

Complexity Dimensions and Learnability

S.H. Nienhuys-Cheng and M. Polman

Dept. of Computer Science,
Erasmus University of Rotterdam,
The Netherlands,
Email: cheng@cs.few.eur.nl

Abstract. In a discussion of the *Vapnik Chervonenkis* (VC) *dimension* ([7]), which is closely related to the learnability of concept classes in Valiant's PAC-model ([6]), we will give an algorithm to compute it. Furthermore, we will take Natarajan's equivalent dimension for *well-ordered* classes into a more general scheme, by showing that these well-ordered classes happen to satisfy some general condition, which makes it possible to construct for a class a number of equivalent dimensions. We will give this condition, as well as a relatively efficient algorithm for the calculation of one such dimension for well-ordered classes.

1 Introduction

The PAC-model is concerned with learning concepts f (sets of strings from some domain X), grouped together in a concept class F . F is called PAC-learnable if, globally, an algorithm exists that reads in examples (pairs x, y where $x \in X$ and $y = 0$ if $x \notin f$ and $y = 1$ if $x \in f$) for some *target* concept $f \in F$, and outputs a concept $g \in F$ that is with tunable probability tunably close to f . We are interested in the number of examples needed to PAC-learn F ; we call F *polynomial sample learnable* (PSL) if, globally, the number of examples needed to learn any $f \in F$ is bounded by a polynomial in the input parameters of a PAC-algorithm, among which is n , the maximum length of the example strings. For formal definitions of the above notions, we refer to [2] and [4]. For the proofs omitted in this article, regarding a number of results, we refer to our research report ([4]).

2 Shattering and the VC-Dimension

An important notion in PAC-learning is *shattering*.

Definition. A class of concepts F on X *shatters* a set $S \subseteq X$ if the set given by $\{f \cap S | f \in F\}$ is the power set of S (denoted by 2^S).

Definition. The *VC-dimension* of a concept class F on X is the greatest integer d such that there exists a set $S \subseteq X$ of cardinality d that is shattered by F . It is denoted by $D_{vc}(F)$. If no greatest d exists, $D_{vc}(F)$ is infinite.

We can think of a set S , shattered by class F , *partitioning* F : $f, g \in F$ are equivalent if $f \cap S = g \cap S$, which gives a total of $2^{|S|}$ equivalence classes.

Our objective is to find an algorithm for the computation of $D_{vc}(F)$, where $|\cup F|$ is finite. Consider the following:

Definition. Let F shatter S . We call $x \in \cup F - S$ an *extending* element of S if for every $A \subseteq S$, we can split the equivalence class (denoted by $F_{S,A}$) into two nonempty subclasses: the concepts that contain x , and the ones that do not.

Lemma 1. Let F shatter S . Then $T = S \cup \{x\}$ is shattered by F iff x is an extending element for S . (If no extending element exists, S is called maximal.)

Notice that for a maximal set S , $|S|$ does not always equal D_{vc} .

The algorithm we constructed computes $D_{vc}(F)$ by expanding in each iteration the shattered sets found in previous iterations. This is done by adding to these sets an extending element, if possible, until no extending element can be found for any generated set. Extending elements for previously found shattered sets are found by splitting the equivalence classes as in the definition.

Algorithm:

1. Let $d = 0$. Start with the empty set \emptyset and its equivalence class $F_{\emptyset, \emptyset} = F$.
2. Suppose $d = n$. Suppose also that we have constructed S_1, \dots, S_k where every S_i is shattered by F and $|S_i| = n$. For all $x \in \cup F - S_i$ check if every $F_{S_i, A}$ (found in the previous iteration) can be divided into $F_{S_i \cup \{x\}, A \cup \{x\}}$ and $F_{S_i \cup \{x\}, A}$, both nonempty. If yes, then let $d = n + 1$. Every $S_i \cup \{x\}$ is a shattered set of $n + 1$ elements.
3. Repeat step 2 until no extending element exists for any S_i : $D_{vc}(F) = d$.

Example 1. Consider class F : $f_1 = \{0, 2, 3\}$, $f_2 = \{0, 3, 4\}$, $f_3 = \{1, 2, 3\}$, $f_4 = \{0, 1, 3, 4\}$, $f_5 = \{0, 1, 2, 3\}$, $f_6 = \{2, 3, 5\}$, $f_7 = \{1, 3, 4\}$, $f_8 = \{3, 4\}$. $D_{vc}(F)$ can be found in the following way (we will only follow one "track"): \emptyset can be extended to $\{0\}$ because F can be divided into: $F_1 = \{f_1, f_2, f_4, f_5\}$ and $F_2 = \{f_3, f_6, f_7, f_8\}$. $\{0\}$ can be extended to $\{0, 1\}$ because F_1 can be divided into: $F_3 = \{f_4, f_5\}$ and $F_4 = \{f_1, f_2\}$. Similarly $\{0, 1\}$ can be extended to $\{0, 1, 2\}$. No 4-element shattered set can be found, so $D_{vc}(F) = 3$.

We proceed with a number of definitions and results, important for the following sections:

Definition. Let X_n be the set of strings in domain X of length $\leq n$. The class of *projections* F_n over X_n is the set given by $\{f \cap X_n | f \in F\}$. A concept class F is said to be of *polynomial* VC-dimension if $D_{vc}(F_n)$ is $O(p(n))$ for some polynomial p .

Theorem 2 [1]. A class of concepts F is PSL if and only if F is of polynomial VC-dimension.

Lemma 3 [3]. Let F be a class of concepts on some finite domain X . Then $2^{d_{vc}} \leq |F| \leq (|X| + 1)^{d_{vc}}$, where $d_{vc} = D_{vc}(F)$.

3 Alternative Dimension

Besides PSL another learnability criterion exists, called PSL with omission-only error from positive examples: only positive examples are given to the algorithm and its output concept should always be a subset of the target concept. For a class F to be thus learnable, it has to be *well-ordered* and have a polynomial VC-dimension or a polynomial *dimension* (see [2]). We will need the following definitions:

Definition. Let $f \in F$. $graph(f)$ is the set of all examples for f . We call f *consistent* with a set of examples S if $S \subseteq graph(f)$. A class F is called *well-ordered* if, for any set of positive examples S for some concept f in F , there exists a concept $g \in F$ such that g is consistent with S and g is a subset of any concept in F consistent with S . (We call g the *least concept consistent with S* , denoted by $M(S)$.)

Definition dimension ([2]). The *dimension* of a well-ordered class F of concepts, $dim(F)$, is the least integer d such that for every concept $f \in F$, there exists a set S_f , $|S_f| \leq d$, such that $f = M(S_f)$.

In [2] it is proved that a well-ordered class is PSL with omission only error iff $dim(F_n)$ is $O(p(n))$ for some polynomial p . To find $dim(F)$ we consider the sets S of elements in every f such that f is $M(S)$. Any such S of minimal cardinality may be chosen as S_f . Then we have: $dim(F) = \max\{|S_f| \mid f \in F\}$. We will proceed with a number of propositions concerning well-ordered classes.

Proposition 4 ([2]). A finite class of concepts F is well-ordered iff for any $f, g \in F$, there exists an $h \in F$ such that $h = f \cap g$.

Proposition 5 ([2]). Let F be well-ordered. Let A and B be two sets of elements. Then $M(A \cup B) = M(M(A) \cup M(B))$

Proposition 6. Let F be well-ordered. Then $f \in F$ is the least concept consistent with a set $S \subset f$ iff there is no $g \in F$ such that $S \subseteq g \subset f$. Also, let f be $M(S)$ and let $S' \supset S$, $S' \subseteq f$. Then $f = M(S')$.

With these propositions the following (essential) theorem can be proved

Theorem 7. Let F be a well-ordered class over domain X . If $f \in F$ and S is any set such that:

1. $f = M(S)$
2. there is no $S' \subset S$ such that $f = M(S')$

Then S is shattered by F and has no extending elements within f . (We call S a minimal set of f .)

Proof. f is not the least concept consistent with any proper subset of S . Suppose F does not shatter S . Then there exists a set $S' \subset S$ such that there is no concept $g \in F$ with $g \cap S = S'$. Let $h = M(S')$ and $T = h \cap S$. Then $T \neq S'$, so $T \cap (S - S')$

is nonempty and $|T| > |S'|$. Since $S = (S - T) \cup S' \cup T$, it follows from Proposition 5 that:

$$M(S) = M(M(S - T) \cup M(S') \cup M(T)) .$$

Now, from Proposition 6 it follows that $M(S')$ is the same concept as $M(T)$. Therefore, and by Proposition 5:

$$M(S) = M(M(S - T) \cup M(S')) = M((S - T) \cup S') .$$

So, $f = M(S) = M((S - T) \cup S')$, but $|S| > |(S - T) \cup S'|$: f is the least concept consistent with a subset of S , which gives us a contradiction: $h \cap S = S'$. We conclude that S is shattered by F . After this it is easily verifiable that S has no extending elements within f .

4 Equivalent Dimensions

We will define a more general property for concept classes than that of well-orderedness; for classes having this property, a number of alternative dimensions (among which *dim*), all equivalent to D_{vc} , can be constructed.

Consider this: let F be a concept class over finite domain X , such that there exists an injective function $\mu : F \rightarrow 2^X$, where $|\mu(f)| \leq D_{vc}(F)$ for every $f \in F$. If we associate with μ a number $D_\mu(F) = \max\{|\mu(f)| \mid f \in F\}$, then

Theorem 8. $d_\mu \leq d_{vc} \leq d_\mu^2 \log(|X| + 1)$, where $d_\mu = D_\mu(F)$ and $d_{vc} = D_{vc}(F)$.

The essence of the above is this: suppose that F is such, that for each $f \in F$, there is a set S , $|S| \leq D_{vc}(F)$, somehow uniquely related to f . Then, if the function μ gives one such set for every f , $D_\mu(F)$ gives us a dimension with the properties of Theorem 8. Now suppose that F is such, that a μ exists, which generates (for every $n \in \mathbb{N}$) for every $f_n \in F_n$ a set of elements from X_n smaller than $D_{vc}(F_n)$. Then our result would change to:

$$D_\mu(F_n) \leq D_{vc}(F_n) \leq D_\mu(F_n)^2 \log(|X_n| + 1) .$$

Since $2 \log(|X_n| + 1)$ grows only polynomially in n , $D_{vc}(F_n)$ is $O(p(n))$ iff $D_\mu(F_n)$ is $O(q(n))$, where p and q are polynomials in n : D_μ is equivalent to D_{vc} .

Thus, the property we were looking for is the existence of a function μ as specified above. With any such μ we can associate a new dimension equivalent to the VC-dimension. Clearly, well-ordered classes are an example of classes having this general property: μ could be such that $\mu(f_n) = S_{f_n}$. Then $D_\mu(F_n) = \text{dim}(F_n)$, which is equivalent to $D_{vc}(F_n)$.

5 An Algorithm For An Alternative Dimension

We have constructed an algorithm to find, for any concept f in a well-ordered class F over domain X , a set R_f , such that f is $M(R_f)$. Any subset R_f^- of R_f , that is a minimal set (as in Theorem 7) of f , is shattered by F . Furthermore, if

we define a new dimension as the cardinality of the greatest R_f^- for all f , then this dimension is equivalent to the VC-dimension (just as $dim(F)$). We need the following definitions:

Definition. \emptyset is said to have 0 layers. $f \in F$ is said to have k layers if every $g \subset f$ has less than k layers and there is at least one $g \subset f$ that has $k - 1$ layers. For every $f \in F$, a representation set R_f is defined as follows:

1. If f is \emptyset , then $R_f = \emptyset$.
2. Suppose R_g is defined for every concept g with less than k layers. Consider the set H , being $\{h \mid \bar{\exists}g, f \supset g \supset h\}$. Let $H = \{h_1, \dots, h_n\}$. If $f \neq \cup h_i$, then pick any $a \in f - \cup h_i$ and let $R_f = \{a\}$. If $f = \cup h_i$, then define $R_f = \cup R_{h_i}$.

Example 2. Consider Fig. 1a: In this figure concepts in a concept class over the integer domain are represented by Venn-diagrams. A representation set for the 3-layer concept $\{1, 2, 3, 4\}$ is $\{2, 3, 4\}$. A representation set for the most outer concept is $\{9\}$.

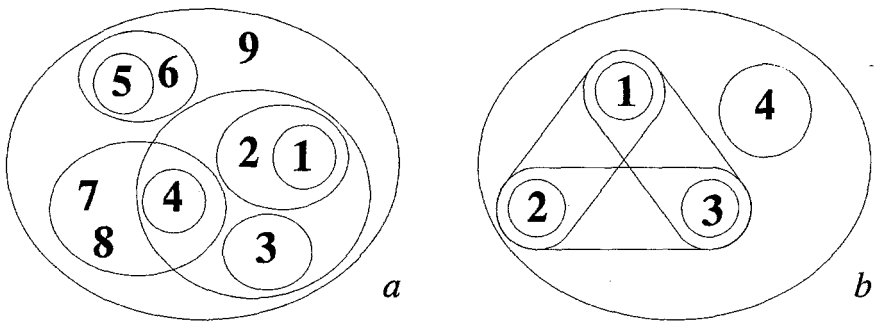


Fig. 1. Concept classes over the integer domain

Proposition 9. For any $f \in F$, $\bar{\exists}g \in F$, with $R_f \subseteq g \subset f$. So f is the least concept consistent with R_f .

From Theorem 7 of Sect. 3, it follows that any minimal set $R_f^- \subseteq R_f$ is shattered by F . Proposition 9 guarantees that such an R_f^- exists. Suppose that F_n is well-ordered for each n . If we choose for every $f_n \in F$ an R_{f_n} and if $\mu(f_n)$ is any $R_{f_n}^-$, then we have:

Theorem 10. Let $D(F_n) = \max\{|R_{f_n}^-| \mid f_n \in F_n\}$ for each n . Then D is a dimension equivalent to D_{vc} .

All of the above leads immediately to an algorithm to find an R_f for every $f \in F$: first construct representative sets of the 1-layer concepts, then the 2-layer concepts, etc. until every $f \in F$ has an R_f . The efficiency lies in the fact that every R_f is built up from at most $|H|$ (as in the definition of R_f) sets of elements,

which are already known by the time R_f is being calculated. Furthermore, the total number of elements involved in the calculation never exceeds $|F|$. The next thing to be done is to find a set $R_f^- \subseteq R_f$, that is a minimal set of f . The largest such R_f^- gives us our dimension.

Example 3. Consider Fig. 1b, where the representative set of the most outer concept is $\{1, 2, 3, 4\}$. Minimal subsets of it are $\{1, 2, 3\}$ and $\{1, 4\}$.

6 Conclusion

If a concept class F fulfils the required properties, we can define new dimensions, equivalent to D_{vc} , with respect to functions over F . After proving that a minimal set for some concept in a well-ordered class is shattered, we can compute such a dimension by a relatively efficient algorithm, using a representation set R_f for every concept f and a minimal subset $R_f^- \subseteq R_f$ of f .

7 Future Work

Since we find a strong correspondence between concepts in PAC-learning and models in first order logic, we are currently working on applying complexity dimension theory to classes of models. More about this can be found in [5].

References

1. Blumer, A., Ehrenfeucht, A., Haussler, D., Warmuth, M.: Learnability and the Vapnik-Chervonenkis Dimension. *Journal of the Association for Computing Machinery* **36** No. 4 (1989) 929-965
2. Natarajan, B.: On Learning Boolean Functions. *Proceedings of the 19th Annual ACM Symposium on Theory of Computation* 269-304
3. Natarajan, B.: *Machine Learning, a Theoretical Approach*. Morgan Kaufman Publishers, Inc (1991)
4. Nienhuys-Cheng, S.H., Polman, M.: *Complexity Dimensions and Learnability*. Research Report EUR-CS-92-06 (1992)
5. Polman, M., Nienhuys-Cheng, S.H.: Some Topics Related to PAC-Learning. *Proceedings of CSN'92, Utrecht, Netherlands, (nov. 1992)*
6. Valiant, L.G.: A theory of the learnable. *Communications of the ACM* **27** No. 11 (1984) 1134-1142
7. Vapnik, V., and Chervonenkis, A.: On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and its Applications* **16** No. 2 (1971) 264-280