BUEI Classes & Bracketing Entropy

Yilun Li

Aug 4th, 2021

Table of Contents

Recall

- For a class of measurable functions $\mathcal{F}$, with envelope F, the uniform entropy integral

$$J(\delta, \mathcal{F}, L_2) = \int_0^\delta \sqrt{\sup_{\mathrm{Q}} \log N(\epsilon ||F||_{\mathrm{Q},2}, \mathcal{F}, L_2(\mathrm{Q}))} d\epsilon, \tag{1}$$

  where the supremum is taken over all finitely discrete probability measures Q with $||F||_{\mathrm{Q},2} > 0$.
- The covering number $N(\epsilon, \mathcal{F}, L_2(\mathrm{Q}))$ is the minimum number of ϵ-balls in $L_2(\mathrm{Q})$ needed to cover $\mathcal{F}$.
- Note: $J(\delta, \mathcal{F}, L_2)$ depends on the choice of envelope F.

Definition - BUEI Classes

A class of function $\mathcal{F}$ has bounded uniform entropy integral (BUEI) with envelope F (or is BUEI with envelope F), if $J(1, \mathcal{F}, L_2) < \infty$ for that particular choice of envelope.

- Recall (Theorem 9.3): There exists a universal constant $K < \infty$ such that, for any VC-class of measurable functions $\mathcal{F}$ with integrable envelope F, any $r \geq 1$, any probability measure Q with $||F||_{\mathrm{Q},r} > 0$, and any $0 < \epsilon < 1$,

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

- Based on this theorem, we can verify that a VC-class $\mathcal{F}$ is BUEI with any envelope.

Building BUEI Classes from Other Known BUEI Classes

The following lemma is useful for building BUEI classes from other BUEI classes, and is also related to Donsker preservation results in later sections.

Lemma 1

Let $\mathcal{F}_1, ..., \mathcal{F}_k$ be BUEI classes with respective envelopes $F_1, ..., F_k$, and let $\phi : \mathbb{R}^k \to \mathbb{R}$ satisfy

$$\left|\phi \circ \mathbf{f}(\mathbf{x}) - \phi \circ \mathbf{g}(\mathbf{x})\right|^2 \leq c^2 \sum_{j=1}^{k} \left[f_j(\mathbf{x}) - g_j(\mathbf{x})\right]^2, \qquad (3)$$

for $\forall \mathbf{f} = (f_1, ..., f_k), \mathbf{g} = (g_1, ..., g_k) \in \mathcal{F}_1 \times ... \times \mathcal{F}_k$ and $\forall \mathbf{x}$ for a constant $0 < c < \infty$. Then the class $\phi \circ (\mathcal{F}_1, ..., \mathcal{F}_k) = \{\phi(f_1, ..., f_k) : (f_1, ..., f_k) \in \mathcal{F}_1 \times ... \times \mathcal{F}_k\}$ is BUEI with envelope $H = |\phi(\mathbf{f}_0)| + c \sum_{j=1}^{k}(|f_{0j}| + F_j)$, where $\mathbf{f}_0 = (f_{01}, ..., f_{0k})$ is any function in $\mathcal{F}_1 \times ... \times \mathcal{F}_k$.

Proof of Lemma 1

- For any fixed $\epsilon > 0$ and any finitely discrete probability measure $\mathbb{Q}$, let $\mathbf{f}, \mathbf{g} \in \mathcal{F}_1 \times ... \times \mathcal{F}_k$ satisfy $||f_j - g_j||_{\mathbb{Q},2} \leq \epsilon ||F_j||_{\mathbb{Q},2}$ $(\forall j = 1, ..., k)$.
- Thus, for any choice of $\mathbf{f}_0$,

$$\begin{aligned} ||\phi \circ \mathbf{f} - \phi \circ \mathbf{g}||_{\mathbb{Q},2} &\leq \left\{ \int_{\mathcal{X}} c^2 \sum_{j=1}^{k} \left[f_j(x) - g_j(x) \right]^2 dx \right\}^{\frac{1}{2}} \\ &= c \sqrt{\sum_{j=1}^{k} ||f_j - g_j||_{\mathbb{Q},2}^2} \\ &\leq c\epsilon \sqrt{\sum_{j=1}^{k} ||F_j||_{\mathbb{Q},2}^2} \leq \epsilon ||H||_{\mathbb{Q},2} \end{aligned} \tag{4}$$

Proof of Lemma 1 (cont.)

- So we can verify that,

$$N\left(\epsilon||H||_{\mathrm{Q},2},\ \phi\circ(\mathcal{F}_1,\ldots,\mathcal{F}_k),\ L_2(\mathrm{Q})\right)\le\prod_{j=1}^{k}N\left(\epsilon||F_j||_{\mathrm{Q},2},\mathcal{F}_j,L_2(\mathrm{Q})\right)\tag{5}$$

$$\sup_{\mathrm{Q}}N\left(\epsilon||H||_{\mathrm{Q},2},\ \phi\circ(\mathcal{F}_1,\ldots,\mathcal{F}_k),\ L_2(\mathrm{Q})\right)\le\prod_{j=1}^{k}\sup_{\mathrm{Q}}N\left(\epsilon||F_j||_{\mathrm{Q},2},\mathcal{F}_j,L_2(\mathrm{Q})\right)\tag{6}$$

- And we can get the desired results by taking logs, square roots, and then integrating both sides with respect to ϵ on [0, 1].

Useful Consequences of Lemma 1

Lemma 2

Let $\mathcal{F}$ and $\mathcal{G}$ be BUEI classes with respective envelopes F and G, and let $\phi_0 : \mathbb{R} \to \mathbb{R}$ be a Lipschitz continuous function with Lipschitz constant $0 < c < \infty$. Then

(1) $\mathcal{F} \wedge \mathcal{G} = \{f \wedge g : f \in \mathcal{F}, g \in \mathcal{G}\}$ is BUEI with envelope $F + G$;

(2) $\mathcal{F} \vee \mathcal{G} = \{f \vee g : f \in \mathcal{F}, g \in \mathcal{G}\}$ is BUEI with envelope $F + G$;

(3) $\mathcal{F} + \mathcal{G} = \{f + g : f \in \mathcal{F}, g \in \mathcal{G}\}$ is BUEI with envelope $F + G$;

(4) $\phi_0(\mathcal{F}) = \{\phi_0(f) : f \in \mathcal{F}\}$ is BUEI with envelope $|\phi_0(f_0)| + c(|f_0| + F)$, provided $f_0 \in \mathcal{F}$.

Proof (Take (1) as an example): In Lemma 1, let $k = 2$, $c = 1$, and let

$$\phi : \mathbb{R}^2 \to \mathbb{R}, \ (x_1, x_2) \mapsto \min\{x_1, x_2\}. \tag{7}$$

Then we can verify the inequality (3) and get the desired result.

Product of BUEI Classes

- Most of the BUEI results in this section have parallel Donsker preservation results, but the product of BUEI classes is an exception, which is a primary justification for the use of BUEI techniques.
- In the theorem below, we prove that the product of two BUEI classes is still BUEI.

Theorem 3

Let $\mathcal{F}$ and $\mathcal{G}$ be BUEI classes with respective envelopes F and G. Then

$$\mathcal{F} \cdot \mathcal{G} = \{fg : f \in \mathcal{F}, g \in \mathcal{G}\}$$

is BUEI with envelope FG.

Proof of Theorem 3

- For any fixed $\epsilon > 0$ and any finitely discrete probability measure $\tilde{\mathrm{Q}}$ with $||FG||_{\tilde{\mathrm{Q}},2} > 0$, let $d\mathrm{Q}^* = \frac{G^2}{||G||^2_{\tilde{\mathrm{Q}},2}} d\tilde{\mathrm{Q}}$.
- So Q^* is a finitely discrete probability measure with $||F||_{\mathrm{Q}^*,2} > 0$.
- Let $f_1, f_2 \in \mathcal{F}$ satisfy $||f_1 - f_2||_{\mathrm{Q}^*,2} \leq \epsilon ||F||_{\mathrm{Q}^*,2}$. Then,

$$\epsilon \geq \frac{||f_1 - f_2||_{\mathrm{Q}^*,2}}{||F||_{\mathrm{Q}^*,2}} = \frac{||(f_1 - f_2)G||_{\tilde{\mathrm{Q}},2}}{||FG||_{\tilde{\mathrm{Q}},2}} \tag{8}$$

- Thus if we consider $\mathcal{F} \cdot G = \{fG : f \in \mathcal{F}\}$, then

$$\begin{aligned} N\left(\epsilon ||FG||_{\tilde{\mathrm{Q}},2},\ \mathcal{F} \cdot G,\ L_2(\tilde{\mathrm{Q}})\right) &\leq N\left(\epsilon ||F||_{\mathrm{Q}^*,2},\ \mathcal{F},\ L_2(\mathrm{Q}^*)\right) \\ &\leq \sup_{\mathrm{Q}} N\left(\epsilon ||F||_{\mathrm{Q},2},\ \mathcal{F},\ L_2(\mathrm{Q})\right), \end{aligned} \tag{9}$$

where the supremum is taken over all finitely discrete probability measures Q for which $||F||_{\mathrm{Q},2} > 0$.

Proof of Theorem 3 (cont.)

- Since $\tilde{\mathrm{Q}}$ in the inequality (9) is arbitrary, we have

$$\sup_{\mathrm{Q}} N\left(\epsilon||FG||_{\mathrm{Q},2},\ \mathcal{F}\cdot G,\ L_2(\mathrm{Q})\right) \leq \sup_{\mathrm{Q}} N\left(\epsilon||F||_{\mathrm{Q},2},\ \mathcal{F},\ L_2(\mathrm{Q})\right), \tag{10}$$

where the supremum on the LHS is taken over all finitely discrete probability measures Q with $||FG||_{\mathrm{Q},2} > 0$, and the supremum on the RHS is taken over all finitely discrete probability measures Q with $||F||_{\mathrm{Q},2} > 0$.

- Similarly,

$$\sup_{\mathrm{Q}} N\left(\epsilon||FG||_{\mathrm{Q},2},\ F\cdot \mathcal{G},\ L_2(\mathrm{Q})\right) \leq \sup_{\mathrm{Q}} N\left(\epsilon||G||_{\mathrm{Q},2},\ \mathcal{G},\ L_2(\mathrm{Q})\right), \tag{11}$$

where $F\cdot\mathcal{G} = \{Fg : g \in \mathcal{G}\}$, and the supremum on the RHS is taken over all finitely discrete probability measures Q with $||G||_{\mathrm{Q},2} > 0$.

Proof of Theorem 3 (cont.)

- Note that for any $f_1, f_2 \in \mathcal{F}$ and any $g_1, g_2 \in \mathcal{G}$,

$$\begin{aligned} ||f_1 g_1 - f_2 g_2||_{\mathrm{Q},2} &\leq ||f_1 g_1 - f_2 g_1||_{\mathrm{Q},2} + ||f_2 g_1 - f_2 g_2||_{\mathrm{Q},2} \\ &\leq ||(f_1 - f_2) G||_{\mathrm{Q},2} + ||F(g_1 - g_2)||_{\mathrm{Q},2} \end{aligned} \tag{12}$$

- So we can prove that

$$\begin{aligned} &\sup_{\mathrm{Q}} N\left(\epsilon ||FG||_{\mathrm{Q},2},\ \mathcal{F} \cdot \mathcal{G},\ L_2(\mathrm{Q})\right) \\ &\leq \sup_{\mathrm{Q}} N\left(\frac{\epsilon}{2} ||FG||_{\mathrm{Q},2},\ \mathcal{F} \cdot G,\ L_2(\mathrm{Q})\right) * \sup_{\mathrm{Q}} N\left(\frac{\epsilon}{2} ||FG||_{\mathrm{Q},2},\ F \cdot \mathcal{G},\ L_2(\mathrm{Q})\right) \\ &\leq \sup_{\mathrm{Q}} N\left(\frac{\epsilon}{2} ||F||_{\mathrm{Q},2},\ \mathcal{F},\ L_2(\mathrm{Q})\right) * \sup_{\mathrm{Q}} N\left(\frac{\epsilon}{2} ||G||_{\mathrm{Q},2},\ \mathcal{G},\ L_2(\mathrm{Q})\right) \end{aligned} \tag{13}$$

- Since $\mathcal{F}$ and $\mathcal{G}$ are BUEI with respective envelops F and G, we can get the desired results by taking logs, square roots, and then integrating both sides with respect to ϵ on [0, 1].

Point-wise Measurability (PM) and BUEI

- Conclusions about PM and BUEI can help to derive Donsker results in later sections.
- Recall: A class $\mathcal{F}$ of measurable functions is point-wise measurable if there exists a countable subset $\mathcal{G} \subseteq \mathcal{F}$ s.t. for $\forall f \in \mathcal{F}$, $\exists$ a sequence $\{g_m\} \subseteq G$ with $g_m(x) \to f(x)$ $(\forall x \in \mathcal{X})$.
- Recall (Lemma 8.10): Let $\mathcal{F}_1, ..., \mathcal{F}_k$ be PM classes of real functions on $\mathcal{X}$, and let $\phi : \mathbb{R}^k \to \mathbb{R}$ be continuous. Then the class $\phi \circ (\mathcal{F}_1, ..., \mathcal{F}_k) = \{\phi(f_1, ..., f_k) : (f_1, ..., f_k) \in \mathcal{F}_1 \times ... \times \mathcal{F}_k\}$ is PM.

Point-wise Measurability (PM) and BUEI

Theorem 4

Let the classes $\mathcal{F}_1, ..., \mathcal{F}_k$ be both PM and BUEI with respective envelopes $F_1, ..., F_k$, and let $\phi : \mathbb{R}^k \to \mathbb{R}$ satisfy inequality (3)

$$\left|\phi \circ \mathbf{f}(\mathbf{x}) - \phi \circ \mathbf{g}(\mathbf{x})\right|^2 \leq c^2 \sum_{j=1}^{k} \left[f_j(\mathbf{x}) - g_j(\mathbf{x})\right]^2,$$

for $\forall \mathbf{f} = (f_1, ..., f_k), \mathbf{g} = (g_1, ..., g_k) \in \mathcal{F}_1 \times ... \times \mathcal{F}_k$ and $\forall \mathbf{x}$ for a constant $0 < c < \infty$. Then the class $\phi \circ (\mathcal{F}_1, ..., \mathcal{F}_k) = \{\phi(f_1, ..., f_k) : (f_1, ..., f_k) \in \mathcal{F}_1 \times ... \times \mathcal{F}_k\}$ is both PM and BUEI with envelope $H = |\phi(\mathbf{f}_0)| + c \sum_{j=1}^{k}(|f_{0j}| + F_j)$, where $\mathbf{f}_0 = (f_{01}, ..., f_{0k})$ is any function in $\mathcal{F}_1 \times ... \times \mathcal{F}_k$.

Note: This is a result of Lemma 8.10 and Lemma 1.

Point-wise Measurability (PM) and BUEI

Theorem 5

Let $\mathcal{F}$ and $\mathcal{G}$ be both PM and BUEI with respective envelopes F and G, and let $\phi_0 : \mathbb{R} \to \mathbb{R}$ be a Lipschitz continuous function with Lipschitz constant $0 < c < \infty$. Then

(1) $\mathcal{F} \cup \mathcal{G}$ is PM and BUEI with envelop $F \vee G$.

(2) $\mathcal{F} \wedge \mathcal{G} = \{f \wedge g : f \in \mathcal{F}, g \in \mathcal{G}\}$ is PM and BUEI with envelope $F + G$;

(3) $\mathcal{F} \vee \mathcal{G} = \{f \vee g : f \in \mathcal{F}, g \in \mathcal{G}\}$ is PM and BUEI with envelope $F + G$;

(4) $\mathcal{F} + \mathcal{G} = \{f + g : f \in \mathcal{F}, g \in \mathcal{G}\}$ is PM and BUEI with envelope $F + G$;

(5) $\mathcal{F} \cdot \mathcal{G} = \{fg : f \in \mathcal{F}, g \in \mathcal{G}\}$ is PM and BUEI with envelope FG;

(6) $\phi_0(\mathcal{F})$ is PM and BUEI with envelope $|\phi_0(f_0)| + c(|f_0| + F)$, provided $f_0 \in \mathcal{F}$.

Proof of Theorem 5:

1)
- Since $\mathcal{F}$ and $\mathcal{G}$ are PM, there exists countable subsets $\mathcal{F}_0 \subseteq \mathcal{F}$, $\mathcal{G}_0 \subseteq \mathcal{G}$ s.t. for $\forall f \in \mathcal{F}, g \in \mathcal{G}$, $\exists$ $\{h_{1m}\}_{m=1}^{\infty} \subseteq \mathcal{F}_0$, $\{h_{2m}\}_{m=1}^{\infty} \subseteq \mathcal{G}_0$ such that $h_{1m}(x) \to f(x)$, $h_{2m}(x) \to g(x)$ for $\forall x \in \mathcal{X}$.
- So for $\mathcal{F} \cup \mathcal{G}$, there exists a countable subset $\mathcal{F}_0 \cup \mathcal{G}_0$ s.t. for $\forall f \in \mathcal{F} \cup \mathcal{G}$, $\exists$ $\{h_m\}_{m=1}^{\infty} \subseteq \mathcal{F}_0 \cup \mathcal{G}_0$ such that $h_m(x) \to f(x)$ for $\forall x \in \mathcal{X}$, which means that $\mathcal{F} \cup \mathcal{G}$ is PM.
- Claim: For any $\epsilon > 0$ and any finitely discrete probability measure $\mathbb{Q}$,
  $N\left(\epsilon||F \vee G||_{\mathbb{Q},2},\ \mathcal{F} \cup \mathcal{G},\ L_2(\mathbb{Q})\right) \leq N\left(\epsilon||F||_{\mathbb{Q},2},\ \mathcal{F},\ L_2(\mathbb{Q})\right) + N\left(\epsilon||G||_{\mathbb{Q},2},\ \mathcal{G},\ L_2(\mathbb{Q})\right)$.
  Proof of claim:
  - Suppose $N_1 = N\left(\epsilon||F||_{\mathbb{Q},2},\ \mathcal{F},\ L_2(\mathbb{Q})\right)$ and $B(f_i, \epsilon||F||_{\mathbb{Q},2})$ $(i = 1, \ldots, N_1)$ are the $\epsilon||F||_{\mathbb{Q},2}$-balls that cover $\mathcal{F}$; $N_2 = N\left(\epsilon||G||_{\mathbb{Q},2},\ \mathcal{G},\ L_2(\mathbb{Q})\right)$ and $B(g_j, \epsilon||G||_{\mathbb{Q},2})$ $(j = 1, \ldots, N_2)$ are the $\epsilon||G||_{\mathbb{Q},2}$-balls that cover $\mathcal{G}$.
  - So $\mathcal{F} \cup \mathcal{G}$ can be covered by $(N_1 + N_2)$ $\epsilon||F \vee G||_{\mathbb{Q},2}$-balls: $B(f_i, \epsilon||F \vee G||_{\mathbb{Q},2})$ $(i = 1, \ldots, N_1)$ and $B(g_j, \epsilon||F \vee G||_{\mathbb{Q},2})$ $(j = 1, \ldots, N_2)$. The claim is proved.
- After taking the supremum over the appropriate subsets, log-transform, square-root transform, and then integrating both sides with respect to ϵ on $[0, 1]$, we can prove that $\mathcal{F} \cup \mathcal{G}$ is BUEI with envelop $F \vee G$.

2)-6) are the results of Lemma 8.10, Lemma 2 and Theorem 3.

Note:

- By (Theorem 8.19), (Proposition 8.11) and the above results, If a class of measurable functions $\mathcal{F}$ is both PM and BUEI with envelope F, then $\mathcal{F}$ is P-Donsker whenever $P^* F^2 < \infty$.

Bracketing Entropy

Recall:

(1) An ϵ-bracket in $L_r(P)$ is formed by a pair of functions $l, u \in L_r(P)$ with $P(l(X) \leq u(X)) = 1$ and $||l - u||_{r,P} \leq \epsilon$.

(2) A function $f \in \mathcal{F}$ lies in the bracket l, u if $P(l(X) \leq f(X) \leq u(X)) = 1$.

(3) The bracketing number $N_{[]}(\epsilon, \mathcal{F}, L_r(P))$ is the minimum number of ϵ-brackets in $L_r(P)$ needed to ensure that every $f \in \mathcal{F}$ lies in at least one bracket.

(4) The logarithm of the bracketing number is the entropy with bracketing.

Bracketing Numbers and Covering Numbers

First we mentioned that bracketing numbers are generally larger than covering numbers

Lemma 6

Let $\mathcal{F}$ be any class of real function on X and $||\cdot||$ be any norm on $\mathcal{F}$. Then for any $\epsilon > 0$,

$N(\epsilon,\ \mathcal{F},\ ||\cdot||) \leq N_{[]}(\epsilon,\ \mathcal{F},\ ||\cdot||)$

Proof:

- For any fixed $\epsilon > 0$, and let $\mathcal{B}$ be the collection of ϵ-brackets that covers $\mathcal{F}$.
- From each bracket $B \in \mathcal{B}$, take a function $g_B \in B \cap \mathcal{F}$ to form a finite collection of functions $\mathcal{G} \subseteq \mathcal{F}$ of the same cardinality as $\mathcal{B}$, consisting of one function from each bracket in $\mathcal{B}$.
- By definition, any $f \in \mathcal{F}$ lies in at least one ϵ-brackets from $\mathcal{B}$, so there exists an ϵ-bracket $B \in \mathcal{B}$ such that $||f - g_B|| \leq \epsilon$.
- So for any $f \in \mathcal{F}$, there exists a $g_B \in \mathcal{G}$ such that $||f - g_B|| \leq \epsilon$.
- Thus, $\mathcal{G}$ is an ϵ-cover of $\mathcal{F}$ of the same cardinality as $\mathcal{B}$, and the conclusion is proved.

Class of Smooth Functions

- We consider the class of smooth functions on a bounded set $\mathcal{X} \subseteq \mathbb{R}^d$.
- For any vector $\mathbf{k} = (k_1, ..., k_d)$ of non-negative integers, define the differential operator

$$D^{\mathbf{k}} = \frac{\partial^{|\mathbf{k}|}}{\partial x_1^{k_1} \cdots \partial x_d^{k_d}},$$

where $|\mathbf{k}| = \sum_{i=1}^{d} k_i$. And when $\mathbf{k} = \mathbf{0}$, define $D^{\mathbf{k}}$ as the identity map.

- For any function $f : \mathcal{X} \to \mathbb{R}$ and $\alpha > 0$, define the norm

$$||f||_\alpha = \max_{\mathbf{k}:|\mathbf{k}|\leq\lfloor\alpha\rfloor} \sup_{\mathbf{x}} |D^{\mathbf{k}} f(\mathbf{x})| + \max_{\mathbf{k}:|\mathbf{k}|=\lfloor\alpha\rfloor} \sup_{\mathbf{x},\mathbf{y}} \frac{|D^{\mathbf{k}} f(\mathbf{x}) - D^{\mathbf{k}} f(\mathbf{y})|}{||\mathbf{x}-\mathbf{y}||^{\alpha-\lfloor\alpha\rfloor}} \tag{14}$$

where the suprema are taken over $\mathbf{x} \neq \mathbf{y}$ in the interior of $\mathcal{X}$.

- Let $C_M^\alpha(\mathcal{X})$ be the set of all continuous functions $f : \mathcal{X} \to \mathbb{R}$ with $||f||_\alpha \leq M$.
- For a set A in a metric space $(\mathcal{X}, d)$, define its diameter as $\text{diam}(A) = \sup_{\mathbf{x},\mathbf{y}\in A} d(\mathbf{x}, \mathbf{y})$.

Class of Smooth Functions

Theorem 7

Let $\mathcal{X} \subseteq \mathbb{R}^d$ be bounded and convex with nonempty interior. There exists a constant $K < \infty$ depending only on α, diam$(\mathcal{X})$, and dimension d such that

$$\log N_{[]}\left(\epsilon, C_1^{\alpha}(\mathcal{X}), L_r(\mathrm{Q})\right) \leq K\epsilon^{-d/\alpha}, \tag{15}$$

for every $r \geq 1$, $\epsilon > 0$, and any probability measure Q on $\mathbb{R}^d$. [a]

[a] Proof can be found in Corollary 2.7.2 from *Weak Convergence and Empirical Processes: With Applications in Statistics.* by van der Vaart, A. W. and Wellner, J. A. (1996) (hereafter abbreviated [VW])

We'll talk about the results for covering numbers based on the uniform norm, and also its relationship with bracketing entropy.

Theorem 8

For a compact, convex subset $C \subseteq \mathbb{R}^d$, let $\mathcal{F}$ be the class of all convex functions $f : C \to [0,1]$ with $|f(\mathbf{x}) - f(\mathbf{y})| \leq L||\mathbf{x} - \mathbf{y}||$ $(\forall \mathbf{x}, \mathbf{y} \in C)$. And for some integer $m \geq 1$, let $\mathcal{G} = \{g : [0,1] \to [0,1] | \int_0^1 |g^{(m)}(x)|^2 dx \leq 1\}$, where $g^{(m)}$ denotes the m_{th} derivative of g. Then

$$\log N(\epsilon, \mathcal{F}, ||\cdot||_\infty) \leq K(L+1)^{d/2}\epsilon^{-d/2} \tag{16}$$

$$\log N(\epsilon, \mathcal{G}, ||\cdot||_\infty) \leq M\epsilon^{-1/m}, \tag{17}$$

where $K < \infty$ is a constant depends only on d and C; and the constant M depends only on m. [a]

[a] Inequality (16) is proved in Corollary 2.7.10 of [VW]; and inequality (17) is proved in Theorem 2.4 of van de Geer (2000).

Lemma 9

For any norm $||\cdot||$ dominated by $||\cdot||_\infty$, any class of functions $\mathcal{F}$, and any $\epsilon > 0$,

$$N_{[]}\left(2\epsilon, \mathcal{F}, ||\cdot||\right) \leq N\left(\epsilon, \mathcal{F}, ||\cdot||_\infty\right). \tag{18}$$

Proof:

- Suppose $B(f_i, \epsilon)$ $(i = 1, ..., N_1)$ is the $N_1 = N\left(\epsilon, \mathcal{F}, ||\cdot||_\infty\right)$ ϵ-balls that cover $(\mathcal{F}, ||\cdot||_\infty)$. Consider $\mathcal{G} = \{[f_i - \epsilon, f_i + \epsilon] : i = 1, ..., N_1\}$.
- By definition, for $\forall f \in \mathcal{F}$, $\exists i$ s.t. $||f - f_i||_\infty \leq \epsilon$.
- So for $\forall \mathbf{x} \in \mathcal{X}$, $f_i(\mathbf{x}) - \epsilon \leq f(\mathbf{x}) \leq f_i(\mathbf{x}) + \epsilon$, which means that f is in the bracket $[f_i(\mathbf{x}) - \epsilon, f_i(\mathbf{x}) + \epsilon]$.
- And the size of bracket $||(f_i(\mathbf{x}) + \epsilon) - (f_i(\mathbf{x}) - \epsilon)|| = ||2\epsilon|| \leq ||2\epsilon||_\infty = 2\epsilon$.
- Thus, elements in $\mathcal{G}$ form a 2ϵ-bracket that covers $(\mathcal{F}, ||\cdot||)$. The proof is done.

- Next, we move on to a more general Lipschitz structure.
- Consider the class of function with a form of $\mathcal{F} = \{f_t : t \in \mathcal{T}\}$ where

$$|f_s(\mathbf{x}) - f_t(\mathbf{x})| \leq d(s,t)F(\mathbf{x}) \tag{19}$$

for some metric d on $\mathcal{T}$, some real function F on the sample space $\mathcal{X}$, and for all $\mathbf{x} \in \mathcal{X}$.

- This kind of Lipschitz structure appears in many settings. For example, consider the LAD regression model:
  - $Y = \boldsymbol{\theta}^T \mathbf{U} + e$, where e has median 0, and $\mathbf{U}$, $\boldsymbol{\theta}$ are constrained to known compact sets $\mathcal{U}, \Theta \subseteq \mathbb{R}^p$.
  - Suppose we have n i.i.d. random vectors $\mathbf{U}_1, \ldots, \mathbf{U}_n \in \mathbb{R}^p$, i.i.d. unobserved random errors $e_1, \ldots, e_n$.
  - The observed data $\{\mathbf{X}_i = (Y_i, \mathbf{U}_i) : i = 1, \ldots, n\}$

- Estimation of the true parameter value $\boldsymbol{\theta}_0$ is obtained by minimizing

$$\frac{1}{n}\sum_{i=1}^{n}|Y_i - \boldsymbol{\theta}^T\mathbf{U}_i| - \frac{1}{n}\sum_{i=1}^{n}|Y_i - \boldsymbol{\theta}_0^T\mathbf{U}_i| = P_n m_{\boldsymbol{\theta}} \tag{20}$$

where $m_{\boldsymbol{\theta}}(\mathbf{X}) = |Y - \boldsymbol{\theta}^T\mathbf{U}| - |Y - \boldsymbol{\theta}_0^T\mathbf{U}|$, and P_n is the empirical measure.

- Consider the class of function $\mathcal{F} = \{m_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$.
- Let $\mathcal{T} = \Theta$, $d(\mathbf{s}, \mathbf{t}) = ||\mathbf{s} - \mathbf{t}||$, $F(\mathbf{x}) = F(y, \mathbf{u}) = ||\mathbf{u}||$, then

$$\begin{aligned} |m_{\boldsymbol{\theta}_1}(\mathbf{x}) - m_{\boldsymbol{\theta}_2}(\mathbf{x})| &= \big||Y - \boldsymbol{\theta}_1^T\mathbf{U}| - |Y - \boldsymbol{\theta}_2^T\mathbf{U}|\big| \\ &\le |(Y - \boldsymbol{\theta}_1^T\mathbf{U}) - (Y - \boldsymbol{\theta}_2^T\mathbf{U})| = |(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2)^T\mathbf{U}| \\ &\le ||\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2||||\mathbf{U}|| = d(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)F(\mathbf{x}) \end{aligned} \tag{21}$$

- So $\mathcal{F} = \{m_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \Theta\}$ is a class of function that satisfies (19).

Class of Lipschitz Functions

The following theorem shows that the bracketing numbers for a general class of function $\mathcal{F}$ satisfying (19) are bounded by the covering numbers for the associated index set $\mathcal{T}$.

Theorem 10

Suppose the class of functions $\mathcal{F} = \{f_t : t \in \mathcal{T}\}$ satisfies (19) for every $\mathbf{s}, \mathbf{t} \in \mathcal{T}$ and some fixed function F. Then for any norm $||\cdot||$,

$$N_{[]}(2\epsilon||F||, \mathcal{F}, ||\cdot||) \leq N(\epsilon, \mathcal{T}, d) \tag{22}$$

Proof:

- Let $B(t_i, \epsilon)$ $(i = 1, ..., N_1)$ be the $N_1 = N(\epsilon, \mathcal{T}, d)$ ϵ-balls that covers $(\mathcal{T}, d)$. Consider $\mathcal{G} = \{[f_{t_i} - \epsilon F, f_{t_i} + \epsilon F] : i = 1, ..., N_1\}$.
- By definition, for $\forall f_t \in \mathcal{F}$, $\exists i$ s.t. $d(t, t_i) \leq \epsilon$.
- Thus by (19), $f_{t_i}(\mathbf{x}) - \epsilon F(\mathbf{x}) \leq f(\mathbf{x}) \leq f_{t_i}(\mathbf{x}) + \epsilon F(\mathbf{x})$ $(\forall \mathbf{x} \in \mathcal{X})$, which means that f is in the bracket $[f_{t_i} - \epsilon F, f_{t_i} + \epsilon F]$.
- The size of the bracket $||(f_{t_i} + \epsilon F) - (f_{t_i} - \epsilon F)|| = 2\epsilon||F||$
- Thus, $(\mathcal{F}, ||\cdot||)$ can be covered by the N_1 $2\epsilon||F||$-brackets in $\mathcal{G}$. The proof is done.

Class of Monotone Functions

We study the bracketing entropy of the class of all monotone functions mapping into $[0,1]$:

Theorem 11

For each $r \in \mathbb{N}_+$, there exists a constant $K < \infty$ such that the class $\mathcal{F}$ of monotone functions $f : \mathbb{R} \to [0,1]$ satisfies

$$\log N_{[]}\left(\epsilon, \mathcal{F}, L_r(\mathbb{Q})\right) \leq \frac{K}{\epsilon}, \tag{23}$$

for all $\epsilon > 0$ and every probability measure $\mathbb{Q}$. [a]

[a] The proof is given in Chapter 2.7 of [VW].

Preservation results for bracketing entropy are rare and below are two such results.

Lemma 11

Let $\mathcal{F}$ and $\mathcal{G}$ be classes of measurable function. Then for any probability measure $\mathbb{Q}$ and any $1 \leq r \leq \infty$,

1. $N_{[]}\left(2\epsilon, \mathcal{F}+\mathcal{G}, L_r(\mathbb{Q})\right) \leq N_{[]}\left(\epsilon, \mathcal{F}, L_r(\mathbb{Q})\right) N_{[]}\left(\epsilon, \mathcal{G}, L_r(\mathbb{Q})\right)$
2. Provided $\mathcal{F}$ and $\mathcal{G}$ are bounded by 1,
   $N_{[]}\left(2\epsilon, \mathcal{F}\cdot\mathcal{G}, L_r(\mathbb{Q})\right) \leq N_{[]}\left(\epsilon, \mathcal{F}, L_r(\mathbb{Q})\right) N_{[]}\left(\epsilon, \mathcal{G}, L_r(\mathbb{Q})\right)$

Proof:

- Suppose $\{[l_{1i}, u_{1i}] : i = 1, ..., N_1\}$ are $N_1 = N_{[]}(\epsilon, \mathcal{F}, L_r(\mathrm{Q}))$ ϵ-bracket that covers $\mathcal{F}$, and $\{[l_{2j}, u_{2j}] : j = 1, ..., N_2\}$ are $N_2 = N_{[]}(\epsilon, \mathcal{G}, L_r(\mathrm{Q}))$ ϵ-bracket that covers $\mathcal{G}$.
- Then we can prove that $\{[l_{1i} + l_{2j}, u_{1i} + u_{2j}] : i = 1, ..., N_1,\ j = 1, ..., N_2\}$ is a collection of 2ϵ-brackets that cover $\mathcal{F} + \mathcal{G}$, so the first inequality holds.
- Also, when $\mathcal{F}$ and $\mathcal{G}$ are bounded by 1, we can prove that $\{[\min\{l_{1i}l_{2j}, l_{1i}u_{2j}, u_{1i}l_{2j}, u_{1i}u_{2j}\}, \max\{l_{1i}l_{2j}, l_{1i}u_{2j}, u_{1i}l_{2j}, u_{1i}u_{2j}\}] : i = 1, ..., N_1,\ j = 1, ..., N_2\}$ is a collection of 2ϵ-brackets that cover $\mathcal{F} \cdot \mathcal{G}$, so the second inequality holds.