ML Theory — Homework 3

your NetID here

Version 0

Instructions. (Same as homework 2.)

- Everyone must submit an individual write-up.
- You may discuss with up to 3 other people. State their NetIDs clearly on the first page. Outside of office hours, you should not discuss with anyone but these three.
- Homework is due **Tuesday, December 18, at 11:59pm**; no late homework accepted.
- Please consider using the provided LaTeX file as a template.

1. **Calisthenics.**

   (a) Let k real-valued functions $\mathcal{F} := (f_1, \ldots, f_k)$ be given, and define

   $$\mathcal{G} := \left\{ x \mapsto \operatorname{sgn}\left(b + \sum_i a_i f_i(x) \right) : a \in \mathbb{R}^k, b \in \mathbb{R} \right\}.$$

   Prove $\textsc{vc}(\mathcal{G}) \leq k+1$.

   Hint. Use the VC-dimension of linear separators from Lecture 22.

   Bonus (ungraded). When is this VC upper bound an equality?

   (b) Let $\mathcal{F} := \{x \mapsto \mathbb{1}[\|x-a\|_2^2 \geq b] : a \in \mathbb{R}^d, b \in \mathbb{R}\}$ denote indicators of balls in $\mathbb{R}^d$. Prove $\textsc{vc}(\mathcal{F}) \leq d+2$.

   Hint. Use the previous part.

   (c) Recall from Lecture 21 the *ramp loss* ℓ_γ (where $\gamma > 0$), defined as

   $$\ell_\gamma(r) := \begin{cases} 0 & r < -\gamma, \\ 1 + r/\gamma & r \in [-\gamma, 0], \\ 1 & r > 0. \end{cases}$$

   Prove that for any convex $\ell : \mathbb{R} \to \mathbb{R}_{\geq 0}$,

   $$\ell_\gamma(r) \leq \frac{\ell(r)}{\ell(0)} \qquad \text{when} \quad 0 < \gamma \leq \frac{\ell(0)}{\ell'(0)}.$$

   Remarks. i. Both squared and logistic losses fare pretty well with this. ii. This allows $\mathcal{R}_\ell$ to be used in place of $\mathcal{R}_\gamma$ in any margin-based generalization bound.

   (d) Prove the final theorem in Lecture 19, the three-part "core Rademacher theorem", via the other lemmas and theorems in Lecture 19. (Your main work is in checking the bounded differences condition, and then applying McDiarmid's inequality to a few other quantities from that lecture.)

Solution.

(Your solution here.)

2. **Covering non-decreasing functions.**

Let $\mathcal{F}$ denote all non-decreasing functions from $\mathbb{R}$ to $[0, 1]$, Let a sample $S = (x_1, \ldots, x_n)$ be given, and as usual let $\mathcal{F}_{|S} \subseteq \mathbb{R}^n$ denote the restriction of $\mathcal{F}$ to the sample S.

(a) Prove $\mathcal{N}(\mathcal{F}_{|S}, \epsilon, \|\cdot\|_2) \leq (1 + n)^{1+\sqrt{n}/\epsilon}$.
Note. The bound has some wiggle room. It's okay if you're a little off.
Hint. If you have n (and not $\sqrt{n}$) in your numerator, then try to shift the focus of your cover to the range rather than the domain...

(b) Using the Pollard bound from Lecture 24, prove

$$\mathrm{URad}(\mathcal{F}_{|S}) \leq 1024(n \ln(1 + n))^{2/3}.$$

Note. 1024 is also wiggle room...

(c) Using the Dudley bound from Lecture 24, prove

$$\mathrm{URad}(\mathcal{F}_{|S}) \leq 1024(n \ln(1 + n))^{1/2}.$$

Solution.

(Your solution here.)

3. **Covering linear functions.**

Throughout, let $S = (x_1, \ldots, x_n)$ denote a sample of size n, and construct matrix $X \in \mathbb{R}^{n \times d}$ with the sample points as rows.

(a) Prove

$$\ln \mathcal{N}\left(\left\{x \mapsto \langle x, w\rangle : w \in \Delta_d\right\}_{|S}, \epsilon, \|\cdot\|_2\right) \le \left\lceil \frac{\|X\|_{2,\infty}^2}{\epsilon^2} \right\rceil \ln(d),$$

where $\Delta_d = \left\{\alpha \in \mathbb{R}^d_{\ge 0} : \sum_i \alpha_i = 1\right\}$ and $\|X\|_{2,\infty} = \max_i \|X\mathbf{e}_i\|_2$.

Hint. Use the Maurey Lemma from Lecture 13.

(b) Prove

$$\ln \mathcal{N}\left(\left\{x \mapsto \langle x, w\rangle : \|w\|_1 \le a\right\}_{|S}, \epsilon, \|\cdot\|_2\right) \le \left\lceil \frac{a^2\|X\|_{2,\infty}^2}{\epsilon^2} \right\rceil \ln(2d).$$

Hint. Use the previous part.

(c) Define

$$\mathcal{F}_2(a) := \left\{x \mapsto \langle x, w\rangle : \|w\|_2 \le a\right\}.$$

Prove

$$\ln \mathcal{N}\left(\mathcal{F}_2(a)_{|S}, \epsilon, \|\cdot\|_2\right) \le \left\lceil \frac{a^2\|X\|_{\mathrm{F}}^2}{\epsilon^2} \right\rceil \ln(2d),$$

where $\|X\|_{\mathrm{F}} = \sqrt{\sum_{i=1}^n \sum_{j=1}^d (x_i)_j^2}$ denotes the Frobenius norm.

Hint. Use the previous part.

(d) Use the Pollard bound from Lecture 26 to prove

$$\mathrm{URad}(\mathcal{F}_2(a)_{|S}) = \widetilde{\mathcal{O}}\left(a\|X\|_{\mathrm{F}} n^{1/4}\right).$$

Remark. Use the $\widetilde{\mathcal{O}}$ to hide polylog factors of a, $\|X\|_{\mathrm{F}}$, n, d; the ceiling in the covering number makes things ugly.

(e) Use the Dudley bound from Lecture 26 to prove

$$\mathrm{URad}(\mathcal{F}_2(a)_{|S}) = \widetilde{\mathcal{O}}\left(a\|X\|_{\mathrm{F}}\right).$$

Remark. The direct Rademacher proof gave $\mathrm{URad}(\mathcal{F}_2(a)_{|S}) \le a\|X\|_{\mathrm{F}}$.

Solution.

(Your solution here.)

4. **Are we still friends?**

Solution.

(Your solution here.)