

# Tight list replicability bounds via a novel sphere covering theorem

**Ari Blondal**

*McGill University*

ARI.BLONDAL@MAIL.MCGILL.CA

**Hamed Hatami**

*McGill University*

HAMED.HATAMI@MCGILL.CA

**Pooya Hatami**

*Ohio State University*

HATAMI.2@OSU.EDU

**Chavdar Lalov**

*Ohio State University*

LALOV.1@OSU.EDU

**Sivan Tretiak**

*Ohio State University*

TRETIK.2@OSU.EDU

**Editors:** Steve Hanneke and Tor Lattimore

## Abstract

In recent years, list replicability has emerged as a framework for formalizing reproducibility in learning theory. A central question is how the required list size relates to the accuracy parameter and natural complexity measures of the hypothesis class.

To achieve sharp bounds on list replicability, we prove a novel topological sphere covering theorem, derived from the Borsuk-Ulam theorem. Specifically, if the  $d$ -sphere is covered by open sets, each of which lies in an open hemisphere, then  $d + 1$  of these sets must have a common intersection. Using this result, we obtain a sharp bound on the relationship between list size and accuracy for VC classes. We also show that for large-margin half-spaces, provided the margin is not too large, the optimal list size equals the ambient dimension. However, when the margin is taken to be very large, we devise a replicable algorithm achieving the minimal list size of  $\lceil d/2 \rceil + 1$ .

**Keywords:** Replicability, PAC learning, Stability, Topology

## 1. Introduction

Randomized learning algorithms can produce different hypotheses across multiple executions, even when trained on data drawn from the same distribution. In many contexts, it is desirable for the algorithm to produce consistent outputs across such runs. This idea is also connected to broader discussions of scientific reproducibility, in which repeated experiments are expected to yield consistent conclusions. Motivated by this perspective, a growing body of work in learning theory has introduced formal notions describing when and how randomized algorithms can be made replicable (Bun et al., 2020; Malliaris and Moran, 2022; Chase et al., 2023; Dixon et al., 2023; Bun et al., 2023; Karbasi et al., 2023; Esfandiari et al., 2023a,b; Moran et al., 2023; Eaton et al., 2023; Kalavasis et al., 2024, 2023).

One such notion is *list replicability*, introduced in Chase et al. (2023); Dixon et al. (2023). Informally, the list replicability number of a concept class is the smallest integer  $L$  for which there exists a learning algorithm whose set of likely output hypotheses has size at most  $L$  under every data distribution. This quantity also characterizes the amount of shared randomness required for replicable learning (Impagliazzo et al., 2022; Hopkins and Moran, 2025), and has found applications in differentially private learning (Alon et al., 2022; Ghazi et al., 2021b).

Although the definition of list replicability is purely algorithmic, there has been widespread success in reformulating it as a topological property of the space of distributions that are realizable in the given learning task (Chase et al., 2023; Dixon et al., 2023; Chase et al., 2024; Chornomaz et al., 2025; Blondal et al., 2025a,b). This perspective has enabled the use of tools from algebraic topology to derive general bounds on list size. In particular, Chase et al. (2024) utilized a local version of the Borsuk-Ulam theorem as a powerful tool to lower-bound list replicability. They showed that for any finite cover of the  $d$ -sphere by antipodal-free open sets, there are at least  $\lceil d/2 \rceil + 1$  sets with a common nonempty intersection. This sphere covering result was used to show that the list size of any list-replicable learner for a concept class of VC dimension  $d$  is at least  $\lceil d/2 \rceil + 1$ , independent of the accuracy parameter  $\epsilon \in (0, 1/2)$ . Furthermore, Blondal et al. (2025a) applied the sphere covering result to the class of large-margin half-spaces, showing that its list-replicability number lies between  $\lceil d/2 \rceil + 1$  and  $d$ .

In the local Borsuk–Ulam theorem of Chase et al. (2024), the factor of  $1/2$  is inherent and cannot be improved. In contrast, in its application to list-replicability, this factor is an artifact of the proof technique. In particular, the lower bound of  $\lceil d/2 \rceil + 1$  on the list size is not tight: when the accuracy parameter satisfies  $\epsilon < 1/d$ , the list size must be at least  $d$  (Chase et al. (2023)). Together with the persistent gap between known lower and upper bounds for the list replicability of large-margin halfspaces, this indicates that the appearance of the  $1/2$  factor is due to limitations of the topological tools used so far, rather than an inherent barrier.

We confirm this hypothesis by introducing a novel sphere covering theorem. Instead of merely assuming antipodal-freeness, we impose the stronger condition that each open set in the cover of the  $d$ -sphere is contained in an open hemisphere. Under this assumption, we show that at least  $d + 1$  sets must have a common nonempty intersection. This result yields both sharper and more general bounds for list replicability across several settings.

Using our new sphere covering theorem, we prove that the list size of any list-replicable learner for a concept class of VC dimension  $d$  is at least  $d$  for any  $\epsilon < 1/2$ . Previously, this bound was only known for  $\epsilon < 1/d$ . As a consequence, the optimal list size for the concept class  $\{\pm 1\}^d$  is exactly  $d$ , independent of the accuracy parameter  $\epsilon$ .

In the setting of large-margin half-spaces, we apply our sphere covering result to show that for margins  $\gamma < 1/\sqrt{2}$ , the list replicability number is exactly  $d$ . In contrast, when  $\gamma$  is very close to 1, we construct an explicit learner showing that the list replicability number is  $\lceil d/2 \rceil + 1$ . Together, these results show that the lower and upper bounds from Blondal et al. (2025a) are tight in the appropriate parameter regimes.

Finally, we study list replicability for large-margin half-spaces where the learner is restricted to outputting linear classifiers. For such algorithms, we show that the optimal list replicability is  $d$ , again matching the upper bound in Blondal et al. (2025a).

**Paper organization** In Section 1.1 we collect the fundamental definitions broadly necessary for our main results. Definitions and theorems useful beyond this context are provided in their respective sections.

Our main results are contained in Section 2. We begin by situating our sphere covering result, Theorem 2, within the broader framework of topological methods for list replicability. With this result in hand, we turn to our improved lower bound on  $\epsilon$ -list replicability of VC classes in Theorem 4. Afterwards, we present our new lower and upper bounds on the list replicability number of

large-margin half-spaces in [Theorem 6](#) and [Theorem 7](#). Finally, an application of our topological result yields a tight lower bound on the list replicability number of linear classifiers in [Theorem 8](#).

As the proofs for [Theorems 4, 6 and 8](#) share many similarities, we have extracted those similarities into a framework theorem in [Section 4](#), and deferred the full proofs to [Appendices A, B and C](#) respectively.

### 1.1. Preliminaries

**Partial concept classes.** The framework of partial concept classes, introduced by Alon, Hanneke, Holzman, and Moran ([Alon et al., 2021](#)), extends classical learning theory to settings in which the data is guaranteed to satisfy additional structural assumptions that enable efficient learning. In particular, partial concept classes provide a natural framework for our study of large-margin half-spaces.

A *partial concept class* over an arbitrary domain  $\mathcal{X}$  is a set  $\mathcal{C} \subseteq \{\pm 1, \star\}^{\mathcal{X}}$ , where each function  $c \in \mathcal{C}$  is called a *partial concept*. We say that a partial concept  $c$  is *undefined* at  $x$  whenever  $c(x)$  takes the star value  $\star$ . The *support* of a partial concept  $c$  is defined to be  $\text{supp}(c) := \{x \in \mathcal{X} : c(x) \neq \star\}$ . Note that a *total* concept class  $\mathcal{C} \subseteq \{\pm 1\}^{\mathcal{X}}$  is a special case of a partial concept class.

**PAC learning.** In the *probably approximately correct (PAC) learning* framework, the learner observes  $n$  independent labeled examples  $S = ((x_1, y_1), \dots, (x_n, y_n))$  sampled from an unknown but fixed distribution  $\mu$  over  $\mathcal{X} \times \{\pm 1\}$ . The objective is to produce a *hypothesis*  $h: \mathcal{X} \rightarrow \{\pm 1\}$  that, with high probability, performs well on the overall distribution, as measured by the *population loss*

$$\text{loss}_{\mu}(h) := \mathbb{P}_{(x,y) \sim \mu}[h(x) \neq y].$$

Note that  $\mu$  is a distribution on  $\mathcal{X} \times \{\pm 1\}$ , and therefore the learner never observes labels equal to  $\star$ . Moreover, the hypothesis produced by the learner should not contain  $\star$  labels, since any such label would automatically be counted as an error.

We define a *learning rule* to be a (possibly randomized) function  $\mathcal{A}$  that maps any sample  $S \in \bigcup_{n=0}^{\infty} (\mathcal{X} \times \{\pm 1\})^n$  to a hypothesis  $\mathcal{A}(S) \in \{\pm 1\}^{\mathcal{X}}$ . Since our primary focus is learnability rather than computational efficiency, we impose no computability constraints on  $\mathcal{A}$ .

A distribution  $\mu$  over  $\mathcal{X} \times \{\pm 1\}$  is *realizable* by a partial concept class  $\mathcal{C}$  if for every  $n \in \mathbb{N}$ , a random sample  $S = ((x_i, y_i))_{i=1}^n \sim \mu^n$  is almost surely realizable by some  $c \in \mathcal{C}$ , that is,  $c(x_i) = y_i$  for all  $i = 1, \dots, n$ .

A partial concept  $\mathcal{C}$  is *PAC learnable* by a learning rule  $\mathcal{A}$  if for any  $\epsilon, \delta > 0$ , there exists a *sample complexity*  $n := n(\epsilon, \delta)$  such that for any realizable distribution  $\mu$ , we have

$$\mathbb{P}_{S \sim \mu^n}[\text{loss}_{\mu}(\mathcal{A}(S)) \leq \epsilon] \geq 1 - \delta.$$

**VC dimension.** Let  $A \subseteq \mathcal{X}$  be a subset and  $\mathcal{C}$  be a partial concept class over  $\mathcal{X}$ . We say  $\mathcal{C}$  *shatters*  $A$  if  $\{\pm 1\}^A \subseteq \{c|_A : c \in \mathcal{C}\}$ . The VC dimension of  $\mathcal{C}$  is defined by

$$\text{VC}(\mathcal{C}) := \sup \{|A| : A \subseteq \mathcal{X} \text{ is shattered by } \mathcal{C}\}.$$

As in the total setting, it is shown in ([Alon et al., 2021](#), Theorem 3) that the finiteness of the VC dimension exactly characterizes PAC learnability for a partial concept class  $\mathcal{C}$ . In fact, there exists a learning rule with sample complexity  $O_{\epsilon, \delta}(\text{VC}(\mathcal{C}))$ .

**Large-margin half-spaces.** One of the most prominent examples of a partial concept class is that of large-margin half-spaces. In this setting, we use undefined labels to formalize the large-margin assumption: data points are guaranteed to be well separated.

In the large-margin half-spaces learning problem, the domain is the unit sphere  $\mathbb{S}^{d-1} \subset \mathbb{R}^d$ . For each  $w \in \mathbb{S}^{d-1}$ , we define a partial concept

$$c_w(x) := \begin{cases} \text{sign}(\langle w, x \rangle) & \text{if } |\langle w, x \rangle| \geq \gamma \\ \star & \text{otherwise} \end{cases},$$

where  $\gamma > 0$  is the margin parameter. Note that  $c_w$  deems every  $x$  within distance  $\gamma$  from the hyperplane defined by  $w$  as undefined. Otherwise, it assigns  $\pm 1$  depending on which side of the hyperplane  $x$  is on. We denote by  $\mathcal{H}_\gamma^d$  the set  $\{c_w : w \in \mathbb{S}^{d-1}\}$  of all  $\gamma$ -margin partial concepts. It is well known from the analysis of the Perceptron algorithm (McCulloch and Pitts (1943); Rosenblatt (1958), see also (Shalev-Shwartz and Ben-David, 2014, Theorem 9.1), and (Alon et al., 2021, Proposition 17)), that the *Littlestone dimension* of  $\mathcal{H}_\gamma^d$  is bounded by  $1/\gamma^2$ . Since the Littlestone dimension is a relaxation of the VC dimension, we also have

$$\text{vc}(\mathcal{H}_\gamma^d) \leq \frac{1}{\gamma^2}$$

which is independent of the dimension  $d$ . This implies that under the large-margin assumption  $\gamma > 0$ , linear classification is efficiently PAC learnable even in high dimensions.

**List replicability.** List replicability was introduced in Chase et al. (2023); Dixon et al. (2023) as a framework for studying the replicability of a learning problem while preserving the guarantees of PAC learning.

**Definition 1 (List replicability)** *A learning rule  $\mathcal{A}$  is an  $(\epsilon, L)$ -list replicable learner for a partial concept class  $\mathcal{C}$  if for every  $\delta > 0$  there exists a sample complexity  $n := n(\delta)$  such that the following holds. For every distribution  $\mu$  realizable by  $\mathcal{C}$ , there exists a list of hypothesis  $h_1, \dots, h_L \in \{\pm 1\}^{\mathcal{X}}$  such that*

$$\text{loss}_\mu(h_i) \leq \epsilon \forall i \text{ and } \mathbb{P}_{S \sim \mu^n}[\mathcal{A}(S) \in \{h_1, \dots, h_L\}] \geq 1 - \delta.$$

The  $\epsilon$ -list replicability number of  $\mathcal{C}$  is

$$\text{LR}(\mathcal{C}, \epsilon) := \min\{L : \exists(\epsilon, L)\text{-list replicable learner for } \mathcal{C}\},$$

with  $\text{LR}(\mathcal{C}, \epsilon) = \infty$  if none exists. The list replicability number of  $\mathcal{C}$  is

$$\text{LR}(\mathcal{C}) := \sup_{\epsilon > 0} \text{LR}(\mathcal{C}, \epsilon).$$

We say  $\mathcal{C}$  is list replicable if  $\text{LR}(\mathcal{C}) < \infty$ .

The goal in list replicability is to produce a PAC learner with the smallest possible list size.

**Space of realizable distributions.** Given a partial concept class  $\mathcal{C}$ , we define the set

$$\Delta_{\mathcal{C}} := \{\mu : \mu \text{ realizable by } \mathcal{C}\}$$

of all realizable distributions. When equipped with the *total variation (TV) distance*,  $\Delta_{\mathcal{C}}$  forms a metric space, which we refer to as *the space of realizable distributions* associated with  $\mathcal{C}$ .

**Topological terminology.** Let  $\mathbb{S}^d \subset \mathbb{R}^{d+1}$  denote the  $d$ -dimensional unit sphere. We say that a set  $U \subseteq \mathbb{S}^d$  is *antipodal-free* if it does not contain both a point  $x$  and its antipode  $-x$ . We say a cover  $\mathcal{U} = \{U_i\}_{i \in I}$  of  $\mathbb{S}^d$  is antipodal-free if  $U_i$  is antipodal-free for all  $i \in I$ .

The *overlap degree* of a cover  $\mathcal{U} = \{U_i\}_{i \in I}$  is the largest integer  $k$  for which there exists  $k$  sets  $U_1, \dots, U_k \in \mathcal{U}$  such that  $\bigcap_{i=1}^k U_i \neq \emptyset$ .

If  $\mathcal{U} = \{U_i\}_{i \in I}$  is an open cover, then a *partition of unity* subordinate to  $\mathcal{U}$  is a collection of continuous maps  $f_i: \mathbb{S}^d \rightarrow [0, 1]$  such that the support of  $f_i$  is contained in  $U_i$  for each  $i \in I$ , and  $f_1 + \dots + f_m = 1$ .

Any point  $p \in \mathbb{S}^d$  defines a unique orthogonal homogeneous hyperplane in  $\mathbb{R}^{d+1}$  given by

$$P = \left\{ x \in \mathbb{R}^{d+1} : \langle p, x \rangle = 0 \right\}.$$

For each point  $w \neq p$  in  $\mathbb{S}^d$ , there is a unique line through  $p$  and  $w$ , which intersects the plane  $P$  in a unique point  $w'$ . The *stereographic projection* through  $p$  is the injective map  $\pi: \mathbb{S}^d \setminus \{p\} \rightarrow P \cong \mathbb{R}^d$  given by taking  $w \in \mathbb{S}^d$  to  $w' \in P$ .

## 2. Main Results

### 2.1. Lower-bounding list replicability via topology

The most well-established strategy to lower-bound the  $\epsilon$ -list replicability number  $\text{LR}(\mathcal{C}, \epsilon)$  of a class  $\mathcal{C}$  hinges on exploiting the topological structure of its space of realizable distributions  $\Delta_{\mathcal{C}}$ .

This method was independently introduced by [Chase, Moran, and Yehudayoff \(2023\)](#) and by [Dixon, Pavan, Woude, and Vinodchandran \(2023\)](#), and was subsequently refined and extended ([Chase et al., 2024](#); [Vander Woude et al., 2024](#); [Blondal et al., 2025a](#); [Chornomaz et al., 2025](#); [Blondal et al., 2025b](#)) to address a broader range of problems. These works employ a variety of topological results, such as the Poincaré-Miranda theorem ([Chase et al., 2023](#)), KKM/Sperner's Lemma ([Dixon et al., 2023](#); [Vander Woude et al., 2024](#)), a Local Borsuk-Ulam theorem ([Chase et al., 2024](#); [Blondal et al., 2025a](#); [Chornomaz et al., 2025](#)), and Lebesgue's covering theorem ([Blondal et al., 2025b](#)), to obtain lower bounds on the list replicability number of a variety of concept classes.

In this vein, our first contribution is a sphere-covering result on the overlap degree of open covers with small diameters.

**Theorem 2** *Let  $A_1, \dots, A_m$  be a finite open cover of the  $d$ -sphere  $\mathbb{S}^d$ , where each  $A_i$  is contained in an open hemisphere. Then  $d + 1$  of these sets have a common nonempty intersection.*

We refer the reader to [Lemma 10](#) for a precise framework detailing the use of [Theorem 2](#) in proving lower bounds for list replicability.

### 2.2. List replicability of finite VC classes

In [Chase et al. \(2024\)](#), Chase, Chornomaz, Moran, and Yehudayoff asked whether concept classes with finite VC dimension can be learned in a list-replicable way with a small list size when the accuracy parameter  $\epsilon$  is large. The following bounds are known:

**Theorem 3 (Chase et al. (2024, 2023))** *Let  $\mathcal{C}$  be a concept class with VC dimension  $d$ . Then*

$$\begin{aligned} &\text{for } \epsilon < \frac{1}{2}, \text{LR}(\mathcal{C}, \epsilon) \geq \frac{d}{2} + 1 \\ &\text{whereas for } \epsilon < \frac{1}{d}, \text{LR}(\mathcal{C}, \epsilon) \geq d. \end{aligned}$$

Our first application of [Theorem 2](#) is to show that  $d$  is the correct lower bound for the full range  $\epsilon \in [0, 1/2)$ .

**Theorem 4** *Let  $\mathcal{C}$  be a (partial) concept class with VC dimension  $d$ .*

$$\text{For any } \epsilon < \frac{1}{2}, \text{LR}(\mathcal{C}, \epsilon) \geq d.$$

Note that for the binary cube  $\mathcal{C} = \{\pm 1\}^d$ , we have  $\text{VC}(\mathcal{C}) = d$  and  $\text{LR}(\mathcal{C}, \epsilon) \leq d$  for all  $\epsilon < 1/2$  [Chase et al. \(2023\)](#). Consequently, [Theorem 4](#) implies that the  $\epsilon$ -list replicability number of the binary cube equals  $d$  for all  $\epsilon < 1/2$ . In other words, increasing the accuracy parameter  $\epsilon$  does not yield any improvement in list replicability for some finite VC classes.

### 2.3. List replicability of large-margin half-spaces

The list replicability number of a concept class provides bounds on various learning theory and communication complexity parameters such as VC dimension [Chase et al. \(2023, 2024\)](#), Littlestone dimension [Alon et al. \(2022\)](#); [Ghazi et al. \(2021a\)](#)<sup>1</sup>, and sign-rank [Blondal et al. \(2025a,b\)](#). In particular, several previously open questions in these areas were resolved by studying the list replicability number of large-margin half-spaces [Blondal et al. \(2025a\)](#). The explicit bounds on the list replicability number of the class depended only on the underlying dimension  $d$  of the problem:

**Theorem 5 (Blondal et al. (2025a))** *For any fixed dimension  $d > 1$ , margin  $\gamma \in (0, 1)$ , and accuracy parameter  $\epsilon \in (0, 1/2)$ ,*

$$\left\lceil \frac{d}{2} \right\rceil + 1 \leq \text{LR}(\mathcal{H}_\gamma^d, \epsilon) \leq d.$$

Hence,  $\left\lceil \frac{d}{2} \right\rceil + 1 \leq \text{LR}(\mathcal{H}_\gamma^d) \leq d$ .

These general bounds left open the question of precisely how the margin  $\gamma$  and accuracy parameter  $\epsilon$  affect the  $\epsilon$ -list replicability number, if at all. We give a partial answer to this question. Our second application of [Theorem 2](#) improves the lower bounds for  $\epsilon$ -list replicability of large-margin half-spaces for a wide range of margins  $\gamma$ . This determines the exact value of the list replicability number in this range.

**Theorem 6** *For any dimension  $d > 1$  and margin  $\gamma \in (0, \frac{1}{\sqrt{2}})$ , there exists an accuracy parameter  $\epsilon := \epsilon(\gamma, d)$  such that*

$$\text{LR}(\mathcal{H}_\gamma^d, \epsilon) \geq d.$$

Hence,  $\text{LR}(\mathcal{H}_\gamma^d) = d$  for all  $\gamma \in (0, 1/\sqrt{2})$ .

<sup>1</sup> The bounds in [Alon et al. \(2022\)](#); [Ghazi et al. \(2021a\)](#) are not explicitly stated but can be found in [Chase et al. \(2023\)](#).

As a complement to this result, we examine the setting in which the margin  $\gamma$  is relatively large. In this setting, we prove a tight upper bound on the  $\epsilon$ -list replicability number.

**Theorem 7** *For any dimension  $d > 1$  and accuracy parameter  $\epsilon \in [0, 1/2)$ , there exists some  $\gamma_0 := \gamma_0(d) \in (0, 1)$  such that for any margin  $\gamma \in (\gamma_0, 1)$  we have*

$$\text{LR}(\mathcal{H}_\gamma^d, \epsilon) \leq \left\lceil \frac{d}{2} \right\rceil + 1.$$

Hence,  $\text{LR}(\mathcal{H}_\gamma^d) = \lceil \frac{d}{2} \rceil + 1$  for all  $\gamma \in (\gamma_0(d), 1)$ .

This proof uses an explicit learning rule which exploits a particular cover originally given by Chase, Chornomaz, Moran, and Yehudayoff (Chase et al. (2024), see Theorem 11 for a description), in conjunction with the Lebesgue Number Lemma (Theorem 12).

### Linear classifiers

Our third and final application of Theorem 2 addresses the question of list-replicably learning the large-margin half-space problem with linear classifiers. A hypothesis on the domain  $\mathbb{S}^{d-1}$  is a *linear classifier* if it is of the form

$$h_w(x) := \begin{cases} 1 & \text{if } \langle w, x \rangle \geq 0, \\ -1 & \text{if } \langle w, x \rangle < 0 \end{cases}$$

for some  $w \in \mathbb{S}^{d-1}$ .

The upper bound  $\text{LR}(\mathcal{H}_\gamma^d, \epsilon) \leq d$  from Theorem 5 uses a learning rule which always outputs a linear classifier. We show that, under that restriction, a list of size  $d$  is best possible.

**Theorem 8** *Fix a dimension  $d > 1$  and a margin  $\gamma \in (0, 1)$ . For any error parameter  $\epsilon \in (0, 1/2)$  and list length  $L > 1$ , if  $\mathcal{A}$  is an  $(\epsilon, L)$ -list replicable learner for  $\mathcal{H}_\gamma^d$  which outputs linear classifiers, then  $L \geq d$ .*

### 2.4. Open questions

In Theorem 4, we remove the optimal lower bound's dependency on  $\epsilon$ , while we get a new dependency on  $\epsilon$  in the lower bound of  $\text{LR}(\mathcal{H}_\gamma^d, \epsilon)$ . We wonder if list replicability varies with  $\epsilon$ .

**Question 1** *Is there a total or partial concept class  $\mathcal{C}$  and two error parameters  $\epsilon_1, \epsilon_2 \in (0, \frac{1}{2})$  with  $\epsilon_1 < \epsilon_2$  such that*

$$\text{LR}(\mathcal{C}, \epsilon_1) > \text{LR}(\mathcal{C}, \epsilon_2)?$$

While we have solved  $\text{LR}(\mathcal{H}_\gamma^d)$  for many of the most applicable regimes, the behaviour of the list replicability number is merely bounded in other regimes. Looking at Figure 1, we ask what the transition between the top and bottom regimes looks like.

**Question 2** *For which parameters can the bound  $\text{LR}(\mathcal{H}_\gamma^d, \epsilon) = d$  be established?*

**Question 3** *Are there values of  $\gamma \in (0, 1), \epsilon \in (0, \frac{1}{2})$  such that  $\text{LR}(\mathcal{H}_\gamma^d, \epsilon)$  takes on every integer value in the range  $[\lceil \frac{d}{2} \rceil + 1, d]$ ?*

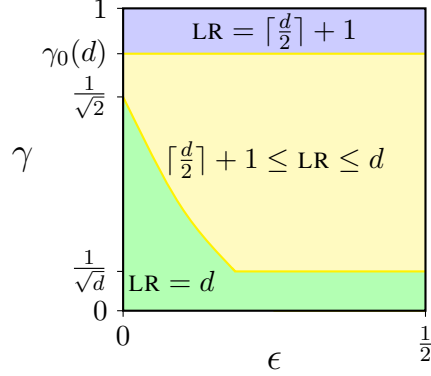


Figure 1: Regimes for  $\text{LR}(\mathcal{H}_\gamma^d, \epsilon)$  when  $d > 1$ . When  $\gamma < \frac{1}{\sqrt{d}}$ , it is well known that  $\text{vc}(\mathcal{H}_\gamma^d) = d$ , so we know from [Theorem 4](#) that  $\text{LR} = d$ .

### 3. Proof of [Theorem 2](#)

In this section, we prove our main topological result.

**Theorem 2** *Let  $A_1, \dots, A_m$  be a finite open cover of the  $d$ -sphere  $\mathbb{S}^d$ , where each  $A_i$  is contained in an open hemisphere. Then  $d + 1$  of these sets have a common nonempty intersection.*

We will use the following classical form of the Borsuk-Ulam theorem.

**Theorem 9** ([\(Matoušek, 2003, Theorem 2.1.1\)](#)) *For every continuous mapping  $f: \mathbb{S}^d \rightarrow \mathbb{R}^d$ , there exists a point  $x \in \mathbb{S}^d$  with  $f(x) = f(-x)$ .*

**Proof of [Theorem 2](#):** Let  $\mathcal{A} = \{A_1, \dots, A_m\}$  be a finite open cover of  $\mathbb{S}^d$  such that there exist  $u_1, \dots, u_m \in \mathbb{S}^d$  with

$$A_i \subseteq \{x \in \mathbb{S}^d : \langle x, u_i \rangle > 0\}.$$

Since  $\mathbb{S}^d$  is a compact Hausdorff space, there is a partition of unity subordinate to  $\mathcal{A}$  ([Rudin, 1987, Theorem 2.13](#)). That is, there exist continuous maps  $f_i: \mathbb{S}^d \rightarrow [0, 1]$  such that the support of  $f_i$  is contained in  $A_i$  for each  $i$ , and  $f_1 + \dots + f_m = 1$ .

Consider the continuous map  $g: \mathbb{S}^d \rightarrow \mathbb{R}^{d+1}$  given by

$$g(x) := \sum_{i=1}^m f_i(x)u_i.$$

Note that  $\langle x, g(x) \rangle > 0$  for all  $x$ , and in particular  $g$  is never zero. It follows that  $h(x) := g(x)/\|g(x)\|$  is a well-defined, continuous map from  $\mathbb{S}^d$  to  $\mathbb{S}^d \subset \mathbb{R}^{d+1}$ . Moreover,  $h$  inherits the property that  $\langle x, h(x) \rangle > 0$  for all  $x$ .

We claim that  $h$  must be surjective. Indeed, assuming otherwise, let  $p \in \mathbb{S}^d$  be a point outside the image of  $h$ , and let  $\pi: \mathbb{S}^d \rightarrow \mathbb{R}^d$  be the stereographic projection through  $p$ . By [Theorem 9](#), the continuous map  $\pi \circ h: \mathbb{S}^d \rightarrow \mathbb{R}^d$  must identify some antipodal pair  $w$  and  $-w$ . Since stereographic

projection is bijective, it follows that  $h(w) = h(-w)$ . This contradicts the fact that  $\langle x, h(x) \rangle > 0$  for all  $x$ .

To complete the proof, we will show that the surjectivity of  $h$  guarantees the overlap of at least  $d + 1$  sets of  $\mathcal{A}$ . The image of  $g$  is a subset of the space  $T$ , composed of the union of a finite number of convex sets.

$$T := \bigcup \left\{ \sigma_S : \text{for all } S \subset [m] \text{ where } \bigcap_{i \in S} A_i \neq \emptyset \right\}$$

$$\text{where } \sigma_S := \text{conv} \{u_i : i \in S\}$$

Since we only consider  $S$  where the corresponding open sets overlap, for each  $S$  there is some  $x_S$  with  $\langle x_S, u_i \rangle > 0$  for all  $i$  in  $S$ . This means that  $\sigma_S$  doesn't intersect 0, and its radial projection onto the sphere lies in a vector space of dimension at most  $|S|$ :

$$\left\{ \frac{x}{|x|} : x \in \sigma_S \right\} \subseteq \mathbb{S}^d \cap \text{span} \{u_i : i \in S\}.$$

Finally, if this vector space has dimension less than  $d + 1$ , its intersection with  $\mathbb{S}^d$  has measure 0 in  $\mathbb{S}^d$ . Therefore, if no  $d + 1$  sets intersect, then the image of  $h$  is contained in a finite union of sets with 0 measure, which itself has 0 measure, contradicting its surjectivity.  $\blacksquare$

#### 4. The topological method in list replicability

In this section, we describe a framework for applying [Theorem 2](#) to obtain lower bounds on list replicability. This framework is applied to prove [Theorems 4, 6 and 8](#), so we explain it in detail, with its applications deferred to [Appendices A, B and C](#) respectively.

The overarching strategy begins by identifying a subset of realizable distributions,  $\Delta$ , that, when endowed with the total variation metric, is homeomorphic to a high-dimensional sphere or ball. The existence of an  $(\epsilon, L)$ -list replicable algorithm  $\mathcal{A}$  then induces an open cover  $\mathcal{U}$  of this subset whose overlap degree lower bounds the list size  $L$ . If  $\mathcal{U}$  satisfies additional structural constraints, such as bounds on the size of the open sets or antipodal freeness, then we invoke a suitable sphere covering theorem to show that any open cover satisfying these properties must exhibit a large overlap, thereby yielding the desired lower bound.

For our purposes, [Theorem 2](#) is the appropriate sphere covering theorem. Our specific realization of the strategy above is described in the following lemma.

**Lemma 10** *Let  $\mathcal{A}$  be an  $(\epsilon, L)$ -list replicable learning algorithm for a concept class  $\mathcal{C} \subseteq \{\pm 1, \star\}^{\mathcal{X}}$  with sample complexity  $n := n(\delta)$ . Fix a dimension  $d \geq 1$  and a confidence parameter  $\delta \leq \frac{1}{2d}$ , and suppose that, for some subset of realizable distributions  $\Delta \subseteq \Delta_{\mathcal{C}}$ , there is a continuous map  $\varphi: \mathbb{S}^{d-1} \rightarrow \Delta$ .*

*If there is an  $\epsilon_0 > \epsilon$  such that, for every output  $h$  of  $\mathcal{A}$ , the set*

$$U_h := \left\{ w \in \mathbb{S}^{d-1} : \mathbb{P}_{S \sim \varphi(w)^n} [\mathcal{A}(S) = h] > \frac{1 - 2\delta}{L} \text{ and } \text{loss}_{\varphi(w)}(h) < \epsilon_0 \right\}$$

*is contained in an open hemisphere of  $\mathbb{S}^{d-1}$ , then  $L \geq d$ . If this holds for any  $\mathcal{A}$  then  $\text{LR}(\mathcal{C}, \epsilon) \geq d$ .*

Each set  $U_h$  collects all points in  $\mathbb{S}^{d-1}$  which, under  $\varphi$ , encode distributions for which  $h$  is a “likely” and “accurate” output of  $\mathcal{A}$ . In application the size of these sets will be controlled with  $\varepsilon$ , whereas the other parameters  $\delta$ ,  $n$ , and  $\mathcal{A}$  are less impactful.

**Proof** First, we will show that the family  $\mathcal{U}$  is open. Let  $X$  be the set of realizable distributions  $x$  for which  $\text{loss}_x(h) < \epsilon_0$ . Since loss is a continuous function,  $X$  is open with respect to TV distance. Similarly,  $\varphi^{-1}(X)$  is in turn open because of the continuity of  $\varphi$ . Note that  $\varphi^{-1}(X)$  is exactly those  $w \in \mathbb{S}^{d-1}$  satisfying  $\text{loss}_{\varphi(w)}(h) < \epsilon_0$ . An analogous argument shows that the set of  $w \in \mathbb{S}^{d-1}$  satisfying  $\mathbb{P}_{S \sim \varphi(w)^n}[\mathcal{A}(S) = h] > \frac{1-2\delta}{L}$  is also open. Therefore, each  $U_h$  is open because it is the intersection of two open sets.

Second, we will show how the  $(\epsilon, L)$ -list replicability of  $\mathcal{A}$  guarantees that  $\mathcal{U}$  is a cover. For any element  $w \in \mathbb{S}^{d-1}$ , the distribution  $\varphi(w)$  is realizable by the definition of  $\varphi$ . The list replicability assumption guarantees some set of hypotheses  $\{h_1, \dots, h_L\}$  for which  $\mathbb{P}_{S \sim \varphi(w)^n}[\mathcal{A}(S) \in \{h_1, \dots, h_L\}] \geq 1 - \delta$  and  $\text{loss}_{\varphi(w)}(h_i) \leq \epsilon$ . It follows by the pigeonhole principle that there is some  $h$  in  $\{h_1, \dots, h_L\}$  such that  $U_h$  contains  $w$ .

Finally, we will show the desired lower bound on  $\text{LR}(\mathcal{C}, \epsilon)$ . By assumption, each  $U_h$  is contained in an open hemisphere. Therefore, [Theorem 2](#) guarantees some  $w \in \mathbb{S}^{d-1}$  is contained in  $d$  sets of  $\mathcal{U}$ . By construction, there are  $d$  distinct hypotheses  $h_1, \dots, h_d$  and a distribution  $\varphi(w)$  such that

$$\mathbb{P}_{S \sim \varphi(w)^n}[\mathcal{A}(S) = h_i] > \frac{1 - 2\delta}{L} \text{ for all } i \in [d].$$

Since the events  $[\mathcal{A}(S) = h_i]$  are disjoint, we have that

$$1 > \frac{d(1 - 2\delta)}{L} \geq \frac{d(1 - 1/d)}{L} = \frac{d-1}{L},$$

since  $\delta \leq 1/2d$ . It follows that  $L \geq d$ . If this is true for arbitrary  $\mathcal{A}$ , we conclude that  $\text{LR}(\mathcal{C}, \epsilon) \geq d$  by the definition of the  $\epsilon$ -list replicability number.  $\blacksquare$

## 5. Proof of [Theorem 7](#)

**Theorem 7** *For any dimension  $d > 1$  and accuracy parameter  $\epsilon \in [0, 1/2]$ , there exists some  $\gamma_0 := \gamma_0(d) \in (0, 1)$  such that for any margin  $\gamma \in (\gamma_0, 1)$  we have*

$$\text{LR}(\mathcal{H}_\gamma^d, \epsilon) \leq \left\lceil \frac{d}{2} \right\rceil + 1.$$

Hence,  $\text{LR}(\mathcal{H}_\gamma^d) = \lceil \frac{d}{2} \rceil + 1$  for all  $\gamma \in (\gamma_0(d), 1)$ .

### 5.1. Notation and toolkit

Before we begin we introduce some basic notation. We endow  $\mathbb{S}^{d-1}$  with the  $\ell_2$  metric and for each  $x \in \mathbb{S}^{d-1}$  and all  $\eta > 0$ , we denote by  $B(\eta, x)$  all points in  $\mathbb{S}^{d-1}$  which are less than  $\eta$  away from  $x$ .

For each realizable distribution  $\mu$ , let  $\text{supp}(\mu)$  denote its support. We define

$$\text{supp}^+(\mu) := \left\{ x \in \mathbb{S}^{d-1} : (x, 1) \in \text{supp}(\mu) \right\}$$

to be the positive label support of  $\mu$ , and likewise for  $\text{supp}^-(\mu)$ .

Next we present some topological results that will be used in the proof.

**Theorem 11** ((Chase et al., 2024, Theorem A) ) *There exists a finite antipodal-free open cover  $\mathcal{U} := \{U_i\}_{i=1}^n$  of  $\mathbb{S}^{d-1}$  such that each point  $x \in \mathbb{S}^{d-1}$  is contained in at most  $\lceil \frac{d}{2} \rceil + 1$  of the sets  $U_i$ .*

We also utilize the following standard result in the theory of compact metric spaces.

**Theorem 12 (Lebesgue Number Lemma)** *Let  $(X, d)$  be a compact metric space, and let  $\mathcal{U}$  be an open cover of  $X$ . Then there exists a number  $\eta > 0$  such that for every  $x \in X$ , the open ball  $B(x, \eta)$  is contained in some  $U \in \mathcal{U}$ .*

**Proof** Since  $X$  is compact, we can assume without loss of generality that the cover  $\mathcal{U}$  is finite. Define the function  $f(x) := \max_{i \in I} d(x, U_i^c)$ . By the continuity of  $f$  and compactness of  $X$ , it follows that  $f$  must achieve its minimum. The minimum cannot be zero as  $\mathcal{U}$  is a cover. Taking  $\eta$  to be the minimum of  $f$  completes the proof.  $\blacksquare$

## 5.2. Proof

Pick any  $\epsilon \in [0, 1/2)$ .

Let  $\mathcal{U}$  be the antipodal-free open cover  $\mathcal{U}$  of  $\mathbb{S}^{d-1}$  from Theorem 11. By Theorem 12, there exists a  $\eta > 0$  such that each ball  $B(\eta, x)$  is contained in some  $U_i$ .

Pick  $\gamma(d) > 0$  to be sufficiently close to 1 so that for any realizable distribution  $\mu$  there exist a point  $w_\mu \in \mathbb{S}^{d-1}$  for which the following containments hold:

$$\text{supp}^+(\mu) \subseteq B(\eta/2, w_\mu) \quad \text{and} \quad \text{supp}^-(\mu) \subseteq B(\eta/2, -w_\mu).$$

Next we construct a  $(0, \lceil \frac{d}{2} \rceil + 1)$ -list replicable algorithm with sample complexity 1.

1. Sample one point  $(x, y)$  from  $\mu$ . Let

$$u(x, y) := \begin{cases} x & \text{for } y = 1 \\ -x & \text{for } y = -1 \end{cases}$$

2. Find the smallest index  $i$  for which  $B(\eta, u) \subseteq U_i$  for some  $U_i \in \mathcal{U}$ . Such an  $i$  is guaranteed to exist by the Lebesgue Number Lemma (Theorem 12).
3. Output the signed indicator hypothesis

$$h_{U_i}(x) := \begin{cases} 1 & \text{for } x \in U_i \\ -1 & \text{for } x \in U_i^c \end{cases}$$

Firstly, we show accuracy. With probability 1 we have the chains of containments

$$\begin{aligned} \text{supp}^+(\mu) \subseteq B(\eta/2, w_\mu) \subseteq B(\eta, u) \subseteq U_i \\ \text{supp}^-(\mu) \subseteq B(\eta/2, -w_\mu) \subseteq B(\eta, -u) \subseteq U_i^c \end{aligned}$$

where the second containment in both chains follows from the fact that  $u \in B(\eta/2, w_\mu)$ , and the third containment in the second chain follows from the property that  $\mathcal{U}$  is antipodal free. We conclude  $h_{U_i}(x)$  achieves zero error with probability 1.

Finally, we check list replicability. With probability 1 we have  $B(\eta/2, w_\mu) \subseteq U_i$  and, in particular,  $w_\mu \in U_i$ . By Theorem 11, there could be at most  $\lceil \frac{d}{2} \rceil + 1$  sets  $U_j \in \mathcal{U}$  containing  $w_\mu$ . Hence, with probability 1, we output from a list of  $\lceil \frac{d}{2} \rceil + 1$  different hypotheses.

## References

- Noga Alon, Steve Hanneke, Ron Holzman, and Shay Moran. A theory of PAC learnability of partial concept classes. In *IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 658–671. IEEE, 2021.
- Noga Alon, Mark Bun, Roi Livni, Maryanthe Malliaris, and Shay Moran. Private and online learnability are equivalent. *J. ACM*, 69(4):Art. 28, 34, 2022.
- Ari Blondal, Hamed Hatami, Pooya Hatami, Chavdar Lalov, and Sivan Tretiak. Borsuk-ulam and replicable learning of large-margin halfspaces, 2025a. URL <https://arxiv.org/abs/2503.15294>.
- Ari Blondal, Hamed Hatami, Pooya Hatami, Chavdar Lalov, and Sivan Tretiak. Simplicial covering dimension of extremal concept classes. *arXiv preprint arXiv:2511.11819*, 2025b.
- Mark Bun, Roi Livni, and Shay Moran. An equivalence between private classification and online prediction. In *IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 389–402, 2020.
- Mark Bun, Marco Gaboardi, Max Hopkins, Russell Impagliazzo, Rex Lei, Toniann Pitassi, Satchit Sivakumar, and Jessica Sorrell. Stability is stable: Connections between replicability, privacy, and adaptive generalization. In *Proceedings of the 55th Annual ACM Symposium on Theory of Computing*, STOC 2023, pages 520–527, 2023.
- Zachary Chase, Shay Moran, and Amir Yehudayoff. Stability and Replicability in Learning. In *IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 2430–2439. IEEE Computer Society, 2023.
- Zachary Chase, Bogdan Chornomaz, Shay Moran, and Amir Yehudayoff. Local borsuk-ulam, stability, and replicability. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, STOC 2024, page 1769–1780, New York, NY, USA, 2024. Association for Computing Machinery. ISBN 9798400703836.
- Bogdan Chornomaz, Shay Moran, and Tom Waknine. Spherical dimension. *arXiv preprint arXiv:2503.10240*, 2025.
- Peter Dixon, Aduri Pavan, Jason Vander Woude, and N. V. Vinodchandran. List and certificate complexities in replicable learning. In *NeurIPS*, 2023.
- Eric Eaton, Marcel Hussing, Michael Kearns, and Jessica Sorrell. Replicable reinforcement learning. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, NeurIPS '23. Curran Associates Inc., 2023.
- Hossein Esfandiari, Alkis Kalavasis, Amin Karbasi, Andreas Krause, Vahab Mirrokni, and Grigoris Velegkas. Replicable bandits. In *The Eleventh International Conference on Learning Representations*, 2023a. URL <https://openreview.net/forum?id=gcD2UtCGMc2>.

- Hossein Esfandiari, Amin Karbasi, Vahab Mirrokni, Grigoris Velegkas, and Felix Zhou. Replicable clustering. In *Advances in Neural Information Processing Systems*, volume 36, pages 39277–39320. Curran Associates, Inc., 2023b. URL [https://proceedings.neurips.cc/paper\\_files/paper/2023/file/7bc3fe234454107149fa9d44faacaa64-Paper-Conference.pdf](https://proceedings.neurips.cc/paper_files/paper/2023/file/7bc3fe234454107149fa9d44faacaa64-Paper-Conference.pdf).
- Badih Ghazi, Noah Golowich, Ravi Kumar, and Pasin Manurangsi. Sample-efficient proper PAC learning with approximate differential privacy. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2021, page 183–196. Association for Computing Machinery, 2021a.
- Badih Ghazi, Ravi Kumar, and Pasin Manurangsi. User-Level Differentially Private Learning via Correlated Sampling. In *Advances in Neural Information Processing Systems*, volume 34, pages 20172–20184. Curran Associates, Inc., 2021b.
- Max Hopkins and Shay Moran. The role of randomness in stability. In *Forty-second International Conference on Machine Learning*, 2025.
- Russell Impagliazzo, Rex Lei, Toniann Pitassi, and Jessica Sorrell. Reproducibility in learning. In Stefano Leonardi and Anupam Gupta, editors, *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2022, page 818–831. Association for Computing Machinery, 2022. ISBN 9781450392648.
- Alkis Kalavasis, Amin Karbasi, Shay Moran, and Grigoris Velegkas. Statistical indistinguishability of learning algorithms. In *Proceedings of the 40th International Conference on Machine Learning*, ICML’23. JMLR.org, 2023.
- Alkis Kalavasis, Amin Karbasi, Kasper Green Larsen, Grigoris Velegkas, and Felix Zhou. Replicable learning of large-margin halfspaces. In *Proceedings of the 41st International Conference on Machine Learning*, ICML’24. JMLR.org, 2024.
- Amin Karbasi, Grigoris Velegkas, Lin Yang, and Felix Zhou. Replicability in reinforcement learning. *Advances in Neural Information Processing Systems*, 36:74702–74735, 2023.
- Maryanthe Malliaris and Shay Moran. The unstable formula theorem revisited via algorithms. *arXiv preprint arXiv:2212.05050*, 2022.
- Jiří Matoušek. *Using the Borsuk-Ulam Theorem: Lectures on Topological Methods in Combinatorics and Geometry*. Universitext. Springer, Berlin/Heidelberg, 2003. ISBN 978-3-540-00362-5.
- Warren S. McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.*, 5:115–133, 1943.
- Shay Moran, Hilla Scheffer, and Jonathan Shafer. The bayesian stability zoo. *Advances in Neural Information Processing Systems*, 36:61725–61746, 2023.
- Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological Review*, 65(6):386, 1958.
- Walter Rudin. *Real and complex analysis*. McGraw-Hill, Inc., 1987.

Shai Shalev-Shwartz and Shai Ben-David. *Understanding machine learning: From theory to algorithms*. Cambridge university press, 2014.

Jason Vander Woude, Peter Dixon, A. Pavan, Jamie Radcliffe, and N. V. Vinodchandran. Replicability in learning: Geometric partitions and kkm-sperner lemma. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang, editors, *Advances in Neural Information Processing Systems*, volume 37, pages 78996–79028. Curran Associates, Inc., 2024. doi: 10.52202/079017-2508.

## Appendix A. Proof of Theorem 4

**Theorem 4** *Let  $\mathcal{C}$  be a (partial) concept class with VC dimension  $d$ .*

$$\text{For any } \epsilon < \frac{1}{2}, \text{LR}(\mathcal{C}, \epsilon) \geq d.$$

**Proof** By definition of VC dimension, the concept class  $\mathcal{C}$  contains the class  $\mathcal{P} := \{\pm 1\}^{[d]}$  as a subclass. It is quick to check that, for every  $\epsilon$ ,

$$\text{LR}(\mathcal{C}, \epsilon) \geq \text{LR}(\mathcal{P}, \epsilon),$$

so we restrict ourselves to the analysis of  $\mathcal{P}$ .

Arguing by way of Lemma 10, first note that the set of realizable distributions  $\Delta_{\mathcal{P}}$  with the metric of total variation distance is homeomorphic to  $\mathbb{S}^{d-1}$  (Here  $\{e_i\}_i$  is the unit vector for the  $i$ th dimension in  $\mathbb{R}^d$ ):

$$p : \Delta_{\mathcal{P}} \rightarrow \mathbb{S}^{d-1}$$

$$\mu \mapsto \frac{\sum_i e_i [\mu(i, +) - \mu(i, -)]}{\sqrt{\sum_i [\mu(i, +) - \mu(i, -)]^2}}$$

Note that this map is a homeomorphism precisely because only one of  $\mu(i, +)$ ,  $\mu(i, -)$  can be non-zero at a time.

What is left is to check that, for any  $\epsilon < \frac{1}{2}$ , each set  $U_h$  is contained in a hemisphere

$$\left\{ x \in \mathbb{S}^{d-1} : \langle x, u_h \rangle > 0 \right\}$$

where  $u_h = (h(1), h(2), \dots, h(d))$

For a realizable distribution  $\mu$  and a hypothesis  $h$ , the loss of  $h$  on  $\mu$  can be calculated as

$$\text{loss}_{\mu}(h) = \frac{1}{2} \left( 1 + \sum_i h(i) [\mu(i, -) - \mu(i, +)] \right).$$

If  $p(\mu) \in U_h$ , then  $\text{loss}_{\mu}(h)$  is less than  $\frac{1}{2}$ , so

$$\sum_i h(i) [\mu(i, +) - \mu(i, -)] > 0.$$

Finally, we get that

$$\langle u_h, p(\mu) \rangle = \frac{\sum_i h(i) \cdot [\mu(i, +) - \mu(i, -)]}{\sqrt{\sum_i [\mu(i, +) - \mu(i, -)]^2}} > 0.$$

■

## Appendix B. Proof of Theorem 6

**Theorem 6** *For any dimension  $d > 1$  and margin  $\gamma \in (0, \frac{1}{\sqrt{2}})$ , there exists an accuracy parameter  $\epsilon := \epsilon(\gamma, d)$  such that*

$$\text{LR}(\mathcal{H}_\gamma^d, \epsilon) \geq d.$$

Hence,  $\text{LR}(\mathcal{H}_\gamma^d) = d$  for all  $\gamma \in (0, 1/\sqrt{2})$ .

Fix  $d$ . Pick any  $\gamma \in (0, 1/\sqrt{2})$ . The lower bound follows from trying to learn the collection of realizable distributions  $\Delta := \{\mu_w : w \in \mathbb{S}^{d-1}\}$ , where  $\mu_w$  is the uniform distribution on

$$\{(x, c_w(x)) : x \in \text{supp}(c_w)\}.$$

The map  $\varphi: \mathbb{S}^{d-1} \rightarrow \Delta$  defined by  $w \mapsto \mu_w$  is a homeomorphism.

Denote by  $\nu$  the spherical measure, that is, the uniformly distributed measure on  $\mathbb{S}^{d-1}$ , normalized so that  $\nu(\mathbb{S}^{d-1}) = 1$ . For each  $x \in \mathbb{S}^{d-1}$ , let  $(x_1, \dots, x_d)$  be its coordinates in  $\mathbb{R}^d$ . Define

$$\epsilon := \nu\left(\left\{x \in \mathbb{S}^{d-1} : x_1 \geq \gamma, x_2 \leq -\gamma\right\}\right) \text{ and } \epsilon_0 := \frac{\nu(\{x \in \mathbb{S}^{d-1} : x_1 \geq \gamma, x_2 \leq -\gamma\})}{\nu(\{x \in \mathbb{S}^{d-1} : |x_1| \geq \gamma\})}.$$

It is not hard to see that both  $\epsilon_0$  and  $\epsilon$  are strictly positive and  $\epsilon_0 > \epsilon$  when  $\gamma \in (0, 1/\sqrt{2})$ .

Now suppose  $\mathcal{A}$  is an  $(\epsilon, L)$ -list replicable learner for  $\mathcal{H}_\gamma^d$ . Let  $T \subseteq \{\pm 1\}^{\mathbb{S}^{d-1}}$  be the collection of all possible hypotheses output by  $\mathcal{A}$ . For each hypothesis  $h \in T$ , define the set  $U_h$

$$U_h = \left\{w \in \mathbb{S}^{d-1} : \mathbb{P}_{S \sim \mu_w^n}[\mathcal{A}(S) = h] > \frac{1 - 2\delta}{L}, \text{loss}_{\mu_w}(h) < \epsilon_0\right\}$$

of distributions on which  $\mathcal{A}$  is likely to output  $h$ .

**Claim 13** *For each  $h \in T$ , if  $u, w \in U_h$ , then  $\langle u, w \rangle > 0$ . In particular,  $U_h$  is contained in an open hemisphere.*

**Proof of Claim:** Let  $u, w \in U_h$ . Define the set on which  $c_u$  and  $c_w$  disagree:

$$\mathcal{E}_{u,w} := \{x \in \text{supp } c_u \cap \text{supp } c_w : c_u(x) \neq c_w(x)\}.$$

We have

$$\begin{aligned} \nu(\mathcal{E}_{u,w}) &= \nu(\text{supp}(c_w)) \frac{\nu(\mathcal{E}_{u,w})}{\nu(\text{supp}(c_w))} \\ &\leq \nu(\text{supp}(c_w))(\text{loss}_{\mu_u}(h) + \text{loss}_{\mu_w}(h)) \\ &< \nu(\text{supp}(c_w))2\epsilon_0 \\ &= 2\epsilon. \end{aligned} \tag{1}$$

The first inequality is implied by the fact that  $h$  is not consistent with either  $\mu_u$  or  $\mu_w$  on each point in  $\mathcal{E}_{u,w}$ . The second inequality follows directly from our definition of  $U_h$ . The last equality is a consequence of  $\nu(\text{supp}(c_w)) = \nu(\{x \in \mathbb{S}^{d-1} : |x_1| \geq \gamma\})$  for any  $w$  and our definitions of  $\epsilon$  and  $\epsilon_0$ .

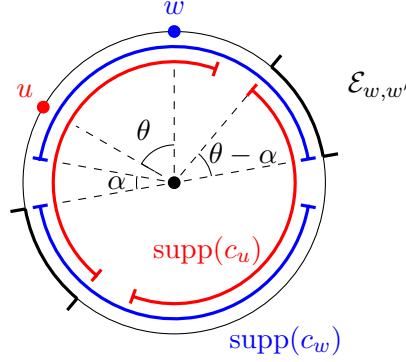


Figure 2: An illustration of the disagreement set  $\mathcal{E}_{u,w}$  between  $c_u$  and  $c_w$ , projected onto the plane common to  $u, w$ , and  $0$ .

Now let  $\theta := \arccos(\langle u, w \rangle)$  be the angle between  $u$  and  $w$ . We have

$$\begin{aligned} \nu(\mathcal{E}_{u,w}) &= \nu(\{x \in \text{supp } c_u \cap \text{supp } c_w : c_u(x) \neq c_w(x)\}) \\ &= 2\nu\left(\left\{x \in \mathbb{S}^{d-1} : \langle u, x \rangle \geq \gamma, \langle w, x \rangle \leq -\gamma\right\}\right) \\ &= 2\nu\left(\left\{x \in \mathbb{S}^{d-1} : x_1 \geq \gamma, \cos(\theta)x_1 + \sin(\theta)x_2 \leq -\gamma\right\}\right) \end{aligned} \quad (2)$$

where the last equality follows from the rotational invariance of the spherical measure.

For the sake of contradiction, assume that  $\theta \in [\pi/2, \pi]$ . Next we show

$$\begin{aligned} \nu\left(\left\{x \in \mathbb{S}^{d-1} : x_1 \geq \gamma, x_2 \leq -\gamma\right\}\right) &\subseteq \\ &\subseteq \nu\left(\left\{x \in \mathbb{S}^{d-1} : x_1 \geq \gamma, \cos(\theta)x_1 + \sin(\theta)x_2 \leq -\gamma\right\}\right). \end{aligned} \quad (3)$$

Indeed, it is enough to check  $\cos(\theta)x_1 + \sin(\theta)x_2 \leq -\gamma$  assuming  $x_1 \geq \gamma$  and  $x_2 \leq -\gamma$ . Since  $\theta \in [\pi/2, \pi]$ , cosine is non-positive and sine is non-negative. Thus, we only need to verify  $\sin(\theta) - \cos(\theta) \geq 1$ . Since, the left hand-side is non-negative, we square both sides to get

$$\sin^2(\theta) - 2\sin(\theta)\cos(\theta) + \cos^2(\theta) \geq 1$$

which holds as  $\theta \in [\pi/2, \pi]$ . We conclude that the inclusion in (3) holds.

Now by (2) and (3), we have

$$\nu(\mathcal{E}_{u,w}) \geq 2\nu\left(\left\{x \in \mathbb{S}^{d-1} : x_1 \geq \gamma, x_2 \leq -\gamma\right\}\right) = 2\epsilon. \quad (4)$$

Finally, by (1) and (4), we have  $2\epsilon \leq \nu(\mathcal{E}_{u,w}) < 2\epsilon$  which is a contradiction. Thus, if  $u, w \in U_h$ , then  $\theta < \frac{\pi}{2}$ , and so  $\langle u, w \rangle > 0$ .  $\blacksquare$

Having proved Claim 13, we have that Lemma 10 implies Theorem 6.

**Remark 14** Examining the relationship between  $\epsilon$  and  $\gamma$  in the above proof, we can see that, for any dimension  $d > 1$ , there exists a continuous decreasing function  $\epsilon: (0, 1/\sqrt{2}) \rightarrow (0, 1/4)$  such that:  $\epsilon(\gamma) \rightarrow 1/4$  as  $\gamma \rightarrow 0^+$ ,  $\epsilon(\gamma) \rightarrow 0$  as  $\gamma \rightarrow (1/\sqrt{2})^-$ , and  $\text{LR}(\mathcal{H}_\gamma^d, \epsilon(\gamma)) \geq d$  for any  $\gamma \in (0, 1/\sqrt{2})$ . This justifies the depiction of the boundary between the lower green and middle yellow regions in [Figure 1](#).

### Appendix C. Proof of [Theorem 8](#)

**Theorem 8** Fix a dimension  $d > 1$  and a margin  $\gamma \in (0, 1)$ . For any error parameter  $\epsilon \in (0, 1/2)$  and list length  $L > 1$ , if  $\mathcal{A}$  is an  $(\epsilon, L)$ -list replicable learner for  $\mathcal{H}_\gamma^d$  which outputs linear classifiers, then  $L \geq d$ .

**Proof** Consider the set of realizable distributions  $\Delta := \{\mu_w : w \in \mathbb{S}^{d-1}\}$ , where  $\mu_w$  is the uniform distribution on

$$\{(x, c_w(x)) : x \in \text{supp}(c_w)\}.$$

The map  $\varphi: \mathbb{S}^{d-1} \rightarrow \Delta$  defined by  $w \mapsto \mu_w$  is a homeomorphism.

By assumption  $\mathcal{A}$  outputs linear classifiers of the form

$$h_w(x) := \begin{cases} 1 & \text{if } \langle w, x \rangle \geq 0, \\ -1 & \text{if } \langle w, x \rangle < 0 \end{cases}$$

for some  $w \in \mathbb{S}^{d-1}$ . Fix  $\epsilon < \epsilon_0 < 1/2$ . For any such output  $h_w$ , we will show that the set

$$U_{h_w} := \left\{ y \in \mathbb{S}^{d-1} : \mathbb{P}_{S \sim \varphi(y)^n}[\mathcal{A}(S) = h_w] > \frac{1 - 2\delta}{d - 1} \text{ and } \text{loss}_{\varphi(y)}(h_w) < \epsilon_0 \right\}$$

is contained in the open hemisphere

$$H_w := \left\{ y \in \mathbb{S}^{d-1} : \langle w, y \rangle > 0 \right\}.$$

Indeed, if  $\langle w, y \rangle \leq 0$ , then  $h_w$  correctly labels no more than half of the support of  $\varphi(y)$ . Since  $\varphi(y)$  is a uniform distribution, it follows that  $\text{loss}_{\varphi(y)}(h_w) \geq 1/2 > \epsilon_0$ . We deduce that  $U_{h_w} \subseteq H_w$ , so applying [Lemma 10](#) completes the proof.  $\blacksquare$