

A generalization of the PAC learning in product probability spaces

Munehiro Kobayashi
Institute of Mathematics
The University of Tsukuba

Abstract

Three notions dependent theory, VC-dimension, and PAC-learnability have been found to be closely related. In addition to the known relation among these notions in model theory, finite combinatorics and probability theory, Chernikov, Palacin, and Takeuchi found a relation between n -dependence and VC_n -dimension, which are generalizations of dependence and VC-dimension respectively. We are now working to find a generalization of PAC-learnability corresponding to the above two generalizations. This attempt is a joint work with Takayuki Kuriyama and Kota Takeuchi. In this article, we see basic definitions and known results as well as some examples.

1 Introduction

It is known that dependent theory, or NIP theory, has close relation to the notions from finite combinatorics and probability theory, VC-dimension and PAC-learnability, respectively. In [1], [2], Shelah introduced a generalized notion of dependence, n -dependence. Recent study [7] of Chernikov, Palacin, and Takeuchi characterized n -dependence by VC_n -dimension, which is a generalization of VC-dimension. However, the corresponding generalization of PAC-learnability is remained to be unknown.

We attempt to find the generalization of PAC-learnability corresponding to the generalization from dependence to n -dependence and from VC-dimension to VC_n -dimension. The attempt is a joint work with Takayuki Kuriyama and Kota Takeuchi. Our main results are specifically presented in another article of ours in this *Kôkyûroku*.

In this article, we first recall the definitions of VC-dimension and PAC-learnability and the equivalence between these notions in section 2. Also, we mention Sauer-Shelah lemma there. In section 3, we see the definition of VC_n -dimension and the corresponding generalization of Sauer-Shelah lemma. After that, we introduce PAC_n -learnability and examine an example in section 4.

2 Preliminaries

We first recall the elementary notions in VC-theory. VC-dimension was introduced by Vapnik and Chervonenkis in [4], but in a different symbol.

Definition 2.1 (Vapnik, Chervonenkis [4]). Let X be a set and $\mathcal{C}$ be a subclass of the power set of X . We identify a subset C of X with the indicator function of C . This is because we need to clarify the domain of C in case we restrict the universal set X to some subset.

1. For a subset A of X , we write $\mathcal{C}|_A$ for the set $\{C|_A \mid C \in \mathcal{C}\}$.
2. A subset A of X is said to be shattered by $\mathcal{C}$ if $\mathcal{C}|_A = 2^A$.
3. We define the VC-dimension of $\mathcal{C}$ by

$$\text{VC}(\mathcal{C}) = \sup \{ |A| \mid A \text{ is a finite subset of } X \text{ shattered by } \mathcal{C} \}.$$

Definition 2.2 (Shatter function). For a class $\mathcal{C}$ of subsets of X , we define $\pi_{\mathcal{C}} : \omega \rightarrow \omega$ the shatter function of $\mathcal{C}$ as follows:

$$\pi_{\mathcal{C}}(m) = \sup \{ |\mathcal{C} \cap A| \mid A \text{ is an } m\text{-element subset of } X \},$$

where $\mathcal{C} \cap A = \{C \cap A \mid C \in \mathcal{C}\}$.

The following lemma is known as Sauer-Shelah lemma. In this article, we just state the asymptotic behavior of shatter functions. For a more specific estimate, see [7].

Lemma 2.3. *Suppose $\text{VC}(\mathcal{C}) = d$. Then, $\log(\pi_{\mathcal{C}}(m)) = O(\log m)$.*

Valiant introduced the notion of PAC-learnable in [3]. Here, for simplicity in measurability arguments, we restrict the universal set X to $\mathbb{R}^k$ or a product space of intervals in $\mathbb{R}$.

Definition 2.4 (Valiant [3]). Let X be $\mathbb{R}^k$ or a product space of intervals in $\mathbb{R}$, $\mathfrak{B}$ be the Borel set of X and $\mathcal{C}$ be a subset of $\mathfrak{B}$.

1. $\mathcal{C}_{\text{fin}} = \{C|_A \mid C \in \mathcal{C}, \text{ and } A \text{ is a finite subset of } X\}$.
2. $D(\bar{a}) = \{a_0, \dots, a_{m-1}\}$ for a tuple $\bar{a} = (a_0, \dots, a_{m-1})$.
3. Let $H : \mathcal{C}_{\text{fin}} \rightarrow \mathfrak{B}$ be a function. $\mathcal{C}$ is said to be PAC-learnable with learning function H if for all $\varepsilon, \delta > 0$, there exists $N \in \omega$ such that for an arbitrary measure $(\mu, \mathfrak{B})$ on X , an arbitrary $C \in \mathcal{C}$, and $m \geq N$,

$$\mu^m(\{ \bar{a} \in X^m \mid \mu(H(C|_{D(\bar{a})}) \Delta C) > \varepsilon \}) \leq \delta.$$

Here, μ^m is the product measure and Δ is the symmetric difference of two sets.

For a concept class $\mathcal{C}$, the finiteness of $\text{VC}(\mathcal{C})$ and the PAC-learnability of $\mathcal{C}$ are equivalent under some condition. This theorem was essentially proved in [4], but in a different notation from here.

Theorem 2.5 (Vapnik, Chervonenkis [4]). *Let X be $\mathbb{R}^k$ or a product space of intervals in $\mathbb{R}$, $\mathfrak{B}$ be the Borel set of X and $\mathcal{C}$ be a well-behaved subclass of $\mathfrak{B}$. The following are equivalent:*

1. $\mathcal{C}$ has finite VC-dimension.
2. $\mathcal{C}$ is PAC-learnable.

We do not look closely into the property of being well-behaved, which is related to measurability. For details, see [5, Appendix A].

3 VC_n -dimension

In this section, we see a generalization of VC-dimension.

Definition 3.6 ([7]). Let $X_0, \dots, X_{n-1}$ be infinite sets, X be the direct product $\prod_{i < n} X_i$, and $\mathcal{C}$ be a subclass of the power set of X .

1. A subset A of X is said to be a box of size m if $A = \prod_{i < n} A_i$ for some subsets A_i of X_i with $|A_i| = m$, $i < n$.
2. We define the VC_n -dimension of $\mathcal{C}$ by

$$\text{VC}_n(\mathcal{C}) = \sup \{ m \mid A \text{ is a box of size } m \text{ that is shattered by } \mathcal{C} \}.$$

Example 3.7. Let $X = [0, 1] \times [0, 1]$, $\mathcal{C}_1$ be “the set of all finite union of subintervals in $[0, 1]$,” and $\mathcal{C} = \{ C_1 \times C_2 \mid C_1, C_2 \in \mathcal{C}_1 \}$. Then, $\text{VC}(\mathcal{C}) = \infty$ and $\text{VC}_2(\mathcal{C}) = 1$.

Indeed, n -element set $\{ (i/n, i/n) \mid i < n \}$ is shattered by $\mathcal{C}$. Hence $\text{VC}(\mathcal{C}) = \infty$ holds. For any box $B = \{ (a_i, b_j) \mid 1 \leq i, j \leq 2 \}$ of size 2, there do not exist C in $\mathcal{C}$ that satisfies $B \cap C = B \setminus \{ (a_2, b_2) \}$. This is the case because for any C in $\mathcal{C}$, $C = C_1 \times C_2$ for some C_1 and C_2 in $\mathcal{C}$ by definition and so $\{ (a_1, b_1), (a_2, b_2) \} \subset C$ implies $B \subset C$. $\square$

Definition 3.8 (Shatter function corresponding to VC_n -dimension). Let $X_0, \dots, X_{n-1}$ be infinite sets, X be the direct product $\prod_{i < n} X_i$, and $\mathcal{C}$ be a subclass of the power set of X . We define $\pi_{\mathcal{C}, n} : \omega \rightarrow \omega$ the shatter function corresponding to VC_n -dimension of $\mathcal{C}$ as follows:

$$\pi_{\mathcal{C}, n}(m) = \sup \{ |\mathcal{C} \cap A| \mid A \text{ is a box of size } m \text{ of } X \},$$

where $\mathcal{C} \cap A = \{ C \cap A \mid C \in \mathcal{C} \}$.

It is known that a generalization of Sauer-Shelah lemma holds for VC_n -dimension. Here, we focus on the asymptotic behavior as above. For a more specific estimate, see [7].

Lemma 3.9 ([7]). *Suppose $\text{VC}_n(\mathcal{C}) = d$. Then, $\log(\pi_{\mathcal{C}, n}(m)) = O(m^{n-\varepsilon} \log m)$, where $\varepsilon = d^{-(n-1)}$.*

4 PAC_n-learnability

We introduce a new notion. We are currently working to figure out if this generalization of PAC-learnability corresponds to that of VC-dimension stated in section 3. As above, for simplicity in measurability arguments, we continue to restrict the universal set X to $\mathbb{R}^k$ or a product space of intervals in $\mathbb{R}$.

Definition 4.10. Let $X_0, \dots, X_{n-1}$ be Euclidian spaces or intervals, X be the product space $\prod_{i < n} X_i$. Also, let $\mathfrak{B}_i$ and $\mathfrak{B}$ be the Borel sets of X_i and X respectively, and $\mathcal{C} \subset \mathfrak{B}$.

1. For $a = (b_0, \dots, b_{n-1})$ in X , we put

$$D_n(a) = \bigcup_{i < n} X_0 \times \dots \times X_{i-1} \times \{b_i\} \times X_{i+1} \times \dots \times X_{n-1}.$$

2. $D_n(\bar{a}) = \bigcup_{i < m} D_n(a_i)$ for a tuple $\bar{a} = (a_0, \dots, a_{m-1})$ in X^m .
3. $\mathcal{C}_{\text{fin}}^n = \{ C|_{D_n(\bar{a})} \mid C \in \mathcal{C}, \bar{a} \in X^m, \text{ and } m \in \omega \}$.
4. Let $H : \mathcal{C}_{\text{fin}}^n \rightarrow \mathfrak{B}$ be a function. $\mathcal{C}$ is said to be PAC_n-learnable with learning function H if for all $\varepsilon, \delta > 0$, there exists $N \in \omega$ such that for arbitrary measures $(\mu_i, \mathfrak{B}_i)$ on X_i , an arbitrary $C \in \mathcal{C}$, and $m \geq N$,

$$\mu^m(\{ \bar{a} \in X^m \mid \mu(H(C|_{D_n(\bar{a})}) \Delta C) > \varepsilon \}) \leq \delta.$$

Here, μ is the product measure $\prod_{i < n} \mu_i$ and Δ is the symmetric difference of two sets.

Example 4.11. Let $X = [0, 1] \times [0, 1]$, and $\mathfrak{B}$ be the Borel set of $[0, 1]$. Also, we put $\mathcal{C}_1 =$ “the set of all finite union of subintervals in $[0, 1]$,” and $\mathcal{C} = \{ C_1 \times C_2 \mid C_1, C_2 \in \mathcal{C}_1 \}$. Then, $\mathcal{C}$ is PAC₂-learnable.

For a finite subset $\bar{a}$ of X , we define a learning function $H : \mathcal{C}_{\text{fin}}^2 \rightarrow \mathfrak{B}$ by

$$H(C|_{D_2(\bar{a})}) = p_1(C|_{D_2(\bar{a})}) \times p_2(C|_{D_2(\bar{a})}).$$

Here, p_1 and p_2 are the projection maps. Observe that we have $H(C|_{D_2(\bar{a})}) = C$ if there are a_1 and a_2 in $\bar{a}$ such that $p_1(a_1) \in p_1(C)$ and $p_2(a_2) \in p_2(C)$.

In order to show that $\mathcal{C}$ is PAC₂-learnable with learning function H , we take arbitrary $\varepsilon, \delta > 0$ and for sufficiently large m with respect to ε and δ , arbitrary measures $(\mu_i, \mathfrak{B}_i)$, and C in $\mathcal{C}$, we estimate $\mu^m(\{ \bar{a} \in X^m \mid \mu(H(C|_{D_2(\bar{a})}) \Delta C) > \varepsilon \})$, where $\mu = \mu_1 \times \mu_2$. Because $H(C|_{D_2(\bar{a})}) \subset C$ holds, if $\mu(C) \leq \varepsilon$ then $\mu^m(\{ \bar{a} \in X^m \mid \mu(H(C|_{D_2(\bar{a})}) \Delta C) > \varepsilon \}) = 0$. We assume $\mu(C) > \varepsilon$. By the above observation, we have

$$\begin{aligned} & \mu^m(\{ \bar{a} \in X^m \mid \mu(H(C|_{D_2(\bar{a})}) \Delta C) > \varepsilon \}) \\ & \leq \mu^m(\{ \bar{a} \in X^m \mid p_1(\bar{a}) \cap p_1(C) = \emptyset \text{ or } p_2(\bar{a}) \cap p_2(C) = \emptyset \}) \\ & \leq \mu^m(\{ \bar{a} \in X^m \mid p_1(\bar{a}) \cap p_1(C) = \emptyset \}) + \mu^m(\{ \bar{a} \in X^m \mid p_2(\bar{a}) \cap p_2(C) = \emptyset \}) \\ & \leq 2(1 - \varepsilon)^m \leq \delta. \end{aligned}$$

The last inequality above is derived from the way we chose m . □

We have reached a result of one side of the equivalence between finiteness of the VC_n -dimension and the PAC_n -learnability. This result was obtained by joint work with Kuriyama and Takeuchi. For the proof, refer to another article of ours in this *Kôkyûroku*.

Theorem 4.12. *Every PAC_n -learnable class has finite VC_n -dimension.*

References

- [1] Shelah, Saharon. “Strongly dependent theories.” arXiv preprint math/0504197 (2005).
- [2] Shelah, Saharon. “Definable groups for dependent and 2-dependent theories.” arXiv preprint math/0703045 (2007).
- [3] Valiant, Leslie G. “A theory of the learnable.” *Communications of the ACM* 27.11 (1984): 1134-1142.
- [4] Vapnik, Vladimir N., and A. Ya Chervonenkis. “On the uniform convergence of relative frequencies of events to their probabilities.” *Theory of Probability & Its Applications* 16.2 (1971): 264-280.
- [5] Blumer, Anselm, Andrzej Ehrenfeucht, David Haussler, and Manfred K. Warmuth. “Learnability and the Vapnik-Chervonenkis dimension.” *Journal of the ACM (JACM)* 36.4 (1989): 929-965.
- [6] Guingona, Vincent. “NIP Theories and Computational Learning Theory.”
- [7] Chernikov, Artem, Daniel Palacin, and Kota Takeuchi. “On n -dependence.” arXiv preprint arXiv:1411.0120 (2014).