THE UNIVERSITY OF TEXAS AT AUSTIN
Department of Statistics and Data Sciences
College of Natural Sciences

SDS 384 11: Theoretical Statistics

Lecture 14: Uniform Law of Large Numbers- Covering number

Purnamrita Sarkar
Department of Statistics and Data Science
The University of Texas at Austin

Definitions

- Recall that a metric space $(\mathcal{T}, \rho)$ consists of a nonempty set $\mathcal{T}$ and a mapping $\rho : \mathcal{T} \times \mathcal{T} \to \mathbb{R}$ that satisfies:
  - Non-negative: $\rho(\theta, \theta') \geq 0$ for all (θ, θ') with equality iff $\theta = \theta'$.
  - Symmetric: $\rho(\theta, \theta') = \rho(\theta', \theta)$ for all pairs (θ', θ), and
  - Triangle ineq holds: $\rho(\theta, \theta') + \rho(\theta', \theta'') \geq \rho(\theta, \theta'')$

- Examples:
  - $\mathcal{T} = \mathbb{R}^d$, $\rho(\theta, \theta') = \|\theta - \theta'\|_2$
  - $\mathcal{T} = \{0, 1\}^d$ with $\rho(\theta, \theta') = \dfrac{1}{d} \sum_i 1(\theta_i \neq \theta'_i)$

Covering numbers

Definition

A δ cover of a set $\mathcal{T}$ w.r.t to a metric ρ is a set $\{\theta^1, \dots, \theta^N\}$ such that for every $\theta \in \mathcal{T}$, $\exists i \in [N]$, s.t. $\rho(\theta, theta^i) \leq \delta$. The δ covering number $N(\delta; \mathcal{T}, \rho)$ is the cardinality of the smallest δ cover.

- We will consider metric spaces which are totally bounded, i.e. $N(\delta; \mathcal{T}, \rho) < \infty$ for all $\delta > 0$.
- The covering number is non-increasing in δ, i.e. $N(\delta) \geq N(\delta')$ for all $\delta < \delta'$
- We are interested in something called Metric entropy, which is the logarithm of the covering number.

Picture

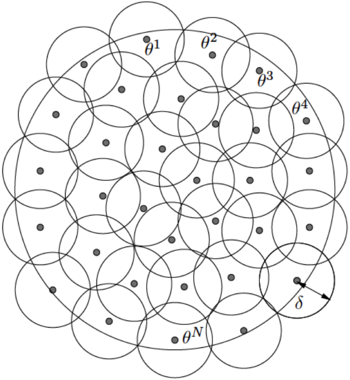

Figure 1: [courtesy: Martin Wainwright's book]

- A δ covering can be thought of as a union of balls with radius δ.

Coevring number of a unit cube

Example

Consider the interval $[-1, 1]$ with $\rho(\theta, \theta') = |\theta - \theta'|$. We have $N(\delta; [-1, 1], |.|) \leq \frac{1}{\delta} + 1$

- Divide the interval into L sub-intervals centered at $\theta^i := -1 + (2i - 1)\delta$ for $i \in [L]$ and each of length at most 2δ.
- By construction this is a δ covering.
- So $L \leq 1 + 1/\delta$

Covering the binary hypercube

Example

Consider a d dimensional binary hypercube $\mathcal{T} = \{0,1\}^d$ with the Hamming metric defined before.

$$\frac{\log N(\delta; \mathcal{T}, \rho)}{\log 2} \leq \lceil d(1-\delta) \rceil$$

- Let $S = \{1, 2, \ldots, \lceil \delta d \rceil\}$
- Consider the set of binary vectors $\mathcal{S}(\delta) := \{\theta \in \mathcal{T} : \theta_j = 0\}$.
- By construction, for every binary vector $\theta' \in \mathcal{T}$, we can find a vector $\theta \in \mathcal{S}(\delta)$ such that $\rho(\theta, \theta') \leq \delta$
- $N(\delta; \mathcal{T}, \rho) \leq |\mathcal{S}(\delta)| = 2^{\lceil d(1-\delta) \rceil}$

Lower bound on Covering number of the binary hypercube

- Let $\delta \in (0, 1/2)$
- If $\{\theta^1, \dots, \theta^N\}$ is a δ covering, then the (unrescaled) Hamming balls of radius $s = \delta d$ around each θ^ℓ must contain all 2^d vectors.
- Let $s = \lfloor \delta d \rfloor$
- For each θ^i there are exactly $\sum_{j=0}^{d} \binom{d}{j}$ vectors within δd distance.
- So $N \sum_{j=0}^{d} \binom{d}{j} \geq 2^d$

Lower bound on Covering number of the binary hypercube

- Let $\delta \in (0, 1/2)$
- So $N \sum_{j=0}^{s} \binom{d}{j} \geq 2^d$
- Now take a Binomial $(d, 1/2)$ random variable X.
- $P(X \leq \delta d) = \sum_{j=0}^{s} \binom{d}{j} / 2^d$
- So $N \geq \dfrac{1}{P(X \leq \delta d)}$
- Using the Hoeffding bound gives: $N \geq \exp(\frac{d}{2}(1/2 - \delta)^2)$
- Using the refined version in your homework gives: $N \geq \exp(dKL(\delta||1/2))$