Mathematical Introduction to Deep Neural Networks

Exercise Sheet 1

Exercise 1 (Supervised learning problem). *This exercise illustrates how the abstract supervised learning framework introduced in the lecture becomes a concrete least–squares problem when the network is linear.*

Consider a simple data generation rule given by

$$y = \mathcal{E}(x) = 2x_1 - 3x_2 + 0.5,$$

where the input variable is $x = (x_1, x_2) \in \mathbb{R}^2$ and the output $y \in \mathbb{R}$ is observed with small noise,

$$y_m = \mathcal{E}(x_m) + \varepsilon_m, \qquad m = 1, \ldots, M.$$

(i) Assume the given training set

$$\mathcal{D}_M = \{(x_m, y_m)\}_{m=1}^{M}, \qquad x_m \in \mathbb{R}^2,\ y_m \in \mathbb{R},$$

and let the neural network $\psi_\theta : \mathbb{R}^2 \to \mathbb{R}$ be the linear map

$$\psi_\theta(x) = w_1 x_1 + w_2 x_2 + b, \qquad \theta = (w_1, w_2, b) \in \mathbb{R}^3.$$

Express the empirical loss function

$$\mathscr{L}(\theta) = \frac{1}{M} \sum_{m=1}^{M} \big|\psi_\theta(x_m) - y_m\big|^2$$

explicitly in terms of w_1, w_2, and b.

(ii) Formulate the supervised learning problem as the minimisation

$$\vartheta = \arg\min_{\theta \in \mathbb{R}^3} \mathscr{L}(\theta).$$

Why can $\psi_\vartheta(x_{M+1})$ be interpreted as an approximation of the unknown output $\mathcal{E}(x_{M+1})$?

Additionally, **Exercises 1.1.1, 1.1.2, 1.1.3, 1.1.4** from the lecture notes.