HOMEWORK 2 - V0

CSC2532 WINTER 2024

University of Toronto

VERSION HISTORY: V0 → V1:

- **Deadline:** Feb 19, by 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. Rademacher complexity.

1.1. *Rademacher complexity of linear functions constrained in ℓ-1 ball [25pts].* For $i = \{1.2, ..., n\}$. we assume that data points z_i, have bounded ℓ_1-norms, i.e., $\max_i \|z_i\|_1 \leq \kappa$ almost surely. Find an upper bound on the Rademacher complexity of linear functions constrained in a ℓ_∞ ball of radius r, i.e.,

$$\mathcal{F} = \{f \ : \ f(z) = \langle \theta, z \rangle, \quad \|\theta\|_\infty \leq r\}. \tag{1.1}$$

Compare your result to the Rademacher complexity of the ℓ_2 ball.

Hint: Region defined by the ℓ_∞ ball has corners at $r(\pm 1, ..., \pm 1)$. Therefore, $\{\theta \ : \ \|\theta\|_\infty \leq r\} = \text{convex-hull}(\cup\{r(\pm 1, ..., \pm 1)\})$.

1.2. *Generalization of binary classification [25pts].* In a binary classification problem, we have a dataset of n iid (x_i, y_i) feature-response pairs where $x_i \in \mathbb{R}^d$ with $\|x_i\|_1 \leq \kappa$ and $y_i \in \{-1, +1\}$. Learning task involves fitting a function of the form $f_\theta(x) = \text{sign}(\langle \theta, x \rangle)$ where $\|\theta\|_\infty \leq r$. We noticed that 0-1 loss function is not smooth; therefore, we use capped-hinge loss as a relaxation which is given as $\ell((y, x), \theta) = \min\{2, \max\{0, 1 - y\langle \theta, x \rangle\}\}$.

Defining $\hat{\theta} = \text{argmin}_\theta \left\{\hat{R}(\theta) = \frac{1}{n}\sum_{i=1}^n \ell((y_i, x_i), \theta)\right\}$ and $\theta_* = \text{argmin}_\theta \left\{R(\theta) = \mathbb{E}[\ell((y, x), \theta)]\right\}$, derive a generalization bound on the excess risk $R(\hat{\theta}) - R(\theta_*)$. State the rate explicitly.

2. VC dimension.

2.1. *Half-planes [10pts].* Let $\mathcal{F}$ be the class of indicator functions of all closed half planes in $\mathbb{R}^2$. Find the VC dimension of $\mathcal{F}$. Show your steps explicitly.

2.2. *Function class with one parameter [15pts].* Let $\mathcal{F} = \{z \to \mathbb{1}_{\{\sin(\theta z) \geq 0\}} \ : \ \theta \in \mathbb{R}\}$. Find the VC dimension of $\mathcal{F}$.

2.3. *Pajor's Lemma [20pts].* Let $\mathcal{F}$ be a class of Boolean functions on a finite set Z. Then, show that

$$|\mathcal{F}| \leq |\{\Lambda \subseteq Z \ : \ \Lambda \text{ is shattered by } \mathcal{F}\}| \tag{2.1}$$

Hint: Prove this by induction on the size of Z.

1. Show it holds for $|Z| = 1$ (count $\Lambda = \emptyset$).
2. Assume the above inequality holds for $|Z| = n$.
3. For $|Z| = n + 1$, chop out a point z_0 from Z and write $Z = \{z_0\} \cup Z_0$.
4. Define $\mathcal{F}_j = \{f \in \mathcal{F} : f(z_0) = j\}$ for $j = 0, 1$, and the counting operator $S(\mathcal{F}) = |\{\Lambda \subset Z : \Lambda \text{ is shattered by } \mathcal{F}\}|$.
5. Show $S(\mathcal{F}) \geq S(\mathcal{F}_0) + S(\mathcal{F}_1)$. Be careful about double counting.