Mathematical Introduction to Deep Neural Networks

Exercise Sheet 6

Exercise 1. *Show that for every $d \in \mathbb{N}$ one can realise the identity function on $\mathbb{R}^d$ with **one hidden layer** ANN using the **softplus**, **swish**, and **GELU** (Gaussian error linear unit) activation functions.*

Exercise 2. *Construct a sum of ANNs of **different** length, i.e. let $\Phi_1, \ldots, \Phi_n \in \mathbf{N}$ satisfy*

$$\mathcal{I}(\Phi_k) = \mathcal{O}(\Phi_k) = 1 \quad \text{for all } k \in \{1, \ldots, n\}.$$

Let $a \in C(\mathbb{R}, \mathbb{R})$, $\mathbb{I} \in \mathbf{N}$ satisfy

$$\mathcal{I}(\mathbb{I}) = \mathcal{O}(\mathbb{I}) = 1, \qquad \mathcal{H}(\mathbb{I}) = 1,$$

and

$$\mathcal{R}_a^{\mathbf{N}}(\mathbb{I})(x) = x \quad \text{for all } x \in \mathbb{R}.$$

Show that there exists $\Psi \in \mathbb{N}$ such that

$$\mathcal{R}_a^{\mathbf{N}}(\Psi) = \sum_{k=1}^{n} \mathcal{R}_a^{\mathbf{N}}(\Phi_k).$$

Exercise 3. *How does the dimension vector (architecture) of Ψ in Exercise 2 above look like in terms of the dimensions of $\Phi_1, \ldots, \Phi_n$?*

Exercise 4. *Let $\Phi_1, \Phi_2 \in \mathbf{N}$ satisfy*

$$\mathcal{I}(\Phi_1) = \mathcal{I}(\Phi_2) = m, \qquad \mathcal{O}(\Phi_1) = \mathcal{O}(\Phi_2) = n,$$

and

$$\mathcal{L}(\Phi_1) = \mathcal{L}(\Phi_2).$$

Let $\mathcal{W}_1, \mathcal{W}_2 \in \mathbb{R}^{n \times n}$ and $\mathcal{B} \in \mathbb{R}^n$. Construct $\Psi \in \mathbf{N}$ such that for all $a \in C(\mathbb{R}, \mathbb{R})$ it holds that

$$\mathcal{R}_a^{\mathbf{N}}(\Psi) = \mathcal{W}_1 \mathcal{R}_a^{\mathbf{N}}(\Phi_1) + \mathcal{W}_2 \mathcal{R}_a^{\mathbf{N}}(\Phi_2) + \mathcal{B}.$$