Homework Assignment 5

Due Apr 27th by midnight

SDS 384-11 Theoretical Statistics

(4+4+4) for Q1. (5+5) for Q2. (2+2+4) for Q3.

1. In this exercise, we explore the connection between VC dimension and metric entropy. Given a set class $\mathcal{S}$ with finite VC dimension ν, we show that the function class $\mathcal{F}_S := 1_S, S \in \mathcal{S}$ of indicator functions has metric entropy at most

$$N(\delta; \mathcal{F}_{\mathcal{S}}, L^1(P)) \le \left(\frac{K \log(3e/\delta)}{\delta}\right)^{\nu} \qquad \text{For a constant } K \tag{1}$$

Let $\{1_{S^1}, \dots, 1_{S^N}\}$ be a maximal delta packing in the $L^1(P)$ norm, so that:

$$\|1_{S_i} - 1_{S_j}\|_1 = E[|1_{S_i}(X) - 1_{S_j}(X)|] > \delta \qquad \text{for all } i \ne j$$

This is an upper bound on the δ covering number.

(a) Suppose that we generate n samples $X_i, i = 1, \dots, n$ drawn i.i.d. from P. Show that the probability that every set S_i picks out a different subset of $\{X_1, \dots, X_n\}$ is at least $1 - \binom{N}{2}(1-\delta)^n$.

We observe that, by a union bound, and applying the above definitions,

$$\begin{aligned}
&1 - \mathbb{P}(\text{every } S_i, i \in [N] \text{ picks different subset of } X_1, \dots, X_n) \\
&= \mathbb{P}(\text{at least two } S_i, S_j, i \ne j \text{ pick same subset}) \\
&= \mathbb{P}\left(\bigcup_{(i,j)\in\binom{[N]}{2}} \{S_i, S_j \text{ pick same subset}\}\right) \\
&\le \binom{N}{2} \mathbb{P}\left(S_i, S_j \text{ pick same subset }\right) \\
&= \binom{N}{2} \mathbb{P}\left(\bigcap_{k=1}^{n} \mathbb{1}_{S_i}(X_k) = \mathbb{1}_{S_j}(X_k)\}\right) \\
&= \binom{N}{2} \mathbb{P}\left(\mathbb{1}_{S_i}(X_k) = \mathbb{1}_{S_j}(X_k)\}\right)^n \\
&= \binom{N}{2} \left(1 - \|\mathbb{1}_{S_i} - \mathbb{1}_{S_j}\|_1\right)^n \\
&\le \binom{N}{2} (1-\delta)^n
\end{aligned}$$

Rearranging terms yields the desired inequality.

(b) Using part (a), show that for $N \geq 2$ and $n = \lceil 2\log N/\delta \rceil$, there exists a set of n points from which $\mathcal{S}$ picks out at least N subsets, and conclude that $N \leq \left(\frac{3e\log N}{\nu\delta}\right)^{\nu}$.

We proceed by the probabilistic method, showing that, for the stated choices of parameters, $\binom{N}{2}(1-\delta)^n < 1$.

We assume without loss of generality that $0 < \delta < 1$. Thus, we have that

$$\binom{N}{2}(1-\delta)^{\lceil 2\log(N)/\delta \rceil} \leq \binom{N}{2}(1-\delta)^{2\log(N)/\delta} \overset{\text{want}}{<} 1$$

Taking log on both sides, it is sufficient to show that

$$\frac{2\log N}{\delta}\log(1-\delta) < -\log\binom{N}{2} \iff \frac{2\log N}{\delta} > \frac{\log(N(N-1)/2)}{\log\frac{1}{1-\delta}}$$

Now, since $N \geq 2$, we have that $N^2 > \binom{N}{2}$ and thus $2\log(N) > \log(N(N-1)/2)$. Finally, using the well-known inequality $\log\frac{1}{1-\delta} > \delta$ when $\delta \in (0,1)$, we conclude that the above inequality is true. Therefore, by the probabilistic method, there exists a set of n points from which $\mathcal{S}$ picks out at least N subsets.

Now, by definition of the growth function, $\Pi_{\mathcal{F}_{\mathcal{S}}}(n) \geq N$. By Sauer's Lemma, we have the following bound on the growth function:

$$\begin{aligned} N &\leq \Pi_{\mathcal{F}_{\mathcal{S}}}(n) \\ &\leq \sum_{i=0}^{\nu}\binom{n}{i} \\ &\leq \left(\frac{en}{\nu}\right)^{\nu} && \text{assuming } n \geq \nu \\ &= \left(\frac{e\lceil 2\log(N)/\delta \rceil}{\nu}\right)^{\nu} \\ &\leq \left(\frac{3e\log(N)}{\nu\delta}\right)^{\nu} \end{aligned}$$

as desired.

(c) Use part (b) to show that Eq (1) holds with $K := 3e^2/(e-1)$. *Hint: Note that you have* $\frac{N^{1/\nu}}{\log N} \leq \frac{3e}{\nu\delta}$. *Let* $g(x) = x/\log x$. *We are solving for* $g(m^{1/\nu}) \leq 3e/\delta$. *Prove that* $g(x) \leq y$ *implies* $x \leq \frac{e}{e-1}y\log y$.

Following the hint, let us suppose that $\frac{x}{\log x} \leq y$. Assume that $y > e$ and $x > 1$. Therefore,

$$\begin{aligned} \frac{e}{e-1}y\log y &\geq \frac{e}{e-1}\frac{x}{\log x}(\log x - \log\log x) \\ &= \frac{e}{e-1}x - \frac{e}{e-1}\frac{x\log\log x}{\log x} \\ &\overset{\text{want}}{\geq} x \end{aligned}$$

Now, the final inequality above is equivalent to

$$\frac{x}{1-e} \geq \frac{e}{e-1}\frac{x\log\log x}{\log x}$$

Now, for $x \in (1, e)$ the above inequality (and thus the claim) is always true, since $\log\log x < 0$. Thus, we may assume that $x \geq e$. In this case, the above is equivalent to

$$\log x \geq e \log\log x$$

Now, since this inequality is satisfied for $x \geq e$, the claim is established.

Given the claim, the desired result is immediate. Indeed, from the previous problem, we have that

$$\begin{aligned}\frac{N^{1/\nu}}{\frac{1}{\nu}\log N} &= g(N^{1/\nu}) \\ &\leq \frac{3e}{\delta}\end{aligned}$$

and thus, by the claim we just proved,

$$\begin{aligned} N^{1/\nu} &\leq \frac{e}{e-1}\frac{3e}{\delta}\log\frac{3e}{\delta} \\ \implies N &\leq \left(\frac{3e^2}{\delta(e-1)}\log\frac{3e}{\delta}\right)^{\nu}\end{aligned}$$

and thus, Equation 1 holds with $K = \frac{3e^2}{e-1}$, as desired.

2. We will find the covering number of ellipses in this problem. Given a collection of positive numbers $\{\mu_j, j = 1\ldots d\}$, consider the ellipse

$$\mathcal{E} = \{\theta \in \mathcal{R}^d : \sum_i \theta_i^2/\mu_i^2 \leq 1\}$$

(a) Show that

$$\log N(\epsilon; \mathcal{E}, \|.\|_2) \geq d\log(1/\epsilon) + \sum_{j=1}^{d}\log\mu_j$$

Suppose that $\{\theta_1, \ldots, \theta_N\}$ is an ϵ-cover of $\mathcal{E}$. Then, by definition, $\mathcal{E} \subset \cup_{i=1}^N \mathcal{B}_\epsilon(\theta_i)$, where $\mathcal{B}_\epsilon(\theta_i) = \{\|\theta - \theta_i\|_2 \leq \epsilon : \theta \in \mathbb{R}^d\}$. Thus, we have that

$$\begin{aligned}\mathrm{Vol}(\mathcal{E}) &\leq \sum_{i=1}^{N}\mathrm{Vol}(\mathcal{B}_\epsilon(\theta_i)) \\ &= N\mathrm{Vol}(\mathcal{B}_\epsilon(\mathbf{0}))\end{aligned}$$

Now, let us consider the change of coordinates from points in the ellipsoid to points in the ball. Given coordinates $\{u_i\}_{i=1}^d$ from the ϵ-ball, we may map these coordinates in a one-to-one manner to points $\{x_i\}_{i=1}^d$ in $\mathcal{E}$ by the formula:

$$x_i = \frac{\mu_i}{\epsilon}u_i$$

Indeed, since by definition $\sum_i u_i^2 \leq \epsilon^2$, and so

$$\epsilon^2 \geq \sum_i u_i^2 = \sum_i \frac{\epsilon^2}{\mu_i^2} x_i^2$$
$$\implies \sum_i \frac{x_i^2}{\mu_i^2} \leq 1$$

as desired. Therefore, we may compute the volume of $\mathcal{E}$ using the change of variable formula

$$\begin{aligned}
\mathrm{Vol}(\mathcal{E}) &= \int_{\mathcal{E}} dx_1, \ldots, x_n \\
&= \int_{\mathcal{B}_\epsilon(\mathbf{0})} \left| \frac{\partial(x_1, \ldots, x_n)}{\partial(u_1, \ldots, u_n)} \right| du_1, \ldots, u_n \\
&= \int_{\mathcal{B}_\epsilon(\mathbf{0})} \left(\prod_{i=1}^{d} \frac{\mu_i}{\epsilon} \right) du_1, \ldots, u_n \\
&= \left(\prod_{i=1}^{d} \frac{\mu_i}{\epsilon} \right) \mathrm{Vol}(\mathcal{B}_\epsilon(\mathbf{0}))
\end{aligned}$$

Hence,

$$\begin{aligned}
\left(\prod_{i=1}^{d} \frac{\mu_i}{\epsilon} \right) \mathrm{Vol}(B_\epsilon(\mathbf{0})) &= \mathrm{Vol}(\mathcal{E}) \\
&\leq N \mathrm{Vol}(\mathcal{B}_\epsilon(\mathbf{0}))
\end{aligned}$$

and thus,

$$N \geq \prod_{i=1}^{d} \frac{\mu_i}{\epsilon}$$
$$\implies \log N \geq d \log \frac{1}{\epsilon} + \sum_{i=1}^{d} \log \mu_i$$

as desired.

(b) Now consider an infinite-dimensional ellipse, specified by the sequence $\mu_j = j^{-2\beta}$ for some parameter $\beta > 1/2$. Show that

$$\log N(\epsilon; \mathcal{E}, \|.\|_2) \geq C \left(\frac{1}{\epsilon} \right)^{1/2\beta},$$

where $\|\theta - \theta'\|_{\ell_2}^2 = \sum_{j=1}^{\infty} (\theta_i - \theta_j)^2$ is the squared ℓ_2-norm on the space of square summable sequences.

Let us denote the ellipse truncated to d dimensions as:

$$\mathcal{E}_d = \{\tilde{\theta} \in \mathbb{R}^d : \theta \in \mathcal{E}, \tilde{\theta}(i) = \theta(i) \forall i \in [d]\}$$

Let $S = \{\theta_1, \ldots, \theta_N\}$ be an ϵ-covering of $\mathcal{E}$. Define S_d as the elements of S truncated to d dimensions, that is, the set of N elements $\tilde{\theta}_i$ such that $\tilde{\theta}_i(j) = \theta_i(j)$ for $j \in [d]$.

Now, we will show that S_d is an ϵ-covering of $\mathcal{E}_d$. Indeed, fix any $\tilde{\theta} \in \mathcal{E}_d$. By definition, there is some θ such that $\tilde{\theta}(j) = \theta(j)$ for every $j \in [d]$. By definition of S, there exists some θ_i satisfying $\|\theta - \theta_i\|_{\ell_2} \leq \epsilon$. Therefore,

$$\begin{aligned}
\epsilon^2 &\geq \|\theta - \theta_i\|_{\ell_2}^2 \\
&= \sum_{j=1}^{d} (\theta(i) - \theta_i(j))^2 + \sum_{j=d+1}^{\infty} (\theta(i) - \theta_i(j))^2 \\
&= \sum_{j=1}^{d} (\tilde{\theta}(i) - \tilde{\theta}_i(j))^2 + \sum_{j=d+1}^{\infty} (\theta(i) - \theta_i(j))^2 \\
&\geq \sum_{j=1}^{d} (\tilde{\theta}(i) - \tilde{\theta}_i(j))^2 + \sum_{j=d+1}^{\infty} (0 - 0)^2 \\
&= \|\tilde{\theta} - \tilde{\theta}_i\|_2^2
\end{aligned}$$

and thus S_d is also an ϵ-cover of $\mathcal{E}_d$. Therefore, we have that

$$\begin{aligned}
\log N(\epsilon; \mathcal{E}, \|\cdot\|_2) &\geq \log N(\epsilon, \mathcal{E}_d, \|\cdot\|_2) \\
&\geq d \log \frac{1}{\epsilon} + \sum_{i=1}^{d} \log \mu_i && \text{by the previous problem} \\
&\geq d \log \frac{1}{\epsilon} - 2\beta \log d! \\
&\geq d \log \frac{1}{\epsilon} - 2\beta \log(d^{d+1/2} e^{-d+1}) && \text{by Sterling's approximation} \\
&= d \log \frac{1}{\epsilon} - 2\beta d \log d + 2\beta \left(d - 1 + \frac{1}{2} \log d\right)
\end{aligned}$$

Now, choose $d = \left\lceil \left(\frac{1}{\epsilon}\right)^{1/2\beta} \right\rceil$. Then the above inequality becomes

$$\begin{aligned}
\log N(\epsilon; \mathcal{E}, \|\cdot\|_2) &\geq d \log \frac{1}{\epsilon} - 2\beta d \underbrace{\log\left(\left(\frac{1}{\epsilon}\right)^{1/2\beta} + 1\right)}_{\leq \frac{1}{2\beta} \log\left(\frac{1}{\epsilon}\right) + \frac{1}{2}} + 2\beta \left(d - 1 + \frac{1}{2} \underbrace{\log d}_{\geq 0}\right) \\
&\geq \beta\,(d - 2) \\
&\geq C\beta d \qquad \text{(for } C < 1 \text{ small enough)} \\
&\geq C\beta \left(\frac{1}{\epsilon}\right)^{1/2\beta}
\end{aligned}$$

as desired.

3. Consider the set $\mathbb{S}^d(s) = \{\theta \in \mathbb{R}^d : \|\theta\|_0 \leq s, \|\theta\|_2 \leq 1\}$ corresponding to all $s-$sparse vectors in the unit Euclidean ball. We will prove that the Gaussian complexity of this class is upper-bounded by

$$\mathcal{G}(\mathbb{S}^d(s)) \leq C\sqrt{s\log(ed/s)} \tag{2}$$

(a) Show that $\mathcal{G}(\mathbb{S}^d(s)) = E[\max_{|S|=s} \|w_S\|_2]$ where $w_S \in \mathbb{R}^{|S|}$ is the sub-vector of $(w_1, \ldots, w_d)$ indexes by $S \subset \{1, \ldots, d\}$.

$$\begin{aligned}
E[\sup_{\theta \in \mathbb{S}^d(s)} w^T\theta] &= E[\sup_{\|\theta\|_2=1, \max_{S:|S|=s}} w(S)^T\theta(S)] \\
&= E[\sup_{\max_{S:|S|=s}} w(S)^T w(S)/\|w(S)\|_2] \\
&= E[\sup_{\max_{S:|S|=s}} \|w(S)\|_2]
\end{aligned}$$

The second line uses the fact that a dot product between two vectors is maximized when they are aligned.

(b) Show that any fixed subset S with $|S| = s$,

$$P(\|w_S\| \geq \sqrt{s} + \delta) \leq \exp(-\delta^2/2).$$

Consider any vector of IID Gaussians $z \in N(0,1)^m$. $\|z\|_2$ is a convex 1 liptschitz function of Gaussians. Using the Gaussian lipschitz theorem,

$$P(\|z\| \geq E[\|z\|] + \delta) \leq \exp(-\delta^2/2)$$

But also note that $m = E\|z\|^2 \geq (E\|z\|)^2$ by Jensen's inequality. This and the previous equation yields the result for a fixed S with $m = s$.

(c) Use part (b) to establish Eq 2.

$$P(\max_{S:|S|=s} \|w\|_S \geq \sqrt{s} + \delta) \leq \binom{d}{s} \exp(-\delta^2/2)$$

Thus the expectation is given by

$$\begin{aligned}
E[X] &\leq \int_t P(X \geq t)dt \\
&= \int_{t \geq \lambda} P(X \geq t)dt + \int_{t \leq \lambda} P(X \geq t)dt \\
&\leq \int_{t \geq \lambda} \binom{d}{s} \exp(-(t-\sqrt{s})^2/2)dt + \lambda \\
&\leq (ed/s)^s e^{-(\lambda-\sqrt{s})^2} + \lambda \\
&\leq e^{s\log(ed/s)-(\lambda-\sqrt{s})^2} + \lambda
\end{aligned}$$

Picking $\lambda = C\sqrt{s\log(ed/s)}$ for some large enough C, we have the desired upper bound in (a).