Metric Entropy

Jiuzhou Miao

School of Statistics and Mathematics, Zhejiang Gongshang University

April 9, 2025

Metric and metric space

Definition 1 (Metric and metric space)

Let $\mathbb{T}$ be a non-empty set. A function $\varrho : \mathbb{T} \times \mathbb{T} \to \mathbb{R}$ is called as a metric on $\mathbb{T}$ if the following conditions hold:

(1) $\varrho(t, t') \geq 0$ for all pairs (t, t') with "$=$" if and only if $t = t'$.

(2) $\varrho(t, t') = \varrho(t', t)$ for all pairs (t, t').

(3) $\varrho(t, t') \leq \varrho(t, t'') + \varrho(t'', t)$ for all pairs (t, t', t'').

The pair $(\mathbb{T}, \varrho)$ is called as a metric space.

Examples of metric space

- The real space $\mathbb{R}^d$ with Euclidean metric

$$\varrho(\boldsymbol{t}, \boldsymbol{t}') = \|\boldsymbol{t} - \boldsymbol{t}'\|_2 = \sqrt{\sum_{i=1}^{d} (t_i - t_i')^2}.$$

Examples of metric space

- The real space $\mathbb{R}^d$ with Euclidean metric

$$\varrho(\boldsymbol{t}, \boldsymbol{t}') = \|\boldsymbol{t} - \boldsymbol{t}'\|_2 = \sqrt{\sum_{i=1}^{d} (t_i - t_i')^2}.$$

- The discrete cube $\{0,1\}^d$ with the rescaled Hamming metric

$$\varrho(\boldsymbol{t}, \boldsymbol{t}') = d^{-1} \sum_{i=1}^{d} \mathbb{I}(t_i \neq t_i').$$

Examples of metric space

- The space $\mathcal{L}^2\big(\mu, [0,1]\big)$ with metric

$$\varrho(f,g) = \|f-g\|_2 = \left[\int_0^1 \{f(x)-g(x)\}^2 \mathrm{d}\mu(x)\right]^{1/2}.$$

Examples of metric space

- The space $\mathcal{L}^2\big(\mu, [0,1]\big)$ with metric

$$\varrho(f,g) = \|f-g\|_2 = \left[\int_0^1 \{f(x)-g(x)\}^2 \mathrm{d}\mu(x)\right]^{1/2}.$$

- The space $\mathcal{C}[0,1]$ with sup-norm metric

$$\varrho(f,g) = \|f-g\|_\infty = \sup_{x\in[0,1]} \big|f(x)-g(x)\big|.$$

Covering number

Definition 2 (Covering number)

A δ-cover of a set $\mathbb{T}$ with respect to a metric ϱ is a set $\{t_1, \ldots, t_N\} \subseteq \mathbb{T}$ such that for each $t \in \mathbb{T}$, there exists some $i \in \{1, \ldots, N\}$ such that $\varrho(t, t_i) \leq \delta$. The δ-covering number $\mathcal{N}(\delta, \mathbb{T}, \varrho)$ is defined by the cardinality of the smallest δ-cover.

- When discussing metric entropy, we restrict our attention to metric spaces $(\mathbb{T}, \varrho)$ that are totally bounded, i.e., the covering number $\mathcal{N}(\delta, \mathbb{T}, \varrho) < \infty$ for all $\delta > 0$.

Covering number

Definition 2 (Covering number)

A δ-cover of a set $\mathbb{T}$ with respect to a metric ϱ is a set $\{t_1, \ldots, t_N\} \subseteq \mathbb{T}$ such that for each $t \in \mathbb{T}$, there exists some $i \in \{1, \ldots, N\}$ such that $\varrho(t, t_i) \leq \delta$. The δ-covering number $\mathcal{N}(\delta, \mathbb{T}, \varrho)$ is defined by the cardinality of the smallest δ-cover.

- When discussing metric entropy, we restrict our attention to metric spaces $(\mathbb{T}, \varrho)$ that are totally bounded, i.e., the covering number $\mathcal{N}(\delta, \mathbb{T}, \varrho) < \infty$ for all $\delta > 0$.
- The quantity $\log \mathcal{N}(\delta, \mathbb{T}, \varrho)$ is called as the metric entropy of the set $\mathbb{T}$ with respect to metric ϱ.

Packing number

Definition 3 (Packing number)

A δ-packing of a set $\mathbb{T}$ with respect to a metric ϱ is a set $\{t_1, \ldots, t_M\} \subseteq \mathbb{T}$ such that $\varrho(t_i, t_j) > \delta$ for all distinct $i, j \in \{1, \ldots, M\}$. The δ-packing number $\mathcal{M}(\delta, \mathbb{T}, \varrho)$ is defined by the cardinality of the largest δ-packing.

Proposition 4 (Packing and covering)

For all $\delta > 0$, one has that

$$\mathcal{M}(2\delta, \mathbb{T}, \varrho) \leq \mathcal{N}(\delta, \mathbb{T}, \varrho) \leq \mathcal{M}(\delta, \mathbb{T}, \varrho).$$

Covering of $[-1, 1]$

- Consider the interval $[-1, 1]$ equipped with the metric $\varrho(t, t') = |t - t'|$. Let $t_i = -1 + 2(i-1)\delta$, $i = 1, \ldots, 1 + \lfloor 1/\delta \rfloor$.

Covering of $[-1, 1]$

- Consider the interval $[-1, 1]$ equipped with the metric $\varrho(t, t') = |t - t'|$. Let $t_i = -1 + 2(i-1)\delta$, $i = 1, \ldots, 1 + \lfloor 1/\delta \rfloor$.
- For any $t \in [-1, 1]$, there exists some $i \in \{1, \ldots, 1 + \lfloor 1/\delta \rfloor\}$ such that $\varrho(t, t_i) \leq 1/\delta$, which shows that

$$\mathcal{N}\left(\delta, [-1, 1], |\cdot|\right) \leq 1 + 1/\delta.$$

Covering of $[-1,1]$

- Consider the interval $[-1,1]$ equipped with the metric $\varrho(t,t') = |t-t'|$. Let $t_i = -1 + 2(i-1)\delta$, $i = 1, \ldots, 1 + \lfloor 1/\delta \rfloor$.
- For any $t \in [-1,1]$, there exists some $i \in \{1, \ldots, 1 + \lfloor 1/\delta \rfloor\}$ such that $\varrho(t, t_i) \le 1/\delta$, which shows that

$$\mathcal{N}\big(\delta, [-1,1], |\cdot|\big) \le 1 + 1/\delta.$$

- For all $i \neq j$, one has that $\varrho(t_i, t_j) \ge 2\delta > \delta$, which implies that

$$\mathcal{M}\big(2\delta, [-1,1], |\cdot|\big) \ge \lfloor 1/\delta \rfloor.$$

Covering of $[-1,1]$

- Consider the interval $[-1,1]$ equipped with the metric $\varrho(t,t') = |t-t'|$. Let $t_i = -1 + 2(i-1)\delta$, $i = 1, \ldots, 1 + \lfloor 1/\delta \rfloor$.
- For any $t \in [-1,1]$, there exists some $i \in \{1, \ldots, 1 + \lfloor 1/\delta \rfloor\}$ such that $\varrho(t,t_i) \le 1/\delta$, which shows that

$$\mathcal{N}\big(\delta, [-1,1], |\cdot|\big) \le 1 + 1/\delta.$$

- For all $i \ne j$, one has that $\varrho(t_i,t_j) \ge 2\delta > \delta$, which implies that

$$\mathcal{M}\big(2\delta, [-1,1], |\cdot|\big) \ge \lfloor 1/\delta \rfloor.$$

- By Proposition 4, one has that

$$\log \mathcal{N}\big(\delta, [-1,1], |\cdot|\big) \asymp \log\left(1/\delta\right).$$

Volume ratios and metric entropy

Proposition 5 (Volume ratios)

Consider a pair of norms $\|\cdot\|$ and $\|\cdot\|'$ on $\mathbb{R}^d$ and let B and B' be their corresponding unit balls. Then the δ-covering number of B in the norm $\|\cdot\|'$-norm obeys the bounds

$$\delta^{-d}\frac{\mathrm{Vol}(B)}{\mathrm{Vol}(B')} \le \mathcal{N}\big(\delta, B, \|\cdot\|'\big) \le \frac{\mathrm{Vol}(2\delta^{-1}B + B')}{\mathrm{Vol}(B')},$$

where $A+B=\{a+b : a\in A, b\in B\}$ is the Minkowski sum of A and B and $\mathrm{Vol}(A)$ is the volume of the set B.

Volume ratios and metric entropy

- When $B' \subseteq B$, then one has that

$$\mathrm{Vol}(2\delta^{-1}B + B') \le \mathrm{Vol}\big((1+2/\delta)B\big) = (1+2/\delta)^d \mathrm{Vol}(B).$$

Volume ratios and metric entropy

- When $B' \subseteq B$, then one has that

$$\mathrm{Vol}(2\delta^{-1}B + B') \leq \mathrm{Vol}\big((1+2/\delta)B\big) = (1+2/\delta)^d \mathrm{Vol}(B).$$

- When $B' = B$, then one has that

$$d \log(1/\delta) \leq \log \mathcal{N}\big(\delta, B, \|\cdot\|\big) \leq d \log(1+2/\delta).$$

Some examples

- Consider the function class $\mathcal{F}_L$ of $f : [0,1] \to \mathbb{R}$ such that f is L-Lipschitz with $f(0) = 0$. Then

$$\log \mathcal{N}(\delta, \mathcal{F}_L, \|\cdot\|_\infty) \asymp (1/\delta).$$

Some examples

- Consider the function class $\mathcal{F}_L$ of $f : [0,1] \to \mathbb{R}$ such that f is L-Lipschitz with $f(0) = 0$. Then

$$\log \mathcal{N}(\delta, \mathcal{F}_L, \|\cdot\|_\infty) \asymp (1/\delta).$$

- For the case that f defined on $[0,1]^d$ and the function class $\mathcal{F}_{L,d}$ similarly defined with $\mathcal{F}_L$, one has that

$$\log \mathcal{N}(\delta, \mathcal{F}_{L,d}, \|\cdot\|_\infty) \asymp (1/\delta)^d.$$

Some examples

- For some integer α and parameter $\gamma \in (0,1]$, consider the class of functions $f : [0,1] \to \mathbb{R}$ such that

$$\left|f^{(j)}(x)\right| \leq C_j$$

for all $x \in [0,1]$ and $j = 0, \ldots, \alpha$ and

$$\left|f^{(\alpha)}(x) - f^{(\alpha)}(x')\right| \leq L|x - x'|^{\gamma}$$

for all $x, x' \in [0,1]$.

Some examples

- For some integer α and parameter $\gamma \in (0,1]$, consider the class of functions $f : [0,1] \to \mathbb{R}$ such that

$$\left|f^{(j)}(x)\right| \le C_j$$

for all $x \in [0,1]$ and $j = 0, \dots, \alpha$ and

$$\left|f^{(\alpha)}(x) - f^{(\alpha)}(x')\right| \le L|x - x'|^\gamma$$

for all $x, x' \in [0,1]$.

- Denote this class by $\mathcal{F}_{\alpha,\gamma}$. One has that

$$\log \mathcal{N}\big(\delta, \mathcal{F}_{\alpha,\gamma}, \|\cdot\|_\infty\big) \asymp (1/\delta)^{1/(\alpha+\gamma)}.$$

Some examples

- Given a sequence of non-negative real numbers $\{\mu_j\}_{j=1}^{\infty}$ such that $\sum_{j=1}^{\infty} \mu_j < \infty$, consider the ellipsoid

$$\mathcal{E} = \left\{ \{\theta_j\}_{j=1}^{\infty} : \sum_{j=1}^{\infty} \theta_j^2/\mu_j \leq 1 \right\} \subseteq l^2(\mathbb{N}).$$

  Such ellipsoids play an important role in the discussion of reproducing kernel Hilbert spaces.

Some examples

- Given a sequence of non-negative real numbers $\{\mu_j\}_{j=1}^{\infty}$ such that $\sum_{j=1}^{\infty} \mu_j < \infty$, consider the ellipsoid

$$\mathcal{E} = \left\{ \{\theta_j\}_{j=1}^{\infty} : \sum_{j=1}^{\infty} \theta_j^2 / \mu_j \leq 1 \right\} \subseteq l^2(\mathbb{N}).$$

  Such ellipsoids play an important role in the discussion of reproducing kernel Hilbert spaces.

- We set $\mu_j = j^{-2\alpha}$ for some $\alpha > 1/2$. Then

$$\log \mathcal{N}\big(\delta, \mathcal{E}, \|\cdot\|_{l^2(\mathbb{N})}\big) \asymp (1/\delta)^{1/\alpha}.$$

l_q-Euclidean balls

- For $q \in [1, \infty]$, define the l_q-norm of $\boldsymbol{x} \in \mathbb{R}^d$ by

$$\|\boldsymbol{x}\|_q = \begin{cases} \left(\sum_{i=1}^d |x_i|^q\right)^{1/q}, & 1 \leq q < \infty \\ \max_{1 \leq i \leq d} |x_i|, & q = \infty \end{cases}.$$

l_q-Euclidean balls

- For $q \in [1, \infty]$, define the l_q-norm of $\boldsymbol{x} \in \mathbb{R}^d$ by

$$\|\boldsymbol{x}\|_q = \begin{cases} \left(\sum_{i=1}^d |x_i|^q\right)^{1/q}, & 1 \leq q < \infty \\ \max_{1 \leq i \leq d} |x_i|, & q = \infty \end{cases}.$$

- The l_q-Euclidean balls $B_q^d(r)$ are defined by

$$B_q^d(r) = \left\{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\|_q \leq r\right\}.$$

l_q-Euclidean balls

- For $q \in [1, \infty]$, define the l_q-norm of $\boldsymbol{x} \in \mathbb{R}^d$ by

$$\|\boldsymbol{x}\|_q = \begin{cases} \left(\sum_{i=1}^d |x_i|^q\right)^{1/q}, & 1 \le q < \infty \\ \max_{1\le i\le d} |x_i|, & q = \infty \end{cases}.$$

- The l_q-Euclidean balls $B_q^d(r)$ are defined by

$$B_q^d(r) = \left\{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\|_q \le r\right\}.$$

- When $r = 1$, the l_q-Euclidean balls $B_q^d(1)$ are denoted by B_q^d for simplicity.

Gaussian and Rademacher complexity

- Let $\mathbf{Z} = (Z_1, \ldots, Z_d)^\top$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_d)^\top$, where $\{Z_i\}_{i=1}^d$ and $\{\varepsilon_i\}_{i=1}^d$ are i.i.d. standard normal and Rademacher random variables respectively.

Gaussian and Rademacher complexity

- Let $\mathbf{Z} = (Z_1, \dots, Z_d)^\mathsf{T}$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_d)^\mathsf{T}$, where $\{Z_i\}_{i=1}^d$ and $\{\varepsilon_i\}_{i=1}^d$ are i.i.d. standard normal and Rademacher random variables respectively.
- Given $\mathbb{T} \subseteq \mathbb{R}^d$, the family of random variables $\{G_{\boldsymbol{t}}, \boldsymbol{t} \in \mathbb{T}\}$ and $\{R_{\boldsymbol{t}}, \boldsymbol{t} \in \mathbb{T}\}$, where

$$G_{\boldsymbol{t}} = \sum_{i=1}^d t_i Z_i = \langle \boldsymbol{t}, \mathbf{Z} \rangle,$$

$$R_{\boldsymbol{t}} = \sum_{i=1}^d t_i R_i = \langle \boldsymbol{t}, \boldsymbol{\varepsilon} \rangle,$$

  are known as canonical Gaussian process and Rademacher process associated with $\mathbb{T}$.

Gaussian and Rademacher complexity

- The quantities $\mathcal{G}(\mathbb{T}) = \mathbb{E}\big(\sup_{t\in\mathbb{T}} G_t\big)$ and $\mathcal{R}(\mathbb{T}) = \mathbb{E}\big(\sup_{t\in\mathbb{T}} R_t\big)$ are known as the Gaussian complexity and Rademacher complexity of $\mathbb{T}$.

Proposition 6 (Gaussian and Rademacher complexity)

For any $\mathbb{T} \subseteq \mathbb{R}^d$, one has that

$$\mathcal{R}(\mathbb{T}) \leq \sqrt{\pi/2}\mathcal{G}(\mathbb{T}).$$

Gaussian and Rademacher complexity of B_2^d

- Recall that

$$B_2^d = \{\boldsymbol{t} \in \mathbb{R}^d : \|\boldsymbol{t}\|_2 \leq 1\}.$$

Gaussian and Rademacher complexity of B_2^d

- Recall that

$$B_2^d = \{\boldsymbol{t} \in \mathbb{R}^d : \|\boldsymbol{t}\|_2 \leq 1\}.$$

- Computing the Rademacher complexity is straightforward:

$$\mathcal{R}(B_2^d) = \mathbb{E}\Big\{\sup_{\|\boldsymbol{t}\|_2 \leq 1} \langle \boldsymbol{t}, \boldsymbol{\varepsilon} \rangle\Big\} = \sqrt{d}.$$

Gaussian and Rademacher complexity of B_2^d

- Recall that

$$B_2^d = \{\boldsymbol{t} \in \mathbb{R}^d : \|\boldsymbol{t}\|_2 \le 1\}.$$

- Computing the Rademacher complexity is straightforward:

$$\mathcal{R}(B_2^d) = \mathbb{E}\Big\{\sup_{\|\boldsymbol{t}\|_2 \le 1} \langle \boldsymbol{t}, \boldsymbol{\varepsilon} \rangle\Big\} = \sqrt{d}.$$

- By replacing ε with $\mathbf{Z}$ and using Jensen's inequality, one can obtain that

$$\mathbb{E}\big(\|\mathbf{Z}\|_2\big) \le \sqrt{\mathbb{E}\big(\|\mathbf{Z}\|_2^2\big)} = \sqrt{d}.$$

Gaussian and Rademacher complexity of B_2^d

- On the other hand, it can be shown that

$$\mathbb{E}(\|\mathbf{Z}\|_2) \geq \sqrt{d}(1 - o(1)).$$

Gaussian and Rademacher complexity of B_2^d

- On the other hand, it can be shown that

$$\mathbb{E}\big(\|\mathbf{Z}\|_2\big) \geq \sqrt{d}\big(1 - o(1)\big).$$

- Combine these upper and lower bounds, one has that

$$\mathcal{R}(B_2^d)/\sqrt{d} = 1 - o(1).$$

Gaussian complexity of function classes

- Let $\mathcal{F}$ be a function class. For any collection $x_1^n = \{x_1, \ldots, x_n\}$, consider the subset of $\mathbb{R}^n$ given by

$$\mathcal{F}(x_1^n) = \left\{ \left(f(x_1), \ldots, f(x_n) \right)^{\mathsf{T}} : f \in \mathcal{F} \right\}.$$

Gaussian complexity of function classes

- Let $\mathcal{F}$ be a function class. For any collection $x_1^n = \{x_1, \ldots, x_n\}$, consider the subset of $\mathbb{R}^n$ given by

$$\mathcal{F}(x_1^n) = \left\{ \Big(f(x_1), \ldots, f(x_n)\Big)^\mathsf{T} : f \in \mathcal{F} \right\}.$$

- Bounding the Gaussian complexity of this subset yields a measure of the complexity of $\mathcal{F}$ at scale n, which plays an important role in the analysis of nonparametric least squares.

Gaussian complexity of function classes

- Let $\mathcal{F}$ be a function class. For any collection $x_1^n = \{x_1, \ldots, x_n\}$, consider the subset of $\mathbb{R}^n$ given by

$$\mathcal{F}(x_1^n) = \left\{ \big(f(x_1), \ldots, f(x_n)\big)^{\mathsf{T}} : f \in \mathcal{F} \right\}.$$

- Bounding the Gaussian complexity of this subset yields a measure of the complexity of $\mathcal{F}$ at scale n, which plays an important role in the analysis of nonparametric least squares.
- It is most natural to analyze a version of the set $\mathcal{F}(x_1^n)$ is rescaled, either by $n^{-1/2}$ or n^{-1}.

Gaussian complexity of function classes

- It is useful to observe that the Euclidean metric on the rescaled set $\mathcal{F}(x_1^n)/\sqrt{n}$ corresponds to the empirical $\mathcal{L}^2(\mathbb{P}_n)$-metric on the function space $\mathcal{F}$, i.e.,

$$\|f - g\|_n = \sqrt{n^{-1} \sum_{i=1}^{n} \{f(x_i) - g(x_i)\}^2}.$$

Gaussian complexity of function classes

- It is useful to observe that the Euclidean metric on the rescaled set $\mathcal{F}(x_1^n)/\sqrt{n}$ corresponds to the empirical $\mathcal{L}^2(\mathbb{P}_n)$-metric on the function space $\mathcal{F}$, i.e.,

$$\|f-g\|_n = \sqrt{n^{-1}\sum_{i=1}^{n}\{f(x_i)-g(x_i)\}^2}.$$

- If $\mathcal{F}$ is b-uniformly bounded, then $\|f|_n \leq b$. In this case, we have the trivial upper bound

$$\mathcal{G}\big(\mathcal{F}(x_1^n)/n\big) \leq bn^{-1/2}\mathbb{E}\big(\|\mathbf{Z}\|_2\big) \leq b.$$

Sub-Gaussian processes

Definition 7 (Sub-Gaussian processes)

A collection of zero-mean random variables $\{X_t, t \in \mathbb{T}\}$ is a sub-Gaussian process with respect to a metric (or a pseudo metric) ϱ_X on $\mathbb{T}$ if

$$\mathbb{E}\left\{e^{\lambda(X_t - X_{t'})}\right\} \leq e^{\lambda^2 \varrho_X^2(t,t')}$$

for all $t, t' \in \mathbb{T}$ and $\lambda \in \mathbb{R}$.

- The canonical Gaussian and Rademacher processes are both sub-Gaussian with respect to the Euclidean metric $\|\boldsymbol{t} - \boldsymbol{t}'\|_2$.

Sub-Gaussian processes

Definition 7 (Sub-Gaussian processes)

A collection of zero-mean random variables $\{X_t, t \in \mathbb{T}\}$ is a sub-Gaussian process with respect to a metric (or a pseudo metric) ϱ_X on $\mathbb{T}$ if

$$\mathbb{E}\left\{e^{\lambda(X_t - X_{t'})}\right\} \leq e^{\lambda^2 \varrho_X^2(t,t')}$$

for all $t, t' \in \mathbb{T}$ and $\lambda \in \mathbb{R}$.

- The canonical Gaussian and Rademacher processes are both sub-Gaussian with respect to the Euclidean metric $\|\boldsymbol{t} - \boldsymbol{t}'\|_2$.
- For all $\epsilon > 0$,

$$\mathbb{P}\left(|X_t - X_{t'}| > \epsilon\right) \leq 2e^{-\frac{\epsilon^2}{2\varrho_X^2(t,t')}}.$$

Upper bound by one-step discretization

Proposition 8 (One-step discretization bound)

Let $\{X_t, t \in \mathbb{T}\}$ be a zero-mean sub-Gaussian process with respect to the metric ϱ_X. Define $D = \sup_{t,t' \in \mathbb{T}} \varrho_X(t,t')$. Then for any $\delta \in [0, D]$ such that $\mathcal{N}(\delta, \mathbb{T}, \varrho_X) \geq 10$, one has that

$$\mathbb{E}\Big\{\sup_{t,t' \in \mathbb{T}} (X_t - X_{t'})\Big\} \leq 2\mathbb{E}\Big\{\sup_{\gamma,\gamma' \in \mathbb{T}, \varrho_X(\gamma,\gamma') \leq \delta} (X_\gamma - X_{\gamma'})\Big\} + 4D\sqrt{\log \mathcal{N}(\delta, \mathbb{T}, \varrho_X)}$$

- The Proposition above always implies an upper bound on $\mathbb{E}(\sup_{t \in \mathbb{T}} X_t)$ by

$$\mathbb{E}\big(\sup_{t \in \mathbb{T}} X_t\big) = \mathbb{E}\big\{\sup_{t \in \mathbb{T}} (X_t - X_{t_0})\big\} \leq \mathbb{E}\Big\{\sup_{t,t' \in \mathbb{T}} (X_t - X_{t'})\Big\}$$

Bound Gaussian and Rademacher complexity

- Let $\mathbb{T} \subseteq \mathbb{R}^d$. Denote

$$\tilde{\mathbb{T}}_\delta = \{\gamma - \gamma' : \gamma, \gamma' \in \mathbb{T}, \|\gamma - \gamma'\|_2 \le \delta\}.$$

Bound Gaussian and Rademacher complexity

- Let $\mathbb{T} \subseteq \mathbb{R}^d$. Denote

$$\tilde{\mathbb{T}}_\delta = \{\gamma - \gamma' : \gamma, \gamma' \in \mathbb{T}, \|\gamma - \gamma'\|_2 \leq \delta\}.$$

- Then Proposition 8 implies that

$$\begin{aligned}
\mathcal{G}(\mathbb{T}) &\leq \inf_{\delta \in [0,D]} \left\{ 2\mathcal{G}(\tilde{\mathbb{T}}_\delta) + 4D\sqrt{\log \mathcal{N}\big(\delta, \mathbb{T}, \|\cdot\|_2\big)} \right\} \\
&\leq \inf_{\delta \in [0,D]} \left\{ 2\delta\sqrt{d} + 4D\sqrt{\log \mathcal{N}\big(\delta, \mathbb{T}, \|\cdot\|_2\big)} \right\}, \\
\mathcal{R}(\mathbb{T}) &\leq \inf_{\delta \in [0,D]} \left\{ 2\mathcal{R}(\tilde{\mathbb{T}}_\delta) + 4D\sqrt{\log \mathcal{N}\big(\delta, \mathbb{T}, \|\cdot\|_2\big)} \right\} \\
&= \inf_{\delta \in [0,D]} \left\{ 2\delta\sqrt{d} + 4D\sqrt{\log \mathcal{N}\big(\delta, \mathbb{T}, \|\cdot\|_2\big)} \right\}.
\end{aligned}$$

Gaussian complexity for smoothness classes

- Recall the class $\mathcal{F}_L$ and its metric entropy

$$\log \mathcal{N}(\delta, \mathcal{F}_L, \|\cdot\|_\infty) \asymp (1/\delta).$$

Assume that the functions in $\mathcal{F}_L$ are uniformly bounded by 1.

Gaussian complexity for smoothness classes

- Recall the class $\mathcal{F}_L$ and its metric entropy

$$\log \mathcal{N}\big(\delta, \mathcal{F}_L, \|\cdot\|_\infty\big) \asymp (1/\delta).$$

Assume that the functions in $\mathcal{F}_L$ are uniformly bounded by 1.

- Let $\mathbb{T} = \mathcal{F}(x_1^n)/\sqrt{n}$. One has that

$$D = \sup_{\boldsymbol{\gamma}, \boldsymbol{\gamma}' \in \mathbb{T}} \|\boldsymbol{\gamma} - \boldsymbol{\gamma}'\|_2 \le 2$$

and

$$\log \mathcal{N}\big(\delta, \mathbb{T}, \|\cdot\|_2\big) = \log \mathcal{N}\big(\delta, \mathcal{F}_L, \|\cdot\|_n\big) \le \log \mathcal{N}\big(\delta, \mathcal{F}_L, \|\cdot\|_\infty\big).$$

Gaussian complexity for $\mathcal{F}_L$

- Then one has that

$$\mathcal{G}\big(\mathcal{F}_L(x_1^n)/n\big) = n^{-1/2}\mathcal{G}(\mathbb{T}) \leq n^{-1/2} \inf_{\delta\in[0,2]}\Big\{2\delta\sqrt{n} + 8c\delta^{-1/2}\Big\}$$

for some positive c independent of n.

Gaussian complexity for $\mathcal{F}_L$

- Then one has that

$$\mathcal{G}\big(\mathcal{F}_L(x_1^n)/n\big) = n^{-1/2}\mathcal{G}(\mathbb{T}) \le n^{-1/2} \inf_{\delta\in[0,2]} \left\{2\delta\sqrt{n} + 8c\delta^{-1/2}\right\}$$

  for some positive c independent of n.

- By taking $\delta \asymp n^{-1/3}$, one has that

$$\mathcal{G}\big(\mathcal{F}_L(x_1^n)/n\big) \lesssim n^{-1/3}.$$

Dudley's entropy integral

Theorem 9 (Dudley's entropy integral bound)

Let $\{X_t, t \in \mathbb{T}\}$ be a zero-mean sub-Gaussian process with respect to the pseudo metric ϱ_X. Define $D = \sup_{t,t' \in \mathbb{T}} \varrho_X(t,t')$ and

$$\mathcal{J}(\delta, D) = \int_\delta^D \sqrt{\log \mathcal{N}(u, \mathbb{T}, \varrho_X)}\mathrm{d}u.$$

Then for any $\delta \in [0, D]$, one has that

$$\mathbb{E}\Big\{\sup_{t,t' \in \mathbb{T}} (X_t - X_{t'})\Big\} \le 2\mathbb{E}\Big\{\sup_{\gamma,\gamma' \in \mathbb{T}, \varrho_X(\gamma,\gamma') \le \delta} (X_\gamma - X_{\gamma'})\Big\} + 32\mathcal{J}(\delta/4, D).$$

Bounds for Vapnik-Chervonenkis classes

- Let $\mathcal{F}$ be a b-uniformly bounded class of functions with finite VC dimension ν. We will bound the Rademacher complexity

$$\mathbb{E}_{\boldsymbol{\varepsilon}}\left\{\sup_{f\in\mathcal{F}}\left|n^{-1}\sum_{i=1}^{n}\varepsilon_i f(x_i)\right|\right\} = n^{-1/2}\mathbb{E}_{\boldsymbol{\varepsilon}}\left(\sup_{f\in\mathcal{F}}|Z_f|\right).$$

Bounds for Vapnik-Chervonenkis classes

- Let $\mathcal{F}$ be a b-uniformly bounded class of functions with finite VC dimension ν. We will bound the Rademacher complexity

$$\mathbb{E}_{\boldsymbol{\varepsilon}}\left\{\sup_{f\in\mathcal{F}}\left|n^{-1}\sum_{i=1}^{n}\varepsilon_i f(x_i)\right|\right\} = n^{-1/2}\mathbb{E}_{\boldsymbol{\varepsilon}}\left(\sup_{f\in\mathcal{F}}|Z_f|\right).$$

- For $f, g \in \mathcal{F}$, one can verify that $Z_f - Z_g$ is sub-Gaussian with parameter

$$\|f-g\|_{\mathbb{P}_n} = n^{-1}\sum_{i=1}^{n}\{f(x_i)-g(x_i)\}^2 \leq 2b$$

uniformly for all $f, g \in \mathcal{F}$.

Bounds for Vapnik-Chervonenkis classes

- By using that the known result that

$$\log \mathcal{N}(\delta, \mathcal{F}, \|\cdot\|_{\mathbb{P}_n}) \lesssim \log(1/\delta),$$

one has that

$$\begin{aligned}\mathbb{E}_{\varepsilon}\left\{\sup_{f\in\mathcal{F}}\left|n^{-1}\sum_{i=1}^{n}\varepsilon_i f(x_i)\right|\right\} &\lesssim n^{-1/2}\int_0^{2b}\sqrt{\log(1/\delta)}\mathrm{d}\delta \\ &\lesssim n^{-1/2}.\end{aligned}$$

Orlicz norm

Definition 10 (Orlicz norm)

Let $\psi_q(x) = \exp(x^q) - 1$, $q \in [1,2]$. The ψ_q-Orlicz norm of a zero-mean random variable X is given by

$$\|X\|_{\psi_q} = \inf\Big\{\lambda > 0 : \mathbb{E}\Big\{\psi_q\big(|X|/\lambda\big)\Big\} \leq 1\Big\}.$$

The Orlicz norm is infinite if there is no $\lambda \in \mathbb{R}$ for which the given expectation is finite.

- If $\|X\|_{\psi_q} < \infty$, then one has that for all $t > 0$,

$$\mathbb{P}\big(|X| \geq t\big) \leq \psi_q^{-1}\big(t/\|X\|_{\psi_q}\big).$$

Orlicz processes

Definition 11 (Orlicz processes)

A zero-mean stochastic process $\{X_t, t \in \mathbb{T}\}$ is a ψ_q-Orlicz process with respect to a metric ϱ_X if

$$\|X_t - X_{t'}\|_{\psi_q} \leq \varrho_X(t, t').$$

- If $\|X\|_{\psi_q} < \infty$, then one has that for all $t > 0$,

$$\mathbb{P}\big(|X| \geq t\big) \leq \psi_q^{-1}\big(t/\|X\|_{\psi_q}\big).$$

Concentration of Orlicz processes

Theorem 12 (Concentration of Orlicz processes)

Let $\{X_t, t \in \mathbb{T}\}$ *is a* ψ_q*-Orlicz process with respect to a metric* ϱ_X*. Define* $D = \sup_{t,t' \in \mathbb{T}} \varrho_X(t, t')$ *and*

$$\mathcal{J}_q(\delta, D) = \int_\delta^D \psi_q^{-1}\big(\mathcal{N}(u, \mathbb{T}, \varrho_X)\big)\mathrm{d}u,$$

where ψ_q^{-1} *is the inverse of* ψ_q*. Then there is a universal constant* c *such that for all* $\epsilon > 0$*,*

$$\mathbb{P}\left[\sup_{t,t' \in \mathbb{T}} |X_t - X_{t'}| \geq c\big\{\mathcal{J}_q(0, D) + \epsilon\big\}\right] \leq 1/\psi_q(\epsilon/D).$$

A general comparison result

Theorem 13 (General Gaussian comparison principle)

Let $(X_1, \ldots, X_N)$ and $(Y_1, \ldots, Y_N)$ be a pair of centered Gaussian random vectors and suppose that there exist disjoint subsets A and B of $\{1, \ldots, N\} \times \{1, \ldots, N\}$ such that $\mathbb{E}(X_i X_j) \leq \mathbb{E}(Y_i Y_j)$ for all $(i,j) \in A$, $\mathbb{E}(X_i X_j) \geq \mathbb{E}(Y_i Y_j)$ for all $(i,j) \in B$ and $\mathbb{E}(X_i X_j) = \mathbb{E}(Y_i Y_j)$ for all $(i,j) \notin A \cup B$. Let $F : \mathbb{R}^N \to \mathbb{R}$ be a twice-differentiable function, and suppose that $\frac{\partial^2 F}{\partial u_i \partial u_j}(\boldsymbol{u}) \geq 0$ for all $(i,j) \in A$ and $\frac{\partial^2 F}{\partial u_i \partial u_j}(\boldsymbol{u}) \leq 0$ for all $(i,j) \in B$. Then one has that

$$\mathbb{E}\{F(X_1, \ldots, X_N)\} \leq \mathbb{E}\{F(Y_1, \ldots, Y_N)\}.$$

Slepian's inequality

Theorem 14 (Slepian's inequality)

Let $(X_1, \ldots, X_N)$ and $(Y_1, \ldots, Y_N)$ be a pair of centered Gaussian random vectors such that $\mathbb{E}(X_i X_j) = \mathbb{E}(Y_i Y_j)$ for all $i \neq j$ and $\mathbb{E}(X_i^2) = \mathbb{E}(Y_i^2)$ for all $i = 1, \ldots, N$. Then one has that

$$\mathbb{E}\big(\max_{1 \le i \le N} X_i\big) \le \mathbb{E}\big(\max_{1 \le i \le N} Y_i\big).$$

Sudakov-Fernique comparison

Theorem 15 (Sudakov-Fernique comparison)

Let $(X_1, \ldots, X_N)$ *and* $(Y_1, \ldots, Y_N)$ *be a pair of centered Gaussian random vectors such that*

$$\mathbb{E}\big\{(X_i - X_j)^2\big\} \leq \mathbb{E}\big\{(Y_i - Y_j)^2\big\}$$

for all $(i, j) \in \{1, \ldots, N\} \times \{1, \ldots, N\}$*. Then one has that*

$$\mathbb{E}\big(\max_{1 \leq i \leq N} X_i\big) \leq \mathbb{E}\big(\max_{1 \leq i \leq N} Y_i\big).$$

Gaussian contraction inequality

Proposition 16 (Gaussian contraction inequality)

Let $\{Z_i\}_{i=1}^d$ *be i.i.d. standard normal random variables and* $\phi_i : \mathbb{R} \to \mathbb{R}$, $j = 1, \ldots, d$ *are* 1*-Lipschitz with* $\phi_i(0) = 0$. *Then for any* $\mathbb{T} \subseteq \mathbb{R}^d$, *one has that*

$$\mathbb{E}\left\{\sup_{\boldsymbol{t}\in\mathbb{T}} \sum_{i=1}^d Z_i \phi_i(t_i)\right\} \leq \mathbb{E}\left(\sup_{\boldsymbol{t}\in\mathbb{T}} \sum_{i=1}^d Z_i t_i\right).$$

Sudakov's lower bound

Theorem 17 (Sudakov's minoration)

Let $\{X_t, t \in \mathbb{T}\}$ *be a zero-mean Gaussian process defined on the non-empty set* $\mathbb{T}$. *Then*

$$\mathbb{E}\big(\sup_{t\in\mathbb{T}} X_t\big) \geq \sup_{\delta>0}\Big\{2^{-1}\delta\sqrt{\log \mathcal{M}(\delta, \mathbb{T}, \varrho_X)}\Big\},$$

where $\varrho_X(t, t') = \sqrt{\mathbb{E}\{(X_t - X_{t'})^2\}}$.

Thank You