ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Prof. Dr. Helmut Bölcskei
Chair for Mathematical Information Science

Exam on Neural Network Theory
August 26, 2021

Please note:

- Exam duration: 180 minutes
- Maximum number of points: 100
- You are not allowed to use any printed or handwritten material (i.e., books, lecture and discussion session notes, summaries), computers, tablets, smart phones or other electronic devices.
- Your solutions should be explained in detail and your handwriting needs to be clean and legible.
- Please do not use red or green pens. You may use pencils.
- Please note that the "ETH Zurich Disciplinary Code" (RS 361.1) applies.

Before you start:

1. The problem statements consist of 5 pages including this page. Please verify that you have received all 5 pages.
2. Please fill in your name, student ID card number and sign below.
3. Please place your student ID card at the front of your desk so we can verify your identity.

During the exam:

4. For your solutions, please use only the empty sheets provided by us. Should you need additional sheets, please let us know.
5. Each problem consists of several subproblems. If you do not provide a solution to a subproblem, you may, whenever applicable, nonetheless assume its conclusion in the ensuing subproblems.

After the exam:

6. Please write your name on every solution sheet and prepare all sheets in a pile. All sheets, including those containing problem statements, must be handed in.
7. Please clean up your desk and stay seated and silent until you are allowed to leave the room in a staggered manner row by row.
8. Please avoid crowding and leave the building by the most direct route.

Family name: .... First name: ....

Student ID card No.: ....

Signature: ....

Problem 1 (25 points)

(a) (4 points) Realize the leaky ReLU function $f : \mathbb{R} \mapsto \mathbb{R}$,

$$f(x) = \begin{cases} x, & \text{if } x > 0, \\ 0.001x, & \text{otherwise,} \end{cases}$$

through a ReLU network Ψ (see the Handout for the definition of a ReLU network) with $\mathcal{L}(\Psi) = 2$. Specify the width, weight magnitude, and connectivity of the network Ψ.

(b) (4 points) For $n \in \mathbb{N}$, let S_n be a ReLU network given by

$$S_n(x) = \begin{cases} 2x, & n = 1, \\ \overbrace{2\,\rho(2\,\rho(\ldots 2\,\rho}^{n-1 \text{ times}}(2\,x)\ldots)), & n \geq 2, \end{cases}$$

for $x \in \mathbb{R}$. For all $n \in \mathbb{N}$, specify the width, depth, and weight magnitude of S_n, and show that, for $x \in \mathbb{R}$,

$$S_n(x) = \begin{cases} 2x, & n = 1, \\ 2^n \rho(x), & n \geq 2. \end{cases}$$

(c) (5 points) For $n \in \mathbb{N} \cup \{0\}$, consider the ReLU network $\Phi_n : \mathbb{R} \to \mathbb{R}$ given by

$$\Phi_n(x) = \begin{pmatrix} 2^n & -2^{n+1} & 2^{n+1} & -2^{n+1} & 2^n \end{pmatrix} \circ \rho \circ \left(\begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} x - \begin{pmatrix} -2 \\ -1 \\ 0 \\ 1 \\ 2 \end{pmatrix} \right).$$

For all $n \in \mathbb{N}$, determine the depth, width, connectivity, and weight magnitude of the network Φ_n according to the definition of ReLU networks provided in the Handout. Plot the function $\Phi_0(x)$.

(d) (2 points) For $n \in \mathbb{N}$, find an expression for Φ_n as a function of Φ_0.

(e) (5 points) We now want to realize Φ_n, $n \in \mathbb{N}$, by a wider ReLU network with weight magnitude upper-bounded by 2 but of the same depth as Φ_n. Formally, show that, for all $n \in \mathbb{N}$, there exists a ReLU network R_n such that $R_n(x) = \Phi_n(x)$, for all $x \in \mathbb{R}$, with $\mathcal{L}(R_n) = 2$ and $\mathcal{B}(R_n) \leq 2$. Specify R_n and $\mathcal{W}(R_n)$, $n \in \mathbb{N}$.

Hint: *Use the result from subproblem (d).*

(f) (5 points) Next, we want to realize Φ_n, $n \in \mathbb{N}$, by a deeper network with weight magnitude no greater than 2 but of the same width as Φ_n. Formally, show that, for all $n \in \mathbb{N}$, there exists a ReLU network T_n such that $T_n(x) = \Phi_n(x)$, for all $x \in \mathbb{R}$, with $\mathcal{W}(T_n) \leq 5$ and $\mathcal{B}(T_n) \leq 2$. Specify T_n and $\mathcal{L}(T_n)$, for all $n \in \mathbb{N}$.

Hint: *Use the results from subproblem (b) and (d) and Lemma 1 in the Handout.*

Problem 2 (25 points)

(a) (13 points) Let $(\mathcal{X}, \rho)$ be a metric space, $\mathcal{C}$ a compact set in $\mathcal{X}$, and $\varepsilon \in \mathbb{R}_+$.

(i) (3 points) State the definition of an ε-covering of $\mathcal{C}$ and of the ε-covering number $N(\varepsilon; \mathcal{C}, \rho)$.

(ii) (3 points) State the definition of an ε-packing of $\mathcal{C}$ and of the ε-packing number $M(\varepsilon; \mathcal{C}, \rho)$.

(iii) (3 points) Order the following quantities without proof

$$N(\varepsilon; \mathcal{C}, \rho), N(2\varepsilon; \mathcal{C}, \rho), M(\varepsilon; \mathcal{C}, \rho), M(2\varepsilon; \mathcal{C}, \rho).$$

(iv) (4 points) Suppose that $\mathcal{D}$ is a subset of $\mathcal{C}$. Show that $M(\varepsilon; \mathcal{D}, \rho) \leq M(\varepsilon; \mathcal{C}, \rho)$.

(b) (12 points) Consider the following parametric class of functions

$$\mathcal{F} = \{f_\theta : [0,1] \to \mathbb{R} \mid \theta \in [0,1]\},$$

where for $\theta \in [0,1]$, we set $f_\theta(x) := \sin(2\pi(x-\theta)), x \in [0,1]$. Consider the metric ρ induced by the L^1-norm of functions defined on $[0,1]$ according to

$$\rho(f,g) = \|f-g\|_{L^1} = \int_0^1 |f(x) - g(x)|\, dx.$$

(i) (6 points) For $\varepsilon < 1/2$, construct an ε-covering $A(\varepsilon)$ for the class $\mathcal{F}$ equipped with the metric ρ as follows: Set $T = \left\lfloor \frac{2}{\varepsilon} \right\rfloor$, and for $i = 0, 1, \ldots, T$, define $\theta_i = \frac{\varepsilon i}{2}$. We also add the point $\theta_{T+1} = 1$, thereby forming a collection $\{\theta_0, \ldots, \theta_T, \theta_{T+1}\}$ contained in $[0,1]$. Show that the associated functions $\{f_{\theta_0}, \ldots, f_{\theta_T}, f_{\theta_{T+1}}\}$ constitute an ε-covering of $\mathcal{F}$ with respect to the metric ρ. Find an upper bound on the ε-covering number $N(\varepsilon; \mathcal{F}, \rho)$ in terms of ε.

Hint: *You can use without proof that* $sin(x) - sin(y) = 2\cos(\frac{x+y}{2})\sin(\frac{x-y}{2})$, $x, y \in \mathbb{R}$, $|sin(x)| \leq |x|$, $\forall x \in \mathbb{R}$, *and* $\int_0^1 |\cos(2\pi(x-\phi))|\, dx = \frac{2}{\pi}$, $\forall \phi \in \mathbb{R}$.

(ii) (6 points) For $\varepsilon < 1/2$, construct an ε-packing $P(\varepsilon)$ for the class $\mathcal{F}$ with respect to the metric ρ, and show that there exists a positive constant c not depending on ε such that $\log M(\varepsilon; \mathcal{F}, \rho) \geq c \log_2(\frac{1}{\varepsilon})$, for all $\varepsilon < 1/2$.

Hint: *Use the Hint in (i). Further, you can use without proof that* $|\sin(x)| \geq \frac{2}{\pi}|x|$, $\forall x \in [-\frac{\pi}{2}, \frac{\pi}{2}]$.

Problem 3 (25 points)

(a) (4 points) Let X_1 be a finite subset of $\mathbb{R}^d$, $d \in \mathbb{N}$, let $\{X_1^+, X_1^-\}$ be a dichotomy of X_1, and consider the mapping $\phi : \mathbb{R}^d \mapsto \mathbb{R}^m$, $m \in \mathbb{N}$. State the definition for the dichotomy $\{X_1^+, X_1^-\}$ to be homogeneously linearly separable and the definition for it to be ϕ-separable.

(b) (6 points) Consider the set $X_2 = \{(1,0), (-1,-1), (0,1)\} \subset \mathbb{R}^2$. List all homogeneously linearly separable dichotomies $\left\{X_2^+, X_2^-\right\}$ along with the corresponding separating w-vectors.

(c) (5 points) Consider $X_3 = \{1,2,3,4\}$, its dichotomy

$$\{X_3^+ = \{1,3\}, X_3^- = \{2,4\}\},$$

and $\phi_1(x) = (x, 1)$. Show that the dichotomy $\{X_3^+, X_3^-\}$ is not ϕ_1-separable.

(d) (5 points) Find a ReLU network $\Phi_1 : \mathbb{R} \mapsto \mathbb{R}$ with $\mathcal{L}(\Phi_1) = 2$ such that the dichotomy $\{X_3^+, X_3^-\}$ as defined in (c) is Φ_1-separable.

(e) (5 points) Does there exist a ReLU network $\Phi_2 : \mathbb{R} \mapsto \mathbb{R}$ such that every dichotomy of X_3 as defined in (c) is Φ_2-separable? If yes, provide an example of such a Φ_2. If no, explain why.

Problem 4 (25 points)

For a function $f : \mathbb{R}^d \mapsto \mathbb{R}^k$, the Lipschitz constant is defined as

$$|f|_{Lip} := \sup_{\substack{x,y\in\mathbb{R}^d \\ x\neq y}} \frac{\|f(x)-f(y)\|_\infty}{\|x-y\|_\infty}, \tag{1}$$

and f is said to be Lipschitz-continuous if $|f|_{Lip}$ is finite. In this problem, we investigate the Lipschitz-continuity of ReLU networks with bounded weight magnitude and all bias vectors equal to zero.

(a) (5 points) Show that for the linear mapping $f : \mathbb{R}^d \mapsto \mathbb{R}^k$ given by $f(x) = Ax$, where $A \in \mathbb{R}^{k\times d}$, the Lipschitz constant of f is upper-bounded by $d\,\|A\|_\infty$.

Hint: *You can use without proof that* $\|Ax\|_\infty \leq n\big\|A\big\|_\infty \|x\|_\infty$*, for every* $A \in \mathbb{R}^{m\times n}$*, for all* $x \in \mathbb{R}^n$*,* $m, n \in \mathbb{N}$*.*

(b) (5 points) Show that for a single-hidden-layer ReLU network $\Phi_1 : \mathbb{R}^d \mapsto \mathbb{R}^\ell$ given by $\Phi_1(x) = A_2\,\rho(A_1x)$, where $A_1 \in \mathbb{R}^{k\times d}$, $A_2 \in \mathbb{R}^{\ell\times k}$, $d, k, \ell \in \mathbb{N}$, and ρ is the ReLU activation function defined in the Handout,

$$|\Phi_1|_{Lip} \leq dk\|A_1\|_\infty\|A_2\|_\infty. \tag{2}$$

Hint: *Use the Hint from (a). Further, use, without proof, that* $\|\rho(x)-\rho(y)\|_\infty \leq \|x-y\|_\infty$*,* $x, y \in \mathbb{R}^n$*,* $n \in \mathbb{N}$*.*

(c) (7 points) Let $n \in \mathbb{N}_{\geq 2}$ and $W \in \mathbb{N}$. Further, let $A_1 \in \mathbb{R}^{W\times 1}$, $A_2 \in \mathbb{R}^{1\times W}$, and consider the single-hidden-layer ReLU network

$$\Psi_n(x) = A_2(\rho(A_1x)), \quad x \in \mathbb{R},$$

with $\mathcal{B}(\Psi_n) \leq 2$ and all bias vectors equal to 0. Show that if such a Ψ_n is to satisfy

$$\Psi_n(x) = 2^n\,\rho(x), \quad \text{for all } x \in \mathbb{R},$$

we necessarily have

$$W \geq 2^{n-2}.$$

Hint: *Consider the Lipschitz constant of* Ψ_n *and use* (2) *from subproblem (b).*

(d) (8 points) Consider a ReLU network with all bias vectors equal to zero, given by

$$\Phi(x) = A_n(\rho\,(A_{n-1}(\ldots\rho\,(A_1x)\ldots))), \quad x \in R^{N_0},$$

where $A_\ell \in \mathbb{R}^{N_\ell\times N_{\ell-1}}$, $\ell = 1, \ldots, n$. Show that

$$|\Phi(x)|_{Lip} \leq (\mathcal{W}(\Phi)\mathcal{B}(\Phi))^n.$$

Hint: *Use the Hints and results from subproblems (a) and (b).*