Stat 6390.001: High Dimensional Probability and Statistics, Fall 2026 **Lecture 16**

Covering and packing (Sec 5.1); Gaussian and Rademacher complexity

Lecturer: Yun Wei *Scribe: Md Farhad Hossain*

October 19, 2026

Recall:

- VC dimension.
- Covering numbers.

Definition (Covering number)**.** A δ-cover of a set T w.r.t a metric ρ is a set $\{\theta^1, \ldots, \theta^N\} \subset T$ s.t. for every $\theta \in T$ there exists $i \in [N]$ with $\rho(\theta, \theta^i) \leq \delta$, equivalently,

$$T \subset \bigcup_{i=1}^{N} B(\theta^i, \delta), \qquad \text{where } B(\theta, \delta) := \{\theta' \in T \mid \rho(\theta, \theta') \leq \delta\}.$$

The δ-covering number $N(\delta; T, \rho)$ is the cardinality of the smallest δ-cover.

The covering number is a quantitative way to measure how compact a set is: T is totally bounded if and only if $N(\delta; T, \rho) < \infty$ for every $\delta > 0$.

Moreover, $N(\delta; T, \rho)$ is non-increasing in δ, and if T is infinite, $N(\delta; T, \rho) \to \infty$ as $\delta \downarrow 0$. The rate at which it grows to ∞ when $\delta \downarrow 0$ measures the size of T.

Definition (Packing number)**.** A δ-packing of a set T w.r.t a metric ρ is a set $\{\theta^1, \ldots, \theta^M\} \subset T$ s.t. $\rho(\theta^i, \theta^j) > \delta, \forall\, i, j \in [M], i \neq j$. The δ-packing number $M(\delta; T, \rho)$ is the cardinality of the largest δ-packing.

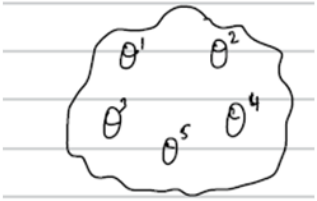


Fig 1: δ-covering of a set

We cannot place infinite points in δ-distance. There should be a maximum number, that is where the δ-packing is coming.

Remark. If N is δ-packing, then the closed $\delta/2$-balls centered in N is disjoint.

Proof. Suppose, $\exists\, i, j$ s.t.

$$A := B(\theta^i, \delta/2) \cap B(\theta^j, \delta/2) \neq \emptyset.$$

Find some $y \in A$.

$$\rho(\theta^i, \theta^j) \leq \rho(\theta^i, y) + \rho(\theta^j, y) \leq \delta. \quad \text{Contradiction.} \qquad \square$$

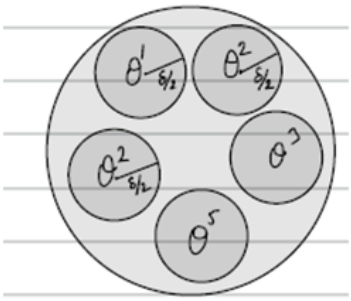


Fig 2: δ-packing of a set

Lemma. *For $\delta > 0$,*

$$M(2\delta; T, \rho) \le N(\delta; T, \rho) \le M(\delta; T, \rho)$$

Proof: Homework.

Recall: q-norm. $x \in \mathbb{R}^d$

$$\|x\|_q = \begin{cases} \Big(\sum_{i=1}^{d} |x_i|^q\Big)^{1/q}, & q \ge 1 \\ \max\limits_{1\le i\le d} |x_i|, & q = \infty \end{cases}$$

and $B_q^d(1) := \{x \in \mathbb{R}^d \mid \|x\|_q \le 1\}$.

Minkowski Sum:

$$A + B := \{a + b \mid a \in A,\ b \in B\}$$

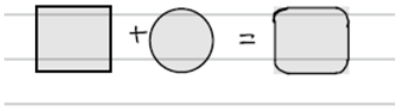


Fig 3: Geometric Illustration of the Minkowski Sum

We take on fixed point in A and connect all the points of B. So the ball can be anywhere.

Lemma. *Consider $K \subset \mathbb{R}^d$. Let $B := \{\theta \in \mathbb{R}^d \mid \|\theta\| \le 1\}$. Then*

$$\Big(\frac{1}{\delta}\Big)^d \frac{\mathrm{Vol}(K)}{\mathrm{Vol}(B)} \le N(\delta; K, \|\cdot\|) \le M(\delta; K, \|\cdot\|) \le \frac{\mathrm{Vol}(K + \frac{\delta}{2}B)}{\mathrm{Vol}(\frac{\delta}{2}B)} = \frac{\mathrm{Vol}(\frac{2}{\delta}K + B)}{\mathrm{Vol}(B)}$$

Proof. Lower bound: Let $N = N(\delta; K, \|\cdot\|)$. Consider a δ-covering $\{\theta^1, \theta^2, \dots, \theta^N\}$,

$$K \subset \bigcup_{i=1}^{N} B(\theta^i, \delta).$$

$$\mathrm{Vol}(K) \le \mathrm{Vol}\Big(\bigcup_{i=1}^{N} B(\theta^i, \delta)\Big) \le \sum_{i=1}^{N} \mathrm{Vol}\big(B(\theta^i, \delta)\big) = N \cdot \mathrm{Vol}\big(B(0, \delta)\big) = N \cdot \delta^d \,\mathrm{Vol}(B)$$

$$\therefore \Big(\frac{1}{\delta}\Big)^d \frac{\mathrm{Vol}(K)}{\mathrm{Vol}(B)} \le N \qquad \square$$

Upper bound: Let $M = M(\delta; K, \|\cdot\|)$. Consider $\{\theta^1, \theta^2, \dots, \theta^M\}$ to be δ-packing. By the remark (mentioned above) the closed balls $B(\theta^i, \delta/2)$ are disjoint.

$$\bigcup_{i=1}^{M} B(\theta^i, \delta/2) \subset K + \tfrac{\delta}{2}B = \bigcup_{\theta \in K} B(\theta, \delta/2)$$

$$\mathrm{Vol}\Big(\bigcup_{i=1}^{M} B(\theta^i, \delta/2)\Big) \le \mathrm{Vol}\big(K + \tfrac{\delta}{2}B\big)$$

$$\Rightarrow \sum_{i=1}^{M} \mathrm{Vol}\big(B(\theta^i, \delta/2)\big) \le \mathrm{Vol}\big(K + \tfrac{\delta}{2}B\big)$$

$$\Rightarrow M\,\mathrm{Vol}\big(\tfrac{\delta}{2}B\big) \le \mathrm{Vol}\big(K + \tfrac{\delta}{2}B\big)$$

$$\therefore M \le \frac{\mathrm{Vol}(K + \frac{\delta}{2}B)}{\mathrm{Vol}(\frac{\delta}{2}B)} = \frac{\cancel{(\delta/2)^d}\,\mathrm{Vol}(\frac{2}{\delta}K + B)}{\cancel{(\delta/2)^d}\,\mathrm{Vol}(B)} = \frac{\mathrm{Vol}(\frac{2}{\delta}K + B)}{\mathrm{Vol}(B)} \qquad \square$$

Remark. If $B \subset K$, $\mathrm{Vol}(\frac{2}{\delta}K + B) \le \mathrm{Vol}(\frac{2}{\delta}K + K) = \mathrm{Vol}\big((\frac{2}{\delta}+1)K\big) = (\frac{2}{\delta}+1)^d \mathrm{Vol}(K)$

$$\Big(\frac{1}{\delta}\Big)^d \boxed{\frac{\mathrm{Vol}(K)}{\mathrm{Vol}(B)}} \le N(\delta; K, \|\cdot\|) \le M(\delta; K, \|\cdot\|) \le \Big(\frac{2}{\delta}+1\Big)^d \boxed{\frac{\mathrm{Vol}(K)}{\mathrm{Vol}(B)}}.$$

behaves lik δ^{-d} when $\delta \downarrow 0$ since $\boxed{\frac{\mathrm{Vol}(K)}{\mathrm{Vol}(B)}}$ does not depends on δ.

Remark. If $K = B$

$$\Big(\frac{1}{\delta}\Big)^d \le N(\delta; B, \|\cdot\|) \le M(\delta; B, \|\cdot\|) \le \Big(\frac{2}{\delta}+1\Big)^d$$

Example.

$$\Big(\frac{1}{\delta}\Big)^d \le N\big(\delta, \underbrace{B_\infty^d(1)}_{\ell_\infty\text{-unit ball}}, \|\cdot\|_\infty\big) \le \Big(\frac{2}{\delta}+1\Big)^d$$

$$\Big(\frac{1}{\delta}\Big) \le N\big(\delta, \underbrace{B_2^d(1)}_{\ell_2\text{-unit ball}}, \|\cdot\|_2\big) \le \Big(\frac{2}{\delta}+1\Big)^d$$

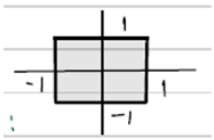
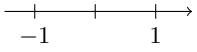


Fig 4: Illustration of the Unit Ball in the l_∞-Norm

Recall: $d = 1$, 

$$N(\delta, [-1,1], |\cdot|) \le \Big\lceil \frac{2}{\delta} \Big\rceil \le \frac{2}{\delta} + 1$$

5.2 Gaussian and Rademacher Complexity

$$T \underset{\text{use random process to study } T.\ \text{(Sec: 5.2)}}{\overset{\text{study random process (Sec 5.3, 5.4, } \dots\text{)}}{\rightleftarrows}} \{X_\theta \mid \theta \in T\}$$

Recall: Let $T \subset \mathbb{R}^d$.

Definition (Rademacher process and Rademacher complexity)**.** Let $\varepsilon = (\varepsilon_1, \dots, \varepsilon_d)$, where ε_i are iid Rademacher r.v., i.e., $\mathbb{P}(\varepsilon_i = 1) = \mathbb{P}(\varepsilon_i = -1) = 1/2$. For each $\theta \in T$, define

$$R_\theta := \langle \varepsilon, \theta \rangle = \sum_{i=1}^d \varepsilon_i \theta_i.$$

The collection $\{R_\theta \mid \theta \in T\}$ is a random process indexed by T, called the Rademacher process. The Rademacher complexity of T is

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

Definition (Canonical Gaussian process and Gaussian complexity)**.** Let $w = (w_1, \dots, w_d)$, where w_i are iid $N(0,1)$. For each $\theta \in T$, define

$$G_\theta := \langle w, \theta \rangle = \sum_{i=1}^d w_i \theta_i.$$

The collection $\{G_\theta \mid \theta \in T\}$ is called the canonical Gaussian process indexed by T. The Gaussian complexity of T is

$$\mathcal{G}(T) := \mathbb{E}\Big[\sup_{\theta \in T} G_\theta\Big].$$

Homework:

$$R(T) \leq \sqrt{\pi/2}\, G(T)$$

Example. $B_2^d(1)$

$$R(B_2^d) = \mathbb{E}\Big[\sup_{\theta \in B_2^d} \langle \theta, \varepsilon \rangle\Big] = \mathbb{E}\big[\|\varepsilon\|_2\big] = \mathbb{E}\Big[\sqrt{\textstyle\sum_{i=1}^d \varepsilon_i^2}\Big] = \sqrt{d}.$$

References

[1] M. J. Wainwright, *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*, Cambridge University Press, 2019. Sec 5.1 and Sec 5.2