One-inclusion Hypergraph Density Revisited

Hans Ulrich Simon and Balázs Szörényi

August 12, 2009

Abstract

In this paper, we show that the density of the one-inclusion hypergraph induced by a family of multi-valued functions is bounded by the pseudo-dimension of this family. (The original proof for this fact, that has been published quite recently, makes use of a wrong claim.) We show furthermore that the well-known graph-dimension is another upper bound on the density (which solves an open problem from [4]).

1 Introduction

Haussler, Littlestone, and Warmuth [2] investigated prediction-strategies for binary classification problems in a setting where the learner receives n random points $n-1$ of which are labeled correctly, and the challenge is to predict the missing label of the n'th point with a small probability for making a mistake. In this model, it is assumed that the "true labels" are assigned according to an (unknown) function f taken from a (known) class $\mathcal{F}$ of functions. In [2], a clever prediction strategy is designed that employs a data structure named "one-inclusion graph", and the following is shown:

- The density (i.e. the number of edges over the number of nodes) of the one-inclusion graph divided by n upper bounds the probability for making a mistake.
- The VC-dimension of $\mathcal{F}$ (or even of $\mathcal{F}$ restricted to the n given points) upper-bounds the density of the one-inclusion graph.

Thus, if d denotes the VC-dimension, one arrives at a mistake-bound (i.e. a bound on the probability for making a mistake) of the form d/n when the one-inclusion prediction strategy is applied.
Rubinstein, Bartlett, and Rubinstein [4] extended these results as follows:

- They proved a mistake-bound for the one-inclusion prediction strategy that is strictly smaller than d/n (although asymptotically approaching d/n when d is fixed and n goes to infinity).
- They generalized the one-inclusion prediction strategy to multiclass classification, thereby dealing with one-inclusion hypergraphs, and proved a mistake-bound of the form d/n where d denotes the pseudo-dimension of the (known) class $\mathcal{F}$ of multi-valued functions.

The proofs of these results are based on a further development of the so-called "shifting-technique" by Haussler [1]. Specifically, the proof of the second result makes use of a theorem, Theorem 52 in [4], stating that the pseudo-dimension of $\mathcal{F}$ upper-bounds the density of the corresponding one-inclusion hypergraph. In the proof of this theorem, the shifting-technique is applied and it is claimed that shifting cannot lead to a larger pseudo-dimension. This claim, however, is provably wrong. (For a counterexample, see Section A.) In this paper, we provide an alternative proof for Theorem 52 in [4]. We show furthermore that the well-known graph-dimension of $\mathcal{F}$ is another upper bound on the density (which solves an open problem from [4]).

2 Definitions, Notations, and Observations

For sake of brevity, let $[k] = \{0, 1, \ldots, k\}$. Throughout this paper, the set R is a discretized hyper-rectangle, i.e., R is of the form $[k_1] \times \cdots \times [k_n]$. Note that n is the dimension and $k_1, \ldots, k_n$ are the side-lengths of R.
In the following definition, "P-dimension" stands for "pseudo-dimension", and G-dimension" stands for "graph-dimension". The other dimensions occurring in the definition are introduced for technical reasons (with the GP-dimension being originally introduced in [3]).

Definition 2.1 (Shattered sets, dimensions) *Let F be a subset of R.*[1]

1. *We say that $I = \{i_1, \ldots, i_d\} \subseteq \{1, \ldots, n\}$ is G-shattered at levels $t_1, \ldots, t_d$ if, for every $b \in \{0,1\}^d$, F has a non-empty intersection with the set*

$$R_b = \{f \in R | \ \forall i \in I : (f_i = t_i \text{ if } b_i = 1) \text{ and } (f_i \neq t_i \text{ if } b_i = 0)\} \ .$$

   The G-dimension of F, denoted as G-dim(F), is the largest d such that there exists a set I of size d that is G-shattered by F (at d properly chosen levels).

2. *We say that $I = \{i_1, \ldots, i_d\} \subseteq \{1, \ldots, n\}$ is G'-shattered at levels $t_1, \ldots, t_d$ if, for every $b \in \{0,1\}^d$, F has a non-empty intersection with the set*

$$R_b = \{f \in R | \ \forall i \in I : (f_i \neq t_i \text{ if } b_i = 1) \text{ and } (f_i = t_i \text{ if } b_i = 0)\} \ .$$

   The G'-dimension of F, denoted as G-dim$'(F)$, is the largest d such that there exists a set I of size d that is G'-shattered by F (at d properly chosen levels).

3. *We say that $I = \{i_1, \ldots, i_d\} \subseteq \{1, \ldots, n\}$ is P-shattered at levels $t_1, \ldots, t_d$ if, for every $b \in \{0,1\}^d$, F has a non-empty intersection with the set*

$$R_b = \{f \in R | \ \forall i \in I : (f_i \geq t_i \text{ if } b_i = 1) \text{ and } (f_i < t_i \text{ if } b_i = 0)\} \ .$$

   The P-dimension of F, denoted as P-dim(F), is the largest d such that there exists a set I of size d that is P-shattered by F (at d properly chosen levels).

[1] You may think of every element of F as a function table for a function $f \in \mathcal{F}$ restricted to the n random points that are given to the learner. Compare with the introduction.

4. *We say that* $I = \{i_1, \ldots, i_d\} \subseteq \{1, \ldots, n\}$ *is GP-shattered at levels* $t_1, \ldots, t_d$ *if, for every* $b \in \{0,1\}^d$, F *has a non-empty intersection with the set*

$$R_b = \{f \in R \mid \forall i \in I : (f_i = t_i \text{ if } b_i = 1) \text{ and } (f_i < t_i \text{ if } b_i = 0)\} \ .$$

The GP-dimension of F*, denoted as GP-dim*(F)*, is the largest* d *such that there exists a set* I *of size* d *that is GP-shattered by* F *(at* d *properly chosen levels).*

5. *We say that* $I = \{i_1, \ldots, i_d\} \subseteq \{1, \ldots, n\}$ *is GP*$'$*-shattered at levels* $t_1, \ldots, t_d$ *if, for every* $b \in \{0,1\}^d$, F *has a non-empty intersection with the set*

$$R_b = \{f \in R \mid \forall i \in I : (f_i > t_i \text{ if } b_i = 1) \text{ and } (f_i = t_i \text{ if } b_i = 0)\} \ .$$

The GP$'$*-dimension of* F*, denoted as GP-dim*$'(F)$*, is the largest* d *such that there exists a set* I *of size* d *that is GP*$'$*-shattered by* F *(at* d *properly chosen levels).*

The following inequalities are easy to verify:

$$\begin{aligned}
\max\{\text{GP-dim}(F), \text{GP-dim}'(F)\} &\leq \text{P-dim}(F) &(1)\\
\text{GP-dim}(F) &\leq \text{G-dim}(F) &(2)\\
\text{GP-dim}'(F) &\leq \text{G-dim}'(F) &(3)
\end{aligned}$$

Definition 2.2 (One-inclusion hypergraph and corresponding graph) *Let* F *be a subset of* R*.*

1. *We define the* hypergraph H_F induced by F *as follows. The set of nodes coincides with* F*. For every* $1 \leq j \leq n$*, and every choice of coordinates* $f_1, \ldots, f_{j-1}, f_{j+1}, \ldots, f_n$*, we have (at most) one hyperedge consisting of all nodes* $f \in F$ *of the form* $f = (f_1, \ldots, f_{j-1}, h, f_{j+1}, \ldots, f_n)$ *for some* $1 \leq h \leq k_j$ *but we exclude those hyperedges that consist of one node only.*

2. *The* graph G_F induced by F *is defined similarly. The set of nodes coincides with* F*. An edge in* G_F *is a pair of (non-identical) nodes that are of the form*

$$(f_1, \ldots, f_{j-1}, h, f_{j+1}, \ldots, f_n) \ , \ (f_1, \ldots, f_{j-1}, \hat{h}, f_{j+1}, \ldots, f_n) \ , \ h < \hat{h}$$

provided that

$$\{(f_1, \ldots, f_{j-1}, t, f_{j+1}, \ldots, f_n) \mid h < t < \hat{h}\} \cap F = \emptyset \ .$$

The *density of a finite (hyper-)graph* is defined as the number of (hyper-)edges divided by the number of nodes. Since the node sets of H_F and G_F coincide, and every hyperedge in H_F is represented by at least one edge in G_F, the following obviously holds:

$$\text{density}(H_F) \leq \text{density}(G_F) \tag{4}$$

3 Main Results

The following lemma is the main technical contribution of our paper. The recursion used in the proof is similar to the recursion in the original proof of Sauer's Lemma [5] (although Sauer's Lemma is not at all concerned with density). It is different from the recursion used in the proof from [3] for generalized versions of Sauer's Lemma (but could be used as well for the purpose of proving these generalized versions).

Lemma 3.1 *Let $F \subseteq [k_1] \times [k_2] \times \cdots \times [k_n]$, and let G_F be the graph induced by F. Then,* $\mathrm{density}(G_F) \le$ *GP-dim*$'(F)$.

Proof Throughout the proof, we assume that $k_1, \ldots, k_n \ge 1$. Note that this can be always achieved by identifying a hyper-rectangle with some dimensions of side-length 0 with the corresponding lower-dimensional hyper-rectangle without dimensions of side-length 0. Let $d := \text{GP-dim}'(F)$. It follows that

$$s := k_1 + \cdots + k_n \ge n \ge d \ .$$

The proof proceeds by induction over $s \ge d$. If $s = d$, then F coincides with $\{0,1\}^d$ so that H_F is the d-dimensional Boolean cube with 2^d nodes and $d2^{d-1}$ edges. In this case,

$$\mathrm{density}(H_F) = \frac{d \cdot 2^{d-1}}{2^d} = \frac{d}{2} < d \ .$$

Let us assume now that $s > d$ and remember that $n \ge d$. We associate with F the following two sets:

$$\begin{aligned} F' &:= \{f \in F |\ f_n \le k_n - 1\} \cup \{(f_1, \ldots, f_{n-1}, k_n - 1) |\ (f_1, \ldots, f_{n-1}, k_n) \in F\} \\ F'' &:= \{(f_1, \ldots, f_{n-1}) |\ (f_1, \ldots, f_{n-1}, k_n - 1), (f_1, \ldots, f_{n-1}, k_n) \in F\} \end{aligned}$$

Let $G_F = (V, E)$, $G_{F'} = (V', E')$ and $G_{F''} = (V'', E'')$. We claim that, with this notation, the following holds:

$$\text{GP-dim}'(F') \le d \tag{5}$$

$$\text{GP-dim}'(F'') \le d - 1 \tag{6}$$

$$|V| = |V'| + |V''| \tag{7}$$

$$|E| \le |E'| + |E''| + |V''| \tag{8}$$

This would accomplish the proof of the lemma because

$$|E| \le |E'| + |E''| + |V''| \overset{(*)}{\le} d \cdot |V'| + (d-1) \cdot |V''| + |V''| = d \cdot |V| \ ,$$

where the inequality marked "$(*)$" follows from (5), (6) and the inductive hypothesis. We still have to prove (5), (6), (7) and (8).

As for (5), assume that $1 \le i_1 < \cdots < i_r \le n$ is GP$'$-shattered by F' at levels $t_1, \ldots, t_r$, respectively. Since F' has side-length $k_n - 1$ in dimension n, we know that either $i_r < n$ or

$i_r = n$ and $t_r \le k_n - 2$. Note that, according to the definition of F', every vector $f \in F'$ belongs to F too or has coordinate $k_n - 1$ in dimension n and will belong to F after lifting-up this coordinate to level k_n. It follows that $i_1, \ldots, i_r$ is GP$'$-shattered by F at the same levels $t_1, \ldots, t_r$ (because, even if $i_r = n$, it is immaterial whether the n'th coordinate of a vector is $k_n - 1$ or k_n).

As for (6), assume that $1 \le i_1 < \cdots < i_r \le n-1$ is GP$'$-shattered by F'' at levels $t_1, \ldots, t_r$, respectively. Then, according to the definition of F'', $i_1, \ldots, i_r, n$ is GP$'$-shattered by $\mathcal{F}$ at levels $t_1, \ldots, t_r, k_n - 1$, respectively. Thus, (6) must hold.

(7) clearly holds because

$$(f_1, \ldots, f_{n-1}, f_n) \mapsto \begin{cases} (f_1, \ldots, f_{n-1}, f_n) & \text{if } f_n \le k_n - 1 \\ (f_1, \ldots, f_{n-1}, k_n - 1) & \text{if } f_n = k_n \text{ and } (f_1, \ldots, f_{n-1}, k_n - 1) \notin F \\ (f_1, \ldots, f_{n-1}) & \text{if } f_n = k_n \text{ and } (f_1, \ldots, f_{n-1}, k_n - 1) \in F \end{cases} \tag{9}$$

is a bijection between V and $V' \cup V''$.

We finally have to verify (8). To this end, we will apply a counting argument that proceeds "line-wise". Let us first fix coordinates $f_1, \ldots, f_{n-1}$ and consider the line segment L between $(f_1, \ldots, f_{n-1}, 0)$ and $(f_1, \ldots, f_{n-1}, k_n)$. The number of edges on a line is always one less then the number of nodes. Thus, since by (9)

$$\begin{aligned} |\{(f_1, \ldots, f_{n-1}, h) \in V : 0 \le h \le k_n\}| = & |\{(f_1, \ldots, f_{n-1}, h) \in V' : 0 \le h \le k_n - 1\}| \\ & + |\{(f_1, \ldots, f_{n-1})\} \cap V''| \ , \end{aligned}$$

it holds that the number of edges in E passing along dimension n equals to the sum of $|V''|$ and the number of edges in E' passing along dimension n.

Consider now edges $e \in E$ that pass along a dimension $1 \le j \le n-1$. By reasons of symmetry, we may assume that $j = 1$. Fix coordinates $f_2, \ldots, f_{n-1}, h$ and consider the line segment L_h between $(0, f_2, \ldots, f_{n-1}, h)$ and $(k_1, f_2, \ldots, f_{n-1}, h)$. We proceed by case analysis (see Figures 1 and 2 for an illustration of the second case):

Case 1: $h \le k_n - 2$.
Then clearly $|E \cap L_h| = |E' \cap L_h|$.

Case 2: $h = k_n - 1, k_n$.
We consider both line segments simultaneously. Furthermore, let L'' be the line segment between $(0, f_2, \ldots, f_{n-1})$ and $(k_1, f_2, \ldots, f_{n-1})$. The bijection in (9) shows that

$$|E \cap (L_{k_n} \cup L_{k_n - 1})| = \begin{cases} |E' \cap (L_{k_n} \cup L_{k_n-1})| + |E'' \cap L''| - 1 & \text{if } V'' \cap L'' = \emptyset \text{ and } V \cap L_{k_n}, V \cap L_{k_n - 1} \ne \emptyset \\ |E' \cap (L_{k_n} \cup L_{k_n-1})| + |E'' \cap L''| & \text{otherwise} \end{cases} .$$

(8) follows from the preceding discussion. •

Corollary 3.2 *Let $F \subseteq [k_1] \times [k_2] \times \cdots \times [k_n]$, and let G_F be the graph induced by F. Then,* density$(G_F) \le$ *GP-dim(F).*

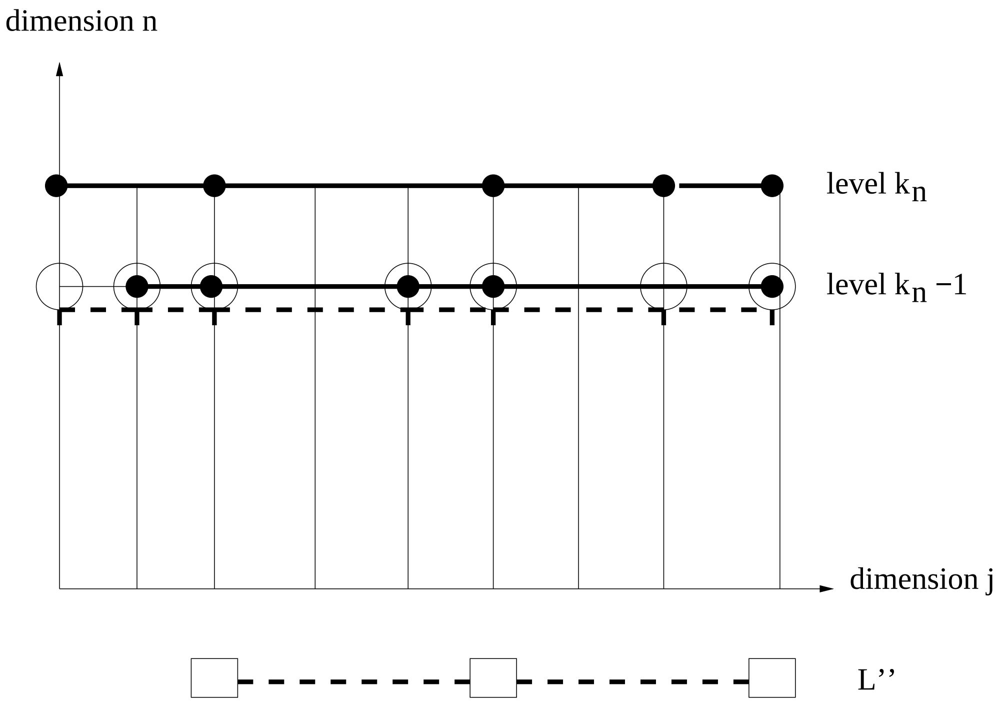

Figure 1: The scenario for edges passing along dimension j at levels k_n or $k_n - 1$, respectively: Solid circles represent nodes from V, hollow circles represent nodes from V', and squares represent nodes from V''. Thick solid lines indicate edges from E and thick dotted lines indicate edges from E' and E''. In this example, $|E \cap L_{k_n}| = |E \cap L_{k_n - 1}| = 4$, $|E' \cap L_{k_n - 1}| = 6$, and $|E'' \cap L''| = 2$.

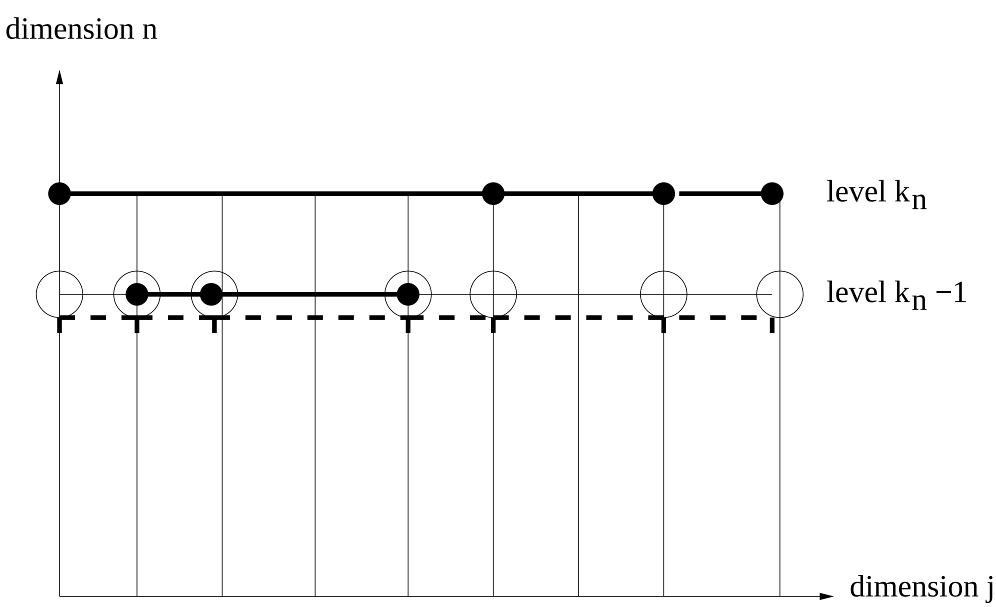

Figure 2: A similar scenario as in Figure 1 but now $V'' \cap L'' = \emptyset$. In this example, $|E \cap L_{k_n}| = 3$, $|E \cap L_{k_n - 1}| = 2$, $|E' \cap L_{k_n - 1}| = 6$, and $|E'' \cap L''| = 0$.

Proof The proof is analogous to the proof of Lemma 3.1, but we have to deal with hyper-rectangles of the form

$$R = \{l_1, \ldots, k_1\} \times \cdots \times \{l_n, \ldots, k_n\} \ , \tag{10}$$

which makes the notation more cumbersome. Initially, $l_1 = \cdots = l_n = 0$, but we are about to apply an inductive argument which leads to higher values of the parameters l_i. Given $F \subseteq R$ where R is a hyper-rectangle of the form (10), the sets F' and F'' are now defined as follows:

$$\begin{aligned} F' &:= \{f \in F |\ f_n \geq l_n + 1\} \cup \{(f_1, \ldots, f_{n-1}, l_n + 1) |\ (f_1, \ldots, f_{n-1}, l_n) \in F\} \\ F'' &:= \{(f_1, \ldots, f_{n-1}) |\ (f_1, \ldots, f_{n-1}, l_n), (f_1, \ldots, f_{n-1}, l_n + 1) \in F\} \end{aligned}$$

We may apply a symmetry argument with levels l_n and $l_n + 1$ in dimension n playing the role that levels $k_n, k_n - 1$ were playing before. •

Combining Lemma 3.1 and Corollary 3.2 with (1), (2), (3), and (4), we arrive at the following result:

Corollary 3.3 *Let $F \subseteq [k_1] \times [k_2] \times \cdots \times [k_n]$, and let H_F be the hypergraph induced by F. Then,*

$$\begin{aligned} \mathrm{density}(H_F) \leq \mathrm{density}(G_F) &\leq \min\{\text{GP-dim}(F), \text{GP-dim}'(F)\} \\ &\leq \min\{\text{P-dim}(F), \text{G-dim}(F), \text{G-dim}'(F)\} \end{aligned}$$

References

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

[2] David Haussler, Nick Littlestone, and Manfred K. Warmuth. Predicting $\{0, 1\}$ functions on randomly drawn points. *Information and Computation*, 115(2):284–293, 1994.

[3] David Haussler and Phil M. Long. A generalization of Sauer's lemma. *Journal of Combinatorial Theory, Series A*, 71(2):219–240, 1995.

[4] Benjamin I. P. Rubinstein, Peter L. Bartlett, and J. Hyam Rubinstein. Shifting: One-inclusion mistake bounds and sample compression. *Journal of Computer and System Sciences*, 75(1):37–59, 2009.

[5] N. Sauer. On the density of families of sets. *Journal of Combinatorial Theory, Series A*, 13(1):145–147, 1972.

A Shifting and Pseudo-dimension

In [4], the following shifting-operators are used:

- Operator $S_{i,t}$ shifts any vector in F that is located at level t in dimension i to the lowest level in dimension i that is not occupied by another node from V (thereby keeping the remaining coordinates fixed).
- Operator S_i is defined as the concatenation $S_{i,k_i} \circ \cdots \circ S_{i,1}$.

It is claimed in [4] that any index set P-shattered by $S_i(F)$ is P-shattered by F too. Clearly, this would imply that the P-dimension of $S_i(F)$ is upper-bounded by the P-dimension of F. The simplest counterexample against this claim is as follows:

$$\begin{aligned} F &= \{(0,0),(0,1),(1,1),(1,2)\} \\ S_2(F) &= \{(0,0),(0,1),(1,0),(1,1)\} \end{aligned}$$

Note that

$$\text{P-dim}(F) = 1 < 2 = \text{P-dim}(S_1(F)) \ .$$

By induction, we can extend this counterexample so as to get a d-dimensional set F_d of P-dimension 1 that can be iteratively shifted to a set $\tilde{F}_d$ of P-dimension d (which shows that shifting can increase the P-dimension as much as we like):

1. $F_1 = \{(0),(1)\}$ and $v_{max}(F_1) = (1)$.
2. $F_d = (\{0\} \times F_{d-1}) \cup (\{1\} \times (v_{max} + F_{d-1}))$ and $v_{max}(F_d) = \{1\} \times (2v_{max}(F_{d-1}))$.

For example, F_2 is the simple counterexample that we had discussed first, $v_{max}(F_2) = (1,2)$, and

$$\begin{aligned} F_3 &= \{(0,0,0),(0,0,1),(0,1,1),(0,1,2),(1,1,2),(1,1,3),(1,2,3),(1,2,4)\} \ , \\ v_{max}(F_3) &= (1,2,4) \ . \end{aligned}$$

It easy to show inductively that the following holds:

- The vectors from F_d form a chain w.r.t. relation $\leq$ (understood component-wise). This implies that $\text{P-dim}(F_d) = 1$.
- Shifting according to $S_2 \circ \cdots \circ S_d$ transforms F_d into $\tilde{F}_d = \{0,1\}^d$. The latter set has P-dimension d.