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 17, 2020**

Problem 1

(a) By definition we have

$$g_{\boldsymbol{B}_{n+1}}(\boldsymbol{x}) = \int_{-\frac{1}{2}}^{\frac{1}{2}} g_{\boldsymbol{B}_n}(\boldsymbol{x} - t\boldsymbol{b}_{n+1})\,\mathrm{d}t.$$

This yields

$$\begin{aligned}
\hat{g}_{\boldsymbol{B}_{n+1}}(\xi) &= \int_{\mathbb{R}^n} g_{\boldsymbol{B}_{n+1}}(\boldsymbol{x}) e^{-i\xi^{\mathrm{T}}\boldsymbol{x}}\,\mathrm{d}\boldsymbol{x} \\
&= \int_{-\frac{1}{2}}^{\frac{1}{2}} \left(\int_{\mathbb{R}^n} g_{\boldsymbol{B}_n}(\boldsymbol{x} - t\boldsymbol{b}_{n+1}) e^{-i\xi^{\mathrm{T}}\boldsymbol{x}}\,\mathrm{d}\boldsymbol{x} \right) \mathrm{d}t \\
&= \int_{-\frac{1}{2}}^{\frac{1}{2}} \left(\int_{\mathbb{R}^n} |\det(\boldsymbol{B}_n)| g_{\boldsymbol{B}_n}(\boldsymbol{B}_n\boldsymbol{z}) e^{-i\xi^{\mathrm{T}}(\boldsymbol{B}_n\boldsymbol{z} + t\boldsymbol{b}_{n+1})}\,\mathrm{d}\boldsymbol{z} \right) \mathrm{d}t \qquad (1) \\
&= \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{-i\xi^{\mathrm{T}}\boldsymbol{b}_{n+1}t}\,\mathrm{d}t \prod_{i=1}^{n} \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{-i\xi^{\mathrm{T}}\boldsymbol{b}_i z_i}\mathrm{d}z_i, \qquad (2)
\end{aligned}$$

where in (1) we changed variables to $\boldsymbol{z} = \boldsymbol{B}_n^{-1}(\boldsymbol{x} - t\boldsymbol{b}_{n+1})$ and (2) follows from the fact that

$$|\det(\boldsymbol{B}_n)| g_{\boldsymbol{B}_n}(\boldsymbol{B}_n\boldsymbol{z}) = \begin{cases} 1, & \text{if } \boldsymbol{z} \in \left[-\frac{1}{2},\, \frac{1}{2}\right]^n \\ 0, & \text{else.} \end{cases}$$

As

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} e^{-iat}\,\mathrm{d}t = \frac{2}{a}\sin\left(\frac{a}{2}\right), \quad \text{for all } a > 0, \qquad (3)$$

we can conclude that

$$\hat{g}_{\boldsymbol{B}_{n+1}}(\xi) = \prod_{i=1}^{n+1} \left(\frac{2}{\xi^{\mathrm{T}}\boldsymbol{b}_i} \sin\left(\frac{\xi^{\mathrm{T}}\boldsymbol{b}_i}{2}\right) \right).$$

(b) The result for $q = n+1$ follows from subproblem (a). For general $q > n$, we proceed by induction as follows. Suppose that the statement holds for $q-1$, i.e.,

$$\hat{g}_{\boldsymbol{B}_{q-1}}(\xi) = \prod_{i=1}^{q-1} \left(\frac{2}{\xi^{\mathrm{T}}\boldsymbol{b}_i} \sin\left(\frac{\xi^{\mathrm{T}}\boldsymbol{b}_i}{2}\right) \right). \qquad (4)$$

By definition we have

$$g_{\boldsymbol{B}_q}(\boldsymbol{x}) = \int_{-\frac{1}{2}}^{\frac{1}{2}} g_{\boldsymbol{B}_{q-1}}(\boldsymbol{x} - t\boldsymbol{b}_q)\,\mathrm{d}t.$$

This yields

$$\begin{aligned}
\hat{g}_{\boldsymbol{B}_q}(\xi) &= \int_{\mathbb{R}^n} g_{\boldsymbol{B}_q}(\boldsymbol{x}) e^{-i\xi^{\mathrm{T}}\boldsymbol{x}}\,\mathrm{d}\boldsymbol{x} \\
&= \int_{-\frac{1}{2}}^{\frac{1}{2}} \left(\int_{\mathbb{R}^n} g_{\boldsymbol{B}_{q-1}}(\boldsymbol{x} - t\boldsymbol{b}_q) e^{-i\xi^{\mathrm{T}}\boldsymbol{x}}\,\mathrm{d}\boldsymbol{x} \right) \mathrm{d}t \\
&= \int_{-\frac{1}{2}}^{\frac{1}{2}} \left(\int_{\mathbb{R}^n} g_{\boldsymbol{B}_{q-1}}(\boldsymbol{z}) e^{-i\xi^{\mathrm{T}}(\boldsymbol{z}+t\boldsymbol{b}_q)}\,\mathrm{d}\boldsymbol{z} \right) \mathrm{d}t \qquad (5) \\
&= \hat{g}_{\boldsymbol{B}_{q-1}}(\xi) \int_{-\frac{1}{2}}^{\frac{1}{2}} e^{-i\xi^{\mathrm{T}}\boldsymbol{b}_q t}\,\mathrm{d}t \\
&= \prod_{i=1}^{q} \left(\frac{2}{\xi^{\mathrm{T}}\boldsymbol{b}_i} \sin\left(\frac{\xi^{\mathrm{T}}\boldsymbol{b}_i}{2} \right) \right), \qquad (6)
\end{aligned}$$

where in (5) we changed variables to $\boldsymbol{z} = \boldsymbol{x} - t\boldsymbol{b}_q$ and (6) follows from (3) and (4).

(c) We first determine the Fourier transform $\widehat{f \star g}(\xi)$ of the convolution $(f \star g)(\boldsymbol{x})$ for two general functions $f, g \colon \mathbb{R}^n \to \mathbb{R}$. Specifically, we have

$$\begin{aligned}
\widehat{f \star g}(\xi) &= \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} f(\boldsymbol{y}) g(\boldsymbol{x} - \boldsymbol{y})\mathrm{d}\boldsymbol{y} \right) e^{-i\xi^{\mathsf{T}}\boldsymbol{x}}\mathrm{d}\boldsymbol{x} \\
&= \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} g(\boldsymbol{x} - \boldsymbol{y}) e^{-i\xi^{\mathsf{T}}\boldsymbol{x}}\mathrm{d}\boldsymbol{x} \right) f(\boldsymbol{y})\mathrm{d}\boldsymbol{y} \\
&= \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^n} g(\boldsymbol{u}) e^{-i\xi^{\mathsf{T}}\boldsymbol{u}}\mathrm{d}\boldsymbol{u} \right) f(\boldsymbol{y}) e^{-i\xi^{\mathsf{T}}\boldsymbol{y}}\mathrm{d}\boldsymbol{y} \qquad (7) \\
&= \hat{f}(\xi)\hat{g}(\xi),
\end{aligned}$$

where in (7) we changed variables to $\boldsymbol{u} = \boldsymbol{x} - \boldsymbol{y}$. Particularizing this result to $f = g = g_{\boldsymbol{B}_q}$ yields

$$\begin{aligned}
\hat{k}(\xi) &= \widehat{g_{\boldsymbol{B}_q} \star g_{\boldsymbol{B}_q}}(\xi) \\
&= (\hat{g}_{\boldsymbol{B}_q})^2(\xi) \\
&= \prod_{i=1}^{q} \left(\frac{2}{\xi^{\mathrm{T}}\boldsymbol{b}_i} \sin\left(\frac{\xi^{\mathrm{T}}\boldsymbol{b}_i}{2} \right) \right)^2, \qquad (8)
\end{aligned}$$

where in (8) we used the explicit expression for $\hat{g}_{\boldsymbol{B}_q}(\xi)$ obtained in subproblem (b).

(d) We have to show that for every $k \in \mathbb{N}$ and all $\boldsymbol{x}_1, \dots, \boldsymbol{x}_k \in \mathbb{R}^n$, the $k \times k$ Gramian matrix

$$\boldsymbol{K}(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k) = \begin{pmatrix} K(\boldsymbol{x}_1, \boldsymbol{x}_1) & K(\boldsymbol{x}_1, \boldsymbol{x}_2) & \dots & K(\boldsymbol{x}_1, \boldsymbol{x}_k) \\ K(\boldsymbol{x}_2, \boldsymbol{x}_1) & K(\boldsymbol{x}_2, \boldsymbol{x}_2) & \dots & K(\boldsymbol{x}_2, \boldsymbol{x}_k) \\ \vdots & \vdots & \dots & \vdots \\ K(\boldsymbol{x}_k, \boldsymbol{x}_1) & K(\boldsymbol{x}_k, \boldsymbol{x}_2) & \dots & K(\boldsymbol{x}_k, \boldsymbol{x}_k) \end{pmatrix}$$

is positive semidefinite. This is effected by noting that, for every $k \in \mathbb{N}$, all $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k \in \mathbb{R}^n$, and all $\boldsymbol{c} = (c_1 \ldots c_n)^\mathsf{T} \in \mathbb{R}^n$, we have

$$\begin{aligned}
\boldsymbol{c}^\mathsf{T} \boldsymbol{K}(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k)\boldsymbol{c} &= \sum_{i,j=1}^{k} c_i c_j K(\boldsymbol{x}_i, \boldsymbol{x}_j) \\
&= \sum_{i,j=1}^{k} c_i c_j k(\boldsymbol{x}_i - \boldsymbol{x}_j) \\
&= \frac{1}{(2\pi)^n} \sum_{i,j=1}^{k} c_i c_j \int \hat{k}(\xi) e^{i\xi^\mathrm{T}(\boldsymbol{x}_i - \boldsymbol{x}_j)} \mathrm{d}\xi \\
&= \frac{1}{(2\pi)^n} \int \hat{k}(\xi) \left| \sum_{i=1}^{k} c_i e^{i\xi^\mathrm{T}\boldsymbol{x}_i} \right|^2 \mathrm{d}\xi \\
&\geq 0,
\end{aligned}$$

which, owing to $\hat{k}(\xi) \geq 0$, for all $\xi \in \mathbb{R}^n$, proves that K is positive semidefinite.

Problem 2

(a) The Lagrange function $L(\boldsymbol{w}, b, \boldsymbol{c})$ is given by

$$L(\boldsymbol{w}, b, \boldsymbol{c}) = \frac{1}{2}\|\boldsymbol{w}\|_2^2 + \sum_{i=1}^{3} c_i(1 - y_i(\langle \boldsymbol{w}, \boldsymbol{x}_i \rangle_2 - b)). \tag{9}$$

(b) Setting $\nabla_{\boldsymbol{w}} L(\boldsymbol{w}, b, \boldsymbol{c}) = 0$ and $\nabla_b L(\boldsymbol{w}, b, \boldsymbol{c}) = 0$ yields

$$\boldsymbol{w} = \sum_{i=1}^{3} c_i y_i \boldsymbol{x}_i \tag{10}$$

and

$$\sum_{i=1}^{3} c_i y_i = 0, \tag{11}$$

respectively. Using (10) and (11) in (9) results in

$$\begin{aligned} g(\boldsymbol{c}) &= \min_{\boldsymbol{w} \in \mathbb{R}^2, b \in \mathbb{R}} L(\boldsymbol{w}, b, \boldsymbol{c}) \\ &= \sum_{i=1}^{3} c_i - \frac{1}{2} \sum_{i,j=1}^{3} c_i y_i c_j y_j \langle \boldsymbol{x}_i, \boldsymbol{x}_j \rangle_2 \\ &= \boldsymbol{a}^\mathsf{T} \boldsymbol{c} - \frac{1}{2} \boldsymbol{c}^\mathsf{T} \boldsymbol{A}\, \boldsymbol{c} \end{aligned}$$

with $\boldsymbol{a} = (1 \;\; 1 \;\; 1)^\mathsf{T}$ and

$$\boldsymbol{A} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & -\lambda \\ 0 & -\lambda & 1+\lambda^2 \end{pmatrix}.$$

(c) It follows from subproblem (b) that the Lagrange dual function $g(\boldsymbol{c})$ can be written as

$$g(\boldsymbol{c}) = c_1 + c_2 + c_3 - \frac{1}{2}(c_2^2 + (1+\lambda^2)c_3^2 - 2\lambda c_2 c_3).$$

Since $-g(\boldsymbol{c})$ is convex, we can, as mentioned in the hint, consider the Lagrange function

$$\tilde{L}(\boldsymbol{c}, \boldsymbol{\mu}, \gamma) = -g(\boldsymbol{c}) - \boldsymbol{\mu}^\mathsf{T} \boldsymbol{c} + \gamma(c_1 + c_2 - c_3),$$

where $\boldsymbol{\mu} = (\mu_1, \mu_2, \mu_3)^\mathsf{T} \in \mathbb{R}^3$ and $\gamma \in \mathbb{R}$ are Lagrange multipliers. The KKT conditions now read as follows:

$$c_i \geq 0, \quad \text{for } i = 1,2,3 \tag{12}$$

$$c_1 + c_2 - c_3 = 0, \tag{13}$$

$$\mu_i \geq 0, \quad \text{for } i = 1,2,3 \tag{14}$$

$$c_i \mu_i = 0, \quad \text{for } i = 1,2,3 \tag{15}$$

$$\frac{\partial \tilde{L}(\boldsymbol{c}, \boldsymbol{\mu}, \gamma)}{\partial c_i} = 0, \quad \text{for } i = 1,2,3. \tag{16}$$

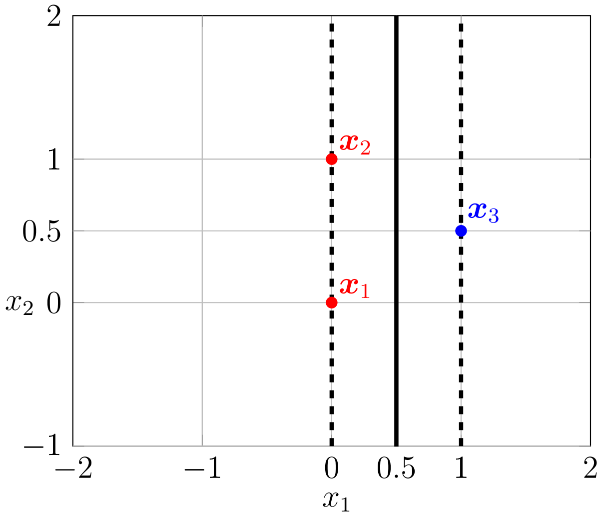
Evaluating the derivatives in (16) yields

$$-1 - \mu_1 + \gamma = 0 \tag{17}$$

$$-1 + c_2 - \lambda c_3 - \mu_2 + \gamma = 0 \tag{18}$$

$$-1 + (1 + \lambda^2)c_3 - \lambda c_2 - \mu_3 - \gamma = 0. \tag{19}$$

Suppose first that $\lambda \in (0, 1)$. The following figure depicts the separating straight line between $\{\boldsymbol{x}_1, \boldsymbol{x}_2\}$ and $\{\boldsymbol{x}_3\}$ of largest possible margins for $\lambda = 1/2$:

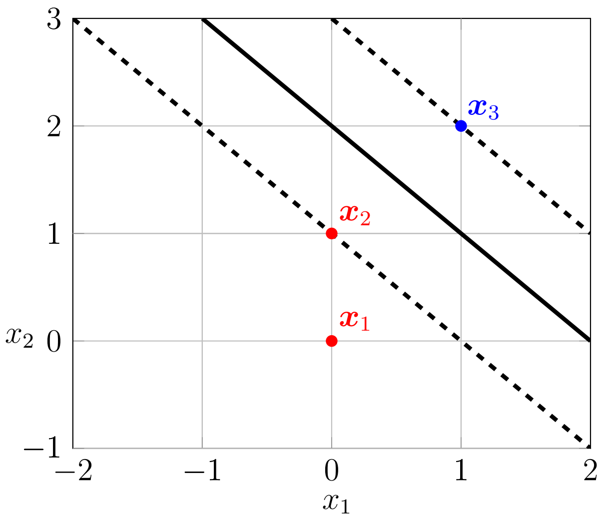


From the figure we can see that $\boldsymbol{x}_1, \boldsymbol{x}_2$, and $\boldsymbol{x}_3$ are all support vectors. We therefore make the ansatz $c_i > 0$ for $i = 1, 2, 3$. Then, (15) implies that $\mu_1 = \mu_2 = \mu_3 = 0$. Using this in (17)–(19) yields $c_2 = 2\lambda$ and $c_3 = 2$. Finally, (13) implies that $c_1 = 2 - 2\lambda$. We can therefore conclude that

$$\boldsymbol{c}^* = (2 - 2\lambda \quad 2\lambda \quad 2)^\mathsf{T}$$

is a solution of the Lagrange dual problem for $\lambda \in (0, 1)$.

Next, suppose that $\lambda > 1$. The following figure depicts the separating straight line between $\{\boldsymbol{x}_1, \boldsymbol{x}_2\}$ and $\{\boldsymbol{x}_3\}$ of largest possible margins for $\lambda = 2$:



From the figure we can see that now $\boldsymbol{x}_2$ and $\boldsymbol{x}_3$ are the support vectors. This leads to the ansatz $c_2 > 0$, $c_3 > 0$, and $c_1 = 0$. Then, (15) implies that $\mu_2 = \mu_3 = 0$ and (13) yields $c_2 = c_3$. Using this in (17)–(19), we obtain

$$\begin{aligned} c_2(\lambda - 1) &= \mu_1 \\ c_2(1 + \lambda^2 - \lambda) &= \mu_1 + 2, \end{aligned}$$

which yields

$$c_2 = \frac{2}{2 - 2\lambda + \lambda^2}.$$

We can therefore conclude that

$$\boldsymbol{c}^* = \frac{2}{2 - 2\lambda + \lambda^2} \begin{pmatrix} 0 & 1 & 1 \end{pmatrix}^\mathsf{T}$$

is a solution of the Lagrange dual problem for $\lambda \geq 1$.

(d) By assumption we have $\lambda \in (0, 1)$. Using (10) and $\boldsymbol{c}^* = (2 - 2\lambda \quad 2\lambda \quad 2)^\mathsf{T}$ from subproblem (c), we obtain the following solution $\boldsymbol{w}^*$ of the optimization problem in subproblem (a):

$$\boldsymbol{w}^* = 2\lambda \boldsymbol{x}_2 - 2\boldsymbol{x}_3 = -\begin{pmatrix} 2 \\ 0 \end{pmatrix}.$$

The solution b^* can be obtained as follows. Since $c_1^* > 0$, i.e., $\boldsymbol{x}_1$ is a support vector, the corresponding inequality constraint $y_1(\langle \tilde{\boldsymbol{w}}, \boldsymbol{x}_1 \rangle_2 - b^*) \geq 1$ must be satisfied with equality. This yields

$$\begin{aligned} b^* &= \langle \tilde{\boldsymbol{w}}, \boldsymbol{x}_1 \rangle_2 - y_1 \\ &= -1. \end{aligned}$$

The hard margin binary classifier $g_{\mathrm{hm}}(\boldsymbol{x})$ is therefore given by

$$\begin{aligned} g_{\mathrm{hm}}(\boldsymbol{x}) &= (\langle \tilde{\boldsymbol{w}}, \boldsymbol{x} \rangle_2 - b^*) \\ &= -2 \begin{pmatrix} 1 & 0 \end{pmatrix} \boldsymbol{x} + 1. \end{aligned}$$

Problem 3

(a) Denote the ReLU function as $\rho(x) = \max\{0, x\}$. The function $f(x)$ can be realized according to

$$f(x) = \rho(4x) - \rho(8x - 1.6) + \rho(14x - 5.6) - \rho(20x - 10) + \rho(14x - 8.4) \\ - \rho(8x - 6.4) + \rho(4x - 4).$$

This function can, in turn, be realized through a depth-2 ReLU network according to

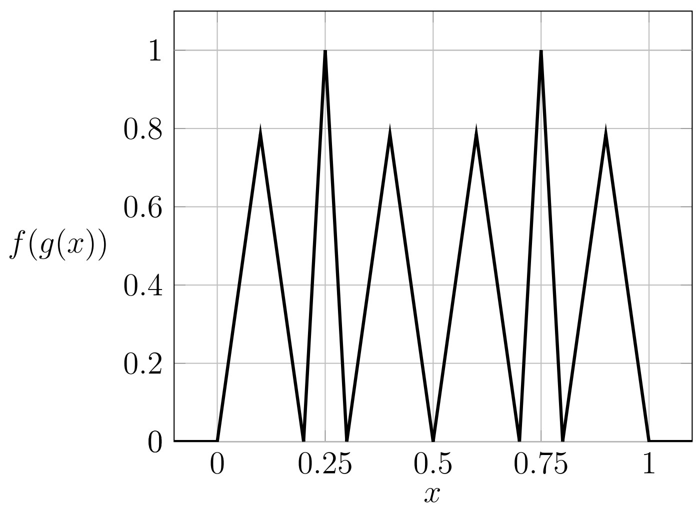
$$\Phi(x) = W_2(\rho(W_1(x)))$$

with

$$W_1(x) = \begin{pmatrix} 4 \\ 8 \\ 14 \\ 20 \\ 14 \\ 8 \\ 4 \end{pmatrix} x + \begin{pmatrix} 0 \\ -1.6 \\ -5.6 \\ -10 \\ -8.4 \\ -6.4 \\ -4 \end{pmatrix}, \quad W_2(x) = \begin{pmatrix} 1 & -1 & 1 & -1 & 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \end{pmatrix}.$$

The network Φ has depth 2, width 7, and connectivity 20.

(b) The following figure depicts $f(g(x))$. The axis of symmetry is at $x = 0.5$.



(c) An alternative way to realize $f(x)$ is obtained by first noting that $f(x)$ is symmetric around $x = 0.5$ and can thus be written as $f(x) = h(g(x))$, where $h(x) = f(0.5x)$, for $x \in [0, 1]$, and $g(x) = \rho(2x) - \rho(4x - 2) + \rho(2x - 2) = W_2^g(\rho(W_1^g(x)))$ is the sawtooth function from subproblem (b). Here,

$$W_1^g(x) = \begin{pmatrix} 2 \\ 4 \\ 2 \end{pmatrix} x + \begin{pmatrix} 0 \\ -2 \\ -2 \end{pmatrix}, \quad W_2^g(x) = \begin{pmatrix} 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

The function $h(x) = \rho(2x) - \rho(4x - 1.6) + \rho(7x - 5.6) = W_2(\rho(W_1(x)))$ satisfies $h(x) = f(0.5x)$, for $x \in [0, 1]$. Here,

$$W_1(x) = \begin{pmatrix} 2 \\ 4 \\ 7 \end{pmatrix} x + \begin{pmatrix} 0 \\ -1.6 \\ -5.6 \end{pmatrix}, \quad W_2(x) = \begin{pmatrix} 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

Therefore, $f(x) = h(g(x))$ can be realized through the network

$$\Phi_2(x) = W_2(\rho(W_1(W_2^g(\rho(W_1^g(x)))))) = W_3'(\rho(W_2'(\rho(W_1'(x)))))$$

with

$$W_1'(x) = W_1^g(x) = \begin{pmatrix} 2 \\ 4 \\ 2 \end{pmatrix} x + \begin{pmatrix} 0 \\ -2 \\ -2 \end{pmatrix}, \quad W_3'(x) = W_2(x) = \begin{pmatrix} 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix},$$

$$\begin{aligned} W_2'(x) = W_1(W_2^g(x)) &= \begin{pmatrix} 2 \\ 4 \\ 7 \end{pmatrix} \begin{pmatrix} 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + \begin{pmatrix} 0 \\ -1.6 \\ -5.6 \end{pmatrix} \\ &= \begin{pmatrix} 2 & -2 & 2 \\ 4 & -4 & 4 \\ 7 & -7 & 7 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} + \begin{pmatrix} 0 \\ -1.6 \\ -5.6 \end{pmatrix}. \end{aligned}$$

The network Φ_2 has depth 3, width 3, and connectivity 19.

Problem 4

(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 $i \in [1, N]$ so that $\rho(x, x_i) \le \epsilon$. The ϵ-covering number $N(\epsilon; \mathcal{C}, \rho)$ 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 $\rho(x_i, x_j) > \epsilon$, for all distinct i, j. The ϵ-packing number $M(\epsilon; \mathcal{C}, \rho)$ is the cardinality of a largest ϵ-packing of $\mathcal{C}$.

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

(b) (i) Let $\epsilon < 2^{-n}$ be arbitrary but fixed, take $K = \log_2(1/\epsilon)$, and divide the interval $I_n := [-2^{-n}, 2^{-n}]$ into $L := \lceil 2^{K-n} \rceil \ge 2$ sub-intervals of equal length, centered at the points $\theta_i = -2^{-n} + \frac{(2i-1)2^{-n}}{L}$, for $i \in \{1, 2, \ldots, L\}$, and each of length $\frac{2^{1-n}}{L} \le 2^{1-K} = 2\epsilon$. By construction, for every $\theta \in I_n$, there is hence a $j \in \{1, 2, \ldots, L\}$ such that $|\theta - \theta_j| \le \epsilon$, which proves that $A_n(\epsilon) = \{\theta_i; i \in \{1, 2, \ldots, L\}\}$ is an ϵ-covering.

(ii) For the construction of the 2ϵ-packing, take the points $\theta_i' = -2^{-n} + \frac{2(i-1)2^{-n}}{L-1}$, for $i \in \{1, 2, \ldots, L\}$. As for every neighboring pair $\theta_i', \theta_j' \in I_n$, it holds that $|\theta_i' - \theta_j'| = \frac{2^{1-n}}{L-1} > 2\epsilon$, we have established that $P_n(\epsilon) = \{\theta_i'; i \in \{1, 2, \ldots, L\}\}$ constitutes a 2ϵ-packing.We finally note that $|P_n(\epsilon)| = |A_n(\epsilon)| = \lceil 2^{K-n} \rceil$.

(iii) By (b.i) we have

$$N(\epsilon; I_n, \rho_1) \le |A_n(\epsilon)| = \lceil 2^{K-n} \rceil$$

and by (b.ii),

$$M(2\epsilon; I_n, \rho_1) \ge |P_n(\epsilon)| = \lceil 2^{K-n} \rceil.$$

Combining this with $M(2\epsilon; \mathcal{C}, \rho) \le N(\epsilon; \mathcal{C}, \rho)$ for every compact set $\mathcal{C}$ in the metric space $(\mathcal{X}, \rho)$, yields $N(\epsilon; I_n, \rho_1) = \lceil 2^{K-n} \rceil$.

(c) (i) An ϵ-covering of $\mathcal{C}$ can be obtained by forming the Cartesian product, across $n \in \mathbb{N}$, of the ϵ-coverings of $[-2^{-n}, 2^{-n}]$ according to $A(\epsilon) = \{f : \mathbb{N} \to \mathbb{R}; f(n) \in A_n(\epsilon), \forall n \in \mathbb{N}\}$, where, for $\epsilon < 2^{-n}$, $A_n(\epsilon)$ is the ϵ-covering of $[-2^{-n}, 2^{-n}]$ constructed in subproblem (b.i) and $A_n(\epsilon) = \{0\}$, for $\epsilon \ge 2^{-n}$.

(ii) Fix $\epsilon \le 1/2$ and take $K = \log_2(1/\epsilon)$. By subproblems (b.iii) and (c.i), we have

$$\begin{aligned} N(\epsilon; \mathcal{C}, \rho_2) \le |A(\epsilon)| = \prod_{n=1}^{\infty} |A_n(\epsilon)| \le \prod_{n=1}^{\lceil K \rceil - 1} 2^{\lceil K \rceil - n} = 2^{(\lceil K \rceil - 1)\lceil K \rceil / 2} \\ \le \left(\frac{1}{\epsilon}\right)^{\lceil K \rceil / 2} \le \left(\frac{1}{\epsilon}\right)^{\frac{1}{2}\log_2(1/\epsilon) + C}, \end{aligned}$$

for some $C > 0$ that is independent of ϵ.