Chapter 9. Verifying the RIP through the JL Lemma

support set S with $|S| \le k$ $\qquad$ $\mathcal{X}_S$ ... set of vectors supported on S

$$y = \underset{m\times n}{\Phi} \underset{n\times 1}{x} + n$$

set of vectors with (fixed) support set S is a k-dim. lin. space

count no. of k-dim. spaces

Lemma 9.1. Let $\Phi \in \mathbb{R}^{m\times n}$ be an i.i.d. $\mathcal{N}(0, 1/m)$ matrix. Then, for every set S with $|S| = k < m$ and every $\delta \in (0,1)$, we have

$$(1-\delta)\,\|\underline{x}\| \le \|\Phi \underline{x}\| \le (1+\delta)\,\|\underline{x}\|, \quad \forall x \in \mathcal{X}_S$$

w.p. $\qquad$ Lem. 8.2.

$$\ge 1 - 2(12/\delta)^k e^{-c_0(\delta/2)m}$$

where $c_0(x) = \frac{1}{4}(x^2 - x^3)$.

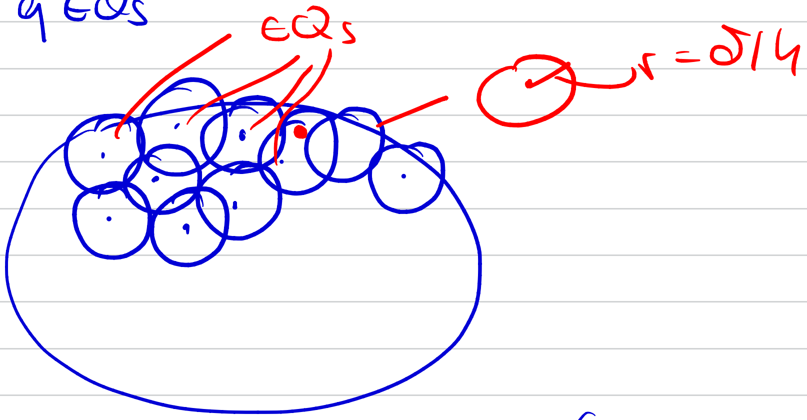
Proof. It suffices to prove the statement for $\|x\| = 1$.

choose a set of points Q_S s.t.

i) $Q_S \subseteq \mathcal{X}_S$, $\quad \|q\| = 1 \quad \forall q \in Q_S$

ii) $x \in \mathcal{X}_S$ with $\|x\| = 1$, we have

$$\min_{q \in Q_S} \|x - q\| \le \delta/4.$$



$$|Q_S| \le (12/\delta)^k \qquad \Big(\delta^{-\text{dim.}^{=k}} \Big)$$

can use Lem. 8.2. (with $\varepsilon = \delta/2$)

$$\Rightarrow (1-\delta/2)\|q\|^2 \le \|\Phi q\|^2 \le (1+\delta/2)\|q\|^2, \quad \forall q \in Q_S$$

holds w.p. $\qquad$ from Lem. 8.2.

$$\ge 1 - 2(12/\delta)^k e^{-c_0(\delta/2)m}$$

$$c_0(\varepsilon) = \frac{x^2 - x^3}{4}$$

define A as the smallest no. s.t.

$$\|\Phi x\| \le (1+A)\|x\| \quad , \forall x \in \mathcal{X}_S$$

goal : $A \le \delta$.

for every $x \in \mathcal{X}_S$ with $\|x\| = 1$ $\exists$ $q \in Q_S$ s.t. $\|x - q\| \le \delta/4$.

$$\|\Phi x\| = \|\Phi(x - q + q)\| \overset{\Delta}{=} \|\Phi q\| + \|\Phi(x-q)\|$$

$$\|\Phi(x-q)\| \le (1+A)\|x-q\| \le (1+\delta)\frac{\delta}{4}$$

$$\|\Phi q\|^2 \le (1+\delta/2)\|q\|^2$$

$$\|\Phi q\|^2 \le \sqrt{1+\delta/2}\,\|q\| \le (1+\delta/2)\|q\|$$

$$\le (1+\delta/2)\|q\|$$

$$\le (1+\delta/2)\underbrace{\|q\|}_{1} + (1+A)\frac{\delta}{4}$$

$$= 1 + \delta/2 + (1+A)\frac{\delta}{4}$$

as A is the smallest no. s.t. $\|\Phi x\| \le (1+A)\|x\|$

$$A \le \delta/2 + (1+A)\frac{\delta}{4}$$

$$A \le \delta$$

have established that

$$\|\Phi x\| \le (1+\delta)\|x\|$$

$$\|\Phi x\| \ge (1-\delta)\|x\|$$

$$\|\Phi x\| \overset{\text{rev.}\,\Delta}{\ge} \|\Phi q\| - \underbrace{\|\Phi(x-q)\|}_{(1+\delta)\frac{\delta}{4}} \ge 1 - \delta . \ \square$$

$$\ge (1-\delta/2)\underbrace{\|q\|}_{1}$$

finishing up

Thm. 9.2. Suppose that m, n and $\delta \in (0,1)$ are given. If the pdf generating Φ satisfies the concentration of measure inequality in Lem. 8.2., then $\exists$ $c_1, c_2 > 0$ depending only on δ s.t. RIP holds for Φ with the prescribed δ and every $k \leq c_1 m / \log(n/k)$ w.p. $\geq 1 - 2e^{-c_2 m}$.

Proof. We know that for each k-dim. space X_S, Φ fails to satisfy

$$(1-\delta)\|x\| \leq \|\Phi x\| \leq (1+\delta)\|x\|, \quad \forall x \in X_S \qquad (*)$$

w.p.
$$\leq 2\,(12/\delta)^k \, e^{-c_0(\delta/2)m}.$$

counting: there are $\binom{n}{k} \leq \left(\frac{en}{k}\right)^k$ subspaces of dim. k in n-dim. ambient space

$(*)$ fails to hold w.p.

$$\leq 2(12/\delta)^k \, e^{-c_0(\delta/2)m} \left(\frac{en}{k}\right)^k$$

show that for a fixed $c_1 > 0$, whenever

$$k \leq \frac{c_1 m}{\log(n/k)} \Rightarrow \boxed{m \geq C\, s \log(n)}$$

$$= \cancel{2}\, e^{-c_0(\delta/2)m + k[\log(en/k) + \log(12/\delta)]}$$

$$\leq \cancel{2}\, e^{-c_2 m}$$

$$e^{-c_0(\delta/2)m + k[\log(en/k) + \log(12/\delta)]} \leq e^{-c_2 m}$$

$$\left(k \leq \frac{c_1 m}{\log(n/k)} \right)$$

or just like this

$$e^{-m\left(c_0(\delta/2) - c_1 \frac{\log(en/\ell) + \log(1/2\delta)}{\log(n/\ell)}\right)} \leq e^{-c_2 m}$$

and hence

$$c_2 \leq c_0(\delta/2) - c_1\left(1 + \frac{1+\log(1/2\delta)}{\log(n/\ell)}\right). \square$$

Interpretation: CS: 1. $m \geq s^2$ Γ-bottleneck

2. $m \geq 2s$ (ESPRIT, FRI, multi-band sampling)

3. $m \sim s \log n$ prob. regime.

Chapter 10. Approximation Theory

10.1. Min-Max (Kolmogorov) Rate-Distortion Theory

$\underbrace{C \subset L^2(\Omega)}_{\text{function class}}$, $d \in \mathbb{N}$, $\Omega \subset \mathbb{R}^d$

For each $\ell \in \mathbb{N}$, we denote by

$$\mathcal{E}^\ell := \{E : C \to \{0,1\}^\ell\}$$

the set of binary encoders of C of length ℓ, and we let

$$\mathcal{D}^\ell := \{D : \{0,1\}^\ell \to L^2(\Omega)\}$$

be the set of binary decoders of length ℓ.

An encoder-decoder pair $(E,D) \in \mathcal{E}^\ell \times \mathcal{D}^\ell$ is said to achieve uniform error ε over C, if

$$\sup_{f \in C} \| D(E(f)) - f \|_{L^2(\Omega)} \leq \varepsilon.$$

Def. 10.1. Let $d \in \mathbb{N}$, $\Omega \subset \mathbb{R}^d$, and $C \subset L^2(\Omega)$. Then, for $\varepsilon > 0$, the minimax code length $L(\varepsilon, C)$ is

$$L(\varepsilon, C) := \min\{\ell \in \mathbb{N} : \exists (E,D) \in \mathcal{E}^\ell \times \mathcal{D}^\ell :$$

$$\sup_{f\in\mathcal{C}} \| \mathcal{D}(\mathcal{E}(f)) - f \|_{L^2(\Omega)} \leq \varepsilon \}$$

Moreover, the optimal exponent $\gamma^*(\mathcal{C})$ is defined as

$$\gamma^*(\mathcal{C}) := \sup\{ \gamma \in \mathbb{R} : L(\varepsilon, \mathcal{C}) \in \mathcal{O}(\varepsilon^{-1/\gamma}), \varepsilon \to 0 \}.$$

$$f \in \mathcal{O}(g) \Rightarrow |f(x)| \leq C |g(x)|$$

$$\frac{1}{\varepsilon^{1/\gamma}} \Rightarrow \gamma \downarrow \Rightarrow \frac{1}{\varepsilon^{1/\gamma}} \uparrow$$

|| Larger $\gamma^*(\mathcal{C})$ results in smaller growth rate and hence smaller memory requirements for storing signals $f \in \mathcal{C}$.

10.2. Metric entropy, covering, and packing

Idea: quantify $\gamma^*(\mathcal{C})$ as a function of $\mathcal{C}$ or, equiv., quantify the description complexity of a given function class $\mathcal{C}$.

$(\mathcal{X}, \rho)$ is a metric space

set, dist. measure

$\rho : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$

i) $\rho(x, x') \geq 0$, $\forall x, x' \in \mathcal{X}$, with $=$ iff $x = x'$

ii) $\rho(x, x') = \rho(x', x)$, $\forall x, x' \in \mathcal{X}$

iii) $\rho(x, \tilde{x}) \leq \rho(x, x') + \rho(x', \tilde{x})$, $x, x', \tilde{x} \in \mathcal{X}$

Exs.

$$x = \mathbb{R}^d, \quad \rho(x, x') = \sqrt{\sum_{j=1}^{d} (x_j - x_j')^2}$$

$X = \text{cube}\{0,1\}^d$

$$\rho_H(x, x') = \frac{1}{d} \sum_{j=1}^{d} \mathbb{I}(x_j \neq x_j')$$

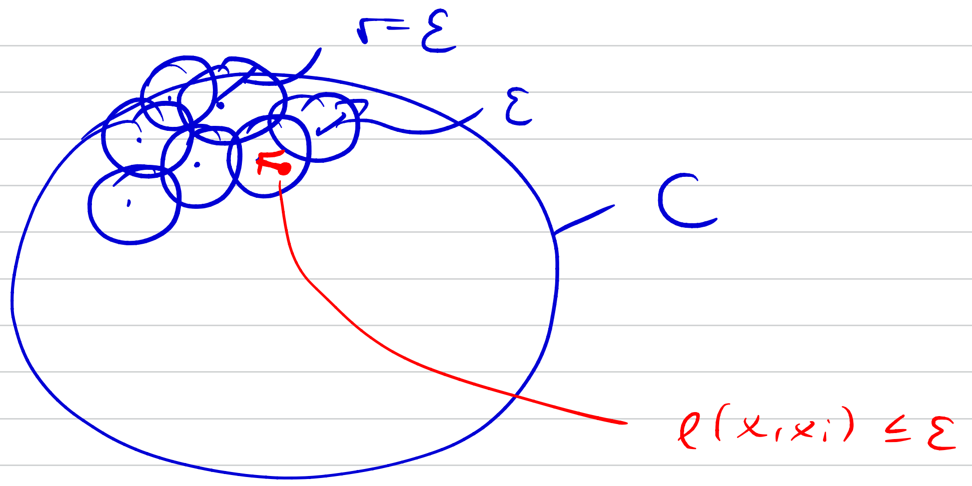
$L^2([0,1])$

$$\|f-g\|_2 = \left[\int_0^1 |f(x)-g(x)|^2 dx\right]^{1/2}$$

$C([0,1])$

$$\|f-g\|_\infty = \sup_{x\in[0,1]} |f(x)-g(x)| .$$

Def. 10.2. Let $(\mathcal{X},\rho)$ be a metric space. An ε-covering of a compact set $C \subseteq \mathcal{X}$ w.r.t. the metric ρ is a set $\{x_1,\dots,x_N\} \subset C$ s.t. for every $x\in C$, $\exists$ $i\in[1,N]$ s.t. $\rho(x,x_i)\le\varepsilon$. The ε-covering no. $N(\varepsilon;C,\rho)$ is the cardinality of a smallest ε-covering.



$$C \subset \bigcup_{i=1}^{N} \underbrace{B(x_i,\varepsilon)}_{\text{ball}}$$

ball, radius, ball center

$B(x_i,\varepsilon)$ is a ball in the metric ρ of radius ε centered at x_i

cov. no. is non-increasing, i.e., $N(\varepsilon)\ge N(\varepsilon')$, $\varepsilon\le\varepsilon'$

when C is not finite $\Rightarrow$ cov. no. $\to\infty$ as $\varepsilon\to 0$ $\Rightarrow$ will be interested in the growth rate

in particular, we will be interested in

$$\log_2 N(\varepsilon; C,\rho) \quad \text{(in bits)}$$

$\Uparrow$

metric entropy (Kolmogorov–Tikhomirov metric entropy) specifying an encoder-decoder pair

$x \in C$, want to encode it into a bit string of finite length s.t. the resulting decoding error $\le \varepsilon$.

Q: What is the min. no. of bits needed to represent any $x \in C$ with error — measured in ρ — of at most ε?

A: It is $\lceil \log_2 N(\varepsilon; C, \rho) \rceil$.

Why? What are the corr. E & D?

$x \in \mathcal{X}$, $E(x)$ maps to the closest — in ρ — covering ball center x_i

output of encoder is a bit string labeling the chosen cov. ball center

decoder: simply outputs x_i corr. to the bit string it receives

$$x \in C \Rightarrow E(x) \Rightarrow b_1 \ldots b_\ell \Rightarrow D(b_1 \ldots b_\ell) \Rightarrow x_i$$

$$\rho(x, x_i) \le \varepsilon$$

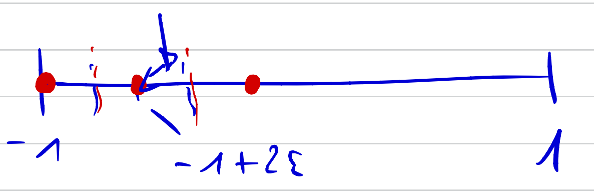
$(\mathcal{X}, \rho) = (\mathbb{R}, |\cdot|)$

$C = [-1, 1]$, $\rho(x, x') = |x - x'|$

compute covering no. of C

divide C up into intervals of length 2ε by setting
$x_i = -1 + 2(i-1)\varepsilon$, $i \in [1, L]$.

$$\frac{2}{2\varepsilon} = \frac{1}{\varepsilon} \Rightarrow L = \lfloor \tfrac{1}{\varepsilon} \rfloor + 1$$



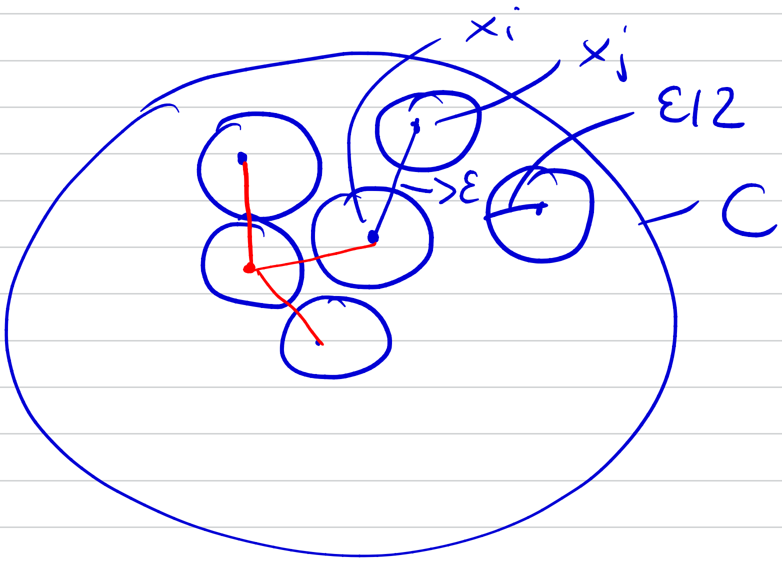
$$N(\varepsilon; C, \rho) \le \overbrace{\lfloor \tfrac{1}{\varepsilon} \rfloor + 1}^{L} \le \frac{1}{\varepsilon} + 1$$

$\log_2 N(\varepsilon; C, \rho) \leq \log_2\left(\frac{1}{\varepsilon}+1\right) \asymp \log_2(\varepsilon^{-1})$

can generalize to d-dim. unit cube : $\log_2 N(\varepsilon; C, \rho) \leq d \log_2(1/\varepsilon + 1) \asymp d \log_2(\varepsilon^{-1})$

need to introduce the notion of the packing no. of a set C in $(\mathcal{X}, \rho)$

Def. 10.3. Let $(\mathcal{X}, \rho)$ be a metric space. An ε-packing of a compact set $C \subset \mathcal{X}$ w.r.t. the metric ρ is a set $\{x_1, \dots, x_n\} \subset C$ s.t. $\rho(x_i, x_j) > \varepsilon$, $\forall$ distinct i,j. The ε-packing no. $M(\varepsilon; \mathcal{X}, \rho)$ is the cardinality of a largest ε-packing.

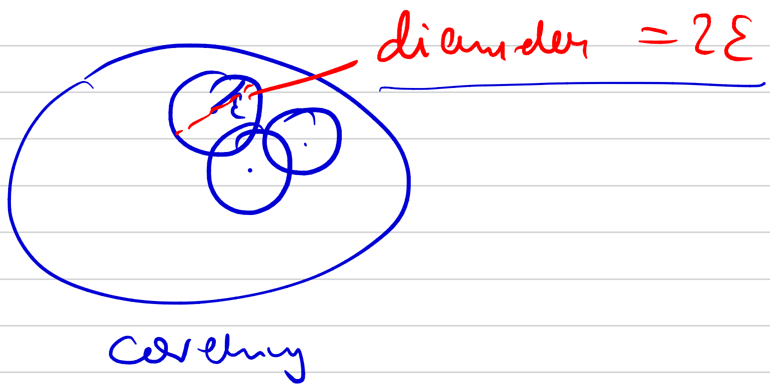


Lemma 10.4. Let $(\mathcal{X}, \rho)$ be a metric space and C a compact set in $\mathcal{X}$. For all $\varepsilon > 0$, the packing & the cov. no. are related according to

$$M(2\varepsilon; C, \rho) \leq N(\varepsilon; C, \rho) \leq M(\varepsilon; C, \rho).$$

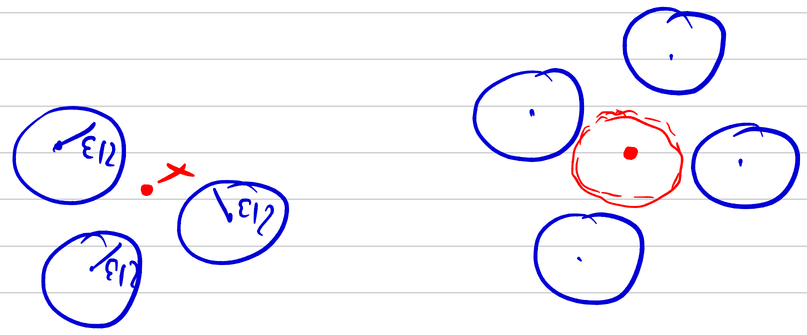
Proof. Choose a minimal ε-covering and a maximal (2ε)-packing of C.

i) no two centers of the 2ε-packing can lie in the same ball of the ε-covering



$$\Rightarrow M(2\varepsilon; C, \rho) \leq N(\varepsilon; C, \rho)$$

ii) to establish $N(\varepsilon; C, \rho) \leq M(\varepsilon; C, \rho)$, note that, given a maximum packing $M(\varepsilon; C, \rho)$, for every $x \in C$, we have the center of at least one of the balls in the packing within distance less than ε.



$\Longrightarrow$ this maximal packing also provides an ε-covering and since $N(\varepsilon; C, \rho)$ is a minimal covering, we must have

$$N(\varepsilon; C, \rho) \leq M(\varepsilon; C, \rho). \qquad \square$$

finishing up the example:

$$M(2\varepsilon; C, \rho) \leq N(\varepsilon; C, \rho) \leq M(\varepsilon; C, \rho)$$

$$N(\varepsilon; C, \rho) \leq \log_2\left(\frac{1}{\varepsilon} + 1\right) \asymp \log_2(\varepsilon^{-1})$$

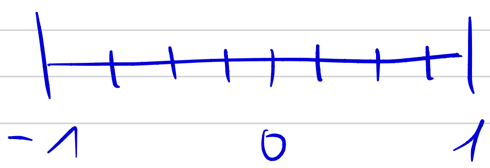
the points x_i, x_j are separated according to $|x_i - x_j| = 2\varepsilon > \varepsilon$

$$M(\varepsilon; C, |\cdot|) \geq L = \left\lfloor \frac{1}{\varepsilon} \right\rfloor + 1 \geq \frac{1}{\varepsilon}$$

$$\log_2(\varepsilon^{-1}) \asymp \log_2\left(\frac{1}{2\varepsilon}\right) \leq \log_2 M(2\varepsilon; C, \rho) \leq \log_2 N(\varepsilon; C, \rho) \leq \log_2\left(\frac{1}{\varepsilon} + 1\right) \asymp \log_2(\varepsilon^{-1})$$

$$\Longrightarrow \log_2 N(\varepsilon; C, \rho) \asymp \log_2(\varepsilon^{-1})$$

$$\frac{2}{1/4} = 8$$



$$\log_2\left(\frac{2}{\varepsilon}\right) \asymp \log(1/\varepsilon)$$

$$\|x\|_q = \begin{cases} \left(\sum_{i=1}^{d} |x_i|^q\right)^{1/q}, & q \in (1, \infty) \\ \max_{i=1 \ldots d} |x_i|, & q = \infty \end{cases}$$

Lemma 10.5. Consider a pair of norms $\|\cdot\|$ and $\|\cdot\|'$ on $\mathbb{R}^d$, and let B and B' be their corresponding unit balls, i.e., $B = \{x \in \mathbb{R}^d : \|x\| \le 1\}$ and $B' = \{x \in \mathbb{R}^d \mid \|x\|' \le 1\}$. Then, the ε-covering number of B in the $\|\cdot\|'$-norm satisfies

$$\left(\frac{1}{\varepsilon}\right)^d \frac{vol(B)}{vol(B')} \le N(\varepsilon; B, \|\cdot\|') \le \frac{vol\left(\frac{2}{\varepsilon}B + B'\right)}{vol(B')}.$$

Proof. Let $\{x_1, \dots, x_{N(\varepsilon; B, \|\cdot\|')}\}$ be an ε-covering of B in the $\|\cdot\|'$-norm. Then, we have

$$B \subseteq \bigcup_{j=1}^{N(\varepsilon; B, \|\cdot\|')} \{x_j + \varepsilon B'\}$$

$$\Rightarrow vol(B) \le N(\varepsilon; B, \|\cdot\|')\, \varepsilon^d\, vol(B')$$

$\Rightarrow$ lower bound ✓

For the upper bound: maximal ε-packing $\{x_1, \dots, x_{M(\varepsilon; B, \|\cdot\|')}\}$ of B in the $\|\cdot\|'$-norm.

the balls $\{x_j + \frac{\varepsilon}{2}B', \ j = 1, \dots, M(\varepsilon; B, \|\cdot\|')\}$ are all disjoint

taking volumes $\Rightarrow$

$$\sum_{j=1}^{M(\varepsilon; B, \|\cdot\|')} vol\left(x_j + \frac{\varepsilon}{2}B'\right) \le \underline{vol\left(B + \frac{\varepsilon}{2}B'\right)}.$$

$$vol\left(\frac{\varepsilon}{2}B'\right) = \left(\frac{\varepsilon}{2}\right)^d vol(B')$$

$$vol\left(B + \frac{\varepsilon}{2}B'\right) = \left(\frac{\varepsilon}{2}\right)^d vol\left(\frac{2}{\varepsilon}B + B'\right)$$

$$\underbrace{M(\varepsilon; B, \|\cdot\|')}_{\ge\, N(\varepsilon; B, \|\cdot\|')} \left(\frac{\varepsilon}{2}\right)^d vol(B') \le \left(\frac{\varepsilon}{2}\right)^d vol\left(\frac{2}{\varepsilon}B + B'\right)$$

$\square$

Interpretation: $B = B'$

$$\mathrm{vol}\left(\tfrac{2}{\varepsilon}B + B'\right) = \mathrm{vol}\left(\left(\tfrac{2}{\varepsilon}+1\right)B\right) = \left(\tfrac{2}{\varepsilon}+1\right)^d \mathrm{vol}(B)$$

$$\varepsilon^{-d} \leq N(\varepsilon; B, \|\cdot\|) \leq \left(\tfrac{2}{\varepsilon}+1\right)^d$$

$$\Downarrow$$

$$\underline{\underline{N(\varepsilon; B, \|\cdot\|) \asymp \varepsilon^{-d}}}$$

$$\log_2 N(\varepsilon; B, \|\cdot\|) = d \log(1/\varepsilon).$$

$B_\infty^d = [-1,1]^d$, $\|\cdot\|_\infty$ $\nearrow$