13. (Aside) Dudley's Chaining (covering numbers to R.C.)

Friday, February 9, 2024 3:16 PM

Bounding (empirical) Rademacher Complexity $\hat{R}(A) = \mathbb{E}_\sigma \sup_{a \in A} \frac{1}{m}\langle \sigma, a\rangle$

Suppose A is bounded, eg. $\forall a \in A$, $\|a\|_2 \le c_2$

(and recall $\hat{R}$ is shift-invariant) Via Cauchy-Schwarz, $\langle \sigma, a\rangle \le \underbrace{\|\sigma\|_2}_{=\sqrt{m}} \cdot \underbrace{\|a\|_2}_{\le c}$

so $\hat{R}(A) \le \frac{c_2}{\sqrt{m}}$.

(or, if $\|a\|_1 \le c_1$ then $\langle \sigma, a\rangle \le \|\sigma\|_\infty \cdot \|a\|_1 \le 1 \cdot c_1$ so $\hat{R}(A) \le \frac{c_1}{m}$.

Can we do better?

Recall covering number $N_2(\varepsilon, A)$.

Lemma 27.1 [SS]: if $\|a\|_2 \le c_2$ $\forall a \in A \subseteq \mathbb{R}^m$, $A \subseteq$ d-dim subspace, then $N_2(\varepsilon, A) \le \left(2 \cdot c_2 \frac{\sqrt{d}}{\varepsilon}\right)^d$

(we proved a variant of this)

Thm (Dudley '67, '87 / Lemma 27.4 [SS]) Dudley's Chaining

Let $c = \min_{\bar{a}} \max_{a \in A} \|a - \bar{a}\|$ be the radius of $A \subseteq \mathbb{R}^m$ (all norms are Euclidean)

then $\forall M \in \mathbb{N}$,

$$\hat{R}(A) \le c \cdot \left(\frac{2^{-M}}{\sqrt{m}} + \frac{6}{m} \cdot \sum_{k=1}^{M} 2^{-k} \sqrt{\log(N_2(c \cdot 2^{-k}, A))} \right)$$

tune M to balance terms

proof wlog let $\bar{a} = 0$ since both $\hat{R}$ and N_2 are shift-invariant.

Also, wlog let $c = 1$ since $\hat{R}(A) \le c \cdot \hat{R}(c^{-1}A)$, and $N_2(\varepsilon, cA) \le N_2(c\varepsilon, A)$

Key idea: Don't form an ε-net and do union bound: too many points!

Instead, very cleverly re-use points.

$k=0$: define $B_0 = \{0\}$, an $\varepsilon = 1$ cover of A (recall A is centered w/ radius 1)

$k \ge 1$:

let B_k be a set corresponding to a minimal 2^{-k} cover

so $|B_k| = N_2(2^{-k}, A)$

Now, recall $\hat{R}(A) := \frac{1}{m} \mathbb{E}_\sigma \underbrace{\sup_{a \in A} \langle \sigma, a\rangle}$ fix σ for now, let $a^* \in \operatorname{argmax}_{a \in A} \langle \sigma, a\rangle$

(if A isn't closed, pick an almost-optimal one ...)

13a. Dudley's Chaining

Thursday, February 15, 2024 1:29 PM

Let $b^{(k)} \in B_k$ be the closest pt. in B_k to a^*,

so... $\|a^* - b^{(k)}\| \le 2^{-k}$

$b^{(k)} = \mathrm{proj}_{B_k}(a^*)$

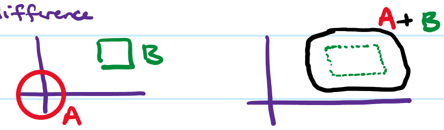
and $b^{(0)} = 0$.

Then
$$\begin{aligned} \|b^{(k)} - b^{(k-1)}\| &\le \|b^{(k)} - a^*\| + \|a^* - b^{(k-1)}\| && \text{triangle inequality} \\ &\le 2^{-k} + 2^{-(k-1)} && \text{by definition of covering set} \\ &= 3 \cdot 2^{-k} \end{aligned}$$

motivates

For $k \in [M]$ define
$$\bar{B}_k = \{ \underbrace{b - b'}_{\text{Minkowski sum / difference}} : b \in B_k,\ b' \in B_{k-1} \text{ st } \|b - b'\| \le 3 \cdot 2^{-k} \}$$

(which does <u>not</u> depend on σ)



How large is $\bar{B}_k$?

$$|\bar{B}_k| \le |B_k| \cdot |B_{k-1}| \le |B_k|^2 = N_2(2^{-k}, A)^2$$

Altogether:

really $a^* = a^*(\sigma)$

$$\hat{R}(A) := \frac{1}{m} \mathop{\mathbb{E}}_{\sigma} \sup_{a \in A} \langle \sigma, a \rangle = \frac{1}{m} \mathop{\mathbb{E}}_{\sigma} \langle \sigma, a^* \rangle$$

$$= \frac{1}{m} \mathop{\mathbb{E}}_{\sigma} \Big(\langle \sigma, a^* - b^{(M)} \rangle + \overbrace{\sum_{k=1}^{M} \langle \sigma, b^{(k)} - b^{(k-1)} \rangle}^{\text{telescopes}} \Big)$$

$$\le \frac{1}{m} \underbrace{\|\sigma\|_2}_{=\sqrt{m}} \cdot \underbrace{\|a^* - b^{(M)}\|_2}_{\substack{\le 2^{-M} \\ \text{(small)}}} + \sum_{k=1}^{M} \frac{1}{m} \mathop{\mathbb{E}}_{\sigma} \langle \sigma, \underbrace{b^{(k)} - b^{(k-1)}}_{} \rangle$$

not necessarily small enough.
We don't know what it is,
but it's in $\bar{B}_k$

$$\le \frac{1}{\sqrt{m}} 2^{-M} + \sum_{k=1}^{M} \frac{1}{m} \mathop{\mathbb{E}}_{\sigma} \sup_{b \in \bar{B}_k} \langle \sigma, b \rangle$$

$$= \frac{1}{\sqrt{m}} 2^{-M} + \sum_{k=1}^{M} \underbrace{\hat{R}(\bar{B}_k)}_{}$$

bound via Massart's Lemma

$$\le \underbrace{\text{radius}}_{3 \cdot 2^{-k}} \cdot \frac{1}{\underbrace{\text{dimension}}_{=m}} \sqrt{2 \cdot \log(\#\text{points})}$$

$$\le 3 \cdot 2^{-k} \cdot \frac{1}{m} \sqrt{2 \cdot \log(N_2(2^{-k}, A)^2)}$$

$$= 6 \cdot 2^{-k} \cdot \frac{1}{m} \sqrt{\log(N_2(2^{-k}, A))} \qquad \square$$