HOMEWORK 1 - V0

CSC2532 WINTER 2024

University of Toronto

VERSION HISTORY: V0 → V1:

- **Deadline:** Feb 4, at 23:59.
- **Submission:** You need to submit your solutions through Crowdmark, including all your derivations, plots, and your code. You can produce the file however you like (e.g. LaTeX, Microsoft Word, etc), as long as it is readable. Points will be deducted if we have a hard time reading your solutions or understanding the structure of your code.

1. ε-Nets and Covering.

1.1. *ϵ-net over sphere [20pts].* Let $\mathcal{N}_\epsilon$ be an *epsilon*-net over the d-dimensional unit Euclidean sphere $\mathcal{S}^{d-1}$ equipped with the Euclidean metric. Note that this is the surface of a unit ball.

(a) Show that there exists a net such that $\forall \epsilon > 0$, $|\mathcal{N}_\epsilon| \leq (1 + 2/\epsilon)^d$.
(b) For $x \in \mathbb{R}^d$, show that $\max_{u \in \mathcal{N}_\epsilon} \langle u, x \rangle \geq (1 - \epsilon)\|x\|_2$.

1.2. *Vector concentration [20pts].* Assume that $x_1, x_2, ..., x_n$ are independent random vectors satisfying $\mathbb{E}[x_i] = \mu \in \mathbb{R}^d$ and $\|x_i\|_2 \leq \kappa$ almost surely. Denoting by $\hat{\mu}$ their sample mean, i.e., $\hat{\mu} = \frac{1}{n}\sum_{i=1}^n x_i$, show that $\hat{\mu}$ enjoys sub-gaussian concentration around its expectation. That is, find an upper bound on the following probability

$$\mathbb{P}\left(\|\hat{\mu} - \mu\|_2 \geq ?\sqrt{\frac{d}{n}} \right) \leq ?$$

All steps, conditions, and constants must be stated explicitly. Hint: you may find the previous question part b helpful.

1.3. *More Concentration [20pts].* In an estimation problem, assume that the features $x \in \mathbb{R}^d$ satisfy $\mathbb{E}[x] = \mu$ and the response $y \in \mathbb{R}$ is given as $y = f(\langle x, \mu \rangle) + \text{noise}$, where noise has 0 expectation. Here, f is known, uniformly bounded by $|f| \leq B$ and L-Lipschitz, and the response is not observed. Our objective is to estimate $\mathbb{E}[y]$.

For independent observations $x_1, x_2, ..., x_n$ satisfying $\|x_i\|_2 \leq \kappa$ almost surely, we will use

$$\hat{\xi} = \frac{1}{n}\sum_{i=1}^n f(\langle x_i, \hat{\mu} \rangle)$$

for this task where $\hat{\mu}$ denotes the sample mean, i.e., $\hat{\mu} = \frac{1}{n}\sum_{i=1}^n x_i$.

Establish a concentration result for $\hat{\xi}$ around $\mathbb{E}[y]$ using an ϵ-net argument.. Show your work explicitly. Constants not depending on n and d do not need to be explicit.

2. Rademacher Complexity.

2.1. *Properties [20pts].* Prove the following properties of the Rademacher complexity:

- For two sets $\mathcal{F}_1$, $\mathcal{F}_2$ and $\mathcal{F}_1 + \mathcal{F}_2 = \{f_1 + f_2 \ : \ f_1 \in \mathcal{F}_1 \text{ and } f_2 \in \mathcal{F}_2\}$. Show that

$$\mathcal{R}_n(\mathcal{F}_1 + \mathcal{F}_2) = \mathcal{R}_n(\mathcal{F}_1) + \mathcal{R}_n(\mathcal{F}_2)$$

- For a bounded function g, let $\mathcal{F} + g = \{f + g \ : \ f \in \mathcal{F}\}$. Show that

$$\mathcal{R}_n(\mathcal{F} + g) = \mathcal{R}_n(\mathcal{F}).$$

2.2. *Gaussian Complexity [20pts].* For a function class $\mathcal{F}$, let $\mathcal{G}_n(\mathcal{F})$ denote the Gaussian complexity of $\mathcal{F}$, which is defined as

$$\mathcal{G}_n(\mathcal{F}) := \mathbb{E}\left[\sup_{f\in\mathcal{F}} \frac{1}{n}\sum_{i=1}^{n} g_i f(z_i)\right] \tag{2.1}$$

where $g_i \sim \mathcal{N}(0,1)$ i.i.d. random variables independent from z_i's. Rademacher and Gaussian complexities are equivalent up to absolute constants. In this question, you will prove one side of this equivalence relation, i.e., show that

$$\sqrt{\frac{\pi}{2}}\mathcal{G}_n(\mathcal{F}) \geq \mathcal{R}_n(\mathcal{F}). \tag{2.2}$$

Remark: the other direction of the equivalence (which we do not show in this question) involves a $\log(n)$-factor.