ETH
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 29, 2022**

Problem 1

(a) Let

$$\begin{aligned} A(x) &:= \begin{pmatrix} 1 \\ 1 \end{pmatrix} x + \begin{pmatrix} -1/4 \\ -3/4 \end{pmatrix}, && x \in \mathbb{R} \\ B(y) &:= \begin{pmatrix} 2 & -2 \end{pmatrix} y, && y \in \mathbb{R}^2. \end{aligned}$$

Then, $f_1 = \Phi_1 := B \circ \rho \circ A$ and $\mathcal{L}(\Phi_1) = 2$, $\mathcal{W}(\Phi_1) = 2$, $\mathcal{M}(\Phi_1) = 6$, and $\mathcal{B}(\Phi_1) = 2$.

(b) Let

$$\Phi_3 := B \circ \rho \circ A \circ B \circ \rho \circ A \circ B \circ \rho \circ A.$$

We have $\Phi_3 = \Phi_1 \circ \Phi_1 \circ \Phi_1 = f_1 \circ f_1 \circ f_1 = f_3$ and

$$(A \circ B)(y) = \begin{pmatrix} 2 & -2 \\ 2 & -2 \end{pmatrix} y + \begin{pmatrix} -1/4 \\ -3/4 \end{pmatrix}, \qquad y \in \mathbb{R}^2,$$

which satisfies $\mathcal{B}(\Phi_3) \leq 2$ as required. Moreover, $\mathcal{L}(\Phi_1) = 4$, $\mathcal{W}(\Phi_1) = 2$, and $\mathcal{M}(\Phi_1) = 18$.

(c) We will prove the claim by natural induction. It holds for $n = 1$ by definition of f_1. For the induction step $n \to n+1$ we first note that, owing to the induction assumption,

$$\begin{aligned} f_n(\tfrac{1}{2} - 2^{-(n+2)}) &= 2^n(\tfrac{1}{2} - 2^{-(n+2)} - (\tfrac{1}{2} - 2^{-(n+1)})) \\ &= 2^n(2^{-n-1} - 2^{-n-2}) \\ &= \tfrac{1}{4} \end{aligned} \tag{1}$$

and

$$\begin{aligned} f_n(\tfrac{1}{2} + 2^{-(n+2)}) &= 2^n(\tfrac{1}{2} + 2^{-(n+2)} - (\tfrac{1}{2} - 2^{-(n+1)})) \\ &= 2^n(2^{-n-1} + 2^{-n-2}) \\ &= \tfrac{3}{4}. \end{aligned} \tag{2}$$

Next, note that f_1 is a monotonously increasing function which implies that f_n is monotonously increasing by virtue of being a composition of monotonously increasing functions. Moreover, observe that f_1, and therefore f_n, maps from $[0, 1]$ into $[0, 1]$, which means that $f_n(x) \leq \frac{1}{4}$ ensures $f_n(x) \in [0, \frac{1}{4}]$ and, likewise, $f_n(x) \geq \frac{3}{4}$ ensures $f_n(x) \in [\frac{3}{4}, 1]$. Combined with (1) and (2) this implies

- for $x \in [0, \frac{1}{2} - 2^{-(n+2)}]$, that $f_n(x) \in [0, \frac{1}{4}]$ and thus $f_{n+1} = f_1(f_n(x)) = 0$,

- for $x \in [\frac{1}{2} - 2^{-(n+2)}, \frac{1}{2} + 2^{-(n+2)}]$, that $f_n(x) \in [\frac{1}{4}, \frac{3}{4}]$ and thus

$$\begin{aligned} f_{n+1} &= f_1(f_n(x)) \\ &= f_1(2^n(x - (\tfrac{1}{2} - 2^{-(n+1)}))) \\ &= 2(2^n(x - (\tfrac{1}{2} - 2^{-(n+1)})) - \tfrac{1}{4}) \\ &= 2^{n+1}((x - (\tfrac{1}{2} - 2^{-(n+1)})) - 2^{-(n+2)}) \\ &= 2^{n+1}(x - (\tfrac{1}{2} - 2^{-(n+1)} + 2^{-(n+2)})) \\ &= 2^{n+1}(x - (\tfrac{1}{2} - 2^{-(n+2)})), \end{aligned}$$

- for $x \in [\frac{1}{2} + 2^{-(n+2)}, 1]$, that $f_n(x) \in [\frac{3}{4}, 1]$ and thus $f_{n+1} = f_1(f_n(x)) = 1$.

(d) As $|H(x) - f_n(x)| \leq 1$, for all $x \in [0, 1]$, it follows that

$$\begin{aligned} \|H - f_n\|^2_{L^2([0,1])} &= \int_0^1 |H(x) - f_n(x)|^2 \mathrm{d}x \\ &\leq \int_{\frac{1}{2}-2^{-(n+1)}}^{\frac{1}{2}+2^{-(n+1)}} 1 \, \mathrm{d}x \\ &= 2^{-n}, \end{aligned}$$

and hence $\|H - f_n\|_{L^2([0,1])} \leq 2^{-\frac{n}{2}}$. We can now take Φ_1 realizing f_1 from subproblem (a), define $\Phi_n := \Phi_1 \circ \Phi_{n-1}$, for $n \geq 2$, $n \in \mathbb{N}$, and note that $\Phi_n = f_n$, for $n \in \mathbb{N}$. For $\varepsilon \in (0, \frac{1}{2})$, the ReLU neural network $\Psi_\varepsilon := \Phi_{\lceil 2 \log(\varepsilon^{-1}) \rceil}$ satisfies the desired properties.

(e) For $x \leq 0$ we have $\rho(x) + \rho(-x) = -x = |x|$, and for $x \geq 0$, it holds that $\rho(x) + \rho(-x) = x = |x|$.

(f) Note that $g(x) = \|x\|_1 = \sum_{j=1}^d |x_j|$. We thus have $g = \Gamma := C \circ \rho \circ D$ with

$$D = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ -1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 0 & \cdots & 0 \\ & & \vdots & & \\ 0 & \cdots & 0 & 1 & 0 \\ 0 & \cdots & 0 & -1 & 0 \\ 0 & \cdots & 0 & 0 & 1 \\ 0 & \cdots & 0 & 0 & -1 \end{pmatrix} \in \mathbb{R}^{2d \times d}$$

and

$$C := \begin{pmatrix} 1 & 1 & \cdots & 1 & 1 \end{pmatrix} \in \mathbb{R}^{1 \times 2d}.$$

We observe that $\mathcal{L}(\Gamma) = 2$, $\mathcal{W}(\Gamma) = 2d$, $\mathcal{M}(\Gamma) = 4d$, and $\mathcal{B}(\Gamma) = 1$.

Problem 2

(a) Let $n, d \in \mathbb{N}$. For $k, k' \in \{0, \dots, n-1\}^d$ with $k \neq k'$, the functions $\mathbb{I}_k$ and $\mathbb{I}_{k'}$ are disjointly supported and hence

$$\rho_\infty(\mathbb{I}_k, \mathbb{I}_{k'}) = \|\mathbb{I}_k - \mathbb{I}_{k'}\|_{L^\infty(\mathbb{R}^d)} = \max\{\|\mathbb{I}_k\|_{L^\infty(\mathbb{R}^d)}, \|\mathbb{I}_{k'}\|_{L^\infty(\mathbb{R}^d)}\} = 1. \tag{3}$$

(b) Let $k \in \{0, \dots, n-1\}^d$. Note that, owing to (3), the ε-ball around $\mathbb{I}_k$ contains every element in $\mathcal{C}_{n,d}$, if $\varepsilon \geq 1$. Consequently, the singleton set $\{\mathbb{I}_k\}$ constitutes an ε-covering of $\mathcal{C}_{n,d}$. In contrast, for $\varepsilon < 1$, the ε-ball around $\mathbb{I}_k$ contains $\mathbb{I}_k$ only and therefore we need to take all the elements of $\mathcal{C}_{n,d}$ in order to cover $\mathcal{C}_{n,d}$. In summary, noting that $|\mathcal{C}_{n,d}| = n^d$, we established

$$N(\varepsilon; \mathcal{C}_{n,d}, \rho_\infty) = \begin{cases} 1, & \varepsilon \geq 1 \\ n^d, & \varepsilon < 1 \end{cases}.$$

(c) Note that the set

$$B_{\varepsilon,n,d} := \{j\varepsilon\mathbb{I}_k \colon k \in \{0, \dots, n-1\}^d, j \in \{0, \dots, \lfloor \varepsilon^{-1} \rfloor\}\}$$

is an ε-covering of $\mathcal{C}^*_{n,d}$, as, for every $k \in \{0, \dots, n-1\}^d$ and $\alpha \in [0,1]$, we have $\lfloor \alpha\varepsilon^{-1} \rfloor \varepsilon \mathbb{I}_k \in B_{\varepsilon,n,d}$ and

$$\begin{aligned} \|\alpha\mathbb{I}_k - \lfloor \alpha\varepsilon^{-1} \rfloor \varepsilon \mathbb{I}_k\|_{L^\infty(\mathbb{R}^d)} &= |\alpha - \lfloor \alpha\varepsilon^{-1} \rfloor \varepsilon| \\ &= \alpha - \lfloor \alpha\varepsilon^{-1} \rfloor \varepsilon \\ &\geq \alpha - (\alpha\varepsilon^{-1} - 1)\varepsilon \\ &= \alpha - \alpha + \varepsilon \\ &= \varepsilon, \end{aligned}$$

owing to $\lfloor \alpha\varepsilon^{-1} \rfloor \varepsilon \leq \alpha$ and $\lfloor \alpha\varepsilon^{-1} \rfloor \geq \alpha\varepsilon^{-1} - 1$. Moreover, it holds that

$$|B_{\varepsilon,n,d}| = |\{0, \dots, n-1\}^d| \cdot |\{0, \dots, \lfloor \varepsilon^{-1} \rfloor\}| = n^d(\lfloor \varepsilon^{-1} \rfloor + 1) \leq 2n^d\varepsilon^{-1}.$$

We thus get

$$N(\varepsilon; \mathcal{C}^*_{n,d}, \rho_\infty) \leq 2n^d\varepsilon^{-1},$$

which establishes the claim with $b = 2$.

(d) Note that

$$A_{\varepsilon,n,d} := \{j(\lceil \varepsilon^{-1} \rceil - 1)^{-1}\mathbb{I}_k \colon k \in \{0, \dots, n-1\}^d, j \in \{1, \dots, \lceil \varepsilon^{-1} \rceil - 1\}\}$$

is a subset of $\mathcal{C}^*_{n,d}$ as $j(\lceil \varepsilon^{-1} \rceil - 1)^{-1} \in [0,1]$, for $j \in \{1, \dots, \lceil \varepsilon^{-1} \rceil - 1\}$. The set $A_{\varepsilon,n,d}$ is furthermore an ε-packing of $\mathcal{C}^*_{n,d}$ as, for $k, k' \in \{0, \dots, n-1\}^d$, with $k \neq k'$, and $j, j' \in \{1, \dots, \lceil \varepsilon^{-1} \rceil - 1\}$, we have

$$\begin{aligned} \|j(\lceil \varepsilon^{-1} \rceil - 1)^{-1}\mathbb{I}_k - j'(\lceil \varepsilon^{-1} \rceil - 1)^{-1}\mathbb{I}_{k'}\|_{L^\infty} &= \max\{j(\lceil \varepsilon^{-1} \rceil - 1)^{-1}, j'(\lceil \varepsilon^{-1} \rceil - 1)^{-1}\} \\ &\geq (\lceil \varepsilon^{-1} \rceil - 1)^{-1} \\ &> \varepsilon, \end{aligned}$$

and, for $k \in \{0, \ldots, n-1\}^d$ and $j, j' \in \{1, \ldots, \lceil \varepsilon^{-1} \rceil - 1\}$, with $j \neq j'$, we have

$$\begin{aligned} \|j(\lceil \varepsilon^{-1} \rceil - 1)^{-1}\mathbb{I}_k - j'(\lceil \varepsilon^{-1} \rceil - 1)^{-1}\mathbb{I}_k\|_{L^\infty} &= |(j-j')(\lceil \varepsilon^{-1} \rceil - 1)^{-1}| \\ &\geq (\lceil \varepsilon^{-1} \rceil - 1)^{-1} \\ &> \varepsilon. \end{aligned}$$

Moreover, it holds that

$$|A_{\varepsilon,n,d}| = |\{0, \ldots, n-1\}^d| \cdot |\{1, \ldots, \lceil \varepsilon^{-1} \rceil - 1\}| = n^d(\lceil \varepsilon^{-1} \rceil - 1) \geq \tfrac{1}{2} n^d \varepsilon^{-1},$$

where we used $\varepsilon \in (0, \frac{1}{2})$. We thus get

$$M(\varepsilon; \mathcal{C}^*_{n,d}, \rho_\infty) \geq \tfrac{1}{2} n^d \varepsilon^{-1},$$

which proves the claim with $a = \frac{1}{2}$.

Problem 3

(a) The dichotomy $\left\{X_1^+, X_1^-\right\}$ 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) Suppose, for the sake of contradiction, that $\{X_2^+, X_2^-\}$ is homogeneously linearly separable. Then, there would exist a nonzero vector $w = (u, v)$ so that

$$\langle x, (u,v)\rangle > 0, \text{ for all } x \in X_2^+,$$
$$\langle x, (u,v)\rangle < 0, \text{ for all } x \in X_2^-,$$

which corresponds to

$$\langle(-1,0),(u,v)\rangle = -u > 0, \tag{4}$$
$$\langle(1,0),(u,v)\rangle = u > 0, \tag{5}$$
$$\langle(0,1),(u,v)\rangle = v < 0, \tag{6}$$
$$\langle(0,-1),(u,v)\rangle = -v < 0. \tag{7}$$

Relations (4)-(5) can not hold simultaneously, which establishes the desired contradiction. Let $\phi(x_1,x_2) = x_1^2 - x_2^2$, $(x_1,x_2) \in \mathbb{R}^2$, and take $w = 1$. We then have

$$\langle\phi(-1,0),1\rangle = 1 > 0, \tag{8}$$
$$\langle\phi(1,0),1\rangle = 1 > 0, \tag{9}$$
$$\langle\phi(0,1),1\rangle = -1 < 0, \tag{10}$$
$$\langle\phi(0,-1),1\rangle = -1 < 0, \tag{11}$$

and therefore the dichotomy $\{X_2^+, X_2^-\}$ is ϕ-separable.

(c) Let $f_i \in \mathcal{F}$, $i = 1,2,3,4$, be given by

$$f_1(x) = \text{sgn}\left(\sin\left(-\frac{\pi}{2}\right)\right), \qquad x \in \mathbb{R}, \tag{12}$$
$$f_2(x) = \text{sgn}\left(\sin\left(\pi x - \frac{\pi}{2}\right)\right), \qquad x \in \mathbb{R}, \tag{13}$$
$$f_3(x) = \text{sgn}\left(\sin\left(-\pi x + \frac{\pi}{2}\right)\right), \qquad x \in \mathbb{R}, \tag{14}$$
$$f_4(x) = \text{sgn}\left(\sin\left(\frac{\pi}{2}\right)\right), \qquad x \in \mathbb{R}, \tag{15}$$

such that

$$f_1(0) = \text{sgn}\left(\sin\left(-\frac{\pi}{2}\right)\right) = 0,\ f_1(1) = \text{sgn}\left(\sin\left(-\frac{\pi}{2}\right)\right) = 0, \tag{16}$$

$$f_2(0) = \text{sgn}\left(\sin\left(-\frac{\pi}{2}\right)\right) = 0,\ f_2(1) = \text{sgn}\left(\sin\left(\frac{\pi}{2}\right)\right) = 1, \tag{17}$$

$$f_3(0) = \text{sgn}\left(\sin\left(\frac{\pi}{2}\right)\right) = 1,\ f_3(1) = \text{sgn}\left(\sin\left(-\frac{\pi}{2}\right)\right) = 0, \tag{18}$$

$$f_4(0) = \text{sgn}\left(\sin\left(\frac{\pi}{2}\right)\right) = 1,\ f_4(1) = \text{sgn}\left(\sin\left(\frac{\pi}{2}\right)\right) = 1. \tag{19}$$

In summary, for $(x_1, x_2) = (0, 1)$ and every $(y_1, y_2) \in \{0, 1\}^2$, there exists an $f \in \mathcal{F}$ such that $f(x_i) = y_i$, $i = 1, 2$. By Lemma 1 in the Handout, $\mathcal{F}$ hence shatters $\{0, 1\}$.

(d) We shall first show that $\Pi_{\mathcal{G}}(1) = 1$ implies $|\mathcal{G}| = 1$. Suppose that $\Pi_{\mathcal{G}}(1) = 1$ and assume, for the sake of contradiction, that $|\mathcal{G}| \geq 2$. Then, there would exist $f_1, f_2 \in \mathcal{G}$ and $a \in \mathbb{R}$ such that

$$f_1(a) \neq f_2(a), \tag{20}$$

which, in turn, implies

$$\begin{aligned}
\Pi_{\mathcal{G}}(1) &= \max\{|\mathcal{G}_{|X}| : X \subseteq \mathbb{R}, |X| = 1\} && (21)\\
&\geq |\mathcal{G}_{|\{a\}}| && (22)\\
&\geq |\{f_1|_{\{a\}}, f_2|_{\{a\}}\}| && (23)\\
&= 2, && (24)
\end{aligned}$$

where (21) follows from the definition of the growth function (See Definition 5 in the Handout), and in (23) we used that $f_1|_{\{a\}}, f_2|_{\{a\}} \in \mathcal{G}_{|\{a\}}$ are distinct according to (20). The resulting inequality $\Pi_{\mathcal{G}}(1) \geq 2$ contradicts the assumption $\Pi_{\mathcal{G}}(1) = 1$, and therefore we must have $|\mathcal{G}| = 1$. Since $|\mathcal{G}| = 1$, it follows that $|\mathcal{G}_{|X}| = 1$, for all $X \subseteq \mathbb{R}$, and therefore $\Pi_{\mathcal{G}}(N) = \max\{|\mathcal{G}_{|X}| : X \subseteq \mathbb{R}, |X| = N\} = 1$, for all $N \in \mathbb{N}$.

(e) The proof is effected by establishing that $\text{VC}(\mathcal{F}) \geq n$, for all $n \in \mathbb{N}$. Fix $n \in \mathbb{N}$ and set $X = (x_i)_{i=1}^n = (4^i)_{i=1}^n$. For $y = (y_i)_{i=1}^n \in \{0, 1\}^n$, let $f_y \in \mathcal{F}$ be given by

$$f_y(x) = \text{sgn}\left(\sin\left(\left(\sum_{j=1}^n 4^{-j} y_j\right)\pi x - \frac{\pi}{2}\right)\right), \quad x \in \mathbb{R}.$$

We shall show that $f_y(x_i) = \text{sgn}(\sin((\sum_{j=1}^n 4^{-j} y_j)\pi x_i - \frac{\pi}{2})) = y_i$, for $i = 1, \ldots, n$. To this end, we fix $i \in \{1, \ldots, n\}$ and note that

$$\left(\sum_{j=1}^n 4^{-j} y_j\right)\pi x_i - \frac{\pi}{2} \tag{25}$$

$$= \left(\sum_{j=1}^n 4^{-j} y_j\right)\pi 4^i - \frac{\pi}{2} \tag{26}$$

$$= \left(\sum_{\substack{j\in\{1,\ldots,n\}\\ j<i}} 4^{-j+i} y_j\right)\pi + \left(y_i - \frac{1}{2}\right)\pi + \left(\sum_{\substack{j\in\{1,\ldots,n\}\\ j>i}} 4^{-j+i} y_j\right)\pi. \tag{27}$$

Since $\sum_{\substack{j\in\{1,\ldots,n\}\\ j<i}} 4^{-j+i}y_j$ is an integer multiple of 2, we can write

$$\left(\sum_{\substack{j\in\{1,\ldots,n\}\\ j<i}} 4^{-j+i}y_j\right)\pi = 2k\pi, \tag{28}$$

for some $k \in \mathbb{N}$ (depending on y and i). Moreover, we have

$$\left|\left(\sum_{\substack{j\in\{1,\ldots,n\}\\ j>i}} 4^{-j+i}y_j\right)\pi\right| \leq \sum_{k=1}^{\infty} 4^{-k}\pi \leq \frac{\pi}{3}, \tag{29}$$

where we used $|y_j| \leq 1$, for $j = 1, \ldots, n$, as $y_j \in \{0, 1\}$. Substituting (28) and (29) into (25)-(27) yields

$$2k\pi + \left(y_i - \frac{1}{2}\right)\pi - \frac{\pi}{3} \leq \left(\sum_{j=1}^{n} 4^{-j}y_j\right)\pi x_i - \frac{\pi}{2} \leq 2k\pi + \left(y_i - \frac{1}{2}\right)\pi + \frac{\pi}{3},$$

which, in turn, implies

$$\left(\sum_{j=1}^{n} 4^{-j}y_j\right)\pi x_i - \frac{\pi}{2} \in \begin{cases} \left[2k\pi + \frac{\pi}{2} - \frac{\pi}{3}, 2k\pi + \frac{\pi}{2} + \frac{\pi}{3}\right], & \text{if } y_i = 1, \\ \left[2k\pi - \frac{\pi}{2} - \frac{\pi}{3}, 2k\pi - \frac{\pi}{2} + \frac{\pi}{3}\right], & \text{if } y_i = 0, \end{cases}$$

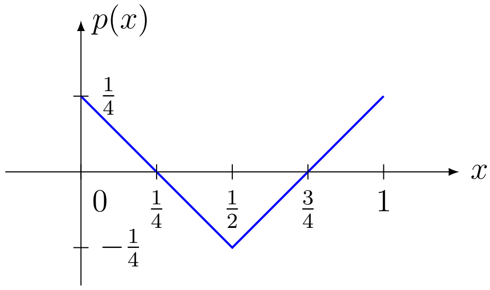
and hence

$$f_y(x_i) = \text{sgn}\left(\sin\left(\left(\sum_{j=1}^{n} 4^{-j}y_j\right)\pi x_i - \frac{\pi}{2}\right)\right) = y_i.$$

In summary, for every $y = (y_i)_{i=1}^{n} \in \{0, 1\}^n$, there exists a function $f_y \in \mathcal{F}$ such that $f_y(x_i) = y_i$, for $i = 1, \ldots, n$. Hence, by Lemma 1 in the Handout, $\mathcal{F}$ shatters $\{x_i\}_{i=1}^{n}$, which, in turn, implies $\text{VC}(\mathcal{F}) \geq n$. As $\text{VC}(\mathcal{F}) \geq n$ holds for all $n \in \mathbb{N}$, we must have $\text{VC}(\mathcal{F}) = \infty$.

Problem 4

(a) The plot of the function p is given below.



The function p is continuous on $[0, \frac{1}{2})$ and $(\frac{1}{2}, 1]$, by the definition of p, continuous at $\frac{1}{2}$ as

$$\lim_{\varepsilon \to 0^+} p\left(\frac{1}{2} + \varepsilon\right) = \lim_{\varepsilon \to 0^-} p\left(\frac{1}{2} + \varepsilon\right) = p\left(\frac{1}{2}\right) = -\frac{1}{4},$$

and therefore continuous on its entire domain $[0, 1]$. Moreover, p is differentiable on $[0, 1]\backslash\{1/2\}$ with $|p'(x)| = 1$, for $x \in [0, 1]\backslash\{1/2\}$. For $x, y \in [0, 1]$ with $x \le y$, we have

$$|p(x) - p(y)| = \left| \int_{[x,y]\backslash\{1/2\}} p'(z)\, dz \right| \le \int_{[x,y]\backslash\{1/2\}} |p'(z)|\, dz = y - x = |x - y|,$$

which establishes that $p \in H^1([0, 1])$.

(b) For $n \in \mathbb{N}$ and $y = (y_0, \ldots, y_n) \in \{0, 1\}^{n+1}$, let $h_y : [0, 1] \mapsto \mathbb{R}$ be given by

$$h_y(x) = \begin{cases} \dfrac{2y_i - 1}{2n}, & \text{for } x = \dfrac{i}{n},\ i = 0, \ldots, n, \\ \dfrac{2y_i - 1}{2n} + (y_{i+1} - y_i)\left(x - \dfrac{i}{n}\right), & \text{for } x \in \left(\dfrac{i}{n}, \dfrac{i+1}{n}\right),\ i = 0, \ldots, n-1. \end{cases}$$

Hence, the first requirement, namely,

$$h_y\left(\frac{i}{n}\right) = \frac{2y_i - 1}{2n}, \text{ for } i = 0, \ldots, n, \tag{30}$$

is trivially met. Applying the function sgn to both sides of (30) yields

$$\operatorname{sgn}\left(h_y\left(\frac{i}{n}\right)\right) = y_i, \text{ for } i = 0, \ldots, n, \tag{31}$$

as desired. It remains to show that $h_y \in H^1([0, 1])$. For $i = 0, \ldots, n-1$, we have

$$h_y(x) = \frac{2y_i - 1}{2n} + (y_{i+1} - y_i)\left(x - \frac{i}{n}\right), \text{ for } x \in \left[\frac{i}{n}, \frac{i+1}{n}\right],$$

by the definition of h_y, which implies that h_y is continuous on $[\frac{i}{n}, \frac{i+1}{n}]$. Therefore, h_y is continuous on $\cup_{i=0}^{n-1}[\frac{i}{n}, \frac{i+1}{n}] = [0, 1]$, which is the entire domain of h_y. Moreover, the function h_y is differentiable on $[0, 1]\backslash\{\frac{i}{n}\}_{i=0}^{n}$ with $|h_y'(x)| \le$

$\sup_{i=0,\dots,n-1} |y_{i+1} - y_i| \leq 1$, for $x \in [0,1]\backslash\{\frac{i}{n}\}_{i=0}^n$. Then, for $x, y \in [0,1]$ with $x \leq y$, we have

$$|h_y(x) - h_y(y)| = \left| \int_{[x,y]\backslash\{\frac{i}{n}\}_{i=0}^n} h_y'(z)\,dz \right| \leq \int_{[x,y]\backslash\{\frac{i}{n}\}_{i=0}^n} |h_y'(z)|\,dz \leq y - x = |x - y|,$$

which, combined with the continuity of h_y, implies $h_y \in H^1([0,1])$.

(c) For $n \in \mathbb{N}$, $y = (y_0, \dots, y_n) \in \{0,1\}^{n+1}$, and $g : [0,1] \mapsto \mathbb{R}$, suppose that

$$\sup_{x\in[0,1]} |h_y(x) - g(x)| \leq \frac{1}{4n}.$$

Then, for $i = 0, \dots, n$, we have

$$g\left(\frac{i}{n}\right) \geq h_y\left(\frac{i}{n}\right) - \left|h_y\left(\frac{i}{n}\right) - g\left(\frac{i}{n}\right)\right| \geq \frac{1}{2n} - \frac{1}{4n} \quad > 0, \text{ if } y_i = 1, \tag{32}$$

$$g\left(\frac{i}{n}\right) \leq h_y\left(\frac{i}{n}\right) + \left|h_y\left(\frac{i}{n}\right) - g\left(\frac{i}{n}\right)\right| \leq -\frac{1}{2n} + \frac{1}{4n} < 0, \text{ if } y_i = 0, \tag{33}$$

where we used $h_y(\frac{i}{n}) = \frac{2y_i - 1}{2n}$ and the assumption $\sup_{x\in[0,1]} |h_y(x) - g(x)| \leq \frac{1}{4n}$. From (32) and (33), we can now conclude that

$$\text{sgn}\Big(g\Big(\frac{i}{n}\Big)\Big) = y_i.$$

(d) Suppose, for the sake of contradiction that, there is no $h \in H^1([0,1])$ such that

$$\sup_{x\in[0,1]} |h(x) - \Phi(x)| > \frac{1}{4CW^2L^2(\log(W) + \log(L))}, \text{ for all } \Phi \in \mathcal{N}(W, L). \tag{34}$$

This would then imply that, for every $h \in H^1([0,1])$, there exists a $\Phi \in \mathcal{N}(W, L)$ (depending on h) so that

$$\sup_{x\in[0,1]} |h(x) - \Phi(x)| \leq \frac{1}{4CW^2L^2(\log(W) + \log(L))}. \tag{35}$$

Let $n = \text{VC}(\text{sgn}(\mathcal{N}(W, L)))$ and fix $y = (y_0, \dots, y_n) \in \{0,1\}^{n+1}$. Then, by subproblem (b), there exists an $h_y \in H^1([0,1])$ satisfying

$$h_y\Big(\frac{i}{n}\Big) = \frac{2y_i - 1}{2n}, \text{ for } i = 0, \dots, n, \tag{36}$$

and

$$\text{sgn}\Big(h_y\Big(\frac{i}{n}\Big)\Big) = y_i, \text{ for } i = 0, \dots, n. \tag{37}$$

By the contradictory assumption, there exists a $\Phi_y \in \mathcal{N}(W, L)$, depending on h_y, such that

$$\sup_{x\in[0,1]} |h_y(x) - \Phi_y(x)| \leq \frac{1}{4CW^2L^2(\log(W) + \log(L))}, \tag{38}$$

which, combined with the VC dimension upper bound $\text{VC}(\text{sgn}(\mathcal{N}(W, L))) \leq CW^2L^2(\log(W) + \log(L))$ and $n = \text{VC}(\text{sgn}(\mathcal{N}(W, L)))$, yields

$$\sup_{x\in[0,1]} |h_y(x) - \Phi_y(x)| \leq \frac{1}{4n}. \tag{39}$$

Application of the result in subproblem (c), with the requisite condition satisfied thanks to (39), yields

$$\operatorname{sgn}\left(\Phi_y\left(\frac{i}{N}\right)\right) = y_i, \text{ for } i = 0, \dots, n.$$

Since the choice of y was arbitrary, we, indeed, have shown that, for $X = (x_i)_{i=0}^n = \left(\frac{i}{n}\right)_{i=0}^n$ and every $y = (y_i)_{i=0}^n \in \{0,1\}^{n+1}$, there exists a function $\operatorname{sgn} \circ \Phi_y \in \operatorname{sgn}(\mathcal{N}(W, L))$, depending on y, so that

$$(\operatorname{sgn} \circ \Phi_y)(x_i) = y_i, \text{ for } i = 0, \dots, n.$$

Finally, application of Lemma 1 in the Handout yields that $\operatorname{sgn}(\mathcal{N}(W, L))$ shatters $X = (x_i)_{i=0}^n$ and hence

$$\mathrm{VC}(\operatorname{sgn}(\mathcal{N}(W, L))) = n + 1 > n = \mathrm{VC}(\operatorname{sgn}(\mathcal{N}(W, L))),$$

which establishes the desired contradiction.