Lecture 13: Packing and Covering Numbers (October 17)

Lecturer: Csaba Szepesvári *Scribes: Kushagra Chandak*

Let us recall the definition of a cover. For the remainder of this section, let (V, d) be a pseudometric space.

Definition 13.1 (Cover). Let $\mathcal{G} \subseteq V$. $G(\varepsilon) \subseteq V$ is an ε-cover of $\mathcal{G}$ if for all $g \in \mathcal{G}$ there exitst $g' \in \mathcal{G}(\varepsilon)$ such that $d(g, g') \leq \varepsilon$.

If $\mathcal{G}(\varepsilon) \subseteq \mathcal{G}$, then it is called the **inside cover**. Inside covers can be useful in the cases when functions in $\mathcal{G}$ have some special properties, e.g., to use Bernstein inequality. The **covering number** of $\mathcal{G}$ at scale ε is given by

$$\mathcal{N}(\varepsilon, \mathcal{G}, d) = \min \{|\mathcal{G}(\varepsilon)| \,:\, \mathcal{G}(\varepsilon) \text{ is an } \varepsilon\text{-cover of } \mathcal{G}\} \,.$$

Note that $\mathcal{N}_{\text{ins}}(\varepsilon) \geq \mathcal{N}(\varepsilon)$ (dropping dependencies of covering numbers on $\mathcal{G}$ and d).

Definition 13.2 (Bracket). $G(\varepsilon) = \{(g_1^L, g_1^U), \dots, (g_m^L, g_m^U)\}$ is an ε-bracket of $\mathcal{G}$ if for all $g \in \mathcal{G}$ there exists $i \in [m]$ such that $g_i^L \leq g \leq g_i^U$ and $d(g_i^U, g_i^L) \leq \varepsilon$.

The bracketing number is given by

$$\mathcal{N}_{[]} = \min \{|\mathcal{G}(\varepsilon)| \,:\, \mathcal{G}(\varepsilon) \text{ is an } \varepsilon\text{-bracket}\} \,.$$

For covering, note that $\mathcal{G}$ may not be a function class but for bracketing $\mathcal{G}$ is usually a function class.

Covering and bracketing numbers are related via the following inequalities.

$$\begin{aligned} &\mathcal{N}(\varepsilon, \mathcal{G}, L_p(P)) \leq \mathcal{N}_{[]}(2\varepsilon, \mathcal{G}, L_p(P)) \\ &\mathcal{N}_{[]}(\varepsilon, \mathcal{G}, L_\infty(P)) \leq \mathcal{N}(\varepsilon/2, \mathcal{G}, L_\infty(P)) \,. \end{aligned} \tag{13.1}$$

Definition 13.3 (Packing). $\mathcal{G}(\varepsilon) \subseteq \mathcal{G}$ is an ε-packing of $\mathcal{G}$ if for any $g, g' \in \mathcal{G}(\varepsilon)$, $d(g, g') > \varepsilon$.

The packing number of $\mathcal{G}$ at scale ε is defined as

$$\mathcal{M}(\varepsilon, \mathcal{G}, d) = \max \{\mathcal{G}(\varepsilon) \,:\, \mathcal{G}(\varepsilon) \text{ is an } \varepsilon\text{-packing of } \mathcal{G}\} \,.$$

We have the following result stating that covering and packing numbers are almost the same with appropriate scale.

Proposition 13.4. *The following holds for a set $\mathcal{G}$ with a pseudometric d.*

$$\mathcal{N}(\varepsilon) \leq \mathcal{N}_{ins}(\varepsilon) \overset{(1)}{\leq} \mathcal{M}(\varepsilon) \overset{(2)}{\leq} \mathcal{N}(\varepsilon/2) \leq \mathcal{N}_{ins}(\varepsilon/2) \,.$$

Proof. The inequalities $\mathcal{N}(\varepsilon) \leq \mathcal{N}_{\text{ins}}(\varepsilon)$ and $\mathcal{N}(\varepsilon/2) \leq \mathcal{N}_{\text{ins}}(\varepsilon/2)$ directly follow from the fact that if $A \subseteq B$, then $\min A \geq \min B$.

To prove (1), we start by picking a maximal packing $\mathcal{G}(\varepsilon) \subseteq \mathcal{G}$. Therefore, $|\mathcal{G}(\varepsilon)| = \mathcal{M}(\varepsilon)$. Note that $\mathcal{G}(\varepsilon)$ is also an ε-cover (not minimum). [If it was not, then there must be a $g \in \mathcal{G}$ such that for all $g' \in \mathcal{G}(\varepsilon)$, $d(g, g') > \varepsilon$, which means g can be added to the packing $\mathcal{G}(\varepsilon)$ contradicting that it is maximal.] Therefore, the cardinality of the minimum (inside) cover should be less than or equal to $\mathcal{M}(\varepsilon)$, or $\mathcal{N}_{\text{ins}}(\varepsilon) \leq \mathcal{M}(\varepsilon)$.

To prove (2), we start by picking a maximal packing $\mathcal{G}(\varepsilon)$ and *any* cover $\mathcal{G}'(\varepsilon/2)$. Now we show an injective mapping ψ between $\mathcal{G}(\varepsilon)$ and $\mathcal{G}'(\varepsilon/2)$ which will imply that $|\mathcal{G}(\varepsilon)| \leq |\mathcal{G}'(\varepsilon/2)|$ proving (2).

We map $g_1 \in \mathcal{G}(\varepsilon)$ to its cover element $g' \in \mathcal{G}'(\varepsilon/2)$. That is, $\psi(g_1) = g'$ for any g' such that $d(g_1, g') \leq \varepsilon/2$, since $\mathcal{G}'(\varepsilon/2)$ is a cover. If there are multiple g', pick an arbitrary one. Now we show that ψ is injective, that is, for $g_1, g_2 \in \mathcal{G}(\varepsilon)$ such that $g_1 \neq g_2$, $\psi(g_1) \neq \psi(g_2)$. We start by the definition of packing:

$$\varepsilon < d(g_1, g_2) \leq d(g_1, \psi(g_1)) + d(\psi(g_1), g_2) \leq \frac{\varepsilon}{2} + d(\psi(g_1), g_2)\,.$$

This gives $d(\psi(g_1), g_2) > \varepsilon/2 \geq d(\psi(g_2), g_2)$. If $\psi(g_1) = \psi(g_2)$ then the previous inequality is always false ($d(g', g_2) > \varepsilon/2$ and $d(g', g_2) \leq \varepsilon/2$). Therefore $\psi(g_1) \neq \psi(g_2)$ for $g_1 \neq g_2$ and ψ is injective. □

Next we show upper and lower bounds on covering and packing numbers for a ball of radius r in k dimensions. We will see that both packing and covering numbers are of the order of $\sim (r/\varepsilon)^k$ using some volume arguments and Proposition 13.4. For the pseudometric, we will just denote the norm that induces it.

Proposition 13.5. *Let $B(r)$ ball in $\mathbb{R}^k$ with radius r and center 0. We have the following.*

1. $\mathcal{M}(\varepsilon, B(r), \|\cdot\|) \leq \left(1 + \frac{2r}{\varepsilon}\right)^k$.
2. $\mathcal{N}(\varepsilon, B(r), \|\cdot\|) \geq \left(\frac{r}{\varepsilon}\right)^k$.

Proof. Let m be the Lebesgue measure on R^k. We have that in $\mathbb{R}^k$, $m(B(r)) = c \cdot r^k$ for some $c > 0$.

1. Let $\mathcal{Z} \subseteq B(r)$ be the maximal (countable) ε-packing of $B(r)$. Since in ε-packing, balls for radius ε/2 are disjoint, we have that for $z, z' \in \mathcal{Z}$

$$B\left(z, \frac{\varepsilon}{2}\right) \cap B\left(z', \frac{\varepsilon}{2}\right) = \varnothing\,,$$

which implies

$$m\left(\bigcup_{z\in\mathcal{Z}} B\left(z, \frac{\varepsilon}{2}\right)\right) = \sum_{z\in\mathcal{Z}} m\left(B\left(z, \frac{\varepsilon}{2}\right)\right) = \mathcal{M}(\varepsilon)\cdot c \cdot \left(\frac{\varepsilon}{2}\right)^k\,.$$

Finally, we also have that the volume of a ball with radius $r + \varepsilon/2$ is as large as volume of the cover:

$$m\left(\bigcup_{z\in\mathcal{Z}} B\left(z, \frac{\varepsilon}{2}\right)\right) \leq m\left(B\left(r + \frac{\varepsilon}{2}\right)\right) = c\cdot\left(r + \frac{\varepsilon}{2}\right)^k\,.$$

Therefore,

$$\mathcal{M}(\varepsilon)\cdot c\cdot\left(\frac{\varepsilon}{2}\right)^k \leq c\cdot\left(r + \frac{\varepsilon}{2}\right)^k$$

finishing the proof.

2. Let $\mathcal{Z} \subseteq \mathbb{R}^k$ be any (finite) ε-cover of $B(r)$. We have $B(r) \subseteq \bigcup_{z\in\mathcal{Z}} B(z, \varepsilon)$, which gives

$$\begin{aligned} & m(B(r)) \leq |\mathcal{Z}|\cdot m(B(z,\varepsilon)) \\ \Longrightarrow\ & cr^k \leq |\mathcal{Z}|\cdot c\cdot\varepsilon^k \\ \Longrightarrow\ & |\mathcal{Z}| \geq \left(\frac{r}{\varepsilon}\right)^k\,. \end{aligned}$$

Since the above is true for any ε-cover, it is true for the minimum cover as well.

□

Since $\mathcal{N}(\varepsilon) \leq \mathcal{M}(\varepsilon)$ by Proposition 13.4, we also have that $\mathcal{N}(\varepsilon, B(r), \|\cdot\|) \leq \left(1 + \frac{2r}{\varepsilon}\right)^k$.

Next we see that the bracketing number of a Lipschitz function class is characterized by the Lipschitz constant. Let us recall the definition of a Lipschitz function. For two metric spaces (U, d_u) and (V, d_v), a function $f : U \to V$ is **Lipschitz** if for all $u, u' \in U$, $d_v(f(u), f(u')) \leq L \cdot d_u(u, u')$ for some $L > 0$.

Let $\mathcal{G} = \{g_w : g_w : \mathcal{Z} \to \mathbb{R}\}$ be a function class parameterized by w, where $w \in \mathcal{W} \subseteq \mathbb{R}^k$.

Proposition 13.6. *Let $|g_w - g'_w| \leq \gamma \|w - w'\|$ for all $w, w' \in \mathcal{W} \subseteq B(r)$, where*

$$\gamma : \mathcal{Z} \to [0, \infty) \quad ; \quad \gamma_p = \|\gamma\|_{L_p} .$$

Then

$$\mathcal{N}_{[\,]}(2\varepsilon, \mathcal{G}, L_p) \leq \left(1 + \frac{2\gamma_p r}{\varepsilon}\right)^k .$$