2.0.1 - Introduction

In this chapter we deal with supervised learning.
This refers to the following situation:

- We are looking for a mapping $f : \mathcal{X} \to \mathcal{Y}$, where we denote $\mathcal{X}$ as the input space and $\mathcal{Y}$ as the output space
- We consider $N \in \mathbb{N}$ data points $(x_i, y_i) \in \mathcal{X} \times \mathcal{Y}$ for $i = 1, \ldots, N$.
- We choose a class of hypotheses $\mathcal{H}$, typically a (subset of a) function space
- We determine $\hat{f} \in \mathcal{H}$ such that $\hat{f}(x_i) \approx y_i$ for all $i = 1, \ldots, N$

Remark: it is often assumed that the data points (x_i, y_i) are samples from an (unknown) probability distribution on $\mathcal{X} \times \mathcal{Y}$
Within supervised learning, a distinction is often made between regression problems and classification problems. The distinguishing feature here is that in the case of quantitative output data, we speak of regression, and in the case of qualitative data, we speak of classification. However, the transition between these categories are fluid, and the terminology is not used consistently.