Spring 2018: STA 6448
Advanced Probability and Inference II
Lecture 8

Yun Yang

- Uniform laws of large numbers and metric entropy

Application: Classical Glivenko-Cantelli theorem

Recall the classical Glivenko-Cantelli theorem on the uniform convergence of CDFs:

$$\|\widehat{F}_n - F\|_\infty \xrightarrow{\text{a.s.}} 0,$$

Corollary

Let F be the cdf and $\widehat{F}_n$ the empirical CDF, then

$$\mathbb{P}\left[\|\widehat{F}_n - F\|_\infty \geq \sqrt{\frac{2\log(2(n+1))}{n}} + \delta\right] \leq 2e^{-\frac{n\delta^2}{8}} \quad \textit{for all } \delta > 0,$$

and hence $\|\widehat{F}_n - F\|_\infty \xrightarrow{\text{a.s.}} 0$.

Proof: Take $\mathcal{F} = \{(-\infty, t] : t \in \mathbb{R}\}$, then $\mathcal{F}$ is uniformly bounded by 1, and has polynomial growth of order 1.

The bound is not tight (the $\log(n+1)$ factor can be removed).

Vapnik-Chervonenkis (VC) dimension

Definition

A class $\mathcal{F} \subset \{0,1\}^{\mathcal{X}}$ shatters $(x_1, \ldots, x_d) \subset \mathcal{X}$ means $|\mathcal{F}(x_1^d)| = 2^d$.

The VC-dimension $d_{VC}(\mathcal{F})$ is defined as the largest integer d for which there is some $(x_1, \ldots, x_d) \subset \mathcal{X}$ of d points that can be shattered by $\mathcal{F}$.

Examples

- $\mathcal{F}_{\text{left}} = \{(-\infty, t] : t \in \mathbb{R}\}$ has VC-dim 1. It has polynomial growth of order 1.
- $\mathcal{F}_{\text{two}} = \{(s, t] : s, t \in \mathbb{R}\}$ has VC-dim 2. It has polynomial growth of order 2 (why?).

Vapnik-Chervonenkis (VC) dimension

Theorem (Sauer's Lemma)

If $d_{VC}(\mathcal{F}) \leq d$*, then*

$$\Pi_{\mathcal{F}}(n) \leq \sum_{k=1}^{d} \binom{n}{k} \leq (n+1)^d.$$

Consequently, if $d_{VC}(\mathcal{F}) < \infty$ *(called VC class), then* $\mathcal{F}$ *has polynomial growth of order* $d_{VC}(\mathcal{F})$*.*

Proof: See “Weak convergence and empirical processes: with applications to statistics", Section 2.6.1.

Some useful results on Rademacher complexity

Properties

1. $\mathcal{F}_1 \subset \mathcal{F}_2$ implies $\mathcal{R}_n(\mathcal{F}_1) \leq \mathcal{R}_n(\mathcal{F}_2)$.
2. For any constant $c \in \mathbb{R}$, $\mathcal{R}_n(c\,\mathcal{F}) = |c|\,\mathcal{R}_n(\mathcal{F})$.
3. For any fixed bounded function g (bounded by b), $\left|\mathcal{R}_n(\mathcal{F}+g) - \mathcal{R}_n(\mathcal{F})\right| \leq b\,\sqrt{2\log 2/n}$.
4. $\mathcal{R}_n(\text{conv}(\mathcal{F})) = \mathcal{R}_n(\mathcal{F})$, where $\text{conv}(\mathcal{F})$ is the convex hull of $\mathcal{F}$.
5. If $\phi : \mathbb{R} \to \mathbb{R}$ is 1-Lipschitz continuous and satisfies $\phi(0) = 0$, then $\mathcal{R}(\phi(\mathcal{F})) \leq 2\mathcal{R}(\mathcal{F})$.

For a proof of the last claim, see “Probability in Banach Spaces” by Michel Ledoux and Michel Talagrand, Theorem 4.12.

Covering and packing numbers

A way to measure the "size" of a set with infinitely many elements. Recall:

Definition

A metric space $(\mathbb{T}, \rho)$ consists of a non-empty set $\mathbb{T}$ equipped with a mapping $\rho : \mathbb{T} \times \mathbb{T} \to [0, \infty)$ satisfying:

1. $\rho(\theta, \theta') = 0$ if and only if $\theta = \theta'$;
2. It is symmetric: $\rho(\theta, \theta') = \rho(\theta', \theta)$;
3. Triangle inequality: $\rho(\theta, \theta'') \leq \rho(\theta, \theta') + \rho(\theta', \theta'')$.

If the first property is replaced with $\rho(\theta, \theta) = 0$, then $(\mathbb{T}, \rho)$ is called a pseudometric space.

Examples: Euclidean space $(\mathbb{R}^n, \|\cdot\|_2)$, function space $(L^2[0, 1], \|\cdot\|_\infty)$, function space with pseudometric $\rho(f, g) = \|f - g\|_n = \sqrt{n^{-1} \sum_{i=1}^n [f(x_i) - g(x_i)]^2}$.

Covering number

Definition

An ε-cover of a set $\mathbb{T}$ w.r.t. a metric ρ is a set $\{\theta^1, \ldots, \theta^N\} \subset \mathbb{T}$ such that for each $\theta \in \mathbb{T}$, there exists some $i \in \{1, \ldots, N\}$, $\rho(\theta, \theta^i) \le \varepsilon$. The ε**-covering number** $N(\varepsilon, \mathbb{T}, \rho)$ is the smallest cardinality of all ε-covers.

A set $\mathbb{T}$ is **totally bounded** if for all $\varepsilon > 0$, $N(\varepsilon, \mathbb{T}, \rho) < \infty$ (compact?).

The function $\varepsilon \mapsto \log N(\varepsilon, \mathbb{T}, \rho)$ is the **metric entropy** of $\mathbb{T}$ w.r.t. ρ.

$N(\varepsilon, \mathbb{T}, \rho)$ is non-increasing in ε. Often interested in the growth of metric entropy as $\varepsilon \to 0_+$. If $\lim_{\varepsilon \to 0_+} \log N(\varepsilon)/\log(1/\varepsilon)$ exists, it is called the **metric dimension**.

Example: Covering number of unit cubes

Example

Consider interval $[-1, 1]$ in $\mathbb{R}$, equipped with the Euclidean metric $|\cdot|$. Then we have

$$N(\varepsilon, [-1, 1], |\cdot|) \leq \frac{1}{\varepsilon} + 1, \quad \text{for all } \varepsilon > 0.$$

More generally, for the d-dim cube $[-1, 1]^d$, we have $N(\varepsilon, [-1, 1]^d, \|\cdot\|_\infty) \leq \left(\frac{1}{\varepsilon} + 1\right)^d$, and its metric dimension is d.

Packing number

Definition

An ε-packing of a set $\mathbb{T}$ w.r.t. a metric ρ is a set $\{\theta^1, \ldots, \theta^M\} \subset \mathbb{T}$ such that $\rho(\theta^i, \theta^j) > \varepsilon$ for all distinct pairs $(i, j) \in \{1, \ldots, M\}^2$. The **$\varepsilon$-packing number** $M(\varepsilon, \mathbb{T}, \rho)$ is the largest cardinality of all ε-packings.

Covering and packing relation

Theorem

For all $\varepsilon > 0$, the packing and covering numbers are related by:

$$M(2\varepsilon, \mathbb{T}, \rho) \leq N(\varepsilon, \mathbb{T}, \rho) \leq M(\varepsilon, \mathbb{T}, \rho).$$

Thus, the scalings of the covering and packing numbers are the same.

Example: Packing number of unit cubes

Example

Consider interval $[-1,\ 1]$ in $\mathbb{R}$, equipped with the Euclidean metric $|\cdot|$. Then we have

$$M(2\varepsilon,\ [-1,\ 1],\ |\cdot|) \geq \left\lfloor \frac{1}{\varepsilon} \right\rfloor, \quad \text{for all } \varepsilon > 0.$$

Therefore, from the previous theorem, we can conclude

$$\log N(\varepsilon,\ [-1,\ 1],\ |\cdot|) \asymp \log \frac{1}{\varepsilon}, \quad \text{for all } \varepsilon > 0.$$

More generally, for the d-dim cube $[-1,\ 1]^d$, we have $\log N(\varepsilon,\ [-1,\ 1]^d,\ \|\cdot\|_\infty) \asymp d \log(1/\varepsilon)$.

Volume ratios and metric entropy

Theorem

Consider a pair of norms $\|\cdot\|_1$ and $\|\cdot\|_2$ on $\mathbb{R}^d$, Let $\mathbb{B}_1$ and $\mathbb{B}_2$ be the corresponding unit balls. The the ε-covering number of $\mathbb{B}_1$ in the $\|\cdot\|_2$ norm satisfies

$$\left(\frac{1}{\varepsilon}\right)^d \frac{\textit{vol}\,(\mathbb{B}_1)}{\textit{vol}\,(\mathbb{B}_2)} \leq N(\varepsilon,\, \mathbb{B},\, \|\cdot\|_2) \leq \frac{\textit{vol}\,(\frac{2}{\varepsilon}\,\mathbb{B}_1 + \mathbb{B}_2)}{\textit{vol}\,(\mathbb{B}_2)}.$$

In particular, if $\|\cdot\|_1 = \|\cdot\|_2 = \|\cdot\|$, then

$$\left(\frac{1}{\varepsilon}\right)^d \leq N(\varepsilon,\, \mathbb{B},\, \|\cdot\|) \leq \left(\frac{2}{\varepsilon} + 1\right)^d.$$