Machine Learning II

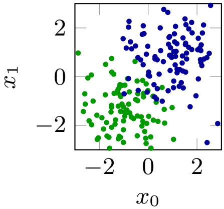
Bjoern Andres

Machine Learning for Computer Vision
TU Dresden

https://mlcv.cs.tu-dresden.de/courses/26-summer/ml2/

Summer Term 2026

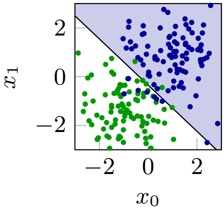
Supervised learning (recap.)



Definition 1. For any finite, non-empty set S, called a set of **samples**, any $X \neq \emptyset$, called a **feature space** and any $x : S \to X$, the tuple (S, X, x) is called **unlabeled data**.

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

Supervised learning (recap.)



Informally, the goal of supervised learning is to find, in a non-empty family $f\colon \Theta \to \mathbb{R}^X$, one function f_θ that optimizes a weighted sum of two objectives:

- f_θ deviates little from labeled data, as quantified by an $L\colon \mathbb{R} \times \{0,1\} \to \mathbb{R}_0^+$, called a **loss function**.
- f_θ has low complexity, as quantified by an $R\colon \Theta \to \mathbb{R}_0^+$, called a **regularizer**.

Definition 2. For any labeled data (S, X, x, y), any $\Theta \neq \varnothing$ and $f : \Theta \to \mathbb{R}^X$, any $L : \mathbb{R} \times \{0, 1\} \to \mathbb{R}_0^+$ called a **loss function**, any $R : \Theta \to \mathbb{R}_0^+$ called a **regularizer**, and any $\lambda \in \mathbb{R}_0^+$:

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

$$\inf_{\theta \in \Theta} \quad \lambda R(\theta) + \sum_{s \in S} L(f_\theta(x_s), y_s) \tag{1}$$

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

$$\inf_{\theta \in \Theta} \quad R(\theta) \tag{2}$$

$$\text{subject to} \quad \forall s \in S : \quad L(f_\theta(x_s), y_s) = 0 \tag{3}$$

– For any $m \in \mathbb{R}_0^+$, given in addition, the instance of the **separability problem** consists in deciding whether there exists a $\theta \in \Theta$ such that

$$R(\theta) \leq m \tag{4}$$

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

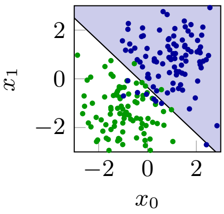
Definition. For any unlabeled data $T = (S, X, x)$, any $\hat{f} : X \to \mathbb{R}$ and any $L : \mathbb{R} \times \{0, 1\} \to \mathbb{R}_0^+$, the instance of the **inference problem** wrt. $T, \hat{f}$ and L is defined as

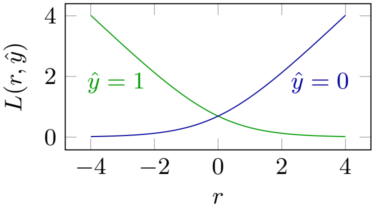
$$\min_{y \in \{0,1\}^S} \sum_{s \in S} L(\hat{f}(x_s), y_s) \quad . \tag{6}$$

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

$$\forall s \in S\colon \quad y_s \in \operatorname*{argmin}_{\hat{y} \in \{0,1\}} L(\hat{f}(x_s), \hat{y}) \quad . \tag{7}$$

Supervised learning (recap.)





Example 1. For

- any labeled data (S, X, x, y) such that $X = \mathbb{R}^n$ and $n \in \mathbb{N}$
- $\Theta = \mathbb{R}^n$ and $f \colon \Theta \to \mathbb{R}^X$ such that $\forall \theta \in \Theta \forall \hat{x} \in X \colon f_\theta(\hat{x}) = \langle \theta, \hat{x} \rangle$
- $L \colon \mathbb{R} \times \{0, 1\} \to \mathbb{R}_0^+$ such that for any $r \in \mathbb{R}$ and any $\hat{y} \in \{0, 1\}$:

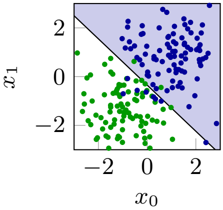
$$L(r, \hat{y}) = -r\hat{y} + \ln(1 + e^r) = \ln(1 + e^{-(2\hat{y}-1)r}) \tag{8}$$

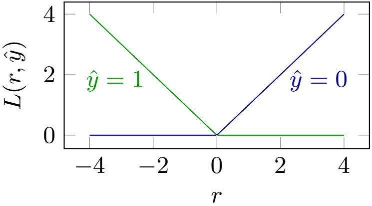
- any $R \colon \Theta \to \mathbb{R}_0^+$ and any $\lambda \in \mathbb{R}_0^+$

the supervised learning problem

$$\inf_{\theta \in \Theta} \quad \lambda R(\theta) + \sum_{s \in S} L(f_\theta(x_s), y_s) \tag{9}$$

is called a **linear logistic regression problem**.





Example 2. For

- any labeled data (S, X, x, y) such that $X = \mathbb{R}^n$ and $n \in \mathbb{N}$
- $\Theta = \mathbb{R}^n$ and $f\colon \Theta \to \mathbb{R}^X$ such that $\forall \theta \in \Theta \forall \hat{x} \in X\colon f_\theta(\hat{x}) = \langle \theta, \hat{x} \rangle$
- $L\colon \mathbb{R} \times \{0, 1\} \to \mathbb{R}_0^+$ such that for any $r \in \mathbb{R}$ and any $\hat{y} \in \{0, 1\}$:

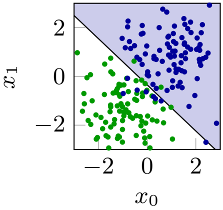
$$L(r, \hat{y}) = \max(0, -(2\hat{y} - 1)r) \tag{10}$$

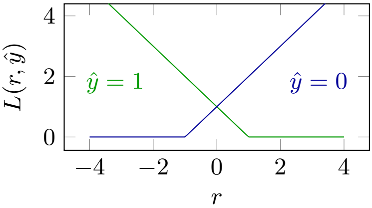
- any $R\colon \Theta \to \mathbb{R}_0^+$ and any $\lambda \in \mathbb{R}_0^+$

the supervised learning problem

$$\inf_{\theta \in \Theta} \quad \lambda R(\theta) + \sum_{s \in S} L(f_\theta(x_s), y_s) \tag{11}$$

is called a **rectified linear unit learning problem**.





Example 3. For

- any labeled data (S, X, x, y) such that $X = \mathbb{R}^n$ and $n \in \mathbb{N}$
- $\Theta = \mathbb{R}^n$ and $f \colon \Theta \to \mathbb{R}^X$ such that $\forall \theta \in \Theta \forall \hat{x} \in X \colon f_\theta(\hat{x}) = \langle \theta, \hat{x} \rangle$
- $L \colon \mathbb{R} \times \{0, 1\} \to \mathbb{R}_0^+$ such that for any $r \in \mathbb{R}$ and any $\hat{y} \in \{0, 1\}$:

$$L(r, \hat{y}) = \max(0, -(2\hat{y} - 1)r + 1) \tag{12}$$

- any $R \colon \Theta \to \mathbb{R}_0^+$ and any $\lambda \in \mathbb{R}_0^+$

the supervised learning problem

$$\inf_{\theta \in \Theta} \quad \lambda R(\theta) + \sum_{s \in S} L(f_\theta(x_s), y_s) \tag{13}$$

is called a **linear support vector machine (SVM) learning problem**.

Lemma 1. For

- any labeled data (S, X, x, y) such that $X = \mathbb{R}^n$ and $n \in \mathbb{N}$
- $\Theta = \mathbb{R}^n$ and $f\colon \Theta \to \mathbb{R}^X$ such that $\forall \theta \in \Theta \forall \hat{x} \in X\colon f_\theta(\hat{x}) = \langle \theta, \hat{x} \rangle$
- $R\colon \Theta \to \mathbb{R}_0^+$ such that $\forall \theta \in \Theta\colon R(\theta) = \|\theta\|_1$.
- any $\lambda \in \mathbb{R}_0^+$
- any $\hat{\theta} \in \Theta$

$\hat{\theta}$ is a solution to the rectified linear unit learning problem wrt. the above iff there exist $\hat{\xi} \in \mathbb{R}^S$ and $\hat{\theta}^+, \hat{\theta}^- \in \mathbb{R}^n$ such that $\hat{\theta} = \hat{\theta}^+ - \hat{\theta}^-$ and $(\hat{\xi}, \hat{\theta}^+, \hat{\theta}^-)$ is a solution to the optimization problem

$$
\begin{aligned}
\min_{\theta \in \Theta, \xi \in \mathbb{R}^S} \quad & \sum_{s \in S} \xi_s + \lambda \sum_{j \in n} \theta_j^+ + \lambda \sum_{j \in n} \theta_j^- && (14) \\
\text{subject to} \quad & \forall s \in S\colon \quad (2y_s - 1)\langle \theta^+ - \theta^-, x_s \rangle + \xi_s \geq 0 && (15) \\
& \forall s \in S\colon \quad \xi_s \geq 0 && (16) \\
& \forall j \in n\colon \quad \theta_j^+ \geq 0 && (17) \\
& \forall j \in n\colon \quad \theta_j^- \geq 0 \ . && (18)
\end{aligned}
$$

Remark 1. (14)–(18) is a linear optimization problem. In particular, the constraints are linear inequalities.

Lemma 2. For

- any labeled data (S, X, x, y) such that $X = \mathbb{R}^n$ and $n \in \mathbb{N}$
- $\Theta = \mathbb{R}^n$ and $f\colon \Theta \to \mathbb{R}^X$ such that $\forall \theta \in \Theta \forall \hat{x} \in X\colon f_\theta(\hat{x}) = \langle \theta, \hat{x} \rangle$
- $R\colon \Theta \to \mathbb{R}_0^+$ such that $\forall \theta \in \Theta\colon R(\theta) = \|\theta\|_1$.
- any $\lambda \in \mathbb{R}_0^+$
- any $\hat{\theta} \in \Theta$

$\hat{\theta}$ is a solution to the linear SVM learning problem wrt. the above iff there exist $\hat{\xi} \in \mathbb{R}^S$ and $\hat{\theta}^+, \hat{\theta}^- \in \mathbb{R}^n$ such that $\hat{\theta} = \hat{\theta}^+ - \hat{\theta}^-$ and $(\hat{\xi}, \hat{\theta}^+, \hat{\theta}^-)$ is a solution to the optimization problem

$$\min_{\theta \in \Theta, \xi \in \mathbb{R}^S} \quad \sum_{s \in S} \xi_s + \lambda \sum_{j \in n} \theta_j^+ + \lambda \sum_{j \in n} \theta_j^- \tag{19}$$

$$\text{subject to} \quad \forall s \in S\colon \quad (2y_s - 1)\langle \theta^+ - \theta^-, x_s \rangle - 1 + \xi_s \geq 0 \tag{20}$$

$$\forall s \in S\colon \quad \xi_s \geq 0 \tag{21}$$

$$\forall j \in n\colon \quad \theta_j^+ \geq 0 \tag{22}$$

$$\forall j \in n\colon \quad \theta_j^- \geq 0 \; . \tag{23}$$

Remark 2. (19)–(23) is a linear optimization problem. In particular, the constraints are linear inequalities.