2.2.1 Linear Classification

In linear classification, we start with the binary case and make the assumption that the sought (to search) sets are half-spaces in $\mathbb{R}^d$ separated by a hyperplane of the form $\{x \in \mathbb{R}^d | \langle w, x \rangle + b\}$
.
We consider the following function $\hat{f} : \mathbb{R}^d \to \{-1, 1\}$, which assigns a label $\hat{f}(x) \in \{-1, 1\}$ to a data point $x \in \mathbb{R}^d$:

2.2.1

$$\hat{f}(x) := \begin{cases} 1 & \text{if } \langle w, x \rangle + b \geq 0 \\ -1 & \text{if } \langle w, x \rangle + b < 0 \end{cases}$$

As with regression problems, we now need to determine the parameters $w \in \mathbb{R}^d$ and $b \in \mathbb{R}$ so that $\hat{f}(x_i) \approx y_i$ for all data points $i = 1, \ldots, N$.
At this point, we emphasize that $\hat{f}(x) = \mathrm{sign}(\langle w, x \rangle + b)$ and thus the activation function $\hat{f}$ is a neural network with one neuron and an activation function $\sigma = \mathrm{sign}$, where

2.2.2

$$\mathrm{sign}(t) := \begin{cases} 1 & \text{if } t \geq 0, \\ -1 & \text{if } t < 0 \end{cases}$$

is the sign function with the convention $\mathrm{sign}(0) = 1$