ML Theory — Homework 3

_your NetID here_

Version 2

**Instructions.** (Same as homework 1.)

- 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 **Friday, December 22, at 3:30pm**; no late homework accepted.

- Please consider using the provided **LaTeX** file as a template.
1. Calisthenics.

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

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

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

**Hint.** Use the VC-dimension of linear separators from lecture.

**Bonus (ungraded).** When is the above an equality?

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

**Hint.** Use the previous part.

(c) Recall from Lecture 23 the ramp loss $\ell_\gamma$ (where $\gamma > 0$), defined as

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

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

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

**Remark.** Both squared and logistic losses fare pretty well with this.

(d) Prove the final theorem in Lecture 20, the three-part “core Rademacher theorem”, via the other lemmas and theorems in Lecture 20. (Your main work is in checking the bounded differences condition, and then applying McDiarmid’s inequality to a few other quantities from that lecture.)

**Note.** The theorem is stated with two typos: the first $f$ in the first line should be $F$, and the right hand side of the third bound should have $\ln(2/\delta)$ not $\ln(1/\delta)$.

**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 26, prove

\[
\text{Rad}(\mathcal{F}|_S) \leq 64(n \ln(n + 1))^{2/3}.
\]

**Note.** 64 is also wiggle room.

(c) Using the Dudley bound from Lecture 26, prove

\[
\text{Rad}(\mathcal{F}|_S) \leq 16(n \ln(n))^{1/2}.
\]

**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
\[
\ln \mathcal{N} \left( \left\{ x \mapsto \langle x, w \rangle : w \in \Delta_d \right\} | S, \epsilon, \| \cdot \|_2 \right) \leq \left\lceil \frac{\|X\|_{2, \infty}^2}{\epsilon^2} \right\rceil \ln(d),
\]
where \( \Delta_d = \left\{ \alpha \in \mathbb{R}^d : \sum_i \alpha_i = 1 \right\} \) and \( \|X\|_{2, \infty} = \max_i \|X e_i\|_2 \).

**Hint.** Use the Maurey Lemma from Lecture 15.

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

**Hint.** Use the previous part.

(c) Define
\[
\mathcal{F}_2(a) := \left\{ x \mapsto \langle x, w \rangle : \|w\|_2 \leq a \right\}.
\]
Prove
\[
\ln \mathcal{N} \left( \mathcal{F}_2(a) | S, \epsilon, \| \cdot \|_2 \right) \leq \left\lceil \frac{a^2 \|X\|_2^2}{\epsilon^2} \right\rceil \ln(2d),
\]
where \( \|X\|_2 = \sqrt{\sum_{i=1}^n \sum_{j=1}^d (x_i j)^2} \) denotes the Frobenius norm.

**Hint.** Use the previous part.

(d) Use the Pollard bound from Lecture 26 to prove
\[
\text{Rad}(\mathcal{F}_2(a) | S) \leq 16a\|X\|_2 (n \ln(2d))^{1/4}.
\]

(e) Use the Dudley bound from Lecture 26 to prove
\[
\text{Rad}(\mathcal{F}_2(a) | S) \leq 16a\|X\|_2 (\ln(n) - \ln(a\|X\|_2)) \ln(2d).
\]

**Remark.** The direct Rademacher proof gave \( \text{Rad}(\mathcal{F}_2(a) | S) \leq a\|X\|_2 \).

**Solution.**
*(Your solution here.)*
4. Are we still friends?

Solution.

(Your solution here.)