On VC_2-dimension and learnability

Kota Takeuchi
Faculty of Pure and Applied Science
University of Tsukuba

Abstract

We give an example of a class with infinite VC_2-dimension which is not PAC_2-learnable defined in [1] or [2]. We give an improved definition of PAC_2-learnability for classes on finite or countable sets.

1 Introduction and Preliminaries

In model theory, it is well known that a formula $\varphi(x, y)$ is NIP (in a structure M) if and only if a class $\mathcal{C} = \{\varphi(M, a) : a \in M\}$ has finite VC-dimension. On the other hand, the finiteness of VC-dimension of class $\mathcal{C}$ coincides with PAC-learnability of $\mathcal{C}$ in statistical learning theory. The author has been considering a generalization of this correspondence.

VC_n-dimension was introduced in [3] to analyze NIP_n-formulas to answer Shelah's question about a number of types in NIP_n-theories, where $n \in \omega$. In the article, a generalization of Sauer-Shelah lemma was shown and they proved that the finiteness of VC_n-dimension of a formula is equivalent that it has NIP_n, hence the definition of VC_n-dimension seems good and to work well.

PAC_n-learning was introduced in [1] or [2] and the correspondence of VC_n-dimension and PAC_n-learnability was studied. In that article, they conjectured that a class $\mathcal{C}$ has finite VC_n-dimension if and only if it is PAC_n-learnable, and gave a proof of "if" direction.

However, recently the author of present article found a counter example for that result. It is considered that the problem is the definition of PAC_n-learnability in [1] is a little strong. In this note we see the counter example and how we fix the definition so that the "if" direction will be true.

In what follows we focus our discussion on the case $n = 2$. Let X_i $(i < 2)$ be any sets and put $X = X_0 \times X_1$.

Definition 1. Let $\mathcal{C}$ be a subset of $2^X := \mathcal{P}(X)$ and $A \subset X$.

1. A is said to be a box of height $m \in \omega$ if A has a form of $A_0 \times A_1$ with $A_i \subset X_i$ and $|A_i| = m$ for $i < 2$.
2. A is said to be shattered by $\mathcal{C}$ if $A \cap \mathcal{C} := \{A \cap C \mid C \in \mathcal{C}\}$ is 2^A.
3. VC$_2$-dimension of $\mathcal{C}$ is defined as $VC_2(\mathcal{C}) = \sup\{m \in \omega \cup \{\infty\} \mid A \subset X$ is a box of height m shattered by $\mathcal{C}\}$.

Now we consider each member of $\mathcal{C} \subset 2^X$ as a function from X to $2 = \{0, 1\}$. The following definitions are basically come from [1] or [2].

Definition 2. Let $x = (x_0, x_1) \in X$.

1. $D(x) = \{(a_0, a_1) \in X \mid x_0 = a_0$ or $x_1 = a_1\}$. We call $D(x)$ a data from x.
2. $D(\bar{x}) = \bigcup_{x \in \bar{x}} D(x)$ for a finite sequence $\bar{x} \in X^n$.
3. Sample space Sample($\mathcal{C}$) is the set $\bigcup_{n \in \omega, \bar{x} \in X^n} \{f|D(\bar{x}) \mid f \in \mathcal{C}\}$. Here, $f|D(\bar{x})$ is considered as a restriction of function.
4. A learning function for $\mathcal{C}$ is a function F : Sample($\mathcal{C}$) $\to \mathcal{C}$.

Definition 3. $\mathcal{C}$ is PAC$_2$-learnable if there is a learning function F : Sample($\mathcal{C}$) $\to \mathcal{C}$ such that for every $\epsilon, \delta > 0$ there is $N_{\epsilon,\delta,\mathcal{C}} \in \omega$, which is called sample complexity, satisfying the following: For every product probability measures $\mu = \mu_0 \times \mu_1$ and for every $f \in \mathcal{C}$,

$$\mathrm{Prov}_{\bar{x} \in X^n}(\mu(F(f|D(\bar{x}))\Delta f) > \epsilon) := \mu^n(\{\bar{x} \in X^n \mid \mu(F(f|D(\bar{x}))\Delta f) > \epsilon\}) < \delta$$

where $|\bar{x}| = n \geq N_{\epsilon,\delta,\mathcal{C}}$. Here we always assume that every element in $\mathcal{C}$ is μ-measurable.

2 The counter example

In this section we give an example of PAC$_2$-learnable $\mathcal{C}$ (in the sense of Definition 3) with infinite VC$_2$-dimension.

Let $X = \mathbb{R}^2$.

Definition 4. Let R be the set of strictly increasing sequences (r_n) of positive real numbers starting with $r_0 = 0$.

1. $f_{(r_n),a} = \{x \in X \mid r_{2n} \le d(a,x) < r_{2n+1}$ for some $n \in \omega\}$.
2. $\mathcal{C}_{\text{circles}} := \{f_{(r_n),a} \mid (r_n) \in R, a \in X\}$.

We first see that $\mathcal{C}_{\text{circles}}$ is very complex in the sense of VC_2-dimension.

Proposition 5. $\mathcal{C}_{\text{circles}}$ has infinite VC_2-dimension.

Proof. Let $A \subset X$ be any box of height $m \in \omega$. We show that A is shattered by X.

Claim A. Let B be any finite subset of X. Then there is $a \in X$ such that for all $b \neq b' \in B$, $d(a,b) \neq d(a,b')$.

Since the set $\{x \in X \mid d(x,b) = d(x,b')\}$ has a form of a line, we can choose a point $a \in X$ avoiding finitely many lines on which a has the same distance from different points in B. (End of Proof of Claim A)

Now, let $a \in X$ be a point such that $d(a,b) \neq d(a,b')$ for all $b \neq b' \in A$. Let $\{b_1, b_2, ..., b_{m^2}\}$ be an enumeration of A in the ascending order of distance from a. For each $A_0 \subset A$, it is easy to find (r_n) such that $A \cap f_{(r_n),a} = A_0$. For example if $A_0 = \{b_2, b_3, b_5, ...\}$, then put $r_1 = d(a,b_1)/2$, $r_2 = (d(a,b_1) + d(a,b_2))/2$, $r_3 = (d(a,b_3) + d(a,b_4))/2$, and so on. □

Next we study the learnability of $\mathcal{C}_{\text{circles}}$.

Lemma 6. Let $f, g \in \mathcal{C}_{\text{circles}}$ and $x \in X$. If $f|D(x) = g|D(x)$, then f and g has the same center point $a \in X$, i.e they has a form of $f_{(r_n),a}$ for some (r_n), and $f = g$ on $X \setminus \{y \in X \mid d(y,a) \ge d(x,a)\}$.

Proof. First we find the point $a \in X$ which is the center point of f and g. Since $D(x)$ is the union of a vertical line and horizontal line, say l_0 and l_1 respectively, $f|D(x) \subset l_0 \cup l_1$. Let see the set $f|D(x) \cap l_0$. It is an infinite union of intervals (or a point), and they are half open and half closed intervals except one interval l_{00} (or a point). Then the center of f must be on the horizontal bisector of l_{00}. The same argument holds for $f|D(x) \cap l_1$, hence we can detect the center point a from $f|D(x)$. It is easy to see the remains of the lemma. □

According the result in [1], $\mathcal{C}_{\text{circles}}$ cannot be PAC_2-learnable sinec it has infinite VC_2-dimension. However, in the following, we see that it is PAC_2-learnable.

Proposition 7. $\mathcal{C}_{\text{circles}}$ is PAC_2-learnable in the sense of Definition 3.

Proof. We first choose a learning function $F : \text{Sample} \to \mathcal{C}_{\text{circles}}$. Let $f = f_{(r_n),a} \in \mathcal{C}_{\text{circles}}$. By the previous lemma, for every $x \in X$, from $f|D(x)$, we can detect the center a and the situation of f outside the disk $D_x = \{y \in X \mid d(y,a) \geq d(x,a)\}$. Put $F(f|D(\bar{x}))$ as any $g \in \mathcal{C}_{\text{circles}}$ such that $f|D(x_0) = g|D(x_0)$ for some $x_0 \in \bar{x}$ with the minimum distance from a. Then $f = g$ on $X \setminus D_{x_0}$.

Let $\epsilon, \delta > 0$ and take $N \in \omega$ large enough (depending only on ϵ, δ and $\mathcal{C}_{\text{circles}}$). Let μ be any product probability measure on X and $f = f_{(r_n),a} \in \mathcal{C}_{\text{circles}}$. Take $n \geq N$ and we will prove that

$$\mu^n(\{\bar{x} \in X^n \mid \mu(F(f|D(\bar{x}))\Delta f) > \epsilon\}) < \delta.$$

Consider a largest open disk D of $\mu(D) \leq \epsilon$ with center a. (D may be empty in the case $\mu(\{a\}) \geq \epsilon$. Since μ is σ-additive, we know that $\mu(\bar{D} \cup \{a\}) \geq \epsilon$. (Here, we add a to the closure for the case that D is empty.) Notice that if $x \in D' := \bar{D} \cup \{a\}$ then $\mu(F(f|D(x))\Delta f) \leq \epsilon$. Therefore,

$$\mu^n(\{\bar{x} \in X^n \mid \mu(F(f|D(\bar{x}))\Delta f) > \epsilon\}) < \text{Prov}(\bar{x} \cap D' \neq \emptyset).$$

Since $\mu(D') \geq \epsilon$ and $|\bar{x}|$ is large enough (depending on ϵ, δ), we have that $\text{Prob}(\bar{x} \cap D' \neq \emptyset) < \delta$. □

Remark 8. In the counter example $\mathcal{C}_{\text{circles}}$, X is uncountable. We can find a similar example with countable $X = \mathbb{Q}^2$.

In the remains of this section, we see some trial of improving the definition of PAC_2-learning. From now on, we assume that X is finite or countable. Moreover, we assume that every single point in X is measurable (so that every set is measurable).

We first define sample spaces in a new way.

Definition 9. Let μ be a measure on X and $x \in X$.

1. $D_\mu(x) := \text{Supp}_\mu(D(x)) = \{y \in D(x) \mid \mu(\{y\}) > 0\}$.
2. $D_\mu(\bar{x}) = \bigcup_{x \in \bar{x}} D_\mu(x)$ for a finite sequence $\bar{x} \in X^n$.
3. Sample space $\text{Sample}^*(\mathcal{C})$ is the set $\bigcup_{n \in \omega, \bar{x} \in X^n} \{f|D_\mu(\bar{x}) \mid f \in \mathcal{C}, \mu \text{ is a product probability measure on } X\}$.
4. A learning function for $\mathcal{C}$ is a function $F : \text{Sample}^*(\mathcal{C}) \to \mathcal{C}$.

Definition 10 (New definition of PAC$_2$-learnability)**.** $\mathcal{C}$ is PAC$_2$-learnable if there is a learning function $F : \text{Sample}^*(\mathcal{C}) \to \mathcal{C}$ such that for every $\epsilon, \delta > 0$ there is $N_{\epsilon,\delta,\mathcal{C}} \in \omega$, which is called sample complexity, satisfying the following: For every product probability measures $\mu = \mu_0 \times \mu_1$ and for every $f \in \mathcal{C}$,

$$\text{Prov}_{\bar{x} \in X^n}(\mu(F(f|D_\mu(\bar{x}))\Delta f) > \epsilon) < \delta$$

where $|\bar{x}| = n \geq N_{\epsilon,\delta,\mathcal{C}}$. Here we always assume that every element in $\mathcal{C}$ is μ-measurable.

Using the above definition, we can prove that if $\mathcal{C}$ is PAC$_2$-learnable then it has finite VC$_2$-dimension. The proof is the same as in [1], indeed, in the article, they prove it by restricting the space X to the support of μ. It doesn't work for their definition, but our new definition works with the proof.

One may notice that if $\mathcal{C}$ is finite, then it is immediate that $\mathcal{C}$ is PAC$_2$-learnable in our definition. So we leave a conjecture which is true for the usual PAC-learnability and VC-dimension. The important point is that the sample complexity does not depend on $\mathcal{C}$ but only on its VC$_2$-dimension (and ϵ, δ).

Problem 11. Show that there is a function $h : \mathbb{R}^2_{>0} \times \omega \to \omega$ such that every $\mathcal{C}$ with $VC_2(\mathcal{C}) = d \in \omega$ is PAC$_2$-learnable with sample complexity $N_{\epsilon,\delta} \leq h(\epsilon, \delta, d)$.

The main idea of the above comes from the following.

Proposition 12. Let $X = \omega^2$. For each $d \in \omega$ there is a universal class $\mathcal{C}_d \subset 2^X$ of VC$_2$-dimension d such that if $\mathcal{C}_d$ is PAC$_2$-learnable with sample complexity $N_{\epsilon,\delta}$, then every class $\mathcal{C}$ (on a finite or countable product set) with dimension at most d is also PAC$_2$-learnable with sample complexity at most $N_{\epsilon,\delta}$.

It's not so difficult to show the above proposition by using amalgamation method. However we leave it for other opportunity.

References

[1] Munehiro Kobayashi, Takayuki Kuriyama and Kota Takeuchi, Higher dimensional PAC_n-learning, Congressus Numerantium Vol. 223 (2015), 227-236.

[2] Takayuki Kuriyama and Kota Takeuchi, On the PAC_n learning, RIMS kokyuroku, Vol. 1938 (2015), 54-58.

[3] Artem Chernikov, Daniel Palacin, and Kota Takeuchi, On n-Dependence, Notre Dame J. Formal Logic, Volume 60, Number 2 (2019), 195-214.