Vapnik-Červonenkis (VC) Classes and Uniform Entropy

Yu Gu

July 28, 2021

Background

- Recall the *uniform entropy condition* in the Donsker theorem

$$\int_0^\infty \sup_Q \sqrt{\log N(\epsilon\|F\|_{Q,2}, \mathcal{F}, L_2(Q))}\, d\epsilon < \infty. \tag{1}$$

- In particular, (1) holds if for some $\delta > 0$,

$$\sup_Q \log N(\epsilon\|F\|_{Q,2}, \mathcal{F}, L_2(Q)) \leq K \left(\frac{1}{\epsilon}\right)^{2-\delta}, \quad 0 < \epsilon < 1.$$

- A much stronger condition is that for some number V,

$$\sup_Q N(\epsilon\|F\|_{Q,2}, \mathcal{F}, L_2(Q)) \leq K \left(\frac{1}{\epsilon}\right)^{V}, \quad 0 < \epsilon < 1. \tag{2}$$

- *Vapnik-Červonenkis (VC) classes* satisfy (2).

Outline

VC classes of sets

Consider an arbitrary collection of n points $\{x_1, \ldots, x_n\}$ in a set $\mathcal{X}$ and a collection $\mathcal{C}$ of subsets of $\mathcal{X}$.

- We say that $\mathcal{C}$ *picks out* a certain subset A of $\{x_1, \ldots, x_n\}$ if $A = C \cap \{x_1, \ldots, x_n\}$ for some $C \in \mathcal{C}$.
- We say that $\mathcal{C}$ *shatters* $\{x_1, \ldots, x_n\}$ if all of the 2^n possible subsets of $\{x_1, \ldots, x_n\}$ are picked out by the sets in $\mathcal{C}$.
- The *VC index* $V(\mathcal{C})$ of the class $\mathcal{C}$ is the smallest n for which no set of size n is shattered by $\mathcal{C}$.
- More formally, VC index is defined through

$$\Delta_n(\mathcal{C}; x_1, \ldots, x_n) = \Big|\Big\{C \cap \{x_1, \ldots, x_n\} : C \in \mathcal{C}\Big\}\Big|,$$

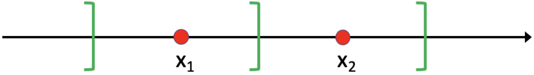
$$V(\mathcal{C}) = \inf\Big\{n : \max_{x_1, \ldots, x_n \in \mathcal{X}} \Delta_n(\mathcal{C}; x_1, \ldots, x_n) < 2^n\Big\}.$$

VC classes of sets (cont.)

- Some books define $V(\mathcal{C})$ as the largest n such that some set of size n is shattered by $\mathcal{C}$. (i.e., $V(\mathcal{C}) - 1$ in our notation).
- If $\mathcal{C}$ shatters sets of arbitrarily large size, we set $V(\mathcal{C}) = \infty$.
- Clearly, the more refined $\mathcal{C}$ is, the higher the VC index.
- We say that $\mathcal{C}$ is a VC class if $V(\mathcal{C}) < \infty$.

Example 1

- Let $\mathcal{X} = \mathbb{R}$ and define the collection of sets $\mathcal{C} = \{(-\infty, c] : c \in \mathbb{R}\}$.
- $\mathcal{C}$ shatters no two-point set $\{x_1, x_2\}$, because it fails to pick out the largest of the two points.

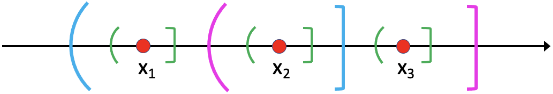


- Thus $V(\mathcal{C}) = 2$ and $\mathcal{C}$ is a VC class.
- When extended to $\mathcal{X} = \mathbb{R}^d$, the VC index of the same type of sets is $d + 1$.

Example 2

- Let $\mathcal{X} = \mathbb{R}$. Now we consider $\mathcal{C} = \{(a, b] : -\infty \leq a < b \leq \infty\}$. This collection shatters every two-point set.

- For any set of three points, $\mathcal{C}$ cannot pick out the subset consisting of the smallest and largest points.



- Thus $V(\mathcal{C}) = 3$ and $\mathcal{C}$ is a VC class.

- With more effort, it can be seen that the VC index of the same type of sets in $\mathbb{R}^d$ is $2d + 1$.

A combinatorial result

Recall that we previously defined $\Delta_n(\mathcal{C}; x_1, \ldots, x_n)$ to be the number of subsets of $\{x_1, \ldots, x_n\}$ picked out by $\mathcal{C}$. The following lemma provides an upper bound for $\Delta_n(\mathcal{C}; x_1, \ldots, x_n)$.

Sauer's lemma

For a VC class of sets $\mathcal{C}$, one has

$$\max_{x_1, \ldots, x_n \in \mathcal{X}} \Delta_n(\mathcal{C}; x_1, \ldots, x_n) \leq \sum_{j=0}^{V(\mathcal{C})-1} \binom{n}{j}.$$

Since the RHS is bounded by $V(\mathcal{C}) n^{V(\mathcal{C})-1}$, the LHS grows polynomially of order at most $O(n^{V(\mathcal{C})-1})$.

Proof of Sauer's lemma

We need to use the following lemma[1]:

Lemma

For any set of n points $\{x_1, \dots, x_n\}$ and any collection of sets $\mathcal{C}$, $\Delta_n(\mathcal{C}; x_1, \dots, x_n)$ is bounded above by the number of subsets of $\{x_1, \dots, x_n\}$ shattered by $\mathcal{C}$.

Sauer's lemma follows immediately from the above lemma, since the size of any shattered set is at most $V(\mathcal{C}) - 1$.

[1] *See Lemma 2.6.2 of VW (pp. 135-136) for the proof.*

Bound on covering number

Let $1\{\mathcal{C}\}$ denote the collection of all indicator functions of sets in the class $\mathcal{C}$. The following theorem[2] gives an upper bound on the L_r covering numbers of $1\{\mathcal{C}\}$:

Theorem 1

There exists a universal constant K such that for any VC class of sets $\mathcal{C}$, any probability measure Q, any $r \geq 1$, and any $0 < \epsilon < 1$,

$$N(\epsilon, \mathbf{1}\{\mathcal{C}\}, L_r(Q)) \leq KV(\mathcal{C})(4e)^{V(\mathcal{C})} \left(\frac{1}{\epsilon}\right)^{r(V(\mathcal{C})-1)}.$$

[2] *See Theorem 2.6.4 of VW (pp. 136-139) for the proof.*

Remarks

Since $F \equiv 1$ serves as an envelope for $1\{\mathcal{C}\}$, it follows immediately from the preceding theorem that the uniform entropy integral

$$\int_0^\infty \sup_Q \sqrt{\log N\left(\epsilon\|F\|_{Q,2}, 1\{\mathcal{C}\}, L_2(Q)\right)}\, d\epsilon$$
$$\lesssim \int_0^1 \sqrt{\log(1/\epsilon)}\, d\epsilon = \int_0^\infty u^{1/2} e^{-u}\, du \leq 1.$$

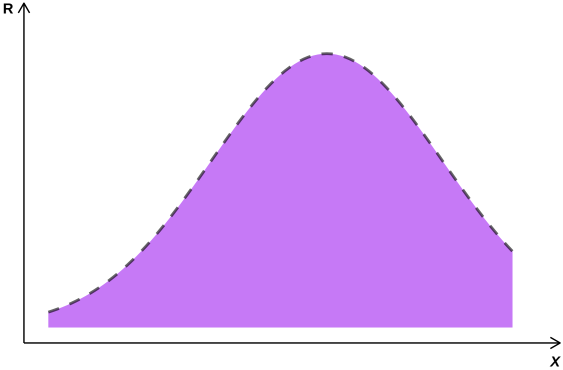
Thus, for any VC class $\mathcal{C}$, $1\{\mathcal{C}\}$ is GC and Donsker, provided the requisite measurability conditions hold.

Outline

Subgraph

The subgraph of a function $f : \mathcal{X} \to \mathbb{R}$ is the subset of $\mathcal{X} \times \mathbb{R}$ given by

$$\{(x, t) : t < f(x)\} .$$



VC classes of functions

- A class $\mathcal{F}$ of measurable real-valued functions on the sample space $\mathcal{X}$ is called a *VC subgraph class*, or just a *VC class*, if the collection of all subgraphs of the functions in $\mathcal{F}$ forms a VC class of sets (in $\mathcal{X} \times \mathbb{R}$).
- Let $V(\mathcal{F})$ denote the VC index of the set of subgraphs of $\mathcal{F}$.
- Just as for sets, the covering numbers of VC classes of functions grow at a polynomial rate.

Bound on covering number

This is more precisely stated in the following theorem:

Theorem 2

For a VC class of functions $\mathcal{F}$ with measurable envelope function F and $r \geq 1$, one has for any probability measure Q with $\|F\|_{Q,r} > 0$,

$$N\left(\epsilon\|F\|_{Q,r}, \mathcal{F}, L_r(Q)\right) \leq KV(\mathcal{F})(4e)^{V(\mathcal{F})}\left(\frac{2}{\epsilon}\right)^{r(V(\mathcal{F})-1)},$$

for a universal constant K and $0 < \epsilon < 1$.

Thus a VC class of functions easily satisfies the uniform entropy condition. Hence, a suitably measurable VC class is Donsker, provided its envelope has a weak second moment.

Proof for $r = 1$

- Let $\mathcal{C}$ be the set of all subgraphs C_f of functions $f \in \mathcal{F}$. By Fubini's theorem, $Q|f - g| = (Q \times \lambda)\left(C_f \,\Delta\, C_g\right)$, where λ is Lebesgue measure on the real line, and $A \,\Delta\, B = A \cup B - A \cap B$ for any two sets A and B.

- We renormalize $Q \times \lambda$ to a probability measure on the set $\{(x, t) : |t| \leq F(x)\}$ by defining $P = (Q \times \lambda)/(2QF)$.

- For any $f, g \in \mathcal{F}$, $\|f - g\|_{Q,1} = 2QF\|1\{C_f\} - 1\{C_g\}\|_{P,1}$.

- By the result for sets in Theorem 1, there exists a universal constant K such that

$$N(\epsilon 2QF, \mathcal{F}, L_1(Q)) = N(\epsilon, 1\{\mathcal{C}\}, L_1(P)) \leq KV(\mathcal{F})(4e)^{V(\mathcal{F})}\left(\frac{1}{\epsilon}\right)^{V(\mathcal{F})-1}.$$

Proof for $r > 1$

- For $r > 1$, define a probability measure R with density F^{r-1}/QF^{r-1} with respect to Q, so that $Rf = Q\{fF^{r-1}\}/QF^{r-1}$.

$$
\begin{aligned}
& Q|f-g|^r \leq Q\left\{|f-g|(2F)^{r-1}\right\} = 2^{r-1}R|f-g|QF^{r-1} \\
\Rightarrow\ & \|f-g\|_{Q,r} \leq 2^{1-1/r}(QF^{r-1})^{1/r}\|f-g\|_{R,1}^{1/r} \\
\Rightarrow\ & \frac{\|f-g\|_{Q,r}}{2\|F\|_{Q,r}} \leq \left(\frac{QF^{r-1}}{2QF^r}\right)^{1/r}\|f-g\|_{R,1}^{1/r} = \left(\frac{\|f-g\|_{R,1}}{2RF}\right)^{1/r}.
\end{aligned}
$$

- Hence, elementary manipulations yield

$$
\begin{aligned}
N\left(\epsilon 2\|F\|_{Q,r}, \mathcal{F}, L_r(Q)\right) &\leq N\left(\epsilon^r 2RF, \mathcal{F}, L_1(R)\right) \\
&\leq KV(\mathcal{F})(4e)^{V(\mathcal{F})}\left(\frac{1}{\epsilon}\right)^{r(V(\mathcal{F})-1)}.
\end{aligned}
$$

Outline

Convex hulls and VC hull classes

- The *symmetric convex hull* of a class of functions $\mathcal{F}$ is defined by

$$\operatorname{sconv}\mathcal{F} := \left\{ \sum_{i=1}^{m} \alpha_i f_i : \sum_{i=1}^{m} |\alpha_i| \leq 1,\ f_i \in \mathcal{F} \right\}.$$

- Similarly, the *convex hull* of $\mathcal{F}$ is defined by

$$\operatorname{conv}\mathcal{F} := \left\{ \sum_{i=1}^{m} \alpha_i f_i : \alpha_i > 0, \sum_{i=1}^{m} |\alpha_i| \leq 1,\ f_i \in \mathcal{F} \right\}.$$

- We use $\overline{\operatorname{conv}}\mathcal{F}$ and $\overline{\operatorname{sconv}}\mathcal{F}$ to denote the pointwise closures of $\operatorname{conv}\mathcal{F}$ and $\operatorname{sconv}\mathcal{F}$, respectively.
- A class of measurable functions $\mathcal{F}$ is called a *VC hull class* if $\mathcal{F} = \overline{\operatorname{sconv}}\mathcal{G}$ for some VC class $\mathcal{G}$.

Bound on entropy

The following theorem[3] gives an upper bound on the entropy of a VC hull class:

Theorem 3

Let Q be a probability measure on $(\mathcal{X}, \mathcal{A})$, and let $\mathcal{F}$ be a class of measurable functions with measurable square integrable envelope F such that $QF^2 < \infty$ and

$$N\left(\epsilon\|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right) \leq C\left(\frac{1}{\epsilon}\right)^V, \quad 0 < \epsilon < 1.$$

Then there exists a constant K depending only on C and V such that

$$\log N\left(\epsilon\|F\|_{Q,2}, \overline{\operatorname{conv}}\mathcal{F}, L_2(Q)\right) \leq K\left(\frac{1}{\epsilon}\right)^{2V/(V+2)}.$$

[3] *See Theorem 2.6.9 of VW (pp. 142-144) for the proof.*

Remarks

- The preceding theorem shows that the entropy of the convex hull of any polynomial class is of lower order than $(1/\epsilon)^r$ for some $r < 2$, which is just enough to ensure that the uniform entropy condition holds.
- Since $\text{sconv}\,\mathcal{F}$ is contained in the convex hull of $\mathcal{F} \cup \{-\mathcal{F}\} \cup \{0\}$, and the covering number of $\mathcal{F} \cup \{-\mathcal{F}\} \cup \{0\}$ is at most twice the covering number of $\mathcal{F}$ plus 1, the bound in Theorem 3 is valid for $\overline{\text{sconv}}\mathcal{F}$ as well.
- Thus, any VC hull class satisfies the uniform entropy condition.
- However, VC hull classes can be considerably larger than VC classes, so we do not have similar results for covering numbers.

Bound for VC hull classes

Finally, we have an easy corollary that gives precise bounds for entropy numbers of VC hull classes:

Corollary 4

For any VC hull class $\mathcal{F}$ of measurable functions and any probability measure Q,

$$\log N\left(\epsilon\|F\|_{Q,2}, \mathcal{F}, L_2(Q)\right) \leq K\left(\frac{1}{\epsilon}\right)^{2-2/V_m(\mathcal{F})}, \quad 0 < \epsilon < 1,$$

for a constant K that depends only on the VC index $V_m(\mathcal{F})$ of the VC subgraph class associated with $\mathcal{F}$.

Proof of the corollary

Proof.

Let $\mathcal{G}$ be the VC class associated with $\mathcal{F}$, i.e., $\mathcal{F} = \overline{\text{sconv}}\mathcal{G}$. Since $\mathcal{G}$ is contained in $\mathcal{F}$, F is also an envelope function for $\mathcal{G}$. By Theorem 2,

$$N\left(\epsilon\|F\|_{Q,2}, \mathcal{G}, L_2(Q)\right) \leq C\left(\frac{1}{\epsilon}\right)^{2(V_m(\mathcal{F})-1)}, \quad 0 < \epsilon < 1.$$

The desired bounds follow immediately from Theorem 3 with $V = 2(V_m(\mathcal{F}) - 1)$. □

Outline

Example: vector space of functions

Lemma 5

Any finite-dimensional vector space $\mathcal{F}$ of measurable functions $f : \mathcal{X} \to \mathbb{R}$ is VC-subgraph with $V(\mathcal{F}) \leq dim(\mathcal{F}) + 2$.

Proof.

Suppose $V(\mathcal{F}) > dim(\mathcal{F}) + 2$, then there exists a collection of $n = dim(\mathcal{F}) + 2$ points $(x_1, t_1), \ldots, (x_n, t_n)$ in $\mathcal{X} \times \mathbb{R}$ that can be shattered by the subgraphs of $\mathcal{F}$. By assumption, the vectors $(f(x_1) - t_1), \ldots, (f(x_n) - t_n)^T$, as f ranges over $\mathcal{F}$, are contained in a $dim(\mathcal{F}) + 1 = (n-1)$-dimensional subspace of $\mathbb{R}^n$.

Proof of Lemma 5 (cont.)

Proof.

There exists a vector a with at least one strictly positive coordinate that is orthogonal to this subspace. Thus,

$$\sum_{i:a_i>0} a_i \left(f\left(x_i\right) - t_i\right) = \sum_{i:a_i<0} \left(-a_i\right)\left(f\left(x_i\right) - t_i\right), \quad \text{for every } f \in \mathcal{F}. \quad (3)$$

Consider the subset $A = \{(x_i, t_i) : \ a_i > 0\}$ and its complement $A^c = \{(x_i, t_i) : \ a_i \leq 0\}$. Since A is picked out by the subgraphs of $\mathcal{F}$, it must be contained in the subgraph of some $f \in \mathcal{F}$, while A^c must be outside the subgraph of this f. Then the LHS of (3) is strictly positive while the RHS is nonpositive (contradiction!) □

Example: translates of monotone function

Lemma 6

The set of all translates $\{\psi(x-h) : h \in \mathbb{R}\}$ *of a fixed monotone function* $\psi : \mathbb{R} \to \mathbb{R}$ *is VC-subgraph of index 2.*

Proof.

Without loss of generality, we assume ψ is nondescreasing. For any $h_1 > h_2$, the subgraph of $x \mapsto \psi(x-h_1)$ is contained in the subgraph of $x \mapsto \psi(x-h_2)$. Any collection of sets with this property shatters no two-point set, thus has VC index 2. □

Example: monotone stochastic process

Lemma 7

Let $\{X(t) : t \in T\}$ be a monotone increasing stochastic process, where $T \subset \mathbb{R}$. Then X is VC-subgraph of index 2.

Proof.

Let $\mathcal{X}$ be the set of all increasing functions mapping T to $\mathbb{R}$. For any $t \in T$, define function $f_t : \mathcal{X} \to \mathbb{R}$ with $f_t(x) = x(t)$. We only need to show that the class of functions $\mathcal{F} = \{f_t : t \in T\}$ is VC of index 2. For any $t_1 < t_2$, the subgraph of f_{t_1} is contained in the subgraph of f_{t_2}. $\square$

Build VC classes from basic VC classes of sets

Lemma 8

Let $\mathcal{C}$ and $\mathcal{D}$ be VC classes of sets in a set $\mathcal{X}$ and $\mathcal{E}$ a VC class of sets in $\mathcal{Y}$. Also, let $\phi : \mathcal{X} \to \mathcal{Y}$ and $\psi : \mathcal{Z} \to \mathcal{X}$ be fixed functions. Then[a]

(i) $\mathcal{C}^c = \{C^c : C \in \mathcal{C}\}$ *is VC with* $V(\mathcal{C}^c) = V(\mathcal{C})$;

(ii) $\mathcal{C} \sqcap \mathcal{D} = \{C \cap D : C \in \mathcal{C}, D \in \mathcal{D}\}$ *is VC with index* $\leq V(\mathcal{C}) + V(\mathcal{D}) - 1$;

(iii) $\mathcal{C} \sqcup \mathcal{D} = \{C \cup D : C \in \mathcal{C}, D \in \mathcal{D}\}$ *is VC with index* $\leq V(\mathcal{C}) + V(\mathcal{D}) - 1$;

(iv) $\mathcal{D} \times \mathcal{E}$ *is VC in* $\mathcal{X} \times \mathcal{Y}$ *with VC index* $\leq V(\mathcal{D}) + V(\mathcal{E}) - 1$;

(v) $\phi(\mathcal{C})$ *is VC with index* $V(\mathcal{C})$ *if* ϕ *is one-to-one;*

(vi) $\psi^{-1}(\mathcal{C})$ *is VC with index* $\leq V(\mathcal{C})$.

[a] *See Lemma 9.7 in Kosorok (pp. 159-160) for the proof.*

Build VC classes from basic VC classes of functions

Lemma 9

Let $\mathcal{F}$ and $\mathcal{G}$ be VC subgraph classes of functions on a set $\mathcal{X}$ and $g : \mathcal{X} \to \mathbb{R}$, $\phi : \mathbb{R} \to \mathbb{R}$, and $\psi : \mathcal{Z} \to \mathcal{X}$ fixed functions. Then[a]

(i) $\mathcal{F} \wedge \mathcal{G} = \{f \wedge g : f \in \mathcal{F}, g \in \mathcal{G}\}$ *is VC-subgraph with index* $\leq V(\mathcal{F}) + V(\mathcal{G}) - 1$*;*

(ii) $\mathcal{F} \vee \mathcal{G} = \{f \vee g : f \in \mathcal{F}, g \in \mathcal{G}\}$ *is VC-subgraph with index* $\leq V(\mathcal{F}) + V(\mathcal{G}) - 1$*;*

(iii) $\{\mathcal{F} > 0\} = \{\{f > 0\} : f \in \mathcal{F}\}$ *is VC with index* $V(\mathcal{F})$*;*

(iv) $-\mathcal{F}$ *is VC-subgraph with index* $V(\mathcal{F})$*;*

(v) $\mathcal{F} + g = \{f + g : f \in \mathcal{F}\}$ *is VC-subgraph with index* $V(\mathcal{F})$*;*

(vi) $\mathcal{F} \cdot g = \{fg : f \in \mathcal{F}\}$ *is VC-subgraph with index* $\leq 2V(\mathcal{F}) - 1$*;*

(vii) $\mathcal{F} \circ \psi = \{f(\psi) : f \in \mathcal{F}\}$ *is VC-subgraph with index* $\leq V(\mathcal{F})$*;*

(viii) $\phi \circ \mathcal{F}$ *is VC-subgraph with index* $\leq V(\mathcal{F})$ *for monotone* ϕ*.*

---

[a] *See Proof of Lemma 9.9 in Kosorok (pp. 174-175) for the proof.*

Relation between $\mathcal{C}$ and $1\{\mathcal{C}\}$

Lemma 10

$1\{\mathcal{C}\}$ is VC-subgraph if and only if $\mathcal{C}$ is a VC class of sets. Moreover, their respective VC indices are equal.

Proof.

Let $\text{sub}(C)$ denote the subgraph of the indicator function $1\{C\}$ for any $C \in \mathcal{C}$. Let $\mathcal{D}$ denote the collection of all $\text{sub}(C)$. We can easily verify that for any $x \in \mathcal{X}$ and $C \in \mathcal{C}$,

$$x \in C \iff (x, 0) \in \text{sub}(C).$$

Thus, for any $n > 0$ and any $x_1, \ldots, x_n \in \mathcal{X}$, $\{x_1, \ldots, x_n\}$ can be shattered by $\mathcal{C}$ if and only if $\{(x_1, 0), \ldots, (x_n, 0)\}$ can be shattered by $\mathcal{D}$. The conclusion then follows. □

Result of monotone functions within $[0, 1]$

Lemma 11

The set $\mathcal{F}$ of all monotone functions $f : \mathbb{R} \to [0, 1]$ satisfies

$$\log N(\epsilon, \mathcal{F}, L_2(Q)) \leq \frac{K}{\epsilon}, \quad 0 < \epsilon < 1$$

for a universal constant K and any probability measure Q.

Proof.

We first consider the set $\mathcal{F}_+$ of all monotone increasing functions $f : \mathbb{R} \to [0, 1]$. Any $f \in \mathcal{F}_+$ is the pointwise limit of the sequence

$$f_m = \sum_{i=1}^{m} \frac{1}{m} 1 \left\{ f > \frac{i}{m} \right\}.$$

Proof of Lemma 11 (cont.)

Proof.

Thus, $\mathcal{F}_+ = \overline{\text{conv}} 1\{\mathcal{G}\}$ for some class $\mathcal{G}$ of sets of the form $\{f > t\}$, with f ranging over $\mathcal{F}_+$ and t over $\mathbb{R}$. Since f is increasing, $\mathcal{G}$ is contained in the collection of intervals $\{(t, +\infty) : t \in \mathbb{R}\} \cup \{[t, +\infty) : t \in \mathbb{R}\}$, which is VC of index 2. Thus, $\mathcal{G}$ is also VC of index 2. By Corollary 4,

$$\log N(\epsilon, \mathcal{F}_+, L_2(Q)) \leq \frac{K_0}{\epsilon}, \quad 0 < \epsilon < 1.$$

Any monotone decreasing function $g : \mathbb{R} \to [0, 1]$ can be written as $1 - f$ for some $f \in \mathcal{F}_+$. Thus, for $0 < \epsilon < 1$,

$$\log N(\epsilon, \mathcal{F}, L_2(Q)) = \log(2) + \log N(\epsilon, \mathcal{F}_+, L_2(Q)) \leq \frac{K}{\epsilon}.$$

□