC dimension at most d, then $|E(\mathcal{H})| \leqslant d \cdot |\mathcal{H}|$.*

Proof. Let us orient the edges of $G_U(\overline{\mathcal{H}})$ from bigger sets to smaller sets. Using Shifting Closure Properties (3), we know that each vertex $\overline{H}$ of $G_U(\overline{\mathcal{H}})$ has exactly $|\overline{H}|$ outgoing edges. Because of Shifting Closure Properties (2) and (3) we know that sets of $\overline{\mathcal{H}}$ have cardinality at most d, which gives a bound of at most d outgoing arrows for each vertex. Counting the total number of edges by summing up outgoing edges for each vertex and using Shifting Closure Properties (4) and (1), we get:

$$|E(\mathcal{H})| \leqslant |E(\overline{\mathcal{H}})| \leqslant d \cdot |\overline{\mathcal{H}}| = d \cdot |\mathcal{H}|$$

□

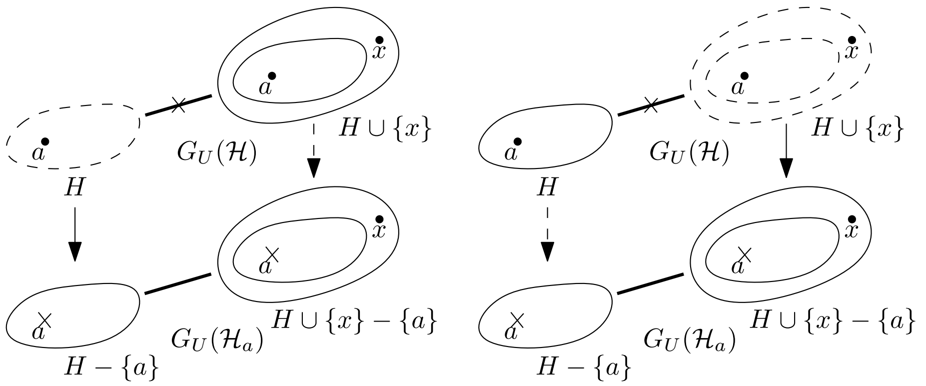
Proof of Shifting Closure Properties (4). Fix an element $a \in X$. We want to prove that the number of edges in the unit distance graph of the shift of $\mathcal{H}$ by a is not any less than that in the unit distance graph of $\mathcal{H}$. Consider an edge $\{H, H \cup \{x\}\}$ in $G_U(\mathcal{H})$. If $a \notin H$, and $x \neq a$ then neither set-endpoint of the edge is affected by shifting by a so the edge also exists in $G_U(\mathcal{H}_a)$. If $a \notin H$ but $x = a$, then this means that shifting $H \cup a$ would create a duplicate so once again both endpoints of the edge are unnafected by shifting by a.

Assume then that $a \in H$. Further assume that only one of the two sets H and $H \cup x$ is shifted (otherwise the edge is preserved). There are two cases.

H shifts but not $H \cup \{x\}$. This case is illustrated on the left diagram. If $H \cup \{x\}$ did not shift, it means $H \cup \{x\} - \{a\}$ already existed in $\mathcal{H}$ and thus also in $\mathcal{H}_a$. This set is a neighbour of H shifted by a which gives us a new edge.

H does not shift but $H \cup \{x\}$ does. This case is illustrated on the right diagram. If H did not shift, it means $H - \{a\}$ already existed in $\mathcal{H}$ and thus also in $\mathcal{H}_a$. This set is a neighbour of $H \cup \{x\}$ shifted by a which gives us a new edge.

Note that in both of these cases the new exhibited edge is uniquely determined by H, a and x and cannot arise from any other edge so they are all distinct.



□

4. Proof of Haussler's Lemma

To derive Haussler's Packing Lemma we will instead prove the following theorem:

Main Theorem. *Let X be a set of n elements and $\mathcal{H} = \{H_1, H_2, \ldots, H_m\}$ be a set system of VC dimension d on X. If $\mathcal{H}$ is ϵ-separated then there exists a subset A of size $\lceil \frac{8dn}{\epsilon} \rceil$ picked uniformly at random without replacement from elements of X such that:*

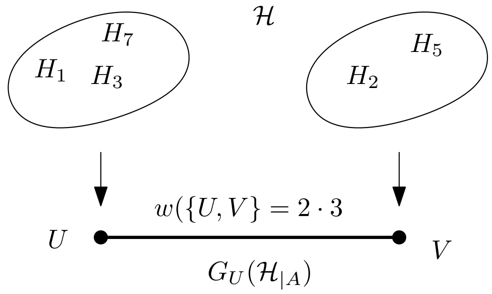
$$|\mathcal{H}| \leqslant 2 \cdot \mathbb{E}[|\mathcal{H}_{|A}|]$$

Note that Haussler's Packing Lemma follows directly from Main Theorem through an application of the Primal Shatter Lemma.

Idea. We want to motivate the intuition that picking a random set of size $\frac{n}{\epsilon}$ is the right idea. Our goal is to somehow ensure that the projection of $\mathcal{H}$ to A is injective. Two sets of $\mathcal{H}$ project to the same set in A if and only if no element of their symmetric difference is in A. But sets are ϵ-separated so their symmetric difference has at least ϵ elements, each of them with a probability of $\frac{1}{\epsilon}$ to end up in A. In expectation there should therefore be one element of A in each symmetric difference to ensure injectivity.

Goal. The proof of Main Theorem centres around a double counting argument for a quantity which measures the failure of injectivity. Namely, we will give an upper bound for the number W of sets in $\mathcal{H}$ that end up at unit distance in $\mathcal{H}_a$.

Upper Bound. The upperbound on W will be quickly derived from a modified version of Lemma A. Indeed, Lemma A is almost what we need: it counts the number of sets of $\mathcal{H}_{|A}$ at unit symmetric difference in $G_U(\mathcal{H}_{|A})$. This is not quite what we want, as many sets in $\mathcal{H}$ might end up at unit distance in $\mathcal{H}_{|A}$.
To account for this, we introduce the *weighted unit distance graph*, which introduces weights to the unit distance graph meant to keep track of the cardinality of the fibers under the projection to A. Namely, we give each vertex a weight corresponding to the cardinality of its fiber under the projection to A. Each edge then has weight equal to the product of the weights of its endpoints. This is exactly the number of pairs of sets in $\mathcal{H}$ that end up as endpoints of this edge in $G_U(\mathcal{H}_{|A})$. So that summing up the weights of all edges exactly yields W.



To simplify the proof, we will instead define the weight of an edge as the minimum of the weights of its endpoints. This is because in the proof, the sum of the weights of the endpoints will be fixed. With this is mind, the minimum and the product are interchangeable and within a factor 2 of each other since $\frac{1}{2}\min\{a,b\} \cdot (a+b) \leqslant ab \leqslant \min\{a,b\} \cdot (a+b)$.

Lemma A'. *If $\mathcal{H}$ has VC dimension at most d, then $W \leqslant 2d \cdot |\mathcal{H}|$.*

Proof. We count the edges of $G_U(\mathcal{H}_{|A})$ by summing the degrees of the vertices and apply Lemma A:

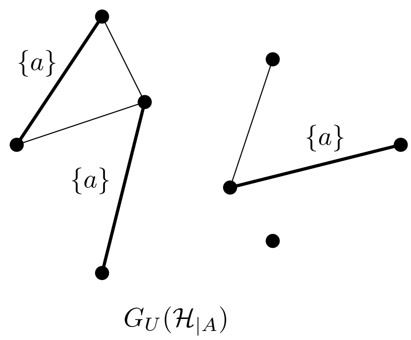
$$\sum_{H' \in \mathcal{H}_{|A}} \deg(H') = 2 \cdot |E(\mathcal{H}_{|A})| \leqslant 2 \cdot d \cdot |\mathcal{H}_{|A}|$$

an application of the pigeon hole principle thus guarantees the existence of a set $H' \in \mathcal{H}_{|A}$ with degree less than or equal to $2d$. By definition, each edge incident to H' weighs less than or equal to the weight of H'. Summing the weight of all the neighbours of H' we thus get a total weight of less than or equal to $2d \cdot w(H')$. Removing H' from $G_U(\mathcal{H}_{|A})$. The removing graph is a unit distance graph on $|\mathcal{H}_{|A}| - 1$ vertices so induction finishes the proof:

$$W \leqslant 2d \sum_{H' \in \mathcal{H}_{|A}} \deg(H') = 2d \cdot |\mathcal{H}|$$

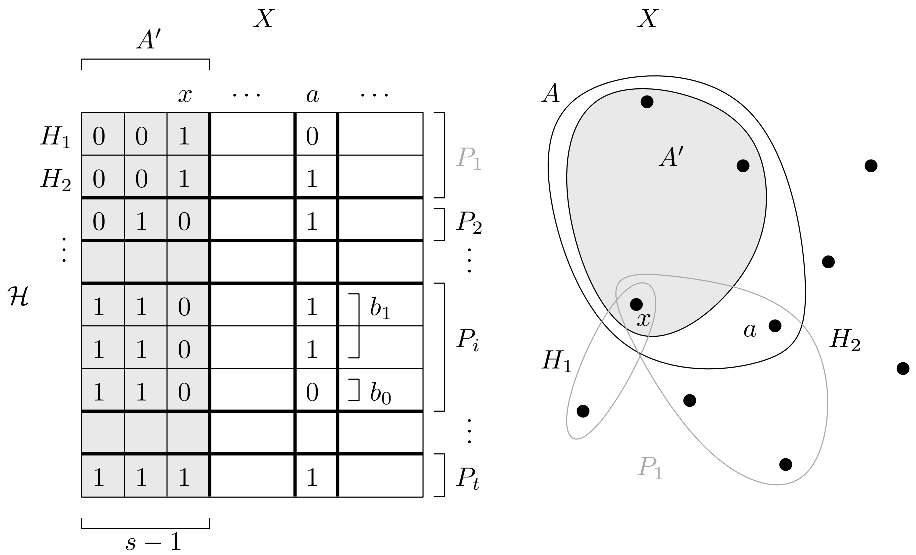
□

Lower Bound. We select a subset $A \subset X$ of size $s = \lceil \frac{8dn}{\epsilon} \rceil$ uniformly at random. A naive starting point would be to find a lower bound on the probability that two sets of $\mathcal{H}$ end up at unit distance after being projected to A and use linearity of expectation to obtain W. This event is equivalent to when exactly one element of the symmetric difference of these two sets of $\mathcal{H}$ is picked into A. This is a problem because this depends on the cardinality of their symmetric difference and seems hard to compute. The trick is that we only care about the sum over all pairs of sets of $\mathcal{H}$ and we can count this sum differently: instead of summing over pairs of sets we will isolate the contribution of a single element $a \in X$ in the unit distance graph and sum over all choices of this element a (see figure below). The choice of this element being arbitrary, by symmetry this contribution is the same for all elements of A. So we will compute the expected sum instead by conditioning on the first $s-1$ elements and looking at the expected number of pairs that end up at unit distance after the last random element of a of A is picked.

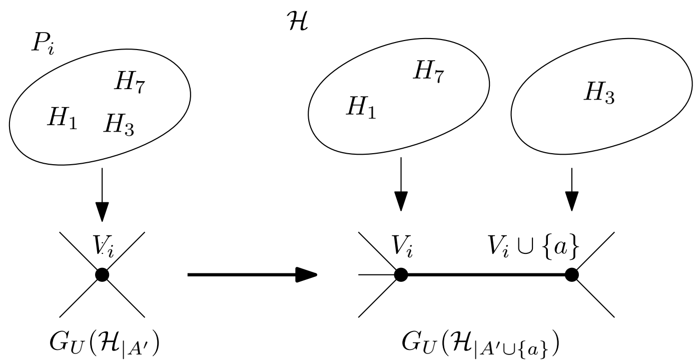


Edges of $G_U(\mathcal{H}_{|A})$ can be partitioned according to the single element that they each correspond to. A single element a may correspond to multiple edges.

Pick then a set of $s-1$ elements $A' \subset X$ uniformly at random, and pick the last s-th a of A at random in $X - A'$ (note that this is equivalent to picking a uniform random subset of s elements). We refer to the boolean matrix view of a set system to visualise this process, where we order the elements of X (the columns) so that the first $s-1$ columns are the elements of A'. We then order the rows of this matrix according to the equivalence classes $P_1, P_2, \dots, P_t$ given by the fibers of the projection of $\mathcal{H}$ to A'. That is, for each prefix V_i of length $s-1$ (i.e. a subset of A'), P_i is the set of sets of $\mathcal{H}$ that project to V_i. Fix some $i \in [t]$ and denote by b the number of sets in P_i. Depending on whether the sets in P_i contain a or not, we further divide them into two classes of size b_0 (if they contain a) and b_1 (otherwise).



Pictorially in terms of unit distance graph, each vertex V_i of $G_U(\mathcal{H}_{|A'})$ corresponds to a unique equivalence class P_i, which may split into two vertices connected by an edge corresponding to $\{a\}$ in the unit distance graph of $\mathcal{H}_{|A}$.



We want to compute the weight W_a of all the edges of $G_U(\mathcal{H}_A)$ that correspond to the element a. For that purpose, we start by computing a lower bound for the expected weight of the edge uniquely associated to P_i (it is 0 if it doesn't exist). By definition, this weight is simply $\min\{b_0, b_1\} \geqslant \frac{b_0 \cdot b_1}{b_0 + b_1} = \frac{b_0 \cdot b_1}{b}$. Since the value of b is independent of the choice of a, we get that:

$$
\begin{aligned}
\mathbb{E}[min(b_0, b_1)] &\geqslant \mathbb{E}\left[\frac{b_0 b_1}{b}\right] \\
&= \frac{\mathbb{E}[b_0 b_1]}{b} \\
&= \frac{(\text{number of pairs of sets of } P_i)\mathbb{P}\{\text{two sets of } P_i \text{ differ on } a\}}{b} \\
&= \frac{\binom{b}{2}\mathbb{P}\{\text{two sets of } P_i \text{ differ on } a\}}{b}
\end{aligned}
$$

Since by assumption all sets of $\mathcal{H}$ are ϵ-separated, and sets inside an equivalence class agree on at least $s-1$ points, the probability that a is in the symmetric difference of two sets of P_i is at least $\frac{\epsilon}{n-(s-1)} \geqslant \frac{\epsilon}{n}$. Thus the expected weight contribution of P_i is at least:

$$
\mathbb{E}\{\text{weight contribution of } P_i\} = \frac{\binom{b}{2}\frac{\epsilon}{n}}{b} = \frac{(b-1)\epsilon}{2n}
$$

We can now sum over all equivalences classes to retrieve $\mathbb{E}\{W_a\}$. Recall that since equivalence classes correspond to sets of $\mathcal{H}_{|A'}$ and $\mathcal{H}_{|A'}$ has VC-dimension at most d, by the Primal Shatter Lemma there are less than $C(\frac{\epsilon}{n})^d$ sets in such a set system, for some constant $C > 0$.

$$
\begin{aligned}
\mathbb{E}\{W_a\} &= \sum_{i=1}^{t} \mathbb{E}\{\text{weight contribution of } P_i\} \\
&\geqslant \sum_{i=1}^{t} \frac{(|P_i| - 1)\epsilon}{2n} \\
&= \frac{\epsilon}{2n}(|\mathcal{H}| - t) \\
&\geqslant \frac{\epsilon}{2n}\left(|\mathcal{H}| - C\left(\frac{\epsilon}{n}\right)^d\right)
\end{aligned}
$$

Setting $s = \lceil \frac{8dn}{\epsilon} \rceil$, we get:

$$\begin{aligned}
\mathbb{E}\{W\} &= s \cdot \mathbb{E}\{W_a\} \\
&\geqslant \frac{s\epsilon}{2n}\left(|\mathcal{H}| - C\left(\frac{n}{\epsilon}\right)^d\right) \\
&= 4d|\mathcal{H}| - 4dC\left(\frac{n}{\epsilon}\right)^d
\end{aligned}$$

Putting lower bound and upper bound together, we finally obtain:

$$\begin{aligned}
4d|\mathcal{H}| - 4dC\left(\frac{n}{\epsilon}\right)^d \leqslant \mathbb{E}\{W\} \leqslant 2d|\mathcal{H}| \\
4d|\mathcal{H}| - 4dC\left(\frac{n}{\epsilon}\right)^d \leqslant 2d|\mathcal{H}| \\
2d|\mathcal{H}| \leqslant 4dC\left(\frac{n}{\epsilon}\right)^d \\
|\mathcal{H}| \leqslant 2C\left(\frac{n}{\epsilon}\right)^d \\
|\mathcal{H}| = O\left(\left(\frac{n}{\epsilon}\right)^d\right)
\end{aligned}$$

References

[1] Peter Frankl and János Pach. On the number of sets in a null t-design. *European Journal of Combinatorics*, 4(1):21–23, 1983. ↑2

[2] Timothy Gowers. Gower's webblog. https://gowers.wordpress.com/2008/07/31/dimension-arguments-in-combinatorics/. Self-published. ↑2

[3] David Haussler. Sphere packing numbers for subsets of the boolean n-cube with bounded vapnik-chervonenkis dimension. *Journal of Combinatorial Theory, Series A*, 69(2):217–232, 1995. ↑2

[4] David Haussler and Emo Welzl. Epsilon-nets and simplex range queries. In *Proceedings of the second annual symposium on Computational geometry*, pages 61–71, 1986. ↑1

[5] Jiri Matousek. *Geometric discrepancy: An illustrated guide*, volume 18. Springer Science & Business Media, 2009. ↑2

[6] Nabil H Mustafa. A simple proof of the shallow packing lemma. *Discrete & Computational Geometry*, 55:739–743, 2016. ↑2

[7] Nabil H Mustafa. *Sampling in combinatorial and geometric set systems*, volume 265. American Mathematical Society, 2022. ↑2

TYPE IN DIKU ADDRESS
Email address: flbr@di.ku.dk
Email address: qpt351@ku.dk