Combinatorial Variability of Vapnik-Chervonenkis Classes

with Applications to Sample Compression Schemes

Shai Ben-David Ami Litman
Computer Science Department, Technion
Haifa 32000, ISRAEL
email: \{shai,litman\}@cs.technion.ac.il

Abstract

We define embeddings between concept classes that are meant to reflect certain aspects of their combinatorial structure. Furthermore, we introduce a notion of universal concept classes – classes into which any member of a given family of classes can be embedded. These universal classes play a role similar to that played in computational complexity by languages that are hard for a given complexity class. We show that classes of half-spaces in $\mathbb{R}^n$ are universal with respect to families of algebraically defined classes.

We present some combinatorial parameters along which the family of classes of a given VC-dimension can be grouped into sub-families. We use these parameters to investigate the existence of embeddings and the scope of universality of classes. We view the formulation of these parameters and the related questions that they raise as a significant component in this work.

A second theme in our work is the notion of Sample Compression Schemes. Intuitively, a class $C$ has a sample compression scheme if for any finite sample, labeled according to a member of $C$, there exists a short sub-sample so that the labels of the full sample can be reconstructed from this sub-sample.

By demonstrating the existence of certain compression schemes for the classes of half-spaces the existence of similar compression schemes for every class embeddable in half-spaces readily follows. We apply this approach to prove existence of compression schemes for all 'geometric concept classes'.

Introduction

This paper introduces a partial order on set systems that comes to reflect certain aspects of their combinatorial richness. Much like the partial order of polynomial reducibility among
computable languages, our combinatorial complexity partial order is based upon a notion of embeddings among set systems. Furthermore, we introduce a notion of universal set systems that plays a role similar to that played in computational complexity by languages that are hard for a given complexity class.

Several parameters defined for a set system may be viewed as reflecting some aspects of the combinatorial richness of the system. Most notable among these parameters is the Vapnik-Chervonenkis dimension. Some other examples originating from computational learning theory are:

- The sample complexity of learning a concept class (in learning theory the term ‘concept class’ is used for a set system, we use both terms interchangeably).

- The size of the minimal compression scheme for a class. (We shall discuss several types of compression schemes, in the spirit of the schemes discussed by Floyd and Warmuth [FW95], each of these schemes gives rise to its own parameter).

- The optimal mistake bound for learning the class online (sometimes called ‘the K-dimension’).

Each of these parameters induces a partial order over the family of concept classes. The partial orders we introduce here refines each of the partial orders induced by these parameters. In other words, if one class is embeddable in another (with respect to the embeddings discussed in this paper) then the value of each of the above parameters for the embedded class is at most the value this parameter has for the class it is embedded in.

Consequently, by demonstrating the existence of a certain compression scheme for a concept class which is universal for a given family of concept classes, the existence of similar compression schemes for all members of this family readily follows. We apply this approach to prove existence of compression schemes for all ‘geometric concept classes’ (a notion that will be defined precisely later).

Our investigation of the embedding relation among concept classes raises some basic questions concerning the structure of classes (or set systems) having a finite VC-dimension. We present some combinatorial parameters along which the family of classes of a given VC-dimension can be grouped into sub-families. We believe that part of the contribution of this paper lies in the formulation of these parameters and in calling attention to the related questions, that seem to be both new and intriguing.

Embeddings of concept classes have been previously introduced in the context of computational learning theory. Pitt and Warmuth [PW90] introduce a notion of ‘prediction-preserving reductions’ for concept classes. Their aim is to define computationally efficient reductions among prediction problems. Consequently, their notion of reduction includes computational constraints. We shift the focus of the discussion from computational complexity issues to the combinatorial structure of classes. Our basic notion of embedding is similar to the Pitt-Warmuth reductions striped off of their computational constraints. We introduce
our notions of embeddings in Section 1 and in Section 2 we prove the universality of certain geometric classes and apply it to get some insight into the structure of algebraically defined classes.

Sample compression schemes were first introduced by Littlestone and Warmuth [LW86] and studied extensively by Floyd [Fl89] and Floyd and Warmuth [FW95]. These papers establish close relationships between compression schemes and learning and provide sample compression schemes for several families of concept classes. For geometric classes, [FW95] prove the existence of finite-size compression schemes for some sub-classes of the class of half-spaces in $\mathbb{R}^n$ and [Fl89] presents compression schemes for the classes of rectangles and triangles in $\mathbb{R}^2$. In Section 3 we define several variants of sample compression schemes and discuss basic properties of such schemes. In Section 4, we introduce strong compression schemes for the (full) classes of half-spaces in $\mathbb{R}^n$, and finite-size compression schemes for any class of geometric objects in $\mathbb{R}^n$.

1 Embeddings of Concept Classes

1.1 Some Preliminaries on Classes and VC-dimension

Commonly, a concept class is defined as a collection of subsets of some domain set. For our purposes it will be instructive to define concept classes in a way that emphasizes the symmetry between ‘concepts’ and domain points.

Definition 1:

- A concept class is a triple $C = (X, Y, R)$ where $X, Y$ are arbitrary sets and $R$ is a binary relation, $R \subseteq X \times Y$, we call $X$ the domain of $C$.
- A concept in such a class $R$ is a subset of $X$ of the form $c_y = \{x \in X : (x, y) \in R\}$, for some $y \in Y$.
- The VC-dimension of a class $C = (X, Y, R)$ is the supremum over all sizes of finite subsets of $X$ shattered by $\{c_y : y \in Y\}$.
- The dual class of a concept class $C = (X, Y, R)$ is the class $C^D = (Y, X, R^D)$, where $R^D \subseteq Y \times X$ is defined by $(y, x) \in R^D \iff (x, y) \in R$. Note that a concept class $C$ may be represented as a matrix over $\{0, 1\}$ whose $(x, y)$’s entry is 1 iff $(x, y) \in R$. If $A_C$ is a the matrix that represents a class $C$ then its transposed matrix, $A_C^T$, represents the dual class, $C^D$.

Definition 2: For a pair of classes, $C = (X, Y, R)$ and $C' = (X', Y', R')$,

- We say that $C'$ is a subclass of $C$ if $X' \subseteq X$, $Y' \subseteq Y$ and $R'$ is the restriction of $R$ to $X' \times Y'$. 

3
• Assuming $X = X'$ and $Y \cap Y' = \emptyset$, we define the union of the classes, $C \cup C' = (X, Y \cup Y', R \cup R')$.

• A class $C$ is finite if both $X$ and $Y$ are finite sets.

• For any $A \subseteq X$, denote by $C_{|A}$ the subclass of $C$ obtained by restricting $C$ to the domain set $A$. I.e., $C_{|A} = (A, Y, R \cap (A \times Y))$.

• Let $F(C)$ denote the representation of $C$ as a class of boolean functions on $X$.

$$F(C) = \{f_y : y \in Y\}$$

where, for any $y \in Y$,

$$f_y(x) = \begin{cases} 1 & \text{if } (x, y) \in R \\ 0 & \text{otherwise} \end{cases}$$

The basic combinatorial result concerning classes of finite VC-dimension is Sauer’s Lemma ([Sa72]):

**Lemma 1**: [Sauer] If a class $C = (X, Y, R)$ has a finite VC-dimension, $d$, then for every finite $A \subseteq X$,

$$|F(C_{|A})| \leq \Sigma_{i=0}^d \binom{|A|}{i}$$

We shall also need the following consequence of Sauer’s lemma,

**Claim 1**: For every pair of classes, $C_0 = (X, Y_0, R_0)$, $C_1 = (X, Y_1, R_1)$ (over the same domain set $X$), $\text{VC-dim}(C_0 \cup C_1) \leq 2\max \{\text{VC-dim}(C_0), \text{VC-dim}(C_1)\} + 1$.

**Proof**: Let $k = \max(\text{VC-dim}(C_0), \text{VC-dim}(C_1))$, and assume, for contradiction, that $\text{VC-dim}(C_0 \cup C_1) > 2k + 1$. Let $A$ be a subset of $X$ of cardinality $2k + 2$ shattered by $C_0 \cup C_1$. This implies: $|F((C_0 \cup C_1)_{|A})| = 2^{2k+2} = \Sigma_{i=0}^{2k+2} \binom{2k+2}{i} = \binom{2k+2}{k+1} + 2 \Sigma_{i=0}^{k} \binom{2k+2}{i}$. Hence, for some $i \in \{0, 1\}$, $|F((C_i)_{|A})| > \Sigma_{i=0}^{k} \binom{2k+2}{i}$. But, by Sauer’s lemma, $|F((C_i)_{|A})| \leq \Sigma_{i=0}^{k} \binom{2k+2}{i}$; a contradiction.

### 1.2 Embeddings of Classes

We turn now to the central tool of this paper – the notions of embeddings among concept classes. We define four variants of embeddings. Our basic definition (1 in the definition below) is similar to the definition of ‘prediction-preserving reductions’ of Pitt and Warmuth [PW90] stripped of its computational constraints. As this paper considers combinatorial structure rather than computational complexity, these constraints are irrelevant to our discussion.

**Definition 3**: Given concept classes, $C = (X, Y, R)$ and $C' = (X', Y', R')$, 

1. An embedding of $C$ into $C'$ is a pair of functions, $(\pi : X \mapsto X', \tau : Y \mapsto Y')$, such that for every $x \in X, y \in Y$, $(x, y) \in R$ iff $(\pi(x), \tau(y)) \in R'$.

2. A generalized embedding of $C$ in $C'$ is a pair of functions, $(\pi : X \mapsto X', \tau : Y \mapsto Y')$, and a function $\sigma : X \mapsto \{0, 1\}$ satisfying, for every $x \in X, y \in Y$,

   \[
   \text{if } \sigma(x) = 0 \text{ then } (x, y) \in R \text{ iff } (\pi(x), \tau(y)) \in R',
   \]

   and

   \[
   \text{if } \sigma(x) = 1 \text{ then } (x, y) \in R \text{ iff } (\pi(x), \tau(y)) \notin R',
   \]

3. A class $C = (X, Y, R)$ is weakly embeddable in a class $C' = (X', Y', R')$ if every finite subclass of $C$ is embeddable in $C'$. This notion applies to both of the above types of embeddings.

   Note that all (four) notions of embeddability that arise from the above definition are reflexive and transitive relations (over a collection of concept classes). Consequently, they induce pre-orderings over concept classes, namely,

**Notation 1:**

1. $C \preceq_{\text{emb}} C'$ iff there exists an embedding of $C$ into $C'$.

2. $C \preceq_{\text{gemb}} C'$ iff there exists a generalized embedding of $C$ into $C'$.

3. $C \preceq_{\text{wemb}} C'$ C is weakly embeddable in $C'$.

4. $C \preceq_{\text{wgemb}} C'$ iff $C$ is weakly embeddable in $C'$ using generalized embeddings.

Note that $C \preceq_{\text{emb}} C'$ implies $C \preceq_{\text{wemb}} C'$, $C \preceq_{\text{gemb}} C'$ implies $C \preceq_{\text{wgemb}} C'$, and also $C \preceq_{\text{em}} C'$ implies $C \preceq_{\text{gem}} C'$ and $C \preceq_{\text{em}} C'$ implies $C \preceq_{\text{gem}} C'$.

**Observation 1:** Let $F_d$ denote the class of all binary vectors over $d$ entries. That is, $F_d = (X, Y, R)$ where $X = \{1, \ldots, d\}$, $Y = \{0, 1\}^X$ and $(x, v) \in R$ iff $v(x) = 1$. The VC-dimension of $F_d$ equals $d$, and for every class $C$ of the same VC-dimension, $F_d \preceq_{\text{em}} C$.

Many combinatorial parameters, reflecting aspects of the combinatorial richness of classes, are invariant under the above notions of embedding. In particular,

**Claim 2:**

1. $C \preceq_{\text{wgemb}} C'$ implies $\text{VC-dim}(C) \leq \text{VC-dim}(C')$.

2. $C \preceq_{\text{wgemb}} C'$ implies that, for any type of compression scheme discussed below, if $C'$ has a compression scheme to some size, $d$, then so does $C$. 


3. $C \leq^{g_{emb}} C'$ implies that, for every $\epsilon, \delta > 0$ and $m \in \mathbb{N}$, if $C'$ is $(\epsilon, \delta)$-PAC-learnable from $m$ examples then so is $C$.

4. $C \leq^{*}_{g_{emb}} C'$ implies that, for every $m \in \mathbb{N}$, if $C'$ is online learnable with at most $m$ many mistakes, then so is $C$.

The only statement in the claim that does not follow directly from the relevant definitions is part 2. The proof of part 2 involves a compactness consideration for compression schemes that will be discussed in Section (4).

Later on in the paper, we shall discuss some more parameters that the partial orders they induce on classes are refined by $\leq^{*}_{g_{emb}}$. Consequently, one may view these ordering relations as reflecting the ‘combinatorial richness’ of classes.

2 Universal Classes for Families of Geometric Objects

The next natural concept to be defined is that of a class being an upper bound for a family of classes. This notion is analogous to that of a language being hard for a given complexity class in computational complexity theory.

**Definition 4:** Let $C$ be a class, $\mathcal{S}$ a family of classes and any partial order on classes, $\preceq$. We say that $C$ is $\preceq$-universal for $\mathcal{S}$ if, for every class $W \in \mathcal{S}$, $W \preceq C$.

In this section we shall present some families of classes and demonstrate the existence of universal classes for them with respect to the partial orders defined above.

2.1 Some Concrete Classes of Geometric Objects

We start our discussion of concrete classes by citing a definition that captures a wide family of ‘natural’ classes. This definition is due to Dudley [Du84] so we call members of this family Dudley classes.

**Definition 5:** Let $\mathcal{F}$ be a collection of real-valued functions over some domain $X$. For a collection $\mathcal{F}$ which is a vector space over the reals (with respect to point-wise addition and scalar multiplication) and any $h : X \mapsto \mathbb{R}$,

A Dudley class is a class of the form $D_{\mathcal{F}, h} = (X, \mathcal{F}, R_{\mathcal{F}, h})$, where

$$R_{\mathcal{F}, h} = \{(x, f) : x \in X, f \in \mathcal{F}, \text{ and } f(x) + h(x) \geq 0\}$$

In other words, letting $\text{pos}(g) \overset{def}{=} \{x \in X : g(x) \geq 0\}$, the concepts of $R_{\mathcal{F}, h}$ are exactly the sets $\{\text{pos}(f + h) : f \in \mathcal{F}\}$.
Dudley [Du84] proves that the VC-dimension of such a class, $\text{pos}(\mathcal{F} + h)$, equals the linear dimension of the vector space $\mathcal{F}$.

The collection of half-spaces in $\mathbb{R}^n$, the collection of all subsets of $\mathbb{R}$ that are unions of at most $k$ many intervals (for any fixed $k$), the collection of all subsets of $\mathbb{R}^{n+1}$ bounded by fixed degree polynomials in $\{x_1, \ldots, x_n\}$, are all examples of Dudley classes. The class of all $n$-dimensional balls, as well as many other natural classes of geometric objects, are subclasses of Dudley classes of the same VC-dimension. All the results obtained in this paper for Dudley classes apply to such classes as well.

We now introduce some concrete classes of geometric objects, these classes will be later shown to be universal w.r.t. families of Dudley classes.

**Notation 2:**

- Let $H S^k$ denote the class of all half-spaces in $\mathbb{R}^k$. I.e.,
  
  $$H S^k = (\mathbb{R}^k, \mathbb{R}^{k+1}, H), \text{ where } H = \{(x_1, \ldots, x_k), (a_1, \ldots, a_{k+1}) : \sum_{i=1}^k a_i x_i + a_{k+1} \geq 0\}$$

- Let $P H S^k$ denote the class of positive half-spaces in $\mathbb{R}^k$. That is, we restrict the class $H S^k$ by replacing $\mathbb{R}^{k+1}$ by $\{(a_1, \ldots, a_{k+1}) : a_1 > 0\}$.

- Let $P H S^k_0$ denote the class of positive half-spaces in $\mathbb{R}^k$ that pass through the origin (the zero vector of $\mathbb{R}^k$). That is, we restrict the set of concepts further by allowing only vectors in which the last component, $a_{k+1}$, is zero.

It should be noted that these classes are all Dudley classes. The following is well known, (see [Fl89]).

**Claim 3:** For every $k \in \mathbb{N}$,

1. $\text{VC-dim}(H S^k) = k + 1$.
2. $\text{VC-dim}(P H S^k) = k$.
3. $\text{VC-dim}((H S^k)^D) = \text{VC-dim}((P H S^k)^D) = k$.
4. $\text{VC-dim}(P H S^k_0) = k - 1$.

### 2.2 Universality Theorems

**Theorem 1:** Let $k$ be any natural number and let $\mathcal{D}_k$ denote the family of Dudley classes of dimension $k$.

1. $P H S^k_0$ is $\leq_{\text{emb}}$-universal for $\mathcal{D}_k$.
2. $(H S^k)^D$ is $\leq_{\text{emb}}$-universal for $\mathcal{D}_k$.
3. $PHS^k$ is $\simeq_{g_{emb}}^{*}$-universal for $D_k$.

4. $(PHS^k)^D$ is $\simeq_{g_{emb}}^{*}$-universal for $D_k$.

Proof:

1. Let $D_{\mathcal{F}, k} = (X, \mathcal{F}, R)$ be a $k$-dimensional Dudley class and let $(f_1, \ldots, f_k)$ be a linear basis for the space $\mathcal{F}$. Now, for $f = \Sigma_{i=1}^{k} \alpha_i f_i$, let

$$\tau(f) = (\alpha_1, \ldots, \alpha_k, 1, 0)$$

and, for any $x \in X$ let

$$\pi(x) = (f_1(x), \ldots, f_k(x), h(x))$$

2. (Using the notation of the previous item) For $f \in \mathcal{F}$ let

$$\tau(f) = (\alpha_1, \ldots, \alpha_k)$$

and, for $x \in X$, let

$$\pi(x) = (f_1(x), \ldots, f_k(x), h(x))$$

3. Applying part 1 of this theorem, it suffices to prove that $PHS_k^{k+1} \simeq_{g_{emb}}^{*} PHS^k$. The proof of this statement is rather technical and is deferred to the Appendix for the sake of readability of this section.

4. Applying part 2 of this theorem, it suffices to prove that $(HS_k^k)^D \simeq_{g_{emb}}^{*} (PHS^k)^D$. Let $(A, B, H)$ be a finite subclass of $(HS_k^k)^D$, due to the finiteness of the class, we may replace each half-space $a \in A$ by a half-space $a'$ that induces the same partition over the points of $B$ and so that no point of $B$ is on its boundary hyper-plane. Now let $\tau$ be the identity function on $B$, for each $a \in A$, let $\pi(a)$ be the positive half-space defined by the boundary of $a'$, and let $\sigma(a)$ be 0 if $a$ is a positive half-space and 1 otherwise.

A natural question that arises at this stage is if the universality of $PHS^k$ and $(PHS^k)^D$ holds with respect to wider classes. By Claim 2, we know that no class whose VC-dimension exceeds $k$ can be embedded in a class having VC-dimension $k$, but is it the case that $PHS^k$ and $(PHS^k)^D$ are universal for the family of all classes of VC-dimension $k$? Alon Frankl and Rodel [AFR85] use a counting argument to show that there exist, for very $n$, concept classes of size $n \times n$ (i.e., both the domain set $X$ and the set of concepts $Y$ have cardinality $n$) that are not $\simeq_{emb}$-embeddable in $PHS^k$ unless $k \geq n/32$. As the VC-dimension of such a class cannot exceed $\log(n)$, it follows that some classes cannot be embedded in $PHS^k$ unless $k$ is exponentially larger than their VC-dimension. The combinatorial parameters discussed in the following subsection will provide several ways to show non-embeddability in $HS^k$ for some concrete and rather simple classes.
2.3 Implications to VC-dimension related parameters

The above results imply that, for many combinatorial parameters, the values of these parameters for the class of positive half-spaces (and its dual class) are upper bounds on the value of these parameters for any Dudley class of the same dimension. In Section 4 below we shall discuss sample compression schemes for such classes. Let us mention here some other parameters for which these considerations yield new results.

The following notion is discussed in [Fl89] and is attributed there to an unpublished manuscript of E. Welzl.

**Definition 6:** A concept class \( C = (X, Y, R) \) having some finite VC-dimension, \( k \), is a maximum class if for every finite \( A \subseteq X \) the cardinality of \( F(C_A) \) meets the upper bound of Sauer’s lemma. (That is, for \( |A| \leq k \), \( |F(C_A)| \leq 2^{[A]} \), and for \( |A| > k \), \( |F(C_A)| = \Sigma_{i=0}^{k} \binom{[A]}{i} \).

Floyd and Warmuth [Fl89],[FW95] have shown that for any set \( X \subseteq \mathbb{R}^k \) whose points are in general position\(^1\), \( (PHS^k)_X \) is a maximum class. Just the same, there are Dudley classes which do not have any nontrivial maximum subclass, namely:

**Claim 4:** Every subclass of \( HS^k \) with VC-dimension \( k+1 \), whose domain set has cardinality greater than \( k+1 \), is not a maximum class.

**Proof:** Clearly, over any finite subset of \( \mathbb{R}^k \), there are at most twice as many half-spaces than positive half-spaces. Let \( A \) be any finite subset of the domain of \( C \) of cardinality \( k+1 \). As \( \text{VC-dim}(PHS^k) = k \), Sauer’s Lemma implies that \( |F(PHS^k_X)| \leq \Sigma_{i=0}^{k} \binom{[A]}{i} \). It follows that \( |F(C_A)| \leq 2\Sigma_{i=0}^{k} \binom{[A]}{i} < \Sigma_{i=0}^{k+1} \binom{[A]}{i} \).

On the other hand, we have:

**Claim 5:** Every Dudley class of dimension \( k \) is \( \preceq_{g_{emb}} \)-embeddable in a maximum class of VC-dimension \( k \).

**Proof:** Let \( X \) be a dense subset of \( \mathbb{R}^k \) in general position. By [Fl89], \( (PHS^k)_X \) is a maximum class. Let \( D \) be a Dudley class of dimension \( k \). By Theorem 1, \( D \) is \( \preceq_{g_{emb}} \)-embeddable in \( PHS^k \). Let \( D' \) be a finite subclass of \( D \). Clearly, \( D' \) is embeddable in a finite subclass of \( PHS^k \). Since \( X \) is dense, \( D' \) is embeddable in a finite subclass of \( (PHS^k)_X \) as well.

The above claim does not hold for general (not Dudley) classes. Floyd and Warmuth [FW95] have presented examples of finite classes that cannot be extended to maximum classes of the same VC-dimension. In our terminology it means that they are not \( \preceq_{emb} \)-embeddable\(^2\) in any maximum class of the same VC-dimension.

Towards gaining a better understanding of the embeddability partial orders, we introduce two more parameters of a class \( C \). These parameters are the VC-dimensions of images of \( C \)

\(^1\)General position, for \( |X| > k \), means that any hyper-plane contains at most \( k \) points of \( X \) and any hyper-plane parallel to the \( x_1 \) axis contains at most \( k - 1 \) point of \( X \).

\(^2\)Note that, for finite classes, the relations \( \preceq_{emb} \) and \( \preceq_{emb} \) coincide.
under two simple mappings of classes – mapping a class $C$ to its dual class $C^D$ and mapping a class to its completion $\overline{C}$, defined as follows.

**Notation 3:** Given a class $C = (X, Y, R)$, let $\overline{R}$ be the relation over $X \times (Y \times \{0, 1\})$ defined by

$$(x, (y, 0)) \in \overline{R} \text{ iff } (x, y) \in R, \text{ and } (x, (y, 1)) \in \overline{R} \text{ iff } (x, y) \notin R.$$ 

We define the completion of a class $C$ as the class $\overline{C} = (X, (Y \times \{0, 1\}), \overline{R})$.

**Claim 6:** For any class $C$: $\log(\text{VC-dim}(C)) - 1 < \text{VC-dim}(C^D) < 2^{\text{VC-dim}(C) + 1}$.

**Proof:** As $(C^D)^D = C$ it suffices to prove only the first inequality. Let $X$ be an arbitrary set of cardinality $k = \lceil \log(\text{VC-dim}(C)) \rceil$; let $Y \triangleq \{X' : X' \subseteq X\}$, $Z \triangleq \{Y' : Y' \subseteq Y\}$, $A \triangleq (X, Y, \varepsilon)$ and $B \triangleq (Y, Z, \varepsilon)$.

The class $A$ satisfies $A \preceq_{\text{emb}} B^D$ via the mapping $\pi(x) \triangleq \{X' \subseteq X : x \in X'\}$ and $\pi(y) = y$. The class $B$ is $\preceq_{\text{emb}}$ than any class of the same or greater VC-dim; hence, $B \preceq_{\text{emb}} C$, which implies $B^D \preceq_{\text{emb}} C^D$. Since $\preceq_{\text{emb}}$ is transitive, $A \preceq_{\text{emb}} C^D$; hence $k = \text{VC-dim}(A) \leq \text{VC-dim}(C^D)$.

Note that the inequalities of Claim 6 are tight, as demonstrated by the above class $B$.

**Claim 7:** For any class $C$: $\text{VC-dim}(\overline{C}) \leq 2(\text{VC-dim}(C)) + 1$.

**Proof:**

Note that $\overline{C}$ is the union of $C_0 = (X, Y \times 0, \overline{R})$ and $C_1 = (X, Y \times 1, \overline{R})$. Since $\text{VC-dim}(C) = \text{VC-dim}(C_0) = \text{VC-dim}(C_1)$, by Claim 1: $\text{VC-dim}(\overline{C}) \leq 2(\text{VC-dim}(C)) + 1$.

The following claim extends Claim 2 to the new parameters. Namely, it relates the embeddings partial orders to those induced by the VC-dimension of the dual of a class and of its completion.

**Claim 8:**

1. $C_1 \preceq_{\text{gemb}}^* C_2$ implies $\overline{C_1} \preceq_{\text{gemb}}^* \overline{C_2}$, and therefore, $\text{VC-dim}(\overline{C_1}) \leq \text{VC-dim}(\overline{C_2})$.

2. $C_1 \preceq_{\text{emb}}^* C_2$ implies $\overline{C_1} \preceq_{\text{emb}}^* \overline{C_2}$, and therefore, $\text{VC-dim}(\overline{C_1}) \leq \text{VC-dim}(\overline{C_2})$.

3. $C_1 \preceq_{\text{gemb}}^* C_2$ implies $\overline{C_1} \preceq_{\text{emb}}^* \overline{C_2}$, and therefore, $\text{VC-dim}(\overline{C_1}) \leq \text{VC-dim}(\overline{C_2})$.

**Corollary 2:** For every finite $k$,

1. If $C$ is a $k$-dimensional Dudley class then $\text{VC-dim}(\overline{C}) \leq k + 1$.

2. If $C$ is a $k$-dimensional Dudley class then $\text{VC-dim}(C^D) \leq k + 1$. 

10
Proof:

1. Applying part 3 of Theorem 1, we know that $C \preceq_{perm}^* PHS^k$. The claim now follows by noting that $\overline{PHS^k} \preceq_{perm}^* HS^k$ and applying part 1 of Claim 8.

2. Applying part 2 of Theorem 1, we have $C \preceq_{perm}^* (HS^k)^D$, by part 2 of Claim 8 we conclude that VC-dim$(C^D) \leq$ VC-dim $((HS^k)^D)^D$. Now just note that, for every class, $((C)^D)^D = C$ and recall that VC-dim$(HS^k) = k + 1$.

In the following examples, we apply Claim 8 to obtain some non-embeddability results for natural classes.

Example 1: Recall that $F_d$ denotes the class of all binary vectors over $d$ entries $(F_d = (X, Y, R)$ where $X = \{1, \ldots, d\}, Y = \{0, 1\}^X$ and $(x, v) \in R$ iff $v(x) = 1$). VC-dim$(F_d) = d$ while VC-dim$((F_d)^D) = \lceil \log(d) \rceil$ (as $|Y| = d$ it cannot exceed $\log(d)$), and as its dual class has VC-dimension $d$, Claim 6 implies that it cannot be smaller than $\lceil \log(d) \rceil$. It follows that $(F_d)^D$ is not $\preceq_{perm}^*$-embeddable in $PHS^k$ for any $k$ less than the exponent of its VC-dimension. Consequently, $(F_2)^D$ is an example of a class whose VC-dimension is $d$ and it is not a subclass of any Dudley class of dimension $< 2^d$.

Example 2: For any $k \leq n \in \mathbb{N}$, let $HW^n_k$ denote the class of all binary vectors of length $n$ having at most $k$ many non-zero entries (i.e., the class whose representing matrix, $A_{HW^n_k}$, has these vectors as its rows).

Note the VC-dim$(HW^n_k) = k$ and, whenever $n \geq 2k + 1$, VC-dim$(\overline{HW^n_k}) = 2k + 1$. It follows that $HW^n_k$ is not $\preceq_{perm}^*$-embeddable in $PHS^d$ for any $d \leq (2k - 1)$.

We therefore have, for every value of $k$, an example of a rather simple class of VC-dimension $k$ that is not embeddable in $PHS^{2k-1}$ and, consequently, cannot be extended to any $d$-dimensional Dudley class, for any $d < 2k$.

On the other hand, the class $HW^n_k$ is $\preceq_{perm}$-embeddable in a $2k$-dimensional Dudley class as follows. Let $I^k$ denote the class of the union of at most $k$ line intervals; i.e., $I^k \triangleq \{ (R, Y, \varepsilon) \mid Y = \{ A \subset \mathbb{R} : A \text{ is the union of at most } k \text{ closed intervals} \} \}$. Clearly, for any $n$, $HW^n_k \preceq_{perm} I^k$. The class $I^k$ is a $2k$-dimensional Dudley class, $D_{x, h}$, were $h(x) = -x^{2k}$ and $\mathcal{F}$ the vector space of polynomials whose degree is at most $2k - 1$.

3 Sample Compression Schemes

Given a class of sets, $C = (X, Y, R)$, a sample-compression scheme for $C$ is a mapping from (input) sequences of elements of $X$ ('examples'), labeled according to the truth value of

---

3The reason we write '$\preceq_{perm}^*$' rather than '$\preceq$' is only formalistic - we have defined $HS^k$ and $PHS^k$ as classes of \textit{closed} half-spaces, so the complement of an element of $PHS^k$ is not a member of $HS^k$. In Section 4 we shall see that this difference can be neglected for our purposes.
$R(x, y)$ for some fixed $y \in Y$, to short sub-sequences. It is required that the labels of the examples in the input sequence can be reconstructed from the short sub-sequence. The size of a sample compression scheme for a class $C$ is the (minimal) upper bound on the length of the sub-sequences in the range of this mapping.

Sample compression schemes where first introduced by Littlestone and Warmuth [I.W86] and studied extensively by Floyd [Fl89] and Floyd and Warmuth [FW95]. These papers establish close relationships between compression schemes and learning and provide sample compression schemes for several families of concept classes.

A simple counting argument shows that, up to a constant factor, the VC-dimension of a class $C$ is a lower bound on the size of compressed sub-sequences of any compression scheme for $C$. The question whether the VC-dimension determines also an upper bound (on the minimal sufficient size of compressed images) is an intriguing open problem.

In this section we shall show that, for all Dudley classes, compression schemes of size equal to their VC-dimension do exist. Furthermore, for all geometric classes we can find finite-size compression schemes.

We shall discuss three variants of this notion, the first two - labeled compression and unlabeled compression - were introduced in the works of Floyd, Littlestone and Warmuth mentioned above, while the third - array compression - is a new variant of the previous notions. For the clarity of representation, let us begin by formally defining the basic version of sample compression schemes.

**Notation 4:**

- Let $g \sqsubseteq f$ denote that $f$ is an extension of $g$ (i.e. $\text{Dom}(g) \subseteq \text{Dom}(f)$ and, for $x \in \text{Dom}(g)$, $g(x) = f(x)$).

- Recall that $\text{F}(C)$ denotes the representation of a class $C$ as a class of boolean functions over its domain $X$ (Definition 2). Let $\text{F}(C)^{<\infty}$ denote the set of finite partial functions of functions in $\text{F}(C)$,

$$\text{F}(C)^{<\infty} = \{g : \text{Dom}(g) \text{ is finite and, for some } f \in \text{F}(C), g \sqsubseteq f\}$$

Finally,

- For any finite $d$, let $\text{F}(C)^{\leq d}$ denote the set of functions in $\text{F}(C)$ whose domain is of size $\leq d$.

**Definition 7:** [[Fl89]] A size-$d$ labeled compression scheme for a class $C$ is a mapping

$$H : \text{F}(C)^{\leq d} \rightarrow \{0, 1\}^X$$

such that, for any $f \in \text{F}(C)^{<\infty}$ there exists some $g \in \text{F}(C)^{\leq d}$, so that

$$g \sqsubseteq f \sqsubseteq H(g)$$
A function $H$ as in the above definition is called a reconstruction function. We have chosen to define compression schemes via their reconstruction functions. It is easy to see that, given a reconstruction function $H$ as above, a converse compression function (mapping $f$ to an appropriate $g$) exists as well.

Let us show a few simple examples:

**Example 3:** For $k \leq n \in \mathbb{N}$, recall that $HW_k^n$ denotes the class of binary vectors of length $n$ whose Hamming weight is at most $k$. The function $H$ defined by

$$H(u)(i) = \begin{cases} 1 & \text{if } u(i) = 1 \\ 0 & \text{otherwise} \end{cases}$$

is a size-$k$ compression scheme for $HW_k^n$.

It turns out that the above compression scheme is a special case of a scheme that applies to a wide family of classes, namely,

**Example 4:** Let $C = (X, Y, R)$ be an intersection-closed class. I.e., for every $y_1, y_2 \in Y$ there exists $y_3 \in Y$ such that $c_{y_3} = c_{y_1} \cap c_{y_2}$. Define a reconstruction function, $H$, by

$$H(g)(x) = 1 \text{ iff } x \in \bigcap \{c_y : g^{-1}(1) \subseteq c_y\}$$

Clearly, for all $g \in F(C)^{<\infty}$, $g \subseteq H(g)$. Natarajan, [Nat89], proves that for intersection closed classes $C$,

$$\forall f \in F(C)^{<\infty}, \exists A \subseteq f^{-1}(1), \text{ so that } |A| \leq \text{VC-dim}(C) \text{ and } f^{-1}(1) \subseteq \bigcap \{c_y : A \subseteq c_y\}$$

It follows that $H$ is a compression scheme of size equal to the VC-dimension of $C$ (provided $C$ is intersection-closed).

Note that in both of the examples above the compression scheme, $H$, is actually stronger than what is required by Definition 7. The function $H$ in these examples can be restricted to positive examples only – it uses only a subsample of $f$’s domain on which $f(x) = 1$.

For a set $X$, let $\mathcal{P}(X)$ denote the power set of $X$ (that is, the set of all subsets of $X$), and let $\mathcal{P}(X)^{\leq d}$ denote the set of all subsets of $X$ whose cardinality is at most $d$.

**Definition 8:** [[FW95]] A size-$d$ unlabeled compression scheme for a class $C$ is a mapping

$$H : \mathcal{P}(X)^{\leq d} \mapsto \{0, 1\}^X$$

such that, for any $f \in F(C)^{<\infty}$ there exists some $A \in \mathcal{P}(X)^{\leq d}$, such that

$$A \subseteq \text{Dom}(f) \text{ and } f \subseteq H(A)$$

In this work we shall discuss yet another variant of compression schemes.
Definition 9: Given a concept class \( C = (X, Y, R) \), let \( b \) be any symbol such that \( b \not\in X \cap Y \). A size-\( d \) array compression scheme for a class \( C \) is a mapping
\[
H : (X \cup \{b\})^d \mapsto \{0, 1\}^X
\]
such that, for any \( f \in F(C)^{<\infty} \) there exists some \( \sigma \in (X \cup \{b\})^d \),
\[
\text{Range}(\sigma) \subseteq \text{Dom}(f) \cup \{b\} \text{ and } f \subseteq H(\sigma)
\]
(where, for a sequence \( \sigma = (1, \ldots, a_d) \), Range(\sigma) is the set \( \{1, \ldots, a_d\} \)).

Array compression may be viewed as reflecting a computer program that stores the information gathered by viewing a long sequence of labeled examples (the function \( f \)) in a \( d \)-size array memory in which every cell may either contain one of the examples seen in \( f \) or may be left blank.

The following lemmata are straightforward.

Lemma 2: If a class \( C \) has an unlabeled-compression scheme of size \( k \) then it has an array-compression scheme of size \( k \).

Lemma 3: If a class \( C \) has an labeled-compression scheme of size \( k \) then it has an array-compression scheme of size \( 2k \).

Array compression schemes may be viewed as a special case of a more general definition of compression. Floyd and Warmuth introduce extended sample compression schemes that are allowed to store, on top of a finite set of sample points from the domain of the compressed function \( f \), a finite set of bits.

It follows that any result concerning these extended schemes is valid for array compression schemes as well. In particular, this applies to the results of Littlestone and Warmuth [IWW86] that the existence of compression schemes of finite size implies PAC learnability, and to the results of Ben-David, Bshouty and Kushilevitz [BBK96], showing that the existence of efficient compression schemes for a class \( C \) imply the existence of efficient Online Learning algorithms for the class of all boolean combinations of members of \( C \).

Floyd [Fl89] and Floyd and Warmuth [FW95] establish the existence of labeled compression schemes for all maximum classes to size equal to their VC-dimension. This implies, as a special case, sample compression of size \( k \) to the class of positive half-spaces over a set of points in general position in \( \mathbb{R}^d \). Floyd [Fl89] also presents size 5 and 6 labeled-sample compression schemes for the class of rectangles in \( \mathbb{R}^2 \) and and for the class of triangles in \( \mathbb{R}^2 \) (respectively). In the next section we shall introduce \( k + 1 \)-size unlabeled compression schemes for the (full) classes \( HS^k \) of half-spaces in \( \mathbb{R}^k \) and finite-size array compression schemes for any class of geometric objects in \( \mathbb{R}^2 \).

Before we do that, we develop in the next two subsections, some basic technical tools concerning compression schemes.
3.1 A Compactness Lemma for Compression Schemes

**Lemma 4:** Let \( C \) be a class, \( n \) an integer and \( \tau \) one of the following types of compression schemes: ‘unlabeled’, ‘labeled’ or ‘array’. Then if any finite subclass of \( C \) admits a type \( \tau \) compression scheme of size \( n \) then so does \( C \).

**Proof:** We consider only the case of array compression. The proof for the other cases is similar. Our proof is based on the Compactness Theorem of Predicate Logic.

Let \( C = \langle X, Y, R \rangle \) be a class s.t. any finite subclass of \( C \) admits an array compression of size \( n \). Assume, w.l.o.g., that \( X \cap Y = \emptyset \) and \( b \notin X \cup Y \). Let \( Z = X \cup \{b\} \).

Consider the two-sorted model \( \mathfrak{A} = \langle Z, Y, X, b, x_{x \in X}, y_{y \in Y} \rangle \). The language of this model has the unary predicates \( \hat{Z}, \hat{Y} \) and \( \hat{X} \), the binary predicates ‘=’ and \( \hat{R} \), the constant symbol \( \hat{b} \), and constant symbols \( \hat{x} \) and \( \hat{y} \) for any \( x \in X \) and \( y \in Y \).

Let \( \hat{H} \) be an \((n+1)\)-ary predicate symbol. For an integer \( k \), let \( \psi_k \) be the sentence:

\[
\psi_k \triangleq \forall y \in \hat{Y} \, \forall x_1, \ldots, x_k \in \hat{X} \, \exists z_1, \ldots, z_n \in \hat{Z} \colon \left[ \land_{i=1}^{n} \lor_{j=1}^{k} (z_i = \hat{b} \lor z_i = x_k) \right] \\
\land \lor_{j=1}^{k} [\hat{H}(z_1, \ldots, z_n, x_j) \iff \hat{R}(x, y)].
\]

Define \( \Gamma = \{ \psi_k : k = 1, 2, \ldots \} \). Note that \( \Gamma \) “says” that the class admits an array compression of size \( n \) via the function \( \overline{\Pi} : Z^n \mapsto \{0,1\}^X \) defined by:

\[
\overline{\Pi}(\vec{z})(x) = 1 \iff H(\vec{z}, x).
\]

Let \( \Delta_{\mathfrak{A}} \) be the diagram of \( \mathfrak{A} \) — the set of all atomic sentences and negation of atomic sentences which hold in \( \mathfrak{A} \). Define \( \Gamma^+ = \Gamma \cup \Delta_{\mathfrak{A}} \). Since any finite subclass of \( C \) has an array compression of size \( n \), it follows that any finite subset of \( \Gamma^+ \) has a model. By the Compactness Theorem, \( \Gamma^+ \) has a model.

Let \( \mathfrak{A}' = \langle Z', Y', X', R', \mathbb{B}', x'_{x \in X'}, y'_{y \in Y'}, H' \rangle \) be a model of \( \Gamma^+ \). Since \( \Delta_{\mathfrak{A}} \subseteq \Gamma^+ \), we may assume that \( \mathfrak{A} \) is a sub-model of \( \mathfrak{A}' \). Consider the class \( C' = \langle X', Y', R' \rangle \). The class \( C' \) admits an array compression of size \( n \), and \( C \) is a subclass of \( C' \). Hence, \( C \) admits such a compression.

A consequence of Lemma 4 is:

**Lemma 5:** Let \( \tau \) be one of the following types of compression schemes: ‘unlabeled’, ‘labeled’ or ‘array’. Assume \( C \) admits a \( \tau \)-compression of size \( n \) and \( C' \preceq_{\tau, \mathfrak{M}} C \). Then \( C' \) admits a \( \tau \)-compression of size \( n \).

3.2 Cartesian Product of Classes

In this section we define the Cartesian Product of classes and show that if all the component classes admit an array compression of a finite size, then so does their product. (It is yet unknown whether the same holds for any other type of compression.) In a later section we use this result to establish that all the Geometric classes admit an array compression.
Let us henceforth assume that the logical value True is 1 and False is 0. We use the following notation to denote sequences and their elements. \( u \) denotes a sequence of elements; \( u_i \) or \( (u)_i \) denotes the \( i \)-th element of \( u \); \( \overline{u} \) denotes a sequence of sequences; \( \overline{u}_i \) or \( (\overline{u})_i \) denote the \( i \)-th sequence of \( \overline{u} \), and \( v_{i,j} \) or \( (v_i)_j \) denotes the \( j \)-th element of the \( i \)-th sequence of \( \overline{v} \).

**Definition 10:** Let \( C_1 = \langle X_1, Y_1, R_1 \rangle, \ldots, C_k = \langle X_k, Y_k, R_k \rangle \) be classes and \( f \) a \( k \)-ary Boolean function (i.e., \( f : \{0,1\}^k \rightarrow \{0,1\} \)). The \( f \)-Cartesian product of \( C_1, \ldots, C_k \), denoted \( (C_1 \times C_2 \times \cdots \times C_k)^f \), is the class \( C = \langle X, Y, R \rangle \) where

\[
X \triangleq X_1 \times X_2 \times \cdots \times X_k, \quad Y \triangleq Y_1 \times Y_2 \times \cdots \times Y_k \quad \text{and} \quad R(\overline{X}, \overline{Y}) = f\left( R_1(x_1, y_1), R_2(x_2, y_2), \ldots, R_k(x_k, y_k) \right)
\]

**Example 5:** Let \( f \) be the ‘and’ function and let \( C \) be the class of line intervals; i.e., \( C = \langle \mathbb{R}, \{[x,y] : x, y \in \mathbb{R} \land x < y, \}, \rangle \). Then \( (C \times C)^f \) is the class of axes parallel rectangles.

**Lemma 6:** Let \( C_1, \ldots, C_k \) and \( C'_1, \ldots, C'_k \) be classes s.t. \( C_i \preceq_{emb} C'_i \) for all \( 1 \leq i \leq k \), and let \( f \) be a \( k \)-ary Boolean function. Then \( (C_1 \times \cdots \times C_k)^f \preceq_{emb} (C'_1 \times \cdots \times C'_k)^f \).

The same holds for \( \preceq_{emb}^* \) instead of \( \preceq_{emb} \).

**Proof:** The proof is immediate. \( \square \)

The above lemma does not hold for \( \preceq_{gemb} \) or \( \preceq_{gemb}^* \). As a counter-example, consider the classes \( C \) and \( C' \) represented by the following matrices.

\[
C = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad C' = \begin{pmatrix} 0 \\ 1 \end{pmatrix}
\]

Clearly, \( C \preceq_{gemb} C' \). Let \( f \) be the ‘and’ function. Then

\[
(C \times C)^f = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad (C' \times C')^f = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}
\]

Clearly, \( (C \times C)^f \) is not \( \preceq_{gemb} \) than \( (C' \times C')^f \).

We next show that if all the components of a Cartesian product admit an array compression, then so does the product.

**Lemma 7:** Let \( C_1, \ldots, C_k \) be classes s.t. each \( C_i \) admits an array compression of size \( n_i \), and let \( f \) be a \( k \)-ary Boolean function. Then \( (C_1, \ldots, C_k)^f \) admits an array compression of size \( \sum_{i=1}^k n_i \).
In this section we show that \( (\text{HS}^n)^D \), the dual class of halfspaces of \( \mathbb{R}^n \), admits an unlabeled compression scheme of size \( n \). This implies that any Dudley class \( C \) admits an unlabeled compression scheme of size \( VC\text{-dim}(C) \).

We consider \( \mathbb{R}^n \) to be an \( n \)-dimension linear space over the real field and a metric space, where the metric is the Euclidean distance \( d \). The dimension \( n \) is fixed throughout this section.

**Definition 11:** We refer to members of \( \mathbb{R}^n \) as **points**. A **\( j \)-dimension coset** is a nonempty subset of \( \mathbb{R}^n \) of the form \( y + q \) where \( y \in \mathbb{R}^n \) and \( q \) is a \( j \)-dimension linear subspace of \( \mathbb{R}^n \). A singleton is a \( 0 \)-dimension coset, a **line** is a \( 1 \)-dimension coset, and a **hyperplane** is an \( (n-1) \)-dimension coset. A **halfspace** is a subset of \( \mathbb{R}^n \) of the form

\[
\{ \alpha \in \mathbb{R} : \alpha \geq \beta \} \cdot y + q
\]

where \( q \) is an \( (n-1) \)-dimension linear subspace, \( y \) is a point not in \( q \), and \( \beta \in \mathbb{R} \). So, a halfspace is a closed convex subset of \( \mathbb{R}^n \) whose boundary is a hyperplane. For any hyperplane there are exactly two halfspaces whose boundaries are that hyperplane; we refer to two such halfspaces as **adjacent** halfspaces. Let \( H^n \) denote the set of all halfspaces of \( \mathbb{R}^n \).

In this section we use the following definition of \( (\text{HS}^n)^D \).

\[
(\text{HS}^n)^D \triangleq (X, Y, R); \quad X \triangleq H^n \cup \{\mathbb{R}^n, \emptyset\}, \quad Y \triangleq \mathbb{R}^n \quad \text{and} \quad R(x,y) \triangleq y \in x.
\]

This \( (\text{HS}^n)^D \) differs from its namesake of Notation 2; however, the two classes are \( \preceq_{\text{emb}} \) one from the other.
We show that \((\text{HS}^n)^D\) admits an unlabeled compression scheme\(^4\) of size \(n\). To this end, we build a small machinery which is based on linear algebra, and use the following elementary lemmas whose proof is omitted.

**Lemma 8:** Let \(q\) and \(p\) be cosets of dimensions \(j\) and \(n-1\). Then \(q \cap p\) is either empty or a coset of dimension \(j\) or \(j-1\).

For \(q \subset \mathbb{R}^n\) let \(B(q)\) denote the boundary of \(q\). For \(Q\), a set of subsets of \(\mathbb{R}^n\), define \(B(Q) \triangleq \{B(q) : q \in Q\}\).

**Lemma 9:** Let \(L\) be a finite set of lines, \(g\) a finite set of points and \(s\) a halfspace. Then there is a halfspace \(s'\) s.t.:
1. \(B(s') \cap l\) is a singleton for any \(l \in L\).
2. \(s' \cap g = s \cap g\).
3. \(B(s') \cap g = \emptyset\).

**Definition 12:** Let \(t\) be a point and \(g\) a closed, convex and nonempty subset of \(\mathbb{R}^n\). We denote by \(m_t(g)\) the unique point \(y\) of \(g\) s.t. \(d(t, y) = \min \{d(t, x) : x \in g\}\).

**Lemma 10:** Let \(q' \subset q\) be cosets of dimensions \(j\) and \((j+1)\). Then \(\{t : m_t(q) \in q'\}\) is a hyperplane.

**Definition 13:** Let \(P\) be a set of hyperplanes. We say that \(P\) is regular if for any \(P' \subset P\):
1. If \(|P'| \leq n\) then \(\bigcap P'\) is an \((n-|P'|)\)-dimension coset.
2. If \(|P'| > n\) then \(\bigcap P' = \emptyset\).

**Definition 14:** Let \(t\) be a point and \(P\) a regular set of hyperplanes. We say that \(t\) separates \(P\) if for any \(P'\) and \(P''\), distinct subsets\(^5\) of \(P\) of cardinality at most \(n\), \(m_t(\bigcap P') \neq m_t(\bigcap P'')\).

**Lemma 11:** Let \(P\) be a regular finite set of hyperplanes. Then there is a point \(t\) that separates \(P\).

**Proof:** By Lemma 10 and since \(\mathbb{R}^n\) is not the union of finitely many hyperplanes, there is a point \(t\) s.t. \(m_t(\bigcap P') \notin p\) for any \(P' \subset P\), \(|P'| \leq n\) and \(p \in P - P'\).

**Definition 15:** For a halfspace \(s\), let \(s^{(0)}\) denote the halfspace adjacent to \(s\), and \(s^{(1)}\) denote the halfspace \(s\). For a hyperplane \(p\) and a point \(y\) not in \(p\), let \(h(p, y)\) denote the halfspace \(s\) s.t. \(B(s) = p\) and \(y \in s\).

Henceforth, let \(P\) be a regular finite set of hyperplanes and \(t\) a point that separates \(P\).

---

\(^4\)As far as compression is concerned, it is clearly irrelevant whether \((\text{HS}^n)^D\) is composed of the open halfspaces, the closed halfspaces or both.

\(^5\)Note that \(P'\) or \(P''\) may be empty; if \(P'\) (say) is empty, we define \(\bigcap P' \triangleq \mathbb{R}^n\).
**Definition 16:** For a point \( y \) define \( P(y) \triangleq \{ p \in P : y \in p \} \). Based on \( P \) and \( t \) we now define, for each point \( y \), a partial function \( \Gamma_y \) from the set of hyperplanes to the set of halfspaces s.t. whenever \( \Gamma_y(p) \) is defined then \( B(\Gamma_y(p)) = p \); i.e., \( \Gamma_y(p) \) chooses one of the two adjacent halfspaces whose boundary is \( p \). \( \Gamma_y(p) \) is defined according to the following three cases:

Case 1, \( y \notin p \): Define \( \Gamma_y(p) \triangleq h(p, y) \).

Case 2, \( y \in p, p \in P \) and \( m_i(\cap P(y)) = y \): Let \( x = m_i(\cap(P(y) - \{p\})) \). Since \( t \) separates \( P \), we have \( x \neq y \) and \( x \notin p \). Define \( \Gamma_y(p) \triangleq (h(p, x))^{[0]} \).

Case 3, otherwise: \( \Gamma_y(p) \) is undefined.

**Definition 17:** Define a cell to be any nonempty subset of \( \mathbb{R}^n \) which is the intersection of finitely many halfspaces. A cell of \( P \) is a cell of the form \( \cap S \) where \( S \) is a finite set of halfspaces, \( B(S) \subset P \), and \( S \) has no adjacent halfspaces.

**Lemma 12:** Let \( q \) be a cell of \( P \) with \( q = \cap S \) and \( B(S) \subset P \), and let \( y = m_i(q) \). Then

1. \( m_i(\cap P(y)) = y \).
2. \( P(y) \subset B(S) \).
3. \( s = \Gamma_y(B(s)) \) for any \( s \in S \).

**Proof:** Assume \( m_i(\cap P(y)) = z \neq y \). Clearly, \( d(t, z) < d(t, y) \) and \( z \notin q \). Since \( \cap P(y) \) is convex, it includes the line interval \([y, z]\). On one hand, \([y, z] \cap q = [y, y]\). On the other hand, for any \( s \in S \): if \( y \in B(s) \) then \([y, z] \cap s = [y, z]\), and if \( y \notin B(s) \) then \([y, z] \cap s = [y, y'] \neq [y, y]\). Therefore \([y, z] \cap q \neq [y, y]; \) a contradiction. This establishes (1).

To establish (2), assume \( m_i(\cap(P(y) \cap B(S))) = z \neq y \). All the arguments of the previous paragraph hold and lead to the same contradiction. Hence, \( m_i(\cap(P(y) \cap B(S))) = y \). Since \( t \) separates \( P \), we have \( P(y) \subset B(S) \).

To establish (3), let \( s \in S \). Say first that \( y \notin B(s) \). By case 1 of \( \Gamma_y \) definition, \( \Gamma_y(B(s)) = h(B(s), y) \). This and \( y \in s \) imply \( \Gamma_y(B(s)) = s \).

Say next that \( y \in B(s) \). By case 2, \( \Gamma_y(B(s)) = (h(p, x))^{[0]} \) with \( x = m_i(\cap(P(y) - \{B(s)\})) \). Consider the line interval \([y, x]\). Since \( d(t, x) < d(t, y) \), we have \([y, x] \cap q = [y, y]\). But for any \( s' \in S - \{s\}, [y, x] \cap s' \neq [y, y]\). Hence, \([y, x] \cap s = [y, y]\); therefore \( x \notin s \); this implies \( \Gamma_y(B(s)) = s \).

**Definition 18:** We say that \( C = \langle S, Z, R \rangle \), a subclass of \((HS^n)^D \), is regular if \( \mathbb{R}^n, \emptyset \notin S \), \( S \) has no adjacent halfspaces, \( B(S) \) is regular and \(^6(\cup B(S)) \cap Z = \emptyset \).

**Lemma 13:** Let \( C = \langle S, Z, R \rangle \) be a finite subclass of \((HS^n)^D \). Then there is a regular subclass of \((HS^n)^D \), \( C' = \langle S', Z, R \rangle \), isomorphic to \( C \).

**Proof:** By induction on \( |S| \) and using lemma 9. \( \square \)

\(^6\)The last requirement, \((\cup B(S)) \cap Z = \emptyset \), is actually not used. We include this requirement to remark that it is irrelevant whether \((HS^n)^D \) is composed of open or closed halfspaces.
Theorem 3: The class $(HS^n)^D$ admits an unlabeled compression of size $n$.

Proof: By lemma 4 it suffices to show that any finite subclass of $(HS^n)^D$ admits such a compression. Let $C = \langle S, Z, R \rangle$ be a finite subclass. By lemma 13 we may assume that $C$ is regular. Set $P \triangleq B(S)$ and let $t$ and $\Gamma$ be as above.

Let $f : S' \to \{0,1\}$ be a sample under $C$ ($S'$ is a subset of $S$). Let $S'' = \{ s^{f(s)} : s \in S' \}$. Pick $z \in Z$ which is consistent with $f$. Since $z \in \cap S''$, we have $q \triangleq \cap S'' \neq \emptyset$; i.e., $q$ is a cell of $P$.

Let $y = m_t(q)$. By lemma 12, $\Gamma_y$ reconstructs the sample $f$ and $y$ is determined from some $S''$, a subset of $S'$ with $|S''| \leq n$, by $y = m_t(\cap B(S''))$.

By Lemma 1, $(HS^n)^D$ is a $\leq_{emb}$-universal class for all Dudley classes whose VC-dim is at most $n$. This implies:

Theorem 4: Let $C$ be a Dudley class. Then $C$ admits an unlabeled compression of size VC-dim($C$).

5 Compression of Geometric Classes

In this section we consider Geometric Classes — classes whose relation $R$ is over the real field and is definable there by first order logic. Geometric Classes have being studied by Goldberg and Jerrum [GJ95]. They have shown that all geometric classes have a finite VC–dimension and established upper bounds on the VC–dimension of a geometric classes as a function of the syntax of their first order formula representation. We show that any geometric class admits an array compression of a finite size. The technique of [GJ95] can be used to establish an upper bound on the size of the array compression. It is yet unknown whether all geometric classes admit a finite-size compression of a stronger type.

Consider the following model:

$$\mathfrak{A} = \langle \mathbb{R}, +, \cdot, \geq, r_{r \in \mathbb{R}} \rangle$$

The language $\mathcal{L}$ of $\mathfrak{A}$ has the binary function symbols ‘+’ and ‘\cdot’, the binary predicate symbols ‘=’ and ‘\geq’ and the constant symbols ‘$r$’ for each $r \in \mathbb{R}$.

Definition 19: Let $R$ be a $j$–ary relation over $\mathbb{R}$. $R$ is definable in $\mathfrak{A}$ if there is a formula $\varphi(v_1, \cdots, v_j)$ of $\mathcal{L}$ s.t. for any $r_1, \cdots, r_j \in \mathbb{R}$:

$$R(r_1, \cdots, r_j) \iff \mathfrak{A} \models \varphi(r_1, \cdots, r_j)$$

A geometric class is a class $C = \langle X, Y, R \rangle$ where $X = \mathbb{R}^l$, $Y = \mathbb{R}^m$ and the $(l + m)$–ary relation $R'(x_1, \cdots, x_l, y_1, \cdots, y_m) = R(x, y)$ is definable in $\mathfrak{A}$.

\footnote{Geometric Classes are called there “Concept Classes Parameterized by Real Numbers”.

20
Tarski and Robinson [CK90] have established that \( \mathfrak{A} \) admits elimination of quantifiers; that is, for any formula \( \varphi(v_1, \cdots, v_j) \) of \( \mathcal{L} \) there is a quantifier free formula \( \hat{\varphi}(v_1, \cdots, v_j) \) s.t.

\[
\mathfrak{A} \models \forall v_1, \cdots, v_j : \varphi(v_1, \cdots, v_j) \iff \hat{\varphi}(v_1, \cdots, v_j).
\]

Note that any quantifier free formula of \( \mathcal{L} \) can be expressed as:

\[
f \left(P_1(v_1, \cdots, v_j), 0, P_2(v_1, \cdots, v_j), 0, \cdots, P_k(v_1, \cdots, v_j), 0\right)
\]

where \( f \) is a \( k \)-ary Boolean function and each \( P_i \) is a polynomial in the variables \( v_1, \cdots, v_j \).

**Theorem 5:** Let \( C \) be a geometric class. Then \( C \) admits an array compression scheme of size \( n \) for some finite \( n \).

**Proof:** Let \( C = \langle X, Y, R \rangle \) with \( X = \mathbb{R}^l \) and \( Y = \mathbb{R}^m \). Let \( R \) be expressed by:

\[
f \left(P_1(x_1, \cdots, x_l, y_1, \cdots, y_m), 0, \cdots, P_k(x_1, \cdots, x_l, y_1, \cdots, y_m), 0\right)
\]

as above.

For each \( 1 \leq i \leq k \), the class \( C_i \triangleq \langle \mathbb{R}^l, \mathbb{R}^m, P_i(x_1, \cdots, x_l, y_1, \cdots, y_m), 0 \rangle \) is embeddable in a Dudley class; hence, by Theorem 4, it has an array compression scheme of a finite size, say \( n_i \). By Lemma 7, the class \( C' = \langle X', Y', R' \rangle = (C_1 \times \cdots \times C_k)' \) has an array compression scheme of size \( n = \sum_{i=1}^{k} n_i \).

Let \( \hat{C} = \langle \hat{X}, \hat{Y}, \hat{R} \rangle \) be the subclass of \( C' \) where \( \hat{X} = \{ x \in X' \mid \forall i, j : x_i = x_j \} \) and \( \hat{Y} = \{ y \in Y' \mid \forall i, j : y_i = y_j \} \). Clearly, \( \hat{C} \) is isomorphic to \( C \). Since \( \hat{C} \) has an array compression scheme of size \( n \), so does \( C \). \( \square \)

**References**


Appendix

PHS\(^n\)\(_0\) \(\preceq_g^{*}\) PHS\(^n\)\(_{-1}\)

We show here that PHS\(^n\)\(_0\), the class of positive half-spaces that pass through the origin, is \(\preceq_g^{*}\) than PHS\(^n\)\(_{-1}\), the class of positive half-spaces in IR\(^{n-1}\). To this end, we use the terminology of Section 4 and define PHS\(^n\)\(_0\) and PHS\(^n\) as follow.

**Definition 20:** Pick a point \(y\) of IR\(^n\) different from the origin \(\vec{0}\). Define:

\[
PHS^n \triangleq (\text{IR}^n, S^p, \in); \quad S^p \triangleq \{s \in H^n | \exists x \in \text{IR}^n : x \not\in s \land x + y \in s\}.
\]

Define: PHS\(^n\)\(_0\) = (\text{IR}^n, S^p_0, \in) as the subclass of PHS\(^n\) where

\[
S^p_0 \triangleq \{s \in S^p : \vec{0} \in B(s)\}.
\]

Note that, up to an isomorphism, the classes PHS\(^n\) and PHS\(^n\)\(_0\) are independent of \(y\).

We use the following elementary lemma.

**Lemma 14:** Let \(l\) be a line and \(L\) a finite set of lines, none of which is parallel to \(l\). Then there is a hyper-plane \(p\) such that \(p \cap \cap l = \emptyset\) and \(p \cap \cap l'\) is a singleton for any \(l' \in L\).

For two distinct points, \(x\) and \(y\), let \(l(x, y)\) denote the line passing through \(x\) and \(y\).

**Lemma 15:** PHS\(^n\)\(_0\) \(\preceq_g^{*}\) PHS\(^n\)\(_{-1}\).
Proof: Let $\text{PHS}_0^n = \langle \mathbb{R}^n, S_0^n, \varepsilon \rangle$, let $y$ be the selected point per Definition 20, and let $C' = \langle Z', Q, \varepsilon \rangle$ be a finite subclass of $\text{PHS}_0^n$.

Since $l(\bar{0}, y) \cap B(s) = \{\bar{0}\}$ for any $s \in Q$, there is a finite subclass of $\text{PHS}_0^n$, $C = \langle Z, Q, \varepsilon \rangle$, isomorphic to $C'$ s.t. $Z \cap l(\bar{0}, y) = \emptyset$. By Lemma 14, there is a hyper-plane $p$ s.t. $p \cap l(\bar{0}, y) = \emptyset$ and $p \cap l(\bar{0}, z)$ is a singleton (say $\{\frac{z}{2}\}$) for any $z \in Z$.

Consider the class $D \triangleq \langle p, G, \varepsilon \rangle$ where $G = \{p \cap s : s \in S_0^n\}$. Clearly, $D$ is isomorphic to $\text{PHS}^{n-1}$.

To establish a generalized embedding of $C$ into $D$, define the functions $\pi : Z \mapsto p$, $\tau : Q \mapsto G$ and $\sigma : X \mapsto \{0, 1\}$ as follows. $\pi(z) \triangleq \frac{z}{2}$, $\tau(s) \triangleq s \cap p$, and $\sigma(z) = 1$ iff $\bar{0}$ is in the line-interval $[z, \frac{z}{2}]$.

Since a half-space and its complement are convex, for any $z \in Z$ and $s \in Q$: $z \in s$ iff $[\pi(z) \in \tau(s) \iff \sigma(z) = 0]$. That is, $(\pi, \tau, \sigma)$ constitute a generalized embedding of $C$ into $D$. \qed