Homework Assignment 5

Due Apr 27th by midnight

SDS 384-11 Theoretical Statistics

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

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

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

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

This is an upper bound on the δ covering number.

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

   (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 $N \leq \left(\frac{3e \log N}{\nu\delta}\right)^{\nu}$.

   (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* $x \leq \frac{e}{e-1} y \log y$.

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

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

   (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.

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 $S \subset \{1, \ldots, d\}$.

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

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

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