CONSTRUCTIVE THEORY OF FUNCTIONS, Varna 2005
(B. Bojanov, Ed.), pp. 183-186
Marin Drinov Academic Publishing House, Sofia, 2006

Inequalities for Pseudo-Dimension Widths *

YURI MALIKHIN †

Given a set I, we denote by $B(I)$ the space of bounded functions $f\colon I \to \mathbb{R}$ with the norm

$$\|f\| = \sup_{x\in I} |f(x)|.$$

Let $W \subset B(I)$ be a nonempty set.

Definition 1. The *pseudo-dimension* $\dim_{\mathcal{P}}(W)$ of W is the maximal natural number n such that there exist points $x_1, \dots, x_n \in I$, and values $y_1, \dots, y_n \in \mathbb{R}$, satisfying the condition: for every $\sigma_1, \dots, \sigma_n \in \{-1, 1\}$ there exists $f \in W$ such that

$$(f(x_i) - y_i)\sigma_i > 0, \quad i = 1, \dots, n. \tag{1}$$

If there are no such n (i.e., W consists of a single element), we put $\dim_{\mathcal{P}} W = 0$. If there are infinitely many such n, we put $\dim_{\mathcal{P}} W = \infty$.

The notion of pseudo-dimension appeared due to the pioneering paper [5]. However, in Approximation Theory similar quantities were studied earlier in [4] and [6].

It is shown in [1] that if W is a linear space of dimension n, then pseudo-dimension of W is equal to n.

Definition 2. The *pseudo-dimensional n-width* of W is defined as

$$\rho_n(W) := \inf\{\sup_{f\in W} \inf_{h\in S} \|f - h\| : \dim_{\mathcal{P}}(S) \le n\}.$$

Note that $\rho_n(W)$ provides the lower bound for the error in approximating W by sets of pseudo-dimension n. On the other hand, for some classes W the well-known n-widths, including the Kolmogorov and entropy n-widths, do not converge to zero, meanwhile pseudo-dimensional n-widths converge to zero. That means, in this case, there are also effective methods of approximation (see [3]).

The following definition is useful for estimation of the pseudo-dimensional n-width.

---

*Presented by S. Konyagin in the Open Problems Session.

†The author was supported by Grants 05-01-00066 from the Russian Foundation for Basic Research and N NSh-3004.2003.1.

Definition 3. The *n-shattering* $s_n(W)$ of W is defined as the supremum of $\varepsilon > 0$ such that there exist points $x_1, \ldots, x_{n+1} \in I$, and values $y_1, \ldots, y_{n+1} \in \mathbb{R}$, satisfying the condition: for every $\sigma_1, \ldots, \sigma_{n+1} \in \{-1, 1\}$ there exists $f \in W$, such that

$$(f(x_i) - y_i)\sigma_i \geq \varepsilon, \quad i = 1, \ldots, n+1. \tag{2}$$

If there are no such ε, we put $s_n(W) = 0$.

It is clear, that for every W we have $\rho_n(W) \geq s_n(W)$.

Example of I, n, W, for which $\rho_n(W) > s_n(W)$.

Let us take $I = [0, 1]$, $n = 2$,

$$W = \{f \in B[0,1] : \mathrm{Var}_0^1 f \leq 1\}$$

(where $\mathrm{Var}_0^1 f$ is the variation of f, i.e., the supremum of sums $\sum_{i=0}^{k-1} |f(x_{i+1}) - f(x_i)|$ over all partitions $0 = x_0 < x_1 < \ldots < x_{k-1} < x_k = 1$). We will show that $s_2(W) = \frac{1}{4}$ and $\rho_2(W) \geq \frac{1}{3}$.

Let us prove that $s_n(W) = \frac{1}{2n}$. If $0 < \varepsilon \leq \frac{1}{2n}$, we can take $x_i = \frac{i-1}{n}, y_i = 0$ $(i = 1, \ldots, n+1)$. Then for every $\sigma_i \in \{-1, 1\}$ we consider the function f, which is linear on each segment $[x_i, x_{i+1}]$, and $f(x_i) = \frac{\sigma_i}{2n}$. It is obvious that $f \in W$ and (2) holds. It remains to prove $s_n(W) \leq \frac{1}{2n}$. If there are such x_i, y_i, that for every $\sigma_i \in \{-1, 1\}$ there exists $f \in W$, such that (2) holds, we can take f_1 for $\sigma_i = (-1)^{i-1}$ and f_2 for $\sigma_i = (-1)^i$. Then

$$f_1(x_{2m+1}) - f_2(x_{2m+1}) \geq 2\varepsilon, \quad f_1(x_{2m}) - f_2(x_{2m}) \leq -2\varepsilon,$$
$$4\varepsilon n \leq \mathrm{Var}_0^1(f_1 - f_2) \leq \mathrm{Var}_0^1(f_1) + \mathrm{Var}_0^1(f_2) \leq 2.$$

So, $\varepsilon \leq \frac{1}{2n}$.

To obtain the inequality $\rho_2(W) \geq \frac{1}{3}$, it sufficies to prove that for $S \subset B(I)$ inequality

$$\sup_{f \in W} \inf_{h \in S} \|f - h\| < \frac{1}{3} - \varepsilon \tag{3}$$

with some $0 < \varepsilon < \frac{1}{3}$ implies $\dim_{\mathcal{P}} S \geq 3$.

Suppose that (3) holds. The function

$$f(x) = \begin{cases} 0, & \text{if} x = 0 \\ 1, & \text{if} 0 < x \leq 1, \end{cases}$$

belongs to W. Hence, there exists $h \in S$, for which $\|f - h\| < \frac{1}{3} - \varepsilon$. We can take x_1, x_2, x_3 satisfying the conditions

$$x_1 = 0, \qquad 0 < x_2 < x_3 \leq 1, \qquad |h(x_2) - h(x_3)| < \varepsilon.$$

Then, we put

$$y_1 = h(x_2) - \frac{1}{3}, \qquad y_2 = h(x_2) - \varepsilon, \qquad y_3 = h(x_2) + \varepsilon.$$

Now, for every $\sigma = (\sigma_1, \sigma_2, \sigma_3) \in \{-1, 1\}^3$ we should find $h_\sigma \in S$ satisfying (1).

- $\sigma = (+1, +1, +1)$. For $f_{+++} \equiv 2$ there exists $h_{+++}(x) \in S$, for which $\|f_{+++} - h_{+++}\| < \frac{1}{3}$. Then $h(x_i) > \frac{5}{3} > \frac{4}{3} > y_i$. The case $\sigma = (-1, -1, -1)$ is analogous.
- $\sigma = (-1, +1, +1)$. Let

$$f_{-++}(0) = f_{-++}(x_1) = y_1 - \frac{1}{3} + \varepsilon,$$
$$f_{-++}(x_2) = y_2 + \frac{1}{3} - \varepsilon,$$
$$f_{-++}(x_3) = f_{-++}(1) = y_3 + \frac{1}{3} - \varepsilon,$$

  and let $f_{-++}(x)$ be linear on each segment $[0, x_1], [x_i, x_{i+1}], [x_3, 1]$. It is clear that $f_{-++} \in W$, i.e., $\mathrm{Var}_0^1 f \le 1$. For the corresponding $h_{-++} \in S$ inequality $\|f_{-++} - h_{-++}\| < \frac{1}{3} - \varepsilon$ implies (1).

  Cases $\sigma = (+1, -1, -1)$, $\sigma = (-1, -1 + 1)$ and $\sigma = (+1, +1, -1)$ are analogous.
- $\sigma = (+1, -1, +1)$. We take f_{+-+} piecewise linear, as before, but with values

$$f_{+-+}(0) = f_{+-+}(x_1) = y_1 + \frac{1}{3} - \varepsilon,$$
$$f_{+-+}(x_2) = y_2 - \frac{1}{3} + \varepsilon,$$
$$f_{+-+}(x_3) = f_{+-+}(1) = y_3 + \frac{1}{3} - \varepsilon.$$

  Arguing as before, we find $h_{+-+} \in S$, $\|f_{+-+} - h_{+-+}\| < \frac{1}{3} - \varepsilon$, so that (1) holds.
- Finally, for $\sigma = (-1, +1, -1)$ we can take $h_{-+-} = h$.

The example gives the answer to the question posed in [2]. However, the following questions are still open.

Question 1. *Is there a constant $C(n)$, which depends only on n, such that for every $I, W \subset B(I)$ the inequality*

$$\rho_n(W) \le C(n) s_n(W)$$

holds?

In particular, I do not know the answer in the simplest case $\#I = 2$, $n = 1$.

Question 2. *Is there an absolute constant C, such that for every $I, W \subset B(I)$ and n the inequality*

$$\rho_n(W) \le C s_n(W)$$

holds?

References

[1] D. Haussler, Decision theoretic generalizations of the PAC model for neural net and other learning applications, *Inform. Comput.* **100** 1992, no. 1, 78-150.

[2] S. V. Konyagin, Lecture notes, Moscow State University.

[3] V. Maiorov, Optimal non-linear approximation using sets of finite pseudo-dimension, East.J.Approx. V.11, N.1, 2005, P.1-19.

[4] H. S. Shapiro, Some negative theorems of approximation theory, *Michigan Math. J.* **11** (1964), 211–217.

[5] V. Vapnik and A. Chervonenkis, Necessary and sufficient conditions for uniform convergence of means to mathematical expectations, *Theory Prob. Applic.* **26** (1971), no. 3, 532-553.

[6] H. E. Warren, Lower bounds for approximation by nonlinear manifolds, *Trans. Amer. Math. Soc.* **133** (1968), no. 1, 167–178.

Yuri Malikhin

Faculty of Mechanics and Mathematics
Moscow State University
Vorobjovy Gory
Moscow 1198999
RUSSIA
E-mail: jura05@yandex.ru