# Error Bounds for Real Function Classes Based on Discretized Vapnik-Chervonenkis Dimensions

Chao Zhang[1] and Dacheng Tao[2] *

[1]

[2] School of Computer Engineering
Nanyang Technological University, 639798, Singapore.
dctao@ntu.edu.sg

**Abstract.** The Vapnik-Chervonenkis (VC) dimension plays an important role in statistical learning theory. In this paper, we propose the discretized VC dimension obtained by discretizing the range of a real function class. Then, we point out that Sauer's Lemma is valid for the discretized VC dimension. We group the real function classes having the infinite VC dimension into four categories by using the discretized VC dimension. As a byproduct, we present the equidistantly discretized VC dimension by introducing an equidistant partition to segmenting the range of a real function class. Finally, we obtain the error bounds for real function classes based on the discretized VC dimensions in the PAC-learning framework.

**Key words:** VC dimension, statistical learning theory, error bound, real function class.

## 1 Introduction

Define $\mathcal{Z} = (\mathcal{X}, \mathcal{Y}) \subseteq \mathrm{R}^{I\times J}$, wherein $\mathcal{X} \subseteq \mathrm{R}^I$ is an input space and $\mathcal{Y} \subseteq \mathcal{R}^J$ is its corresponding output space. It is expected to find a function $T : \mathcal{X} \to \mathcal{Y}$ that, given an $\mathbf{x} \in \mathcal{X}$, can accurately predict the output $\mathbf{y} \in \mathcal{Y}$. In particular, given a loss function $\ell : \mathcal{Y}^2 \to \mathbb{R}$, the target function $T$ minimizes the expected risk

$$\mathrm{E}(\ell(T(\mathbf{x}), \mathbf{y})) = \int \ell(T(\mathbf{x}), \mathbf{y})dP, \tag{1}$$

where $P$ signifies the distribution of $\mathbf{z} = (\mathbf{x}, \mathbf{y}) \in \mathcal{Z}$. Since the distribution $P$ is unknown, the target function $T$ usually cannot be directly obtained by minimizing (1). Therefore, given a function class $\mathcal{G}$ and a sample set $S^N = \{\mathbf{z}_n\}_{n=1}^N \subset \mathcal{Z}$ with $\mathbf{z}_n = (\mathbf{x}_n, \mathbf{y}_n)$, the estimate of $T$ is achieved by minimizing the empirical risk

$$\mathrm{E}_N(\ell(g(\mathbf{x}), \mathbf{y})) = \frac{1}{N}\sum_{n=1}^{N} \ell(g(\mathbf{x}_n), \mathbf{y}_n), \quad g \in \mathcal{G}, \tag{2}$$

which is an approximation to the expected risk (1). The loss function class is defined by

$$\mathcal{F} := \{\mathbf{z} \mapsto \ell(g(\mathbf{x}), \mathbf{y}) : g \in \mathcal{G}\}.$$

and $\mathcal{F}$ is termed as the function class in the rest of the paper. Moreover, we denote, for any $f \in \mathcal{F}$,

$$\mathrm{E}f = \int f(\mathbf{z})dP \quad \text{and} \quad \mathrm{E}_N f = \frac{1}{N}\sum_{n=1}^{N} f(\mathbf{z}_n).$$

It is essential to select specific kinds of function classes to deal with different learning problems. As mentioned in [1], indicator function classes ($\mathcal{F} \subset \{0, 1\}^{\mathcal{Z}}$) and real function classes ($\mathcal{F} \subset [A, B]^{\mathcal{Z}}$) correspond to classification and regression problems, respectively. This paper follows this scenario as well.

One of the major concerns in statistical learning theory is the upper bound of

$$\sup_{f\in\mathcal{F}} \mathrm{E}f - \mathrm{E}_N f,$$

which is called the error bound and measures the probability that a function produced by an algorithm has a sufficiently small error. The error bound can be obtained by incorporating a certain complexity measure of the function class $\mathcal{F}$. For example, Vapnik [1] gave asymptotic error bounds by introducing the annealed VC entropy, the growth function and the VC dimension. Vaart and Wellner [2] exhibited some error bounds based on the Rademacher complexity and the covering number. Bartlett [3] introduced the local Rademacher complexity and presented a sharp error bound for a particular function class $\{f \in \mathcal{F} | \mathrm{E}f^2 < \alpha\mathrm{E}f, \alpha > 0\}$.

The VC dimension of a real function class can be defined as follows ([1]).

* Corresponding author.

**Definition 1.** *Let $\mathcal{F} \subseteq [A,B]^{\mathcal{Z}}$ be a real function class*[3]. *Given a sample set $S^N = \{\mathbf{z}_n\}_{n=1}^N$ drawn from $\mathcal{Z}$, for any $\beta \in (A,B)$, define*

$$\boldsymbol{f}_{S^N}^{\beta} := \big(f(\mathbf{z}_1)-\beta, f(\mathbf{z}_2)-\beta, \cdots, f(\mathbf{z}_N)-\beta\big), \tag{3}$$

*and*

$$I\left(\boldsymbol{f}_{S^N}^{\beta}\right) := \big(\sigma(f(\mathbf{z}_1)-\beta), \sigma(f(\mathbf{z}_2)-\beta), \cdots, \sigma(f(\mathbf{z}_N)-\beta)\big), \tag{4}$$

*where $\sigma$ is a step function*

$$\sigma(x) = \begin{cases} 0, & x<0; \\ 1, & x \geq 0. \end{cases} \tag{5}$$

*Thus, we obtain a set*

$$\mathcal{F}_{|S^N} := \left\{ I\left(\boldsymbol{f}_{S^N}^{\beta}\right) : f \in \mathcal{F},\ \beta \in (A,B) \right\}. \tag{6}$$

*Then, the VC dimension of $\mathcal{F}$ is defined by*

$$VC(\mathcal{F}) := \max\left\{ N>0 : \max_{S^N \in \mathcal{Z}^N} \left\{ \left|\mathcal{F}_{|S^N}\right| \right\} = 2^N \right\}. \tag{7}$$

In Definition 1, $\beta$ traverses the interval $(A,B)$, and thus many real function classes having infinite VC dimensions, *e.g.*, the function class $\{\sin(\alpha x) | \alpha \in \mathbb{R}\}$ shown in [1]. Therefore, the VC dimension is unsuitable to measure the complexities of these real function classes and it is necessary to further investigate them. Ben-David and Gurvits [4] studied the function classes having the infinite VC dimension by using the $\sigma$-ideal and built a relation with the VC dimension and the Lebesgue measure.

In this paper, we propose the discretized VC dimension to measure complexities of real function classes. The discretized VC dimension only requires $\beta$ to be evaluated from a specific partition of $(A,B)$. We then discuss some properties of the discretized VC dimension and show that Sauer's Lemma [5] is valid for the discretized VC dimension. Afterwards, we classify the real function classes having the infinite VC dimension into four categories by using the discretized VC dimension. As a byproduct, we present a special discretized VC dimension - the equidistantly discretized VC dimension, which is generated from a equidistant partition of $[A,B]$, and its properties are discussed as well. Finally, we obtain error bounds for real function classes based on the discretized VC dimensions in the PAC-leaning framework [6].

The rest of this paper is organized as follows. In Section 2, we introduce the discretized VC dimension, discuss its properties and group the real function classes having the infinite VC dimension. Section 3 presents error bounds for real function classes based on the discretized VC dimension in the PAC-learning framework. The proofs of main results are arranged in Section 4 and the last section concludes the paper.

## 2 Discretized Vapnik-Chervonenkis Dimensions

According to Definition 1, the VC dimension of a real function class requires $\beta$ to traverse along the interval $(A,B)$ and this traversal makes the VC dimensions of many real function classes infinite, *e.g.*, the function class $\{\sin(\alpha x) | \alpha \in \mathbb{R}\}$. Therefore, it is valuable to study the real function classes having the infinite VC dimension. For this purpose, we develop a new complexity measure for real function classes based on the partition of the interval $(A,B)$.

Let $\Lambda^M = \{\beta_1, \cdots, \beta_M\}$ be a finite partition of the interval $(A,B)$. By using (3) and (4), define a set

$$\mathcal{F}_{|S^N}^{\Lambda^M} := \left\{ I\left(\boldsymbol{f}_{S^N}^{\beta}\right) : f \in \mathcal{F},\ \beta \in \Lambda^M \right\}. \tag{8}$$

Then, the discretized VC dimension is defined as follows.

**Definition 2.** *Assume that $\mathcal{F} \subseteq [A,B]^{\mathcal{Z}}$ is a real function class. Let $\Lambda^M = \{\beta_1, \cdots, \beta_M\}$ be a finite partition of the interval $(A,B)$. Then, the discretized VC dimension of $\mathcal{F}$ is defined by*

$$DisVC(\mathcal{F}, M) := \max\left\{ N>0 : \max_{\substack{S^N \in \mathcal{Z}^N \\ \Lambda^M \subset (A,B)}} |\mathcal{F}_{|S^N}^{\Lambda^M}| = 2^N \right\}. \tag{9}$$

In the above definition, the discretized VC dimension only requires $\beta$ to be evaluated at the partition $\Lambda^M$ and Sauer's Lemma [5] is valid for the discretized VC dimension.

**Lemma 1.** *Assume that $\mathcal{F} \subseteq [A,B]^{\mathcal{Z}}$ is a function class, wherein $A, B \in \mathbb{R}$. Let $\Lambda^M = \{\beta_1, \beta_2, \cdots, \beta_M\}$ be a partition of the interval $(A,B)$ and $S^N = \{\mathbf{z}_n\}_{n=1}^N \subset \mathcal{Z}$ be an i.i.d. sample set. According to (8), if the discretized VC dimension $DisVC(\mathcal{F}, M) = D$, then we have for any $N \geq D$,*

$$\mathrm{E}\left(\left|\mathcal{F}_{|S^N}^{\Lambda^M}\right|\right) \leq \left(\frac{\mathrm{e}N}{D}\right)^D. \tag{10}$$

Lemma 1 is a direct result of Sauer's Lemma [5]. Moreover, according to Definition 2, the discretized VC dimension has the following properties.

**Theorem 1.** *Assume that $\mathcal{F} \subseteq [A,B]^{\mathcal{Z}}$ is a real function class having the VC dimension $VC(\mathcal{F})$. Let $\Lambda^M = \{\beta_1, \cdots, \beta_M\}$ be a finite partition of the interval $(A,B)$. Then,*

[3] Actually, $A$ and $B$ can be $-\infty$ and $+\infty$, respectively (*cf.* [1]). However, this paper only considers the case that $-\infty < A < B < +\infty$.

(*i*) *for any* $M \in \mathbb{N}$*, we have* $DisVC(\mathcal{F}, M) \leq VC(\mathcal{F})$*;*
(*ii*) *for any* $M \in \mathbb{N}$*, we have* $DisVC(\mathcal{F}, M) \leq VC(\mathcal{F})$*;*

According to Theorem 1, since the discretized VC dimension is smaller than the traditional VC dimension and increases with the increase of the cardinality of the partition, we can classify the real function class having the infinite VC dimension into four categories by using the discretized VC dimension.

**Definition 3.** *Let* $\mathcal{F}$ *be a real function classes having the infinite VC dimension.*

(*i*) *If there exists an* $M_0 \in \mathbb{N}$ *such that* $DisVC(\mathcal{F}, M_0) < \infty$ *and* $DisVC(\mathcal{F}, M) = \infty$ *holds for any* $M_0 < M \in \mathbb{N}$*, then* $\mathcal{F}$ *is said to be of TYPE I.*
(*ii*) *If* $DisVC(\mathcal{F}, M) = \infty$ *holds for any* $M \in \mathbb{N}$*, then the* $\mathcal{F}$ *is said to be of TYPE II.*
(*iii*) *If there exists an increasing sequence* $\{M_k\} \subseteq \mathbb{N}$ *such that* $DisVC(\mathcal{F}, M_k)$ *approaches to the infinity as* $k \to \infty$*, then* $\mathcal{F}$ *is said to be of TYPE III.*
(*iv*) *The others are said to be of TYPE IV.*

In the above definition, we group the real function classes having the infinite VC dimension into four categories. For example, the function class $\{\sin(\alpha x) \,|\, \alpha \in \mathbb{R}\}$ is of TYPE II.

Moreover, there is a special version of the discretized VC dimension, which is defined based on the equidistant partition of the interval $[A, B]$.

**Definition 4.** *Assume that* $\mathcal{F} \subseteq [A, B]^{\mathcal{Z}}$ *is a real function class and let* $A = \beta_0 < \beta_1 < \cdots < \beta_M < \beta_{M+1} = B$ *be an equidistant partition of the interval* $[A, B]$*, i.e., there exists a constant* $\delta > 0$ *such that* $\delta = (\beta_m - \beta_{m-1})$ *holds for any* $1 \leq m \leq M + 1$*. Denote* $\Lambda^\delta = \{\beta_1, \cdots, \beta_M\}$ *and*

$$\mathcal{F}^{\Lambda^\delta}_{|S^N} := \left\{ I\left( \boldsymbol{f}^{\beta}_{S^N} \right) : f \in \mathcal{F}, \; \beta \in \Lambda^\delta \right\}. \tag{11}$$

*Then, the equidistantly discretized VC dimension of* $\mathcal{F}$ *is defined by*

$$DisVC(\mathcal{F}, \delta) := \max\left\{ N > 0 : \max_{S^N \in \mathcal{Z}^N} \left| \mathcal{F}^{\Lambda^\delta}_{|S^N} \right| = 2^N \right\}. \tag{12}$$

In Definition 4, the equidistantly discretized VC dimension $DisVC(\mathcal{F}, \delta)$ is completely determined by the partition $\Lambda^\delta$ of $(A, B)$. In contrast, the discretized VC dimension $DisVC(\mathcal{F}, M)$ needs to select a specific partition to achieve the maximum in (9). Similarly, Sauer's Lemma is valid for the equidistantly discretized VC dimension.

**Lemma 2.** *Assume that* $\mathcal{F} \subseteq [A, B]^{\mathcal{Z}}$ *is a real function class and let* $A = \beta_0 < \beta_1 < \cdots < \beta_M < \beta_{M+1} = B$ *be an equidistant partition of the interval* $[A, B]$*, i.e., there exists a constant* $\delta > 0$ *such that* $\delta = (\beta_m - \beta_{m-1})$ *holds for any* $1 \leq m \leq M + 1$*. Denote* $\Lambda^\delta = \{\beta_1, \cdots, \beta_M\}$*. If the equidistantly discretized VC dimension* $DisVC(\mathcal{F}, \delta) = D$*, then we have for any* $N \geq D$*,*

$$\mathrm{E}\left( \left| \mathcal{F}^{\Lambda^\delta}_{|S^N} \right| \right) \leq \left( \frac{\mathrm{e}N}{D} \right)^D. \tag{13}$$

The equidistantly discretized VC dimension $DisVC(\mathcal{F}, \delta)$ has the following properties.

**Theorem 2.** *Assume that* $\mathcal{F} \subseteq [A, B]^{\mathcal{Z}}$ *is a real function class and let* $A = \beta_0 < \beta_1 < \cdots < \beta_M < \beta_{M+1} = B$ *be an equidistant partition of the interval* $[A, B]$*, i.e., there exists a constant* $\delta > 0$ *such that* $\delta = (\beta_m - \beta_{m-1})$ *holds for any* $1 \leq m \leq M + 1$*. Denote* $\Lambda^\delta = \{\beta_1, \cdots, \beta_M\}$*. Then,*

(*i*) *for any* $\delta > 0$*, we have* $DisVC(\mathcal{F}, \delta) \leq VC(\mathcal{F})$*;*
(*ii*) *for any* $\delta > 0$*, we have* $DisVC(\mathcal{F}, \delta) \leq DisVC(\mathcal{F}, M)$*, wherein* $DisVC(\mathcal{F}, M)$ *is the discretized VC dimension.*

It is worth emphasizing that for some $0 < \delta_1 < \delta_2$, it could be invalid that $DisVC(\mathcal{F}, \delta_1) \geq DisVC(\mathcal{F}, \delta_2)$, *i.e.*, a finer equidistant partition could not provide a larger equidistantly discretized VC dimension, which is different from the case of the discretized VC dimension $DisVC(\mathcal{F}, M)$ shown in Theorem 1.

In the next section, we will give the error bounds based on the discretized VC dimension and the equidistantly discretized VC dimension, respectively.

## 3 Risk Bounds Based on Discretized VC Dimensions

Vapnik [1] obtained error bounds of real function classes for the empirical processes of i.i.d. samples by introducing the VC dimension. In this section, we present the risk bounds based on the discretized VC dimension in the PAC-learning framework.

**Theorem 3.** *Assume that $\mathcal{F} \subseteq [A,B]^{\mathcal{Z}}$ is a real function class. Let $\Lambda^M = \{\beta_1, \beta_2, \cdots, \beta_M\}$ be a partition of the interval $(A,B)$, and*

$$A = \beta_0 < \beta_1 < \beta_2 < \cdots < \beta_M < \beta_{M+1} = B, \tag{14}$$

*with $\Delta\beta_m = \beta_m - \beta_{m-1}$ $(m = 1, 2, \cdots, M+1)$. Then, for an i.i.d. sample set $S^{2N} = \{\mathbf{z}_n\}_{n=1}^{2N}$ drawn from $\mathcal{Z}$, we have with the probability at least $1 - \xi - \sum_{m=1}^{M+1} \eta_m$,*

$$\forall f \in \mathcal{F}, \;\; \mathrm{E}f \leq \mathrm{E}_N f + \Delta\beta^* \left( \frac{1}{N} + \sqrt{\frac{\ln \mathrm{E}\left( \left| \mathcal{F}^{\Lambda^M}_{|S^{2N}} \right| \right) - \ln(\xi/4)}{N}} \right), \tag{15}$$

*where*

$$\eta_m = 2\mathrm{e}^{-2N(\epsilon_m)^2}, \;\; 1 \leq m \leq M, \tag{16}$$

*and $\Delta\beta^* = \max_{1 \leq m \leq M+1}\{\Delta\beta_m\}$. Moreover, if $DisVC(\mathcal{F}, M) = D$, then for any $N \geq D/2$, we have with the probability at least $1 - \xi - \sum_{m=1}^{M+1} \eta_m$,*

$$\forall f \in \mathcal{F}, \;\; \mathrm{E}f \leq \mathrm{E}_N f + \Delta\beta^* \left( \frac{1}{N} + \sqrt{\frac{D(\ln(2\mathrm{e}N) - \ln D) - \ln(\xi/4)}{N}} \right). \tag{17}$$

Similar to Theorem 3, by using Lemma 2, we obtain error bounds of real function classes based on the equidistantly discretized VC dimension as follows.

**Theorem 4.** *Assume that $\mathcal{F} \subseteq [A,B]^{\mathcal{Z}}$ is a real function class having the VC dimension $VC(\mathcal{F})$ and let $A = \beta_0 < \beta_1 < \cdots < \beta_M < \beta_{M+1} = B$ be an equidistant partition of the interval $[A,B]$, i.e., there exists a constant $\delta > 0$ such that $\delta = (\beta_m - \beta_{m-1})$ holds for any $1 \leq m \leq M+1$. Then, for an i.i.d. sample set $S^{2N} = \{\mathbf{z}_n\}_{n=1}^{2N}$ drawn from $\mathcal{Z}$, we have with the probability at least $1 - \xi - \sum_{m=1}^{M+1} \eta_m$,*

$$\forall f \in \mathcal{F}, \;\; \mathrm{E}f \leq \mathrm{E}_N f + \delta \left( \frac{1}{N} + \sqrt{\frac{\ln \mathrm{E}\left( \left| \mathcal{F}^{\Lambda^\delta}_{|S^{2N}} \right| \right) - \ln(\xi/4)}{N}} \right),$$

*where $\eta_m$ $(1 \leq m \leq M)$ is defined in (16). Moreover, if $DisVC(\mathcal{F}, \delta) = D$, then for any $N \geq D/2$, we have with the probability at least $1 - \xi - \sum_{m=1}^{M+1} \eta_m$,*

$$\forall f \in \mathcal{F}, \;\; \mathrm{E}f \leq \mathrm{E}_N f + \delta \left( \frac{1}{N} + \sqrt{\frac{D(\ln(2\mathrm{e}N) - \ln D) - \ln(\xi/4)}{N}} \right).$$

## 4 Proofs of Main Results

In this section, we only prove Theorem 3. Theorem 4 can be directly obtained by combining Theorem 3 and Lemma 2.

*Proof of Theorem 3.* Given a real number $X$ and a function $f$, let $\mathcal{E}(X)$ stand for the event

$$\mathcal{E}(X) = \{\mathbf{z} : f(\mathbf{z}) > X\}. \tag{18}$$

According to the definition of Lebesgue-Stieltjes integrals and based on the partition $\Lambda^M$, for any measurable function $f \in \mathcal{F}$, we can rewrite the expected risk as

$$\begin{aligned} \mathrm{E}(f) &= \int_A^B f(\mathbf{z}) dF(\mathbf{z}) = \sum_{m=1}^{M+1} \int_{\beta_{m-1}}^{\beta_m} f(\mathbf{z}_n) dF(\mathbf{z}) \\ &= \sum_{m=1}^{M+1} \left( \lim_{K \to \infty} \sum_{k=1}^{K} \frac{\Delta\beta_m}{K} \mathrm{P}\left\{ f(\mathbf{z}) > \beta_{m-1} + \frac{k\Delta\beta_m}{K} \right\} \right), \end{aligned} \tag{19}$$

where $\mathrm{P}\left\{ f(\mathbf{z}) > \beta_{m-1} + \frac{k\Delta\beta_m}{K} \right\}$ is the probability of event $\mathcal{E}\left( \beta_{m-1} + \frac{k\Delta\beta_m}{K} \right)$.

Similarly, given a sample set $\{\mathbf{z}_n\}_{n=1}^N$ and a function $f \in \mathcal{F}$, the corresponding empirical risk can be rewritten as

$$\mathrm{E}_N f = \frac{1}{N} \sum_{n=1}^{N} f(\mathbf{z}_n) = \lim_{K \to \infty} \sum_{k=1}^{K} \frac{\Delta\beta_m}{K} \psi \left\{ f(\mathbf{z}) > \beta_{m-1} + \frac{k\Delta\beta_m}{K} \right\}, \tag{20}$$

where $\psi\left\{ f(\mathbf{z}) > \beta_{m-1} + \frac{k\Delta\beta_m}{K} \right\}$ is the frequency of even $\mathcal{E}(\beta_{m-1} + \frac{k\Delta\beta_m}{K})$ with respect to the sample set $\{\mathbf{z}_n\}_{n=1}^N$.

Then, according to (19) and (20), for any $f \in \mathcal{F}$, letting $\Delta\beta^* = \max_{1 \le m \le M+1}\{\Delta\beta_m\}$, we have

$$\begin{aligned}
&\mathrm{E}f - \mathrm{E}_N f \\
=& \sum_{m=1}^{M+1} \lim_{K\to\infty} \sum_{k=1}^{K} \frac{\Delta\beta_m}{K}\left(\mathrm{P}\left\{f(\mathbf{z}) > \beta_{m-1} + \frac{k\Delta\beta_m}{K}\right\} - \psi\left\{f(\mathbf{z}) > \beta_{m-1} + \frac{k\Delta\beta_m}{K}\right\}\right) \\
\le& \sum_{m=1}^{M+1}\left(\lim_{K\to\infty} \sum_{k=1}^{K} \frac{\Delta\beta_m}{K} \sup_{t_m \in [\beta_{m-1},\beta_m)} \left(\mathrm{P}\{f(\mathbf{z}) > t_m\} - \psi\{f(\mathbf{z}) > t_m\}\right)\right) \\
=& \sum_{m=1}^{M+1} \Delta\beta_m \sup_{t_m \in [\beta_{m-1},\beta_m)} \left(\mathrm{P}\{f(\mathbf{z}) > t_m\} - \psi\{f(\mathbf{z}) > t_m\}\right) \\
\le& \Delta\beta^* \sum_{m=1}^{M+1} \sup_{t_m \in [\beta_{m-1},\beta_m)} \left(\int \sigma\{f(\mathbf{z}) - t_m\} dF(\mathbf{z}) - \frac{1}{N}\sum_{n=1}^{N} \sigma\{f(\mathbf{z}) - t_m\}\right).
\end{aligned} \tag{21}$$

For any $m \in \{1, 2, \cdots, M+1\}$, let

$$\epsilon_m = \Big| \int \sigma\{f(\mathbf{z}) - \beta_{m-1}\} dF(\mathbf{z}) - \frac{1}{N}\sum_{n=1}^{N} \sigma\{f(\mathbf{z}) - \beta_{m-1}\}\Big|. \tag{22}$$

According to (21), (22) and Hoeffding's inequality (*cf.* [7]), for any $t_m \in [\beta_{m-1}, \beta_m)$, we then have

$$\mathrm{P}\left\{\Big| \int \sigma\{f(\mathbf{z}) - t_m\} dF(\mathbf{z}) - \frac{1}{N}\sum_{n=1}^{N} \sigma\{f(\mathbf{z}) - t_m\}\Big| > \epsilon_m\right\} < 2\mathrm{e}^{-2N(\epsilon_m)^2}, \tag{23}$$

which implies that with probability at least $1 - 2\mathrm{e}^{-2N(\epsilon_m)^2}$,

$$\Big| \int \sigma\{f(\mathbf{z}) - t_m\} dF(\mathbf{z}) - \frac{1}{N}\sum_{n=1}^{N} \sigma\{f(\mathbf{z}) - t_m\}\Big| \le \epsilon_m. \tag{24}$$

According to (21) and (24), we have with the probability at least $1 - \sum_{m=1}^{M+1} \eta_m$,

$$\begin{aligned}
&\int_A^B f(\mathbf{z}_n) dF(\mathbf{z}) - \frac{1}{N}\sum_{n=1}^{N} f(\mathbf{z}_n) \\
\le& \sum_{m=1}^{M+1} \Big| \int \sigma\{f(\mathbf{z}) - \beta_{m-1}\} dF(\mathbf{z}) - \frac{1}{N}\sum_{n=1}^{N} \sigma\{f(\mathbf{z}) - \beta_{m-1}\}\Big|,
\end{aligned} \tag{25}$$

where $\eta_m = 2\mathrm{e}^{-2N(\epsilon_m)^2}$.

Therefore, by combining (21), (25) and Theorem 4.1 in [1], we have with the probability at least $1 - \sum_{m=1}^{M+1} \eta_m$,

$$\begin{aligned}
&\mathrm{P}\left\{\sup_{f\in\mathcal{F}}\left(\int f(\mathbf{z}) dF(\mathbf{z}) - \frac{1}{N}\sum_{n=1}^{N} f(\mathbf{z}_n)\right) > \varepsilon\right\} \\
\le& \mathrm{P}\left\{\sum_{m=1}^{M+1} \sup_{t_m \in [\beta_{m-1},\beta_m)} \sup_{f\in\mathcal{F}} \left(\int \sigma\{f(\mathbf{z}) - t_m\} dF(\mathbf{z}) - \frac{1}{N}\sum_{n=1}^{N} \sigma\{f(\mathbf{z}) - t_m\}\right) > \frac{\varepsilon}{\Delta\beta^*}\right\} \\
\le& \sum_{m=1}^{M+1} \mathrm{P}\left\{\sup_{t_m \in [\beta_{m-1},\beta_m)} \sup_{f\in\mathcal{F}} \left(\int \sigma\{f(\mathbf{z}) - t_m\} dF(\mathbf{z}) - \frac{1}{N}\sum_{n=1}^{N} \sigma\{f(\mathbf{z}) - \beta_m\}\right) > \frac{\varepsilon}{\Delta\beta^*}\right\} \\
\le& \sum_{m=1}^{M+1} \mathrm{P}\left\{\sup_{f\in\mathcal{F}} \left(\int \sigma\{f(\mathbf{z}) - \beta_{m-1}\} dF(\mathbf{z}) - \frac{1}{N}\sum_{n=1}^{N} \sigma\{f(\mathbf{z}) - \beta_{m-1}\}\right) > \frac{\varepsilon}{\Delta\beta^*}\right\} \\
<& 4\sum_{m=1}^{M+1} \mathrm{E}\left(\mathcal{F}^{\beta^m}_{|S^{2N}}\right) \exp\left\{-\left(\frac{\varepsilon}{\Delta\beta^*} - \frac{1}{N}\right)^2 N\right\} = 4\mathrm{E}\left(\mathcal{F}^{\Lambda^M}_{|S^{2N}}\right) \exp\left\{-\left(\frac{\varepsilon}{\Delta\beta^*} - \frac{1}{N}\right)^2 N\right\} \\
=& 4\exp\left\{\left(\frac{\ln \mathrm{E}\left(\mathcal{F}^{\Lambda^M}_{|S^{2N}}\right)}{N} - \left(\frac{\varepsilon}{\Delta\beta^*} - \frac{1}{N}\right)^2\right) N\right\}.
\end{aligned} \tag{26}$$

Moreover, (17) can be directly resulted from (15) and Lemma 1. This completes the proof. □

## 5 Conclusion

In this paper, we propose the discretized VC dimension for real function classes. The discretized VC dimension is defined based on a specific partition to segment the range of function classes. Then, we discuss the relation between the traditional VC dimension and the discretized VC dimension and show that Sauer's Lemma is valid for the discretized VC dimension as well. By using the discretized VC dimension, we group the real function classes having the infinite VC dimension into four categories. As a byproduct, we give a special version of the discretized VC dimension - the equidistantly discretized VC dimension. Finally, we obtain error bounds for real function classes based on the discretized VC dimension and the equidistantly discretized VC dimension, respectively.

**Acknowledgments.** This work was supported by the Open Project Program of the State Key Lab of CAD&CG (Grant No. A1006), Zhejiang University.

## References

1. Vapnik, V.N.: Statistical Learning Theory. Wiley (1999)
2. van der Vaart, A.W., Wellner, J.: Weak Convergence and Empirical Processes: With Applications to Statistics (Springer Series in Statistics). Springer (1996).
3. Bartlett, P.L., Bousquet, O., Mendelson, S.: Localized Rademacher Complexities. In: 15th Annual Conference on Computational Learning Theory, pp. 44–58. (2002)
4. Ben-David, S., Gurvits, L.: A Note on VC-Dimension and Measure of Sets of Reals. Combinatorics, Probability and Computing. 9, 391–405 (2000)
5. Sauer, N.: On the Density of Families of Sets. Journal of Combinatorial Theory, Series A. 13, 145–147 (1972)
6. Haussler, D.: Probably Approximately Correct Learning. In: Proceedings of the Eighth National Conference on Artificial Intelligence, pp. 1101–1108. (1990)
7. Bousquet, O., Boucheron, S., Lugosi, G.: Introduction to Statistical Learning Theory. In: O. Bousquet *et al.* (eds,) Advanced Lectures on Machine Learning. pp. 169–207. Springer (2004).