

# Homework Assignment 5

## Due in class, Monday April 23rd

SDS 384-11 Theoretical Statistics

1. (VC dimension) Compute the VC dimension of the following function classes
  - (a) Circles in  $\mathcal{R}^2$
  - (b) Axis aligned rectangles in  $\mathcal{R}^2$
  - (c) Axis aligned squares in  $\mathcal{R}^2$
2. In class, you upper bounded the Rademacher complexity of a function class. Now you will derive a lower bound.
  - (a) For function classes  $\mathcal{F}$  with function values in  $[0, 1]$ , prove that  $E\|\hat{P}_n - P\|_{\mathcal{F}} \geq \frac{\mathcal{R}_{\mathcal{F}}}{2} - \sqrt{\frac{\log 2}{2n}}$ . *Hint: may be it is easier to start from  $\mathcal{R}_{\mathcal{F}}$  and show that  $\mathcal{R}_{\mathcal{F}} \leq 2E\|\hat{P}_n - P\|_{\mathcal{F}} + \sqrt{\frac{2\log 2}{n}}$ . In order to do this, you would need to add and subtract  $E[f(X)]$  and then use triangle inequality.*
  - (b) Now prove that  $\|P - \hat{P}_n\|_{\mathcal{F}} \geq E\|P - \hat{P}_n\|_{\mathcal{F}} - \epsilon$  with probability at least  $1 - \exp(-cn\epsilon^2)$  for some constant  $c$ .
  - (c) Recall the class of all subsets with finite size in  $[0, 1]$ ? Prove that then Rademacher complexity of this class is at least  $1/2$ . What does this imply?
3. In this exercise, we explore the connection between VC dimension and metric entropy. Given a set class  $\mathcal{S}$  with finite VC dimension  $\nu$ , we show that the function class  $\mathcal{F}_{\mathcal{S}} := 1_S, S \in \mathcal{S}$  of indicator functions has metric entropy at most

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

Let  $\{1_{S_1}, \dots, 1_{S_N}\}$  be a maximal delta packing in the  $L^1(P)$  norm, so that:

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

This is an upper bound on the  $\delta$  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  $1 - \binom{N}{2}(1 - \delta)^n$ .
- (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  $x \leq \frac{e}{e-1} y \log y$ .*