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

SDS 384 11: Theoretical Statistics

Lecture 15: Uniform Law of Large Numbers-Rademacher and Gaussian Complexity

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

Make a comparison

- Recall that for a L Lipschitz continuous functions supported on $[0, 1]$ with $f(0) = 0$, the metric entropy was L/δ
- Also recall that for a L Lipschitz continuous functions supported on $[0, 1]^d$ with $f(0) = 0$, the metric entropy was $(L/\delta)^d$
- However for a given function class like the last one the metric entropy is $\log(1/\delta)$
- Recall that for Unit hypercubes in d dimensions the metric entropy is $d \log(1 + 1/\delta)$
- Note that for Lipschitz continuous functions the dependence on d is exponential. This is a much richer class of functions, so the size is considerably larger and scales poorly with d.

Application-Random matrix singular value

Theorem

Consider a random matrix $M = (\xi_{ij})_{i,j \in [n]}$ where ξ_{ij} are standard normal random variables.

$$P(\|M\|_{op} \geq A\sqrt{n}) \leq C \exp(-cAn)$$

where c, C are absolute constants and $A \geq C$.

- This works for symmetric wigner ensembles and hermitian matrices as well.

Operator norm

- Let $S_n := \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$

- $\|M\|_{op} := \sup\limits_{x \in \mathbb{R}^n} \|Mx\|$

- First note that we have

$$P(\|Mx\| \geq A\sqrt{n}) \leq C \exp(-cAn)$$

- This is because for each row M_i, we have

$$M_i^T x \sim Subgaussian(1), (M_i^T x)^2 - 1 \sim Subexponential(2, 4)$$

- $\|Mx\|^2 - n \sim Subexponential(2\sqrt{n}, 4)$

Recall sub-exponential random variables?

Theorem

Let X be a sub-exponential random variable with parameters (ν, b). Then,

$$P(X \geq \mu + t) \leq \begin{cases} e^{-\frac{t^2}{2\nu^2}} & \text{if } 0 \leq t \leq \dfrac{\nu^2}{b} \\ e^{-\frac{t}{2b}} & \text{if } t \geq \dfrac{\nu^2}{b} \end{cases}$$

- $P(\|Mx\|^2 - n \geq Cn) \leq e^{-Cn/8}$, $C > 1$.

Can I just use an Union bound?

- Not really.
- But I can form a $1/2$ cover of S_n.
- Find $\mathcal{C} = \{x^1, \ldots, x^N\}$ such that for all $x \in S_n$, $\exists x^i \in S$
  $\|x - x^i\| \leq 1/2$.
- Consider $y \in S$ such that $\|My\| = \|M\|_{op}$. Let x^i be a member of the
  $1/2$ cover s.t. $\|y - x^i\| \leq 1/2$
- So $\|M(y - x^i)\| \leq \|M\|_{op}/2$ and
  $\|M(y - x^i)\| \geq \|My\| - \|Mx^i\| \geq \|M\|_{op} - \|Mx^i\|$.
- Hence $\|Mx^i\| \geq \|M\|_{op}/2$

Using the covering number

$$P(\|M\|_{op} \geq \sqrt{(C+1)n}) \leq P(\exists x^i \in \mathcal{C}, \|Mx^i\| \geq \sqrt{(C+1)n}/2)$$
$$\leq |\mathcal{C}|P(\|Mx^i\| \geq \sqrt{(C+1)n}/4)$$
$$\leq |\mathcal{C}|P(\|Mx^i\|^2 - n \geq (C-3)n/4)$$

$C > 7$ gives $(C-3)n/4 \geq \nu^2/b$ $\qquad \leq |\mathcal{C}|\exp(-(C-3)n/32)$

- ϵ covering number of the unit ball in n dimensions is bounded by $(1 + 2/\epsilon)^n$

$$P(\|M\|_{op} \geq \sqrt{(C+1)n}) \leq 5^n \exp(-(C-3)n/32)$$
$$\leq \exp(-n((C-3)/32 - 1.6))$$

- So C will have to be something like 55!!

Kernel density estimation

Let $X_1, X_2, \ldots, X_n$ be i.i.d. samples of random variable with density f on the real line with support $[0, 1]$. A standard estimate of f is the kernel density estimate

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)$$

where $K : \mathbb{R} \to [0, \infty]$ is a kernel function satisfying $\int_{-\infty}^{\infty} K(t)dt = 1$, and h is a bandwidth parameter. Also assume that $|K(x) - K(y)| \leq L|x - y|$. Let $K(x) \leq K(0)$.

We are interested in the quantity $\sup_{x \in [0,1]} |\hat{f}(x) - E[\hat{f}(x)]|$

Kernel Density Estimation

- First do a ϵ cover of x by $\mathcal{C} := \{x^1, \ldots, x^N\}$.
- Let $\tilde{K}((x - X_i)/h) = K(.) - EK(.)$
- Similarly $\tilde{f}(.) = \hat{f}(.) - E[\hat{f}(.)]$
- The Lipschitz condition gives
$$\left| \tilde{K}\left(\frac{x - X_i}{h}\right) - \tilde{K}\left(\frac{y - X_i}{h}\right) \right| \leq \frac{2L|x - y|}{h}$$
- So $|\tilde{f}(x) - \tilde{f}(x^i)| \leq \dfrac{2L|x - x^i|}{h^2}$
- So this gives a $2L\epsilon/h^2$ cover for the $\tilde{f}$ values.

Kernel Density Estimation

- Let y be the point where $\sup\limits_{x \in [0,1]} |\tilde{f}(x)|$ is achieved.

- There exists a i such that $|\tilde{f}(y) - \tilde{f}(x^i)| \leq 2L\epsilon/h^2$

- So $\exists i, |\tilde{f}(x^i)| \geq \sup\limits_{x \in [0,1]} |\tilde{f}(x)| - 2L\epsilon/h^2$

- Finally

$$P\left(\sup_{x \in [0,1]} |\tilde{f}(x)| \geq \delta\right) \leq P(\exists i \in \mathcal{C}, |\tilde{f}(x^i)| \geq \sup_{x \in [0,1]} |\tilde{f}(x)| - 2L\epsilon/h^2)$$

$$\leq |\mathcal{C}| P\left(|\tilde{f}(x^i)| \geq \delta - 2L\epsilon/h^2\right)$$

- Set $\delta = 4L\epsilon/h^2$, the RHS can be obtained using Hoeffding.

Kernel Density Estimation

- Hoeffding bound gives:

$$P(|\tilde{f}(x^i)| \geq \delta/2) \leq 2 \exp\left(-\frac{nh^2\delta^2}{2}\right)$$

- Also, the covering number of a d dimensional unit sphere is upper bounded by $(1 + 2/\epsilon)^d$.

- Now plug in $\epsilon = \delta h^2/4L$

-

$$P\left(\sum_{x \in [0,1]} |\hat{f}(x) - E[\hat{f}(x)]| \geq \delta\right) \leq 2\left(1 + \frac{8L}{\delta h^2}\right)^d \exp\left(-\frac{nh^2\delta^2}{2}\right)$$