ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Chair for Mathematical Information Science

Prof. Dr. H. Bölcskei
Sternwartstrasse 7
CH-8092 Zürich

Mathematics of Information

Spring semester 2022

Problem Set 12

Problem 1 Sup-norm Continuity. ☕

Which of the following functionals are continuous with respect to the sup-norm?

a) The mean functional $\gamma_1\colon F \mapsto \int x dF(x)$.

b) The Cramér-von Mises functional $\gamma_2\colon F \mapsto \int (F(x) - F_0(x))^2 dF_0(x)$, where F_0 is fixed.

c) The α-quantile functional $\gamma_3\colon F \mapsto \inf\{t \in \mathbb{R} \mid F(t) \geq \alpha\}$.

Problem 2 VC dimension. ☕☕

Given a class $\mathcal{F}$ of binary-valued functions, we say that the set $x_1^n = (x_1, \dots, x_n)$ is shattered by $\mathcal{F}$ if $|\mathcal{F}(x_1^n)| = 2^n$. The VC dimension $\nu(\mathcal{F})$ is the largest integer n for which there is some collection $x_1^n = (x_1, \dots, x_n)$ of n points that is shattered by $\mathcal{F}$.

a) Compute the VC dimension of the following function classes:

i) $\mathcal{F}_1 \coloneqq \{f\colon [0,1] \to \{0,1\} \mid f\colon x \mapsto \mathbb{1}_{x<t}, \quad t \in [0,1]\}$;

ii) $\mathcal{F}_2 \coloneqq \mathcal{F}_1 \cup \{f\colon [0,1] \to \{0,1\} \mid f\colon x \mapsto 1 - \mathbb{1}_{x<t}, \quad t \in [0,1]\}$;

iii) $\mathcal{F}_3 \coloneqq \{f\colon [0,1] \to \{0,1\} \mid f\colon x \mapsto \mathbb{1}_{t_1<x<t_2}, \quad t_1 < t_2 \in [0,1]\}$;

iv) $\mathcal{F}_4 \coloneqq \mathcal{F}_3 \cup \{f\colon [0,1] \to \{0,1\} \mid f\colon x \mapsto 1 - \mathbb{1}_{t_1<x<t_2}, \quad t_1 < t_2 \in [0,1]\}$;

v) $\mathcal{F}_5 \coloneqq \left\{f\colon [0,1] \to \{0,1\} \mid f\colon x \mapsto \sum_{n=1}^{N} \mathbb{1}_{t_{2n-1}<x<t_{2n}}, \quad 0 \leq t_0 < \dots < t_{2N} \leq 1\right\}$;

vi) $\mathcal{F}_6 \coloneqq \{f\colon [-1,1] \to \{-1,1\} \mid f\colon x \mapsto \operatorname{sign}(\sin(tx)), \quad t \in \mathbb{R}\}$.

b) The VC dimension of a class of sets $\mathcal{S}$ is defined as the VC dimension of the class of indicator functions $\{\mathbb{1}_S \mid S \in \mathcal{S}\}$. Let $\mathcal{S}_1$ and $\mathcal{S}_2$ be two classes of sets with finite VC dimensions.

Show that each of the following operations lead to a new set class also with finite dimension.

i) The set class $\mathcal{S}_1^c \coloneqq \{S^c \mid S \in \mathcal{S}_1\}$, where S^c denotes the complement of the set S.

ii) The set class $\mathcal{S}_1 \sqcap \mathcal{S}_2 \coloneqq \{S_1 \cap S_2 \mid S_1 \in \mathcal{S}_1,\ S_2 \in \mathcal{S}_2\}$.

iii) The set class $\mathcal{S}_1 \sqcup \mathcal{S}_2 \coloneqq \{S_1 \cup S_2 \mid S_1 \in \mathcal{S}_1,\ S_2 \in \mathcal{S}_2\}$.

Problem 3 Failure of uniform law.

Let $\mathcal{S}$ be the class of all finite subsets S of $[0,1]$ and consider the function class $\mathcal{F}_{\mathcal{S}} := \{\mathbb{1}_S(\cdot) \mid S \in \mathcal{S}\}$ of indicator functions of such sets. Let $\mathbb{P}$ be a probability distribution over $[0,1]$ that has no atom (i.e., $\mathbb{P}[\{x\}] = 0$ for all $x \in [0,1]$).

a) Show that the function class $\mathcal{F}_{\mathcal{S}}$ is not a Glivenko-Cantelli class for $\mathbb{P}$.

b) Prove the following lower bound on the Rademacher complexity

$$\mathcal{R}_n(\mathcal{F}_{\mathcal{S}}) \geq \frac{1}{2}.$$

Problem 4 VC dim. of rectangles (Winter Exam 2020, Problem 4).

a) Compute the VC dimension of the class

$$\mathcal{H}_1 := \left\{h_{(a,b)} \colon \mathbb{R} \to \{0,1\} \mid a \leq b\right\}$$

of closed intervals of $\mathbb{R}$, with

$$h_{(a,b)}(x) = \begin{cases} 1, & \text{if } a \leq x \leq b, \\ 0, & \text{otherwise.} \end{cases}$$

b) Let

$$\mathcal{H}_2 := \left\{h_{(a_1,a_2,b_1,b_2)} \colon \mathbb{R}^2 \to \{0,1\} \mid a_i \leq b_i, i = 1,2\right\}$$

be the class of axis-aligned rectangles with

$$h_{(a_1,a_2,b_1,b_2)}(x_1,x_2) = \begin{cases} 1, & \text{if } a_i \leq x_i \leq b_i, \ i = 1,2, \\ 0, & \text{otherwise.} \end{cases}$$

i) Provide a set of 4 points shattered by $\mathcal{H}_2$.

ii) Show that there is no $h_{(a_1,a_2,b_1,b_2)} \in \mathcal{H}_2$ such that

$$\begin{cases} h(0,0) = 0, \\ h(-1,0) = h(0,-1) = h(1,0) = h(0,1) = 1. \end{cases}$$

iii) Show that no set of 5 distinct points can be shattered by $\mathcal{H}_2$.

<u>Hint</u>: Taking subproblem b)ii) as an example, show that there is at least one point which cannot be labelled 0 *if all the other points are labelled* 1.

iv) Deduce the VC dimension of the class $\mathcal{H}_2$.

c) Applying the same procedure as in the subproblem (b), compute the VC dimension of the class

$$\mathcal{H}_d := \left\{h_{(a_1,\ldots,a_d,b_1,\ldots,b_d)} \colon \mathbb{R}^d \to \{0,1\} \mid a_i \leq b_i, i = 1,\ldots,d\right\}$$

of axis-aligned d-dimensional rectangles, for $d \geq 1$ a positive integer, where

$$h_{(a_1,\ldots,a_d,b_1,\ldots,b_d)}(x_1,\ldots,x_d) = \begin{cases} 1, & \text{if } a_i \leq x_i \leq b_i,\ i = 1,\ldots,d, \\ 0, & \text{otherwise.} \end{cases}$$

Problem 5 Rademacher complexity (Exam 2020, Problem 3).

We consider a class $\mathcal{F}$ of functions $f\colon \mathbb{R}^d \to \mathbb{R}$ that are b-uniformly bounded, i.e.,

$$\|f\|_\infty \leq b, \quad \forall f \in \mathcal{F}.$$

Given an integer $n \geq 1$, recall the definition of empirical Rademacher complexity

$$\mathcal{R}_n\left(\mathcal{F}\left(x_1^n\right)/n\right) := \frac{1}{n}\,\mathbb{E}_\varepsilon\left[\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^n \varepsilon_i f(x_i)\right|\right],$$

where $x_1^n := \{x_1,\ldots,x_n\}$ with $x_i \in \mathbb{R}^d$, $i = 1,\ldots,n$, is fixed and $\{\varepsilon_i\}_{i=1}^n$ is a sequence of Rademacher random variables, i.e., ε_i takes the values $+1$ and -1 equiprobably, for $i = 1,\ldots,n$. The Rademacher complexity is then defined as

$$\mathcal{R}_n\left(\mathcal{F}\right) := \mathbb{E}_X\left[\mathcal{R}_n\left(\mathcal{F}\left(X_1^n\right)/n\right)\right],$$

where $X_1^n := \{X_1,\ldots,X_n\}$ and the X_i are i.i.d. random variables, for $i = 1,\ldots,n$.

a) Show that

$$\mathcal{R}_n\left(\mathcal{F}\right) \leq \mathcal{R}_n\left(\mathcal{F}\left(X_1^n\right)/n\right) + \sqrt{\frac{2b^2\log(1/\delta)}{n}},$$

with probability at least $1-\delta$.

Hint: Show that the empirical Rademacher complexity satisfies the bounded difference property and use the one sided bounded difference inequality.

b) Consider now two general classes of functions, $\mathcal{F}_1$ and $\mathcal{F}_2$, and define

$$\mathcal{F}_{|\cdot|} := \left\{|f_1 - f_2| \mid f_1 \in \mathcal{F}_1 \text{ and } f_2 \in \mathcal{F}_2\right\}.$$

Derive the following bound:

$$\mathcal{R}_n\left(\mathcal{F}_{|\cdot|}\left(x_1^n\right)/n\right) \leq 2\,\mathcal{R}_n\left(\mathcal{F}_1\left(x_1^n\right)/n\right) + 2\,\mathcal{R}_n\left(\mathcal{F}_2\left(x_1^n\right)/n\right). \tag{1}$$

Hint: Use the Ledoux-Talagrand contraction lemma, i.e., that if $\phi\colon \mathbb{R} \to \mathbb{R}$ is an L-Lipschitz function with $\phi(0) = 0$ and $\mathcal{G}$ a class of functions, then

$$\mathcal{R}_n\left(\left(\phi\circ\mathcal{G}\right)\left(x_1^n\right)/n\right) \leq 2L\mathcal{R}_n\left(\mathcal{G}\left(x_1^n\right)/n\right).$$

c) Define $\mathcal{F}_{\max} := \{f\colon x \mapsto \max\{f_1(x), f_2(x)\} \mid f_1 \in \mathcal{F}_1, f_2 \in \mathcal{F}_2\}$. Using (1), derive the following upper bound

$$\mathcal{R}_n\left(\mathcal{F}_{\max}\left(x_1^n\right)/n\right) \leq \frac{3}{2}\Big(\mathcal{R}_n\left(\mathcal{F}_1\left(x_1^n\right)/n\right) + \mathcal{R}_n\left(\mathcal{F}_2\left(x_1^n\right)/n\right)\Big).$$

Hint: First prove that $\max\{a,b\} = \frac{|a-b|+a+b}{2}$ *for all* $a, b \in \mathbb{R}$.

Problem 6 Metric entropy and VC dim. (Exam 2020, Problem). ☕☕☕

The present problem is concerned with finding a relation between the metric entropy of a uniformly bounded class of functions and its VC dimension. Namely, given a general set $\mathcal{X}$, we consider a class of subsets of $\mathcal{X}$, denoted by $\mathcal{S}$, and we assume that the corresponding class of indicator functions $\mathcal{F}_{\mathcal{S}} := \{\mathbb{1}_S \mid S \in \mathcal{S}\}$ has finite VC dimension ν. Given a probability measure $\mathbb{Q}$ on $\mathcal{X}$, we consider the distance associated with the $L_1(\mathbb{Q})$-norm on $\mathcal{F}_{\mathcal{S}}$, i.e., the distance that assigns to $\mathbb{1}_{S_1} \in \mathcal{F}_{\mathcal{S}}$ and $\mathbb{1}_{S_2} \in \mathcal{F}_{\mathcal{S}}$ the value

$$\|\mathbb{1}_{S_1} - \mathbb{1}_{S_2}\|_{L_1(\mathbb{Q})} := \mathbb{E}_{X\sim\mathbb{Q}}\left[|\mathbb{1}_{S_1}(X) - \mathbb{1}_{S_2}(X)|\right].$$

The aim of this problem is to prove that $\mathcal{F}_{\mathcal{S}}$ has metric entropy $N(\delta; \mathcal{F}_{\mathcal{S}}, L_1(\mathbb{Q}))$ satisfying

$$N(\delta; \mathcal{F}_{\mathcal{S}}, L_1(\mathbb{Q})) \le (2\nu)^{2\nu-1}\left(\frac{3}{\delta}\right)^{2\nu}. \tag{2}$$

In order to establish (2), we fix $\delta > 0$ and take $\{\mathbb{1}_{S_1}, \ldots, \mathbb{1}_{S_M}\}$ to be a maximal δ-packing of size M of $\mathcal{F}_{\mathcal{S}}$ in the $L_1(\mathbb{Q})$-norm, that is

$$\|\mathbb{1}_{S_i} - \mathbb{1}_{S_j}\|_{L_1(\mathbb{Q})} > \delta, \quad \forall i \ne j.$$

We assume in what follows that $\nu \ge 2$ and δ is small enough for

$$3\log(M) > \delta(\nu+1) \tag{3}$$

to hold.

a) Prove that the metric entropy $N(\delta; \mathcal{F}_{\mathcal{S}}, L_1(\mathbb{Q}))$ is upper-bounded according to

$$N(\delta; \mathcal{F}_{\mathcal{S}}, L_1(\mathbb{Q})) \le M.$$

b) Given two distinct sets S_i and S_j, we define their symmetric difference as

$$S_i \triangle S_j := (S_i \cup S_j) \setminus (S_i \cap S_j).$$

Show that a random variable X drawn from $\mathbb{Q}$ satisfies

$$\mathbb{P}\left[X \notin (S_i \triangle S_j)\right] < 1 - \delta.$$

Hint: Note that $\mathbb{1}_{S_i \triangle S_j}(\cdot) = \left|\mathbb{1}_{S_i}(\cdot) - \mathbb{1}_{S_j}(\cdot)\right|$.

c) Suppose that we are given n samples X_k, $k = 1, \ldots, n$, drawn i.i.d. from $\mathbb{Q}$. Show that, given distinct sets S_i and S_j, the sets $S_i \cap \{X_1, \ldots, X_n\}$ and $S_j \cap \{X_1, \ldots, X_n\}$ are distinct if and only if $X_k \in (S_i \triangle S_j)$ for at least one $k \in \{1, \ldots, n\}$.

d) Using the results of subproblems (b) and (c), show that the probability of all the sets $S_i \cap \{X_1, \ldots, X_n\}$, for $i = 1, \ldots, n$, being distinct is at least $1 - \binom{M}{2}(1-\delta)^n$.

Hint: Note that, by subproblem (c), the complementary event of the sets $S_i \cap \{X_1, \ldots, X_n\}$, $i = 1, \ldots, n$, *being distinct is equivalent to the existence of at least one pair of sets* S_i *and* S_j *among the* $\binom{M}{2}$ *possible pairs such that* $X_k \notin (S_i \triangle S_j)$, *for all* $k \in \{1, \ldots, n\}$.

e) From now on, we take $n = \frac{3\log(M)}{\delta} - 1$. Using the inequality $(1-\delta)^n \leq e^{-n\delta}$ together with the assumptions (3) and $\nu \geq 2$, show that

$$1 - \binom{M}{2}(1-\delta)^n \geq 1 - M^2 e^{-n\delta} > 0,$$

and deduce that there must exist a set of n points $\{x_1, \dots, x_n\}$ from which $\mathcal{S}$ picks out at least M subsets, i.e.,

$$M \leq |\mathcal{S}(\{x_1, \dots, x_n\})|,$$

where $\mathcal{S}(\{x_1, \dots, x_n\}) \coloneqq \{(\mathbb{1}_S(x_1), \dots, \mathbb{1}_S(x_n)) \mid S \in \mathcal{S}\}$.

f) By upper-bounding $|\mathcal{S}(\{x_1, \dots, x_n\})|$, prove that

$$M \leq \left(\frac{3\log(M)}{\delta}\right)^{\nu}. \tag{4}$$

<u>Hint</u>: Use the Vapnik-Chervonenkis-Sauer-Shelah lemma. You will also need assumption (3) and $n = \frac{3\log(M)}{\delta} - 1$.

g) Conclude the proof of the desired result (2).

<u>Hint</u>: Using that $te^{-t} \leq 1$, *for* $t \geq 0$, *first prove that*

$$\sup_{t \geq 0} \left(t^{2\nu} e^{-t}\right) \leq (2\nu)^{2\nu - 1},$$

and show that inequality (4) can be rewritten as

$$M \leq \frac{\log(M)^{2\nu}}{M} \left(\frac{3}{\delta}\right)^{2\nu}.$$

You will also need the result established in subproblem (a).