Metric Entropy and its Applications

Presented by Wanteng Ma

Department of Mathematics

Hong Kong University of Science and Technology

March 16, 2022

Metric Entropy

Metric entropy serves as a tool to measure the “size” of a set. It is also a measure of **statistical complexity**.

- First introduced by Kolmogorov.
- Broadly applied to theoretical statistics
  - nonparametric function estimation[1]
  - high-dimensional statistical inference[2]
  - statistical learning theory[3]

Generally, it reveals interesting connections between the complexity of the parameter space and the fundamental difficulty of the statistical problem.

---

[1] Yihong Wu and Pengkun Yang. “Minimax rates of entropy estimation on large alphabets via best polynomial approximation”. In: *IEEE Transactions on Information Theory* 62.6 (2016), pp. 3702–3720.

[2] T Tony Cai, Zongming Ma, and Yihong Wu. “Sparse PCA: Optimal rates and adaptive estimation”. In: *The Annals of Statistics* 41.6 (2013), pp. 3074–3110.

[3] Vladimir Koltchinskii. “Local Rademacher complexities and oracle inequalities in risk minimization”. In: *The Annals of Statistics* 34.6 (2006), pp. 2593–2656.

Definition

Consider a metric space $(\mathbb{T}, \rho)$ with metric $\rho(\theta, \theta')$ satisfying

1. Non-negativity
2. Symmetry
3. Triangular inequality

e.g.

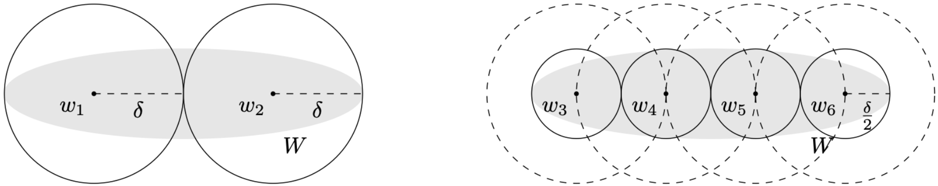
- (Euclidean metric) $\rho(\theta, \theta') = \|\theta - \theta'\|_2$ on $\mathbb{T} = \mathbb{R}^d$
- (Rescaled Hamming metric) $\rho(\theta, \theta') = \frac{1}{d}\sum_{i=1}^{d} \mathbb{1}\left(\theta_i \neq \theta_i'\right)$ on $\mathbb{T} = \{0, 1\}^d$
- Function space $\mathcal{L}^2(\mu, [0, 1])$ with metric

$$\|f - g\|_2 = \left(\int_0^1 (f(x) - g(x))^2 d\mu(x)\right)^{1/2}$$

Definition

Definition (Covering number)

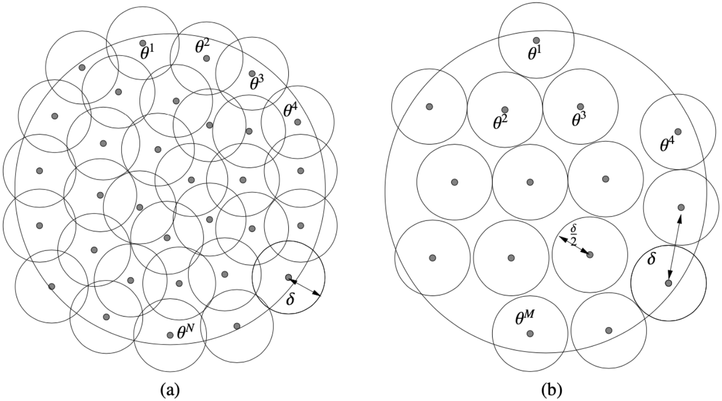
A δ-cover of a set $\mathbb{T}$ with respect to a metric ρ is a set $\{\theta_1, \ldots, \theta_N\} \subseteq \mathbb{T}$ such that for each $\theta \in \mathbb{T}$, there exists some $i \in \{1, \ldots, N\}$ such that $\rho(\theta, \theta_i) \leq \delta$. The δ-covering number $N(\delta; \mathbb{T}, \rho)$ is the cardinality of the smallest δ-cover.



Definition (packing number)

A δ-packing of a set $\mathbb{T}$ with respect to a metric ρ is a set $\{\theta_1, \ldots, \theta_M\} \subseteq \mathbb{T}$ such that for all distinct $i, j \in \{1, \ldots, M\}$, $\rho(\theta_i, \theta_j) > \delta$. The δ-packing number $M(\delta; \mathbb{T}, \rho)$ is the cardinality of the largest δ-packing.

Visualization



From the definition of covering and packing set, we have

1. If M is the cardinality of a δ-packing set, then we need at least M balls with radius $\frac{\delta}{2}$ to cover the space.
2. A δ-packing set with the largest cardinality $M(\delta; \mathbb{T}, \rho)$ is also a δ-cover.

Relation

For all $\delta > 0$, the packing and covering numbers are related as follows:

$$M(2\delta; \mathbb{T}, \rho) \leq N(\delta; \mathbb{T}, \rho) \leq M(\delta; \mathbb{T}, \rho),$$

which indicates

1. The packing and covering numbers exhibit the same scaling behavior as $\delta \to 0$
2. We can find upper/lower bound of covering/packing numbers by constructing cover set or packing set.
3. We can investigate the growth rate of covering/packing numbers by studying *metric entropy* $\log(N(\delta; \mathbb{T}, \rho))$.

Example 1 (Covering and packing of unit cubes)

Consider the interval $[-1, 1]$ in $\mathbb{R}$, with the metric $\rho(\theta, \theta') = |\theta - \theta'|$.

- Divide the interval into $L = \lfloor \frac{1}{\delta} \rfloor + 1$ sub-intervals, with each one centered at $\theta = -1 + (2i - 1)\delta$
- By construction, for any point $\theta \in [-1, 1]$, there is some $j \in [L]$ such that $|\theta_j - \theta| \leq \delta$, i.e., $N(\delta, [-1, 1], |\cdot|) \leq \frac{1}{\delta} + 1$. Generalize this analysis to higher dimensions we can show $N(\delta, [-1, 1]^d, \|\cdot\|_\infty) \leq (\frac{1}{\delta} + 1)^d$.
- Consider $\{\theta_i : i \in [L - 1]\}$. Any two elements θ_i, θ_j in this set have the separation $|\theta_i - \theta_j| \geq 2\delta$, which implies $M(2\delta, [-1, 1], |\cdot|) \geq \lfloor \frac{1}{\delta} \rfloor$. This also means $M(2\delta, [-1, 1]^d, \|\cdot\|_\infty) \geq \lfloor \frac{1}{\delta} \rfloor^d$
- Combined with the relation of packing number and covering number, we have

$$\log N\left(\delta; [0, 1]^d, \|\cdot\|_\infty\right) \asymp d \log(1/\delta)$$

Volume ratios

Lemma (Volume ratios and metric entropy)

Consider a pair of norms $\|\cdot\|$ and $\|\cdot\|'$ on $\mathbb{R}^d$, and let $\mathbb{B}$ and $\mathbb{B}'$ be their corresponding unit balls (i.e., $\mathbb{B} = \{\theta | \ \|\theta\| \leq 1\}$, with $\mathbb{B}'$ similarly defined). Then we have

$$\left(\frac{1}{\delta}\right)^d \frac{\mathrm{vol}(\mathbb{B})}{\mathrm{vol}\left(\mathbb{B}'\right)} \overset{(a)}{\leq} N\left(\delta; B, \|\cdot\|'\right) \overset{(b)}{\leq} \frac{\mathrm{vol}\left(\frac{2}{\delta}\mathbb{B} + \mathbb{B}'\right)}{\mathrm{vol}\left(\mathbb{B}'\right)}.$$

Proof:

- For (a), if $\{\theta_1, ..., \theta_N\}$ a δ-covering of $\mathbb{B}$, then $\mathbb{B} \subseteq \bigcup_{j=1}^N \{\theta_j + \delta\mathbb{B}'\}$, which implies that $\mathrm{vol}(\mathbb{B}) \leq N\,\mathrm{vol}\left(\delta\mathbb{B}'\right) = N\delta^d\,\mathrm{vol}\left(\mathbb{B}'\right)$
- For (b), let $\{\theta_1, ..., \theta_M\}$ be a maximal δ-packing of $\mathbb{B}$ in the $\|\cdot\|'$-norm; by maximality, this set must also be a δ-covering. The balls $\{\theta_j + \frac{\delta}{2}\mathbb{B}' | j \in [M]\}$ are all disjoint and contained within $\mathbb{B} + \frac{\delta}{2}\mathbb{B}'$.

Volume ratios

- This indicates that $M\,\mathrm{vol}(\frac{\delta}{2}\mathbb{B}') \leq \mathrm{vol}(\mathbb{B} + \frac{\delta}{2}\mathbb{B}')$. Taking $\frac{\delta}{2}$ out we have $\mathrm{vol}(\frac{\delta}{2}\mathbb{B}') = (\frac{\delta}{2})^d\,\mathrm{vol}(\mathbb{B}')$, which justifies (b).
- A special case: when $\mathbb{B}' \subseteq \mathbb{B}$, we have:

$$\mathrm{vol}\left(\frac{2}{\delta}\mathbb{B} + \mathbb{B}'\right) \leq \mathrm{vol}\left(\left(\frac{2}{\delta} + 1\right)\mathbb{B}\right) = \left(\frac{2}{\delta} + 1\right)^d \mathrm{vol}(\mathbb{B})$$

- If further we take $\mathbb{B}' = \mathbb{B}$, then we have the following bounds for metric entropy of a unit ball.

$$d\log(1/\delta) \leq \log N(\delta; \mathbb{B}, \|\cdot\|) \leq d\log\left(1 + \frac{2}{\delta}\right)$$

Example 2 (Lipschitz functions on the unit interval)

Now consider the class of Lipschitz functions $\mathcal{F}_L :=$
$\{g : [0,1] \to \mathbb{R} \mid g(0) = 0, \text{and} \left|g(x) - g\left(x'\right)\right| \leq L\left|x - x'\right|, \forall x, x' \in [0,1]\}$.

Lipschitz functions on the unit interval

The metric entropy of $\mathcal{F}_L$ with respect to the sup-norm scales as

$$\log N_{\infty}\left(\delta, \mathcal{F}_L\right) \asymp (L/\delta)$$

Proof:

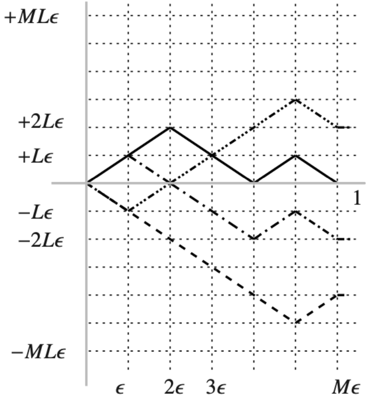
- Prove the lower bound by constructing a sufficiently large packing of the set $\mathcal{F}_L$. For any ϵ, define $M = \lfloor \frac{1}{\epsilon} \rfloor$, consider the points in $[0,1]$ given by

$$x_i = (i-1)\epsilon, \text{ for } i \in [M], \text{ and } x_{M+1} = M\epsilon.$$

- We construct a set of segmented functions that each function is piecewise linear over the intervals $[x_{i-1}, x_i]$ with slope either L or $-L$.

Example 2 (Lipschitz functions on the unit interval)

- Define the unit function: $\phi(u) := \begin{cases} 0 & \text{for } u < 0 \\ u & \text{for } u \in [0,1] \\ 1 & \text{otherwise} \end{cases}$
- For each binary sequence $\beta \in \{-1,1\}^M$, define $f_\beta(y) = \sum_{i=1}^{M} \beta_i L\epsilon\phi\left(\frac{y-x_i}{\epsilon}\right)$.



- We can check that $\{f_\beta \mid \beta \in \{-1,1\}^M\} \subseteq \mathcal{F}_L$

Example 2 (Lipschitz functions on the unit interval)

- From the construction of f_β we know that for any distinct $i, j \in \{-1, 1\}^M$, $\|f_i - f_j\|_\infty \geq 2L\epsilon$, showing that $\{f_\beta | \beta \in \{-1, 1\}^M\}$ is a $2L\epsilon$-packing with cardinality 2^M. Thus

$$\log N\left(\delta; \mathcal{F}_L, \|\cdot\|_\infty\right) \gtrsim L/\delta$$

- For the upper bound, we show that $\{f_\beta \mid \beta \in \{-1, 1\}^M\}$ is also a δ-covering of set $\mathcal{F}_L$, which can be justified by induction.
- An extension of this argument shows that

$$\log N_\infty\left(\delta, \mathcal{F}_L([0, 1]^d)\right) \asymp (L/\delta)^d$$

Gaussian and Rademacher complexity

Metric entropy plays a fundamental role in understanding the behavior of stochastic processes. Define the Gaussian and Rademacher complexity as following:

Definition (Gaussian complexity)

Given a set $\mathbb{T} \in \mathbb{R}^d$, the random variable w follows d-dimensional standard Gaussian distribution, then the Gaussian complexity of $\mathbb{T}$ is defined as

$$\mathcal{G}(\mathbb{T}) := \mathbb{E}\left[\sup_{\theta \in \mathbb{T}} \langle \theta, w \rangle\right].$$

Definition (Rademacher complexity)

Given a set $\mathbb{T} \in \mathbb{R}^d$, ε is a d-dimensional Rademacher variable, then the Rademacher complexity of $\mathbb{T}$ is defined as

$$\mathcal{R}(\mathbb{T}) := \mathbb{E}\left[\sup_{\theta \in \mathbb{T}} \langle \theta, \varepsilon \rangle\right].$$

Gaussian and Rademacher complexity

Trivial relation between Rademacher/Gaussian complexity:

$$\mathcal{R}(\mathbb{T}) \le \sqrt{\frac{\pi}{2}}\mathcal{G}(\mathbb{T})$$

Proof:

- Consider a fixed $\varepsilon \in \{-1,1\}^d$. For $\forall\theta \in \mathbb{T}$, integrate $\langle\theta, w\rangle$ over the corresponding $\frac{1}{2^d}$-half space:

$$\int_{R(\varepsilon)} \langle\theta, w\rangle f(w)dw = \frac{1}{2^{d-1}\sqrt{2\pi}}\langle\theta, \varepsilon\rangle$$

- Taking supremum and Summing all the ε

$$\begin{aligned}\mathbb{E}\left[\sup_{\theta\in\mathbb{T}}\langle\theta,\varepsilon\rangle\right] &= \frac{1}{2^d}\sum_{\varepsilon}\sup_{\theta\in\mathbb{T}}\langle\theta,\varepsilon\rangle \le \sqrt{\frac{\pi}{2}}\sum_{\varepsilon}\sup_{\theta\in\mathbb{T}}\int_{R(\varepsilon)}\langle\theta,w\rangle f(w)dw \\ &\le \sqrt{\frac{\pi}{2}}\int\sup_{\theta\in\mathbb{T}}\langle\theta,w\rangle f(w)dw = \sqrt{\frac{\pi}{2}}\mathcal{G}(\mathbb{T}).\end{aligned}$$

Gaussian and Rademacher complexity

Some prerequisite knowledge on Gaussian variables sequences: Suppose $X \in \mathbb{R}^n$ and $\{X_i\}_{i=1}^n$ is an i.i.d. sequence of $N(0, \sigma^2)$ variables, then we have

- (Concentration bounds on χ^2) The growth of $\|X\|_2$ follows

$$\frac{\mathbb{E}\|X\|_2}{\sigma\sqrt{n}} = 1 - o(1).$$

  with upper bound $\mathbb{E}\|X\|_2 \leq \sigma\sqrt{n}$.

- (Gaussian maxima) The Gaussian maxima $Z_n := \max_{i=1,\ldots,n} |X_i| = \|X\|_\infty$ follows

$$\frac{\mathbb{E}\left[Z_n\right]}{\sqrt{2\sigma^2 \log n}} \to 1 \text{ as } n \to +\infty.$$

  with upper bound $\mathbb{E}Z_n \leq 2\sigma\sqrt{\log n}$.

Example 3 (Rademacher/Gaussian complexity of unit balls)

Compute the Rademacher and Gaussian complexities of the Euclidean ball of unit norm $\mathbb{B}_2^d$.

- Rademacher complexity:

$$\mathcal{R}\left(\mathbb{B}_2^d\right) = \mathbb{E}\left[\sup_{\|\theta\|_2 \leq 1} \langle \theta, \varepsilon \rangle\right] = \mathbb{E}\left[\left(\sum_{i=1}^d \varepsilon_i^2\right)^{1/2}\right] = \sqrt{d}$$

- Gaussian complexity: we have $\mathcal{G}\left(\mathbb{B}_2^d\right) = \mathbb{E}\left[\|w\|_2\right]$, and preliminary knowledge of Gaussian distribution shows

$$\frac{\mathcal{G}\left(\mathbb{B}_2^d\right)}{\sqrt{d}} = 1 - o(1).$$

Example 3 (Rademacher/Gaussian complexity of unit balls)

Compute the Rademacher and Gaussian complexities of the ball of unit l_1 norm $\mathbb{B}_1^d$.

- Rademacher complexity:

$$\mathcal{R}\left(\mathbb{B}_1^d\right) = \mathbb{E}\left[\sup_{\|\theta\|_1 \leq 1} \langle \theta, \varepsilon \rangle\right] = \mathbb{E}\|\varepsilon\|_\infty = 1.$$

- Gaussian complexity: we have $\mathcal{G}\left(\mathbb{B}_1^d\right) = \mathbb{E}\left[\|w\|_\infty\right]$, and preliminary knowledge of Gaussian distribution shows

$$\frac{\mathcal{G}\left(\mathbb{B}_1^d\right)}{\sqrt{2\log d}} = 1 \pm o(1).$$

Metric entropy and sub-Gaussian processes

Definition

A collection of zero-mean random variables $\{X_\theta | \theta \in \mathbb{T}\}$ is a sub-Gaussian process with respect to a metric ρ_X if

$$\mathbb{E}\left[e^{\lambda(X_\theta - X_{\bar{\theta}})}\right] \leq e^{\frac{\lambda^2 \rho_X^2(\theta,\tilde{\theta})}{2}}, \text{ for all } \theta, \widetilde{\theta} \in \mathbb{T}, \text{ and } \lambda \in \mathbb{R}$$

For the sub-Gaussian process, we have the famous Massart's finite class lemma:

Lemma (Massart's)

Let A be some finite set with $\mathrm{Diam}(A) = \sup_{\theta,\theta' \in \mathbb{T}} \rho_X(\theta, \theta') \leq R$, *then*

$$\sup_{\theta \in \mathbb{T}} X_\theta \leq R\sqrt{2 \log |A|}.$$

Massart's finite class lemma

Proof:

- Denote $\mu = \mathbb{E} \sup_{\theta \in \mathbb{T}} (X_\theta - X_{\theta'})$, we have

$$e^{\lambda\mu} \leq \mathbb{E} \sup_{\theta \in \mathbb{T}} e^{\lambda(X_\theta - X_{\theta'})} \leq \mathbb{E} \sum_{\theta \in A} e^{\frac{\lambda^2}{2}\rho_X^2(\theta,\theta')} \leq |A| e^{\frac{\lambda^2}{2}R^2}, \text{ for } \forall\lambda$$

- Taking logarithm we have

$$\mu \leq \frac{1}{\lambda} \log|A| + \frac{\lambda}{2} R^2$$

- Taking minimum value of RHS, we have

$$\sup_{\theta \in \mathbb{T}} X_\theta \leq R\sqrt{2\log|A|}.$$

From this argument we can also derive the Gaussian maxima, together with bound of $\sup_{\theta \in \mathbb{T}} |X_\theta|$ or $\sup_{\theta,\theta' \in \mathbb{T}} (X_\theta - X_{\theta'})$.

Upper bound by one-step discretization

One-step discretization

Consider the zero-mean sub-Gaussian process X_θ with $\mathrm{Diam}(\mathbb{T}) \leq D$. For any appropriate δ, we have

$$\mathbb{E}\left[\sup_{\theta,\widetilde{\theta}\in\mathbb{T}} \left(X_\theta - X_{\widetilde{\theta}}\right)\right] \leq 2\mathbb{E}\left[\sup_{\substack{\gamma,\gamma'\in\mathbb{T}\\ \rho_X(\gamma,\gamma')\leq\delta}} (X_\gamma - X_{\gamma'})\right] + 4\sqrt{D^2 \log N_X(\delta;\mathbb{T})}$$

Proof:

- For a given δ and associated δ-covering $\{\theta_1, ..., \theta_N\}$ with $N = N_X(\theta, \mathbb{T})$. For any θ we can find some θ_j such that $\rho_X(\theta, \theta_j) \leq \delta$. Then

$$\begin{aligned} X_\theta - X_{\theta_1} =& X_\theta - X_{\theta_j} + X_{\theta_j} - X_{\theta_1} \\ \leq& \sup_{\substack{\gamma,\gamma'\in\mathbb{T}\\ \rho_X(\gamma,\gamma')\leq\delta}} (X_\gamma - X_{\gamma'}) + \max_{i=1,2,...,N} |X_{\theta_i} - X_{\theta_1}| \end{aligned}$$

- the same upper bound also holds for $X_{\theta_1} - X_{\widetilde{\theta}}$.

One-step discretization

- By Massart's lemma we have

$$\mathbb{E}\left[\max_{i=1,\ldots,N} |X_{\theta_i} - X_{\theta_1}|\right] \leq 2\sqrt{D^2 \log N},$$

  which yields the claim.

Rmk

Since the zero-mean condition means that

$$\mathbb{E}\left[\sup_{\theta\in\mathbb{T}} X_\theta\right] = \mathbb{E}\left[\sup_{\theta\in\mathbb{T}} (X_\theta - X_{\theta_0})\right] \leq \mathbb{E}\left[\sup_{\theta,\widetilde{\theta}\in\mathbb{T}} \left(X_\theta - X_{\widetilde{\theta}}\right)\right]$$

We can also use this one-step discretization to bound $\mathbb{E}\left[\sup_{\theta\in\mathbb{T}} X_\theta\right]$, e.g., Rademacher/Gaussian complexity. In this case, the factor 2 is not needed because we can fix the $X_{\widetilde{\theta}}$.

One-step discretization

In the l_2 norm, if we define $\widetilde{\mathbb{T}}(\delta) := \{\gamma - \gamma' \mid \gamma, \gamma' \in \mathbb{T}, \|\gamma - \gamma'\|_2 \leq \delta\}$, then the Gaussian complexity follows

$$\mathcal{G}(\mathbb{T}) \leq \min_{\delta \in [0,D]} \left\{ \mathcal{G}(\widetilde{\mathbb{T}}(\delta)) + 2\sqrt{D^2 \log N_2(\delta; \mathbb{T})} \right\}.$$

Here $\mathcal{G}(\widetilde{\mathbb{T}}(\delta)$ is referred to as a localized Gaussian complexity, and probably can be controlled by δ.

If $\mathbb{T}$ is a subset of $\mathbb{R}^d$, we have

$$\mathcal{G}(\widetilde{\mathbb{T}}(\delta)) = \mathbb{E}\left[\sup_{\theta \in \widetilde{\mathbb{T}}(\delta)} \langle \theta, w \rangle\right] \leq \delta \mathbb{E}\left[\|w\|_2\right] \leq \delta\sqrt{d}$$

Thus we have naive discretization bound

$$\mathcal{G}(\mathbb{T}) \leq \min_{\delta \in [0,D]} \left\{ \delta\sqrt{d} + 2\sqrt{D^2 \log N_2(\delta; \mathbb{T})} \right\}.$$

Example 3 (l_2 norm of sub-Gaussian random matrix)

Let $W \in \mathbb{R}^{x\times d}$ be a random matrix with zero-mean i.i.d. entries W_{ij}, each sub-Gaussian with parameter $\sigma = 1$. Then we have

$$\mathbb{E}\left[\|\mathbf{W}\|_2/\sqrt{n}\right] \precsim 1 + \sqrt{\frac{d}{n}}$$

Proof:

- Define matrix class $\mathbb{M}^{n,d}(1) := \left\{\mathbf{\Theta} \in \mathbb{R}^{n\times d} \mid \operatorname{rank}(\Theta) = 1, \|\Theta\|_F = 1\right\}$. First prove the variational representation:

$$\|W\|_2 = \sup_{\Theta\in\mathbb{M}(1)} X_\Theta, \text{ where } X_\Theta = \langle W, \Theta\rangle = \sum_{i,j} W_{ij}\Theta_{ij}$$

- Apply the discretization bound

$$\mathbb{E}\left[\|\mathbf{W}\|_2\right] \leq 2\mathbb{E}\left[\sup_{\substack{\operatorname{mank}(\mathbf{\Gamma})=\operatorname{rank}(\mathbf{\Gamma}')=1\\ \|\mathbf{\Gamma}-\mathbf{\Gamma}'\|_F\leq\delta}} \langle \mathbf{\Gamma} - \mathbf{\Gamma}', \mathbf{W}\rangle\right] + 6\sqrt{\log N_F\left(\delta; M^{n,d}(1)\right)}$$

One-step discretization

- Use SVD of rank 2 matrix to prove the local complexity bound

$$\mathbb{E}\left[\sup_{\substack{\text{rank}(\mathbf{\Gamma})=\text{rank}(\mathbf{\Gamma}')=1\\ \|\mathbf{\Gamma}-\mathbf{\Gamma}'\|_F\leq\delta}}\langle\mathbf{\Gamma}-\mathbf{\Gamma}',\mathbf{W}\rangle\right]\leq\sqrt{2}\delta\mathbb{E}\left[\mathbf{W}\|_2\right].$$

- Use upper bound of metric entropy on unit ball to prove

$$\log N_F\left(\delta;\mathbb{M}^{n,d}(1)\right)\leq(n+d)\log\left(1+\frac{2}{\delta}\right)$$

Recall the metric entropy on Lipschitz function class $\mathcal{F}_L$. Given sample x_1^n, we have $\log N_\infty\left(\delta,\mathcal{F}_L\right)\leq\frac{cL}{\delta}$, thus

$$\mathcal{G}\left(\mathcal{F}_L\left(x_1^n\right)/n\right)\leq\frac{1}{\sqrt{n}}\inf_{\delta\in(0,\delta_0)}\left\{\delta\sqrt{n}+3\sqrt{\frac{cL}{\delta}}\right\}$$

Choose $\delta=\frac{1}{n^3}$, we have

$$\mathcal{G}\left(\mathcal{F}_L(x_1^n)/n\right)\precsim n^{-1/3}$$

Chaining and Dudley's entropy integral

Define the Dudley's entropy integral $\mathcal{J}(\delta; D) := \int_\delta^D \sqrt{\log N_X(u; \mathbb{T})} du$. We have

Dudley's

$$\mathbb{E}\left[\sup_{\theta,\widetilde{\theta}\in\mathbb{T}} \left(X_\theta - X_{\widetilde{\theta}}\right)\right] \leq 2\mathbb{E}\left[\sup_{\substack{\gamma,\gamma'\in\mathbb{T}\\ \rho_X(\gamma,\gamma')\leq\delta}} \left(X_\gamma - X_{\gamma'}\right)\right] + 32\mathcal{J}(\delta/4; D)$$

- we pursue a more refined chaining argument. Denote $\mathbb{U} = \left\{\theta^1, \ldots, \theta^N\right\}$