Lecture Notes

Machine Learning

Winter Semester 24/25

Emma Bach

Lecture by
Dr. Daniele Cattaneo

February 20, 2025

Table of contents

Chapter 1

Introduction

Given a set $\{(x_1, y_1), \ldots, (x_n, y_n)\}$, i.e. a set of features (inputs) $x = \{x_1, \ldots, x_n\}$ and a target (desired output) $y = \{y_1, \ldots, y_n\}$, Machine Learning is the process of algorithmically finding the best possible function $f(x)$ such that

$$\forall x_i [f(x_i) \approx y_i].$$

y is also known as the **ground truth**. In **unsupervised learning**, y is not explicitly given to the algorithm (and may not even exist). The focus of the course, and thus these notes, was **supervised learning**, where y is explicitly given.

In order to find an optimal $f(x)$, we first rephrase the function in terms of **learnable parameters** θ to get a function $f(x, \theta)$. We often write $f(x_i, \theta) = \hat{y}_i$, where the hat ˆ is supposed to show that $\hat{y}_i$ is an approximation of y_i. The full image $\hat{y}$ is often called the **prediction model**.

We now need to define a seperate function that we can use to judge how good our values of θ are. Such a function is known as a **loss function**, which we write as $\mathcal{L}(y, \hat{y})$. The problem of finding the best possible f then comes down to a minimization problem of the form

$$\underset{\theta}{\operatorname{argmin}} \sum_{n=1}^{N} \mathcal{L}(y_n, \hat{y}_n)$$