2.2 Classification

In this chapter, we address classification problems. These differ from regression problems mainly in that the output data is **qualitative** rather than quantitative. Typically, each output is referred to as a label. In the case of two labels, which are usually modeled by $\mathcal{Y} = \{0, 1\}$ or $\mathcal{Y} = \{-1, 1\}$, we speak of **binary classification**.
We consider training data $(x_i, y_i)_{i=1}^N \subset \mathcal{X} \times \mathcal{Y}$, where $\mathcal{Y} = \{l_1, \dots, l_C\}$ denotes the set of all possible labels. The abstract goal is to find disjoint sets $\mathcal{M}_1, \dots, \mathcal{M}_C \subset \mathcal{X}$ with $\mathcal{X} = \bigcup_{c=1}^C \mathcal{M}_c$ ($\mathcal{X}$ is the input set, $\mathcal{M}_c$ should all be disjoint), such that most training points labeled l_c lie in the set $\mathcal{M}_c$. If we then have a new data point $x \in \mathcal{X}$, we determine the unique set $\mathcal{M}_c$ with $x \in \mathcal{M}_c$ and assign the label l_c to x.
We typically consider the case of binary classification with $C = 2$. The general case with $C > 2$ classes can be solved by one of the following approaches:

1. (one-vs one) We compute $\binom{C}{2} = \frac{C!}{2!(C-2)!} = \frac{C(C-1)}{2}$ binary classifiers that compare all pairs of classes. To classify a new data point, we choose the class that is most frequently assigned by these classifiers
2. (one-vs-all) We compute C classifiers that classify the points of a fixed class $c \in \{1, \dots, C\}$ against all other points (the rest). We also assume that this gives us a probability for belonging to the c-th class. To classify a new data point, we choose the class with the highest probability value