

Machine Learning I

Bjoern Andres

Machine Learning for Computer Vision
TU Dresden



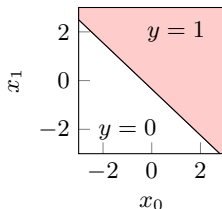
<https://mlcv.cs.tu-dresden.de/courses/26-winter/ml1/>

Winter Term 2026/2027

Supervised learning

Contents. This part of the course introduces the notion of labeled data, the supervised learning problem, the separation problem, the separability problem and the inference problem.

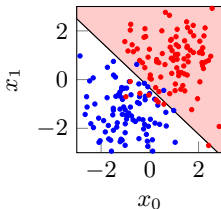
Supervised learning



Example: A medical test with $n \in \mathbb{N}$ design parameters $\theta \in \Theta = \mathbb{R}^n$ measures $m \in \mathbb{N}$ numbers $x \in X = \mathbb{R}^m$ and indicates by $y \in Y = \{0, 1\}$ whether the measurement x is considered to be healthy ($y = 0$) or pathological ($y = 1$).

$$X \xrightarrow{g_\theta} Y$$

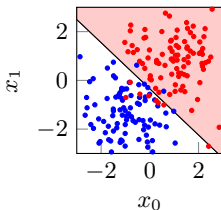
Supervised learning



Example: A medical test with $n \in \mathbb{N}$ design parameters $\theta \in \Theta = \mathbb{R}^n$ measures $m \in \mathbb{N}$ numbers $x \in X = \mathbb{R}^m$ and indicates by $y \in Y = \{0, 1\}$ whether the measurement x is considered to be healthy ($y = 0$) or pathological ($y = 1$).

$$X \xrightarrow{g_\theta} Y$$

Supervised learning



Informally, **supervised learning** is the problem of finding, in a family $g : \Theta \rightarrow Y^X$, one function $g_\theta : X \rightarrow Y$ that minimizes a weighted sum of two objectives:

- ▶ g_θ deviates little from a finite set $\{(x_s, y_s) \mid s \in S\}$ of input-output-pairs, called **labeled data**
- ▶ g_θ has low complexity, as quantified by a function $R : \Theta \rightarrow \mathbb{R}_0^+$, called a **regularizer**

Supervised learning

In this course, we concentrate exclusively on the special case where Y is finite. To begin with, we even concentrate on the case where $Y = \{0, 1\}$, i.e., learning how to make yes/no decisions. Hence, we consider a family $g: \Theta \rightarrow \{0, 1\}^X$.

We allow ourselves to take a detour by not optimizing over a family $g: \Theta \rightarrow \{0, 1\}^X$ directly but instead optimizing over a family $f: \Theta \rightarrow \mathbb{R}^X$ and defining g wrt. f via a function $L: \mathbb{R} \times \{0, 1\} \rightarrow \mathbb{R}_0^+$, called a **loss function**, such that

$$\forall \theta \in \Theta \quad \forall x \in X: \quad g_\theta(x) \in \operatorname{argmin}_{\hat{y} \in \{0, 1\}} L(f_\theta(x), \hat{y}) . \quad (1)$$

Example: 0-1-loss

$$\forall r \in \mathbb{R} \quad \forall \hat{y} \in \{0, 1\}: \quad L(r, \hat{y}) = \begin{cases} 0 & r > 0 \wedge \hat{y} = 1 \\ 0 & r \leq 0 \wedge \hat{y} = 0 \\ 1 & \text{otherwise} \end{cases} . \quad (2)$$

Next, we define the supervised learning problem rigorously.

Definition. For any finite, non-empty set S , called a set of **samples**, any $X \neq \emptyset$, called an **input space** and any $x : S \rightarrow X$, the tuple (S, X, x) is called **unlabeled data**.

For any $y : S \rightarrow \{0, 1\}$, given in addition and called a **labeling**, the tuple (S, X, x, y) is called **labeled data**.

Supervised learning

Definition. For any labeled data $T = (S, X, x, y)$, any $\Theta \neq \emptyset$ and $f : \Theta \rightarrow \mathbb{R}^X$, any $R : \Theta \rightarrow \mathbb{R}_0^+$, called a **regularizer**, any $L : \mathbb{R} \times \{0, 1\} \rightarrow \mathbb{R}_0^+$, called a **loss function**, and any $\lambda \in \mathbb{R}_0^+$:

- The instance of the **supervised learning problem** has the form

$$\inf_{\theta \in \Theta} \lambda R(\theta) + \sum_{s \in S} L(f_\theta(x_s), y_s) \quad (3)$$

- The instance of the **separation problem** has the form

$$\inf_{\theta \in \Theta} R(\theta) \quad (4)$$

$$\text{subject to } \forall s \in S : L(f_\theta(x_s), y_s) = 0 \quad (5)$$

- The instance of the **separability problem** consists in deciding whether there exists a $\theta \in \Theta$ such that

$$R(\theta) \leq m \quad (6)$$

$$\forall s \in S : L(f_\theta(x_s), y_s) = 0 \quad (7)$$

Definition. For any unlabeled data $T = (S, X, x)$, any $\hat{f} : X \rightarrow \mathbb{R}$ and any $L : \mathbb{R} \times \{0, 1\} \rightarrow \mathbb{R}_0^+$, the instance of the **inference problem** wrt. $T, \hat{f}$ and L is defined as

$$\min_{y \in \{0,1\}^S} \sum_{s \in S} L(\hat{f}(x_s), y_s) \quad (8)$$

Lemma. The solutions to the inference problem are the $y : S \rightarrow \{0, 1\}$ such that

$$\forall s \in S: \quad y_s \in \operatorname{argmin}_{\hat{y} \in \{0,1\}} L(\hat{f}(x_s), \hat{y}) \quad (9)$$

Moreover, if $\hat{f}(X) \subseteq \{0, 1\}$ and L is the 0-1-loss, then

$$\forall s \in S: \quad y_s = \hat{f}(x_s) \quad (10)$$

Summary.

- ▶ The **supervised learning problem** is an optimization problem. It consists in finding, in a family of functions, one function that minimizes a weighted sum of two objectives:
 1. It deviates little from given labeled data, as quantified by a loss function
 2. It has low complexity, as quantified by a regularizer.
- ▶ The **separation problem** is an optimization problem. It consists in finding a function in the family that minimizes the regularizer, such that the loss wrt. given labeled data is zero.
- ▶ The **separability problem** is a decision problem. It consists in deciding whether there exists a function in the family with loss zero for which the regularizer does not exceed a given bound.