High Dimensional Statistics **2nd Semester, 2023-2024**

Homework 3 (Deadline: Jun 23)

1. (10 pts) Given a set $T \subset K$. Recall the definition of the covering number: $N(T, d, \varepsilon)$ is the smallest number of points in T which form a ε-covering of T under the metric d. Suppose we are allowed to use the points outside of T to do covering. Then the smallest number of points that are needed to form a ε-covering of T is referred to as the exterior covering number, denoted $N^{ext}(T, d, \varepsilon)$. Show that

$$N^{ext}(T, d, \varepsilon) \leq N(T, d, \varepsilon) \leq N^{ext}(T, d, \varepsilon/2).$$

2. (10pts) Define

$$T^n(s) = \{x \in \mathbb{R}^n : \ \|x\|_0 \leq s, \ \|x\|_2 \leq 1\},$$

where $\|x\|_0$ counts the number of non-zero entries in x. Show that the Gaussian complexity of $T^n(s)$, denoted $\mathcal{G}(T^n(s)) = \mathbb{E}\left[\sup_{x \in T^n(s)} \langle g, x\rangle\right]$, $g \sim \mathcal{N}(0, I_n)$, satisfies

$$\mathcal{G}(T^n(s)) \lesssim \sqrt{s \log\left(\frac{en}{s}\right)}.$$

3. (20 pts) Let $B_1^n = \{x : \ \|x\|_1 \leq 1\}$ be the ℓ_1-norm unit ball. We have already seen that the Gaussian complexity of B_1^n satisfies

$$\mathcal{G}(B_1^n) = \mathbb{E}\left[\sup_{\|x\|_1 \leq 1} \langle g, x\rangle\right] \lesssim \sqrt{\log n}, \quad g \sim \mathcal{N}(0, I_n)$$

based on the duality between ℓ_1-norm and ℓ_∞-norm. In this problem, we attempt to provide a bound for $\mathcal{G}(B_1^n)$ based on the Dudley integral.

- For $\varepsilon > 0$ being sufficiently small, show that the covering of B_1^n under the ℓ_2-norm satisfies

$$\sqrt{\log \mathcal{N}(B_1^n, \|\cdot\|_2, \varepsilon)} \lesssim \min\{\varepsilon^{-1}\sqrt{\log n}, \sqrt{n} \cdot \log(1/\varepsilon)\}.$$

Hint: Volume argument as presented in Lecture 4 may be useful in providing one bound.

- Using the above result and the Dudley integral to provide a bound for $\mathcal{G}(B_1^n)$.

4. (10 pts) Assume $X, Y \in \mathbb{R}^n$ are finite centered Gaussian Processes. Suppose that there exist a pair of index sets $A, B \subset \{1, \cdots, n\}^2$ for which

$$\begin{aligned}
\mathbb{E}[X_i X_j] &\leq \mathbb{E}[Y_i Y_j] \quad \text{for all} \quad (i,,j) \in A;\\
\mathbb{E}[X_i X_j] &\geq \mathbb{E}[Y_i Y_j] \quad \text{for all} \quad (i,,j) \in B;\\
\mathbb{E}[X_i X_j] &= \mathbb{E}[Y_i Y_j] \quad \text{for all} \quad (i,,j) \notin A \cup B.
\end{aligned}$$

Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function whose second derivative satisfies

$$\begin{aligned} \partial_{ij} f \geq 0 \quad & \text{for all} \quad (i,,j) \in A; \\ \partial_{ij} f \leq 0 \quad & \text{for all} \quad (i,,j) \in B. \end{aligned}$$

Show that

$$\mathbb{E}\left[f(X)\right] \leq \mathbb{E}\left[f(Y)\right].$$

5. (10 pts) Let $\phi_j : \mathbb{R} \to \mathbb{R}$ ($j = 1, \cdots, n$) be 1-Lipschitz (i.e., $|\phi_j(t) - \phi_j(s)| \leq |t - s|$). Let w_j ($j = 1, \cdots, n$) be i.i.d $\mathcal{N}(0,1)$ random variables. For any $T \subset \mathbb{R}^n$, show that

$$\mathbb{E}\left[\sup_{t=(t_1,\cdots,t_n)\in T} \sum_{j=1}^n w_j \phi_j(t_j)\right] \leq \mathbb{E}\left[\sup_{t=(t_1,\cdots,t_n)\in T} \sum_{j=1}^n w_j t_j\right].$$

   What does the above result mean in terms of the Gaussian complexity?

6. (10 pts) Show the convex property of KL divergence, i.e., prove that for $0 \leq \alpha \leq 1$, we have
   (a) $D(\alpha\mathbb{P}_1 + (1-\alpha)\mathbb{P}_2 \| \mathbb{Q}) \leq \alpha D(\mathbb{P}_1 \| \mathbb{Q}) + (1-\alpha) D(\mathbb{P}_2 \| \mathbb{Q})$,
   (b) $D(\mathbb{P} \| \alpha\mathbb{Q}_1 + (1-\alpha)\mathbb{Q}_2) \leq \alpha D(\mathbb{P} \| \mathbb{Q}_1) + (1-\alpha) D(\mathbb{P} \| \mathbb{Q}_2)$.

7. (15 pts) Assume X obeys the uniform distribution on $[\theta, \theta+1]$ and the task is to estimate θ from i.i.d observations $X_1, \cdots, X_n$. A natural estimator is the first order statistic

$$X^{(1)} = \min_k X_k.$$

   (a) Prove that

$$\mathbb{E}\left[(X^{(1)} - \theta)^2\right] = \frac{2}{(n+1)(n+2)}.$$

   (b) Use Le Cam method to show that the minimax risk to estimate θ in the squared error is lower bounded by c/n^2 where $c > 0$ is a numerical constant.