Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Prof. Dr. Helmut Bölcskei
Chair for Mathematical Information Science

Solutions to the
Exam on Neural Network Theory
August 26, 2021

Problem 1

(a) One candidate for Ψ is

$$\begin{aligned}\Psi(x) &= \rho(x) - 0.001\rho(-x)\\ &= \begin{pmatrix}1 & -0.001\end{pmatrix} \circ \rho \circ \left(\begin{pmatrix}1\\-1\end{pmatrix} x \right), \; x \in \mathbb{R}.\end{aligned}$$

This network satisfies $\mathcal{W}(\Psi) = 2$, $\mathcal{B}(\Psi) = 1$, and $\mathcal{M}(\Psi) = 4$.

(b) For $n \in \mathbb{N}$, $\mathcal{L}(S_n) = n$, $\mathcal{W}(S_n) = 1$, and $\mathcal{B}(S_n) = 2$. By direct computation we get $S_1(x) = 2x$, $S_2(x) = 2\rho(2x) = 4\rho(x)$. We prove by induction that $S_m(x) = 2^m\rho(x)$, for all $x \in \mathbb{R}$ and $m \geq 2$. The base case $S_2(x) = 4\rho(x)$ was already established. It remains to prove the induction step. To this end, suppose that for some integer $m \geq 2$, we have $S_m = 2^m\rho(x)$, for all $x \in \mathbb{R}$. It then follows from the definition of S_n, $n \geq 2$, in the problem statement that $S_{m+1}(x) = 2\,\rho(S_m(x)) = 2\,\rho(2^m\rho(x)) = 2^{m+1}\rho(x)$, for all $x \in \mathbb{R}$. This establishes the induction step.

(c) For $n \in \mathbb{N} \cup \{0\}$, $\mathcal{L}(\Phi_n) = 2$, $\mathcal{W}(\Phi_n) = 5$, $\mathcal{B}(\Phi_n) = 2^{n+1}$, and $\mathcal{M}(\Phi_n) = 14$. The plot of $\Phi_0(x)$ is given below.

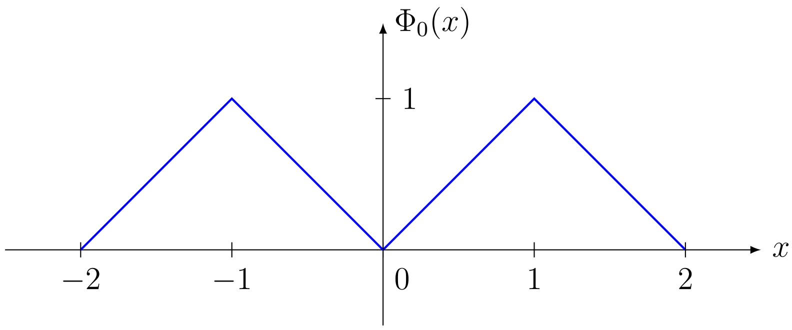

Figure 1: $\Phi_0(x)$.

(d) Let $n \in \mathbb{N}$. We have

$$\begin{aligned}\Phi_n(x) &= 2^n\,\rho(x+2) - 2^{n+1}\,\rho(x+1) + 2^{n+1}\,\rho(x) - 2^{n+1}\,\rho(x-1) + 2^n\,\rho(x-2)\\ &= 2^n(\rho(x+2) - 2\,\rho(x+1) + 2\,\rho(x) - 2\,\rho(x-1) + \rho(x-2))\\ &= 2^n\,\Phi_0(x), \quad x \in \mathbb{R}.\end{aligned}$$

(e) We have, for $x \in \mathbb{R}$,

$$\begin{aligned} \Phi_n(x) &= 2^n \Phi_0(x), & (1) \\ &= 2^n \left(\rho(x+2) - 2\rho(x+1) + 2\rho(x) - 2\rho(x-1) + \rho(x-2)\right) & (2) \\ &= \sum_{i=1}^{2^n} \left(\rho(x+2) - 2\rho(x+1) + 2\rho(x) - 2\rho(x-1) + \rho(x-2)\right), & (3) \end{aligned}$$

where (1) follows from the result of subproblem (d), and in (3) we replaced multiplication by 2^n by summing 2^n copies of the same term. Using the definition of ReLU networks in the Handout, we now see that Φ_n can be realized as a single-hidden-layer ReLU network R_n with $\mathcal{L}(R_n) = 2$, $\mathcal{B}(R_n) = 2$, and $\mathcal{W}(R_n) = 2^n \times 5 = 2^n 5$.

(f) We first show that $(S_n \circ \Phi_0)(x) = (\Phi_n)(x)$, for all $x \in \mathbb{R}$ and $n \in \mathbb{N}$. Suppose first that $n = 1$. For $x \in \mathbb{R}$, $S_n(\Phi_0(x)) = 2\Phi_0(x) = \Phi_1(x)$, where the last equality follows from the result of subproblem (d). Now let $n \in \mathbb{N}$ with $n \geq 2$. Then, for $x \in \mathbb{R}$, $S_n(\Phi_0(x)) = 2^n \rho(\Phi_0(x)) = 2^n \, \Phi_0(x) = \Phi_n(x)$, for all $x \in \mathbb{R}$, where the first equality is by the result from subproblem (b), the second follows from $\Phi_0(x) \geq 0$, $\forall x \in \mathbb{R}$, and the third is by the result from subproblem (d).

Using Lemma 1 in the Handout, $S_n \circ \Phi_0$ can be realized by a ReLU network T_n with $\mathcal{L}\left(T_n\right) = \mathcal{L}(S_n) + \mathcal{L}(\Phi_0) = n + 2$, $\mathcal{W}(T_n) \leq \max\{2, \mathcal{W}(S_n), \mathcal{W}(\Phi_0)\} = 5$, and $\mathcal{B}(T_n) = \max\{\mathcal{B}(S_n), \mathcal{B}(\Phi_0)\} = 2$.

Problem 2

(a) (i) An ε-covering of the compact set $\mathcal{C} \subseteq \mathcal{X}$ with respect to the metric ρ is a set $\{x_1, \ldots, x_N\} \subset \mathcal{C}$ such that for each $x \in \mathcal{C}$, there exists an $i \in \{1, \ldots, N\}$ so that $\rho(x, x_i) \leq \varepsilon$. The ε-covering number $N(\varepsilon; \mathcal{C}, \rho)$ is the cardinality of a smallest ε-covering of $\mathcal{C}$.

(ii) An ε-packing of a compact set $\mathcal{C} \subseteq \mathcal{X}$ with respect to the metric ρ is a set $\{x_1, \ldots, x_N\} \subset \mathcal{C}$ such that $\rho(x_i, x_j) > \varepsilon$, for all distinct i, j. The ε-packing number $M(\varepsilon; \mathcal{C}, \rho)$ is the cardinality of a largest ε-packing of $\mathcal{C}$.

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

(iv) Suppose that $\{x_1, \ldots, x_n\}$ is an ε-packing of $\mathcal{D}$ such that $n = M(\varepsilon; \mathcal{D}, \rho)$. By definition, $\rho(x_i, x_j) > \varepsilon$, for all $i, j \in \{1, \ldots, n\}$ with $i \neq j$. This together with $\mathcal{D} \subset \mathcal{C}$, implies that $\{x_1, \ldots, x_n\} \subset \mathcal{C}$ and is an ε-packing of $\mathcal{C}$, and therefore $M(\varepsilon; \mathcal{D}, \rho) = n \leq M(\varepsilon; \mathcal{C}, \rho)$ by definition of the packing number of $\mathcal{C}$.

(b) (i) For every $f_\theta \in \mathcal{F}$, we can find a θ_i in the set $\{\theta_0, \ldots, \theta_T, \theta_{T+1}\}$ such that $|\theta - \theta_i| \leq \frac{\varepsilon}{4}$. We then have

$$
\begin{aligned}
\|f_{\theta_i} - f_\theta\|_{L^1} &= \int_0^1 |\sin(2\pi(x - \theta_i)) - \sin(2\pi(x - \theta))|\, dx && (4)\\
&= \int_0^1 \left| 2\cos\left(2\pi\left(x - \frac{\theta + \theta_i}{2}\right)\right) \sin(\pi(\theta - \theta_i)) \right| dx && (5)\\
&= 2|\sin(\pi(\theta - \theta_i))| \int_0^1 \left| \cos\left(2\pi\left(x - \frac{\theta + \theta_i}{2}\right)\right) \right| dx && (6)\\
&\leq 2|\pi(\theta - \theta_i)| \frac{2}{\pi} && (7)\\
&\leq \varepsilon, && (8)
\end{aligned}
$$

where (5) and (7) follow from the Hint, and (8) is a consequence of $|\theta - \theta_i| \leq \frac{\varepsilon}{4}$. We can therefore conclude that the set $\{f_{\theta_0}, \ldots, f_{\theta_T}, f_{\theta_{T+1}}\}$ constitutes an ε-covering of $\mathcal{F}$ with respect to the metric ρ. An upper bound on the covering number is hence obtained according to $N(\varepsilon; \mathcal{F}, \rho) \leq T + 2 \leq \frac{2}{\varepsilon} + 2$.

(ii) We construct an explicit packing as follows. Set $T = \left\lfloor \frac{1}{\varepsilon} \right\rfloor$, and for $i = 0, 1, \ldots, T$, define $\theta_i = \frac{\varepsilon i}{2}$. Note that for all i, j with $i \neq j$, we have

$$
\begin{aligned}
\|f_{\theta_i} - f_{\theta_j}\|_{L^1} &= \int_0^1 |\sin(2\pi(x - \theta_i)) - \sin(2\pi(x - \theta_j))| dx && (9)\\
&= \int_0^1 \left| 2\cos\left(2\pi\left(x - \frac{\theta_j + \theta_i}{2}\right)\right) \sin(\pi(\theta_j - \theta_i)) \right| dx && (10)\\
&= 2\,|\sin(\pi(\theta_j - \theta_i))| \int_0^1 \left| \cos\left(2\pi\left(x - \frac{\theta_j + \theta_i}{2}\right)\right) \right| dx && (11)\\
&= 2\,|\sin(\pi(\theta_j - \theta_i))| \frac{2}{\pi} && (12)\\
&\geq 2\,\frac{2}{\pi}|\pi(\theta_j - \theta_i)| \frac{2}{\pi} && (13)\\
&= \frac{8}{\pi}|\theta_j - \theta_i|, && (14)
\end{aligned}
$$

where (10) and (12) follow from the Hint in subproblem (i), (13) from the Hint in this subproblem with $|\pi(\theta_j - \theta_i)| \leq |\pi \frac{\varepsilon T}{2}| \leq \frac{\pi}{2}$. Since $|\theta_j - \theta_i| \geq \frac{\varepsilon}{2}$ for all i, j with $i \neq j$, we have $\|f_{\theta_i} - f_{\theta_j}\|_{L^1} \geq \frac{4}{\pi}\varepsilon > \varepsilon$. We can therefore conclude that $\{f_{\theta_0}, \ldots, f_{\theta_T}\}$ constitutes an ε-packing and the corresponding packing number satisfies $M(\varepsilon; \mathcal{F}, \rho) \geq T + 1 \geq \frac{1}{\varepsilon}$. Hence,

$$\log_2 M(\varepsilon; \mathcal{F}, \rho) \geq \log_2 \left(\frac{1}{\varepsilon}\right),$$

and the solution is concluded by taking $c := 1$.

Problem 3

(a) The dichotomy $\{X_1^+, X_1^-\}$ is said to be homogeneously linearly separable if there exists a nonzero vector $w_1 \in \mathbb{R}^d$ such that

$$\langle x, w_1 \rangle > 0, \text{ for all } x \in X_1^+,$$
$$\langle x, w_1 \rangle < 0, \text{ for all } x \in X_1^-,$$

and it is said to be ϕ-separable if there exists a nonzero vector $w_2 \in \mathbb{R}^m$ such that

$$\langle \phi(x), w_2 \rangle > 0, \text{ for all } x \in X_1^+,$$
$$\langle \phi(x), w_2 \rangle < 0, \text{ for all } x \in X_1^-.$$

(b) The number of homogeneously linearly separable dichotomies is given by

$$\sum_{k=0}^{1} 2\binom{2}{k} = 2 + 4 = 6.$$

These 6 dichotomies $\{X_2^+, X_2^-\}$ of X_2 together with the corresponding w-vectors are as follows

$$\begin{aligned}
&\{X_2^+ = \{(1,0),(-1,-1)\}, X_2^- = \{(0,1)\}\}, w = (1,-2);\\
&\{X_2^+ = \{(0,1)\}, X_2^- = \{(1,0),(-1,-1)\}\}, w = (-1,2);\\
&\{X_2^+ = \{(0,1),(-1,-1)\}, X_2^- = \{(1,0)\}\}, w = (-2,1);\\
&\{X_2^+ = \{(1,0)\}, X_2^- = \{(0,1),(-1,-1)\}\}, w = (2,-1);\\
&\{X_2^+ = \{(0,1),(1,0)\}, X_2^- = \{(-1,-1)\}\}, w = (1,1);\\
&\{X_2^+ = \{(-1,-1)\}, X_2^- = \{(0,1),(1,0)\}\}, w = (-1,-1).
\end{aligned}$$

(c) Suppose for the sake of contradiction that $\{X_3^+, X_3^-\}$ is ϕ_1-separable. Then, there exists a vector $w = (u, v)$ such that

$$\langle \phi_1(x), (u,v) \rangle > 0, \text{ for all } x \in X_3^+,$$
$$\langle \phi_1(x), (u,v) \rangle < 0, \text{ for all } x \in X_3^-,$$

which amounts to

$$u + v > 0, \tag{15}$$
$$3u + v > 0, \tag{16}$$
$$2u + v < 0, \tag{17}$$
$$4u + v < 0. \tag{18}$$

It now follows from (15) and (17) that $u < 0$, while (17) and (16) taken together imply $u > 0$, which establishes the desired contradiction.

(d) Let $\Phi_1(x) := 2\rho(x) - 4\rho(x-1) + 4\rho(x-2) - 4\rho(x-3) - 1$, $x \in \mathbb{R}$, see Fig. 2 for a plot of Φ_1 on $[0,4]$. It follows from the definition in the Handout that $\mathcal{L}(\Phi_1) = 2$. Further, we get by direct inspection $\Phi(1) = \Phi(3) = 1$, $\Phi(2) = \Phi(4) = -1$. Let $w = 1$. It then follows that

$$\langle \Phi_1(x), w \rangle = 1 > 0, \text{ for all } x \in X_3^+,$$
$$\langle \Phi_1(x), w \rangle = -1 < 0, \text{ for all } x \in X_3^-,$$

which establishes the Φ_1-separability of $\{X_3^+, X_3^-\}$.

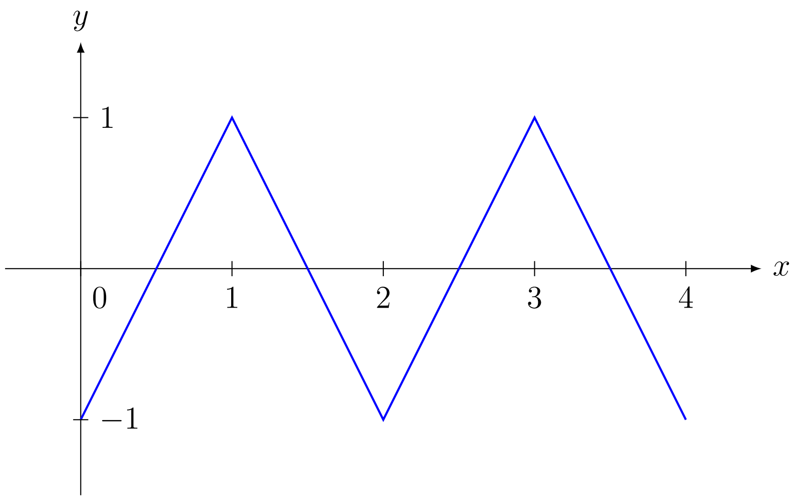

Figure 2: $\Phi_1(x)$ on $[0, 4]$.

(e) No such ReLU network exists. To see this, assume by way of contradiction that Φ_2 is such a ReLU network. If $\{\{1, 2\}, \{3, 4\}\}$ were Φ_2-separable, there would exist a $w_1 \in \mathbb{R}$ such that $\Phi_2(1) \times w_1 > 0$ and $\Phi_2(2) \times w_1 > 0$, which implies that $\Phi_2(1)$ and $\Phi_2(2)$ have the same sign. Next, $\{\{1\}, \{2, 3, 4\}\}$ being Φ_2-separable would imply the existence of a $w_2 \in \mathbb{R}$ such that $\Phi_2(1) \times w_1 > 0$ and $\Phi_2(2) \times w_1 < 0$, which would imply that $\Phi(1)$ and $\Phi(2)$ have different signs. This establishes the desired contradiction.

Problem 4

(a) We have

$$|f|_{Lip} = \sup_{\substack{x,y\in\mathbb{R}^d \\ x\neq y}} \frac{\|f(x)-f(y)\|_\infty}{\|x-y\|_\infty} \tag{19}$$

$$= \sup_{\substack{x,y\in\mathbb{R}^d \\ x\neq y}} \frac{\|A(x-y)\|_\infty}{\|x-y\|_\infty} \tag{20}$$

$$\leq \frac{d\,\|A\|_\infty\,\|x-y\|_\infty}{\|x-y\|_\infty} \tag{21}$$

$$= d\,\|A\|_\infty\,, \tag{22}$$

where (21) follows from the inequality in the Hint.

(b) For all $x, y \in \mathbb{R}^d$ with $x \neq y$, we have

$$\|\Phi_1(x)-\Phi_1(y)\|_\infty = \|A_2\,\rho(A_1x) - A_2\,\rho(A_1y)\|_\infty \tag{23}$$

$$\leq k\,\|A_2\|_\infty\,\|\rho(A_1x)-\rho(A_1x)\| \tag{24}$$

$$\leq k\,\|A_2\|_\infty\,\|A_1x - A_1y\|_\infty \tag{25}$$

$$\leq k\,\|A_2\|_\infty\,d\,\|A_1\|_\infty\,\|x-y\|_\infty\,, \tag{26}$$

where (24) and (26) follow from the Hint in subproblem (a), and (25) is by the inequality in the Hint to subproblem (b).

We therefore have

$$|\Phi_1|_{Lip} = \sup_{\substack{x,y\in\mathbb{R}^d \\ x\neq y}} \frac{\|\Phi_1(x)-\Phi_1(y)\|}{\|x-y\|_\infty}$$

$$\leq \sup_{\substack{x,y\in\mathbb{R}^d \\ x\neq y}} \frac{kd\,\|A_2\|_\infty\,\|A_1\|_\infty\,\|x-y\|_\infty}{\|x-y\|_\infty}$$

$$= dk\,\|A_1\|_\infty\,\|A_2\|_\infty\,.$$

(c) On the one hand, we have

$$|\Psi_n|_{Lip} = \sup_{\substack{x,y\in\mathbb{R} \\ x\neq y}} \frac{|\Psi_n(x)-\Psi_n(y)|}{|x-y|}$$

$$\geq \frac{|\Psi_n(1)-\Psi_n(0)|}{|1-0|} \tag{27}$$

$$= \frac{2^n-0}{1-0}$$

$$= 2^n.$$

On the other hand, it follows by application of the result in subproblem (b) that

$$|\Psi_n|_{Lip} \leq W\,\|A_2\|_\infty\,\|A_1\|_\infty \leq W(\mathcal{B}(\Psi_n))^2 \leq 4W. \tag{28}$$

This upper bound together with the lower bound (27) yields $4W \geq 2^n$, which results in $W \geq 2^{n-2}$.

(d) For all $x, y \in \mathbb{R}^d$ with $x \neq y$, we have

$$
\begin{aligned}
& \|\Phi(x) - \Phi(y)\|_\infty && (29)\\
&= \|A_n(\rho\,(A_{n-1}(\dots\rho\,(A_1 x)\dots))) - A_n(\rho\;(A_{n-1}(\dots\rho\,(A_1 y)\dots)))\|_\infty && (30)\\
&\le N_{n-1}\,\|A_n\|_\infty\,\|\rho\,(A_{n-1}(\dots\rho\,(A_1 x)\dots)) - \rho\,(A_{n-1}(\dots\rho\,(A_1 y)\dots))\|_\infty && (31)\\
&\le N_{n-1}\,\|A_n\|_\infty\,\|A_{n-1}(\dots\rho\,(A_1 x)\dots) - A_{n-1}(\dots\rho\,(A_1 y)\dots)\|_\infty && (32)\\
&\le \mathcal{W}(\Phi)\mathcal{B}(\Phi)\,\|A_{n-1}(\dots\rho\,(A_1 x)\dots) - A_{n-1}(\dots\rho\,(A_1 y)\dots)\|_\infty && (33)\\
&\;\;\vdots\\
&\le (\mathcal{W}(\Phi)\mathcal{B}(\Phi))^n\,\|x - y\|_\infty\,, && (34)
\end{aligned}
$$

where (31) follows from the Hint in subproblem (a), (32) is by the Hint in subproblem (b), and in (33) we used $N_{n-1} \le \mathcal{W}(\Phi)$ and $\|A_n\|_\infty \le \mathcal{B}(\Phi)$.

Therefore, we get

$$
\begin{aligned}
|\Phi|_{Lip} &= \sup_{\substack{x,y\in\mathbb{R}^d\\ x\neq y}} \frac{\|\Phi(x)-\Phi(y)\|_\infty}{\|x-y\|_\infty}\\
&\le \sup_{\substack{x,y\in\mathbb{R}^d\\ x\neq y}} \frac{(\mathcal{W}(\Phi)\mathcal{B}(\Phi))^n\,\|x-y\|_\infty}{\|x-y\|_\infty}\\
&= (\mathcal{W}(\Phi)\mathcal{B}(\Phi))^n
\end{aligned}
$$

as desired.