Machine Learning II

Jannik Irmai, Jannik Presberger, David Stein, Bjoern Andres

Machine Learning for Computer Vision
TU Dresden

https://mlcv.cs.tu-dresden.de/courses/25-summer/ml2/

Summer Term 2025

So far, we have considered

- **learning problems** w.r.t. **labeled data** (S, X, x, y) where, for every $s \in S$, a label $y_s \in \{0, 1\}$ is given
- **inference problems** w.r.t. **unlabeled data** (S, X, x) where no label is given and every combination of labels $y : S \to \{0, 1\}$ is a feasible solution

Next, we will consider learning problems where not every label is given and inference problems where not every combination of labels is feasible. Unlike before, the data we look at in both problems coincides, consisting of tuples $(S, X, x, \mathcal{Y})$ where $\mathcal{Y} \subseteq \{0, 1\}^S$ is a set of feasible labelings. In particular:

- $\mathcal{Y} = \{0, 1\}^S$ is the special case of **unlabeled data**
- $|\mathcal{Y}| = 1$ is the special case of **labeled data**
- Non-trivial choices of $\mathcal{Y}$ will allow us to encode **finite structures** such as maps (for classification), equivalence relations (for clustering) and orders (for ordering).

Definition. For

- any finite, non-empty set S, called a set of **samples**,
- any non-empty set X, called an **feature space**,
- any $x : S \to X$
- any non-empty set $\mathcal{Y} \subseteq \{0,1\}^S$, called a set of **feasible labelings**,

the tuple $T = (S, X, x, \mathcal{Y})$ is called **constrained data**.

Example. We reduce the problem of learning and inferring partitions to the problem of learning and inferring decisions by defining constrained data (S, X, x, Y) such that

$$S = \binom{A}{2} \tag{1}$$

$$\mathcal{Y} = \left\{ y : \binom{A}{2} \to \{0,1\} \;\middle|\; \forall a \in A \; \forall b \in A \setminus \{a\} \; \forall c \in A \setminus \{a,b\} \colon y_{\{a,b\}} + y_{\{b,c\}} - 1 \leq y_{\{a,c\}} \right\} \tag{2}$$

Definition. For

- any **constrained data** $T = (S, X, x, \mathcal{Y})$,
- any $\Theta \neq \varnothing$ and family of functions $f : \Theta \to \mathbb{R}^X$,
- any $R : \Theta \to \mathbb{R}_0^+$, called a **regularizer**,
- any $L : \mathbb{R} \times \{0, 1\} \to \mathbb{R}_0^+$, called a **loss function**
- any $\lambda \in \mathbb{R}_0^+$, called a **regularization parameter**,

the instance of the **learning and inference problem** w.r.t. T, Θ, f, R, L and λ has the form

$$\min_{y \in \mathcal{Y}} \inf_{\theta \in \Theta} \quad \lambda R(\theta) + \frac{1}{|S|} \sum_{s \in S} L(f_\theta(x_s), y_s) \ . \tag{3}$$

The special case of one-elementary $\mathcal{Y} = \{\hat{y}\}$ is called the **supervised learning problem**.

The special case of one-elementary $\Theta = \{\hat{\theta}\}$ written below is called the **inference problem**.

$$\min_{y \in \mathcal{Y}} \quad \sum_{s \in S} L(f_{\hat{\theta}}(x_s), y_s) \tag{4}$$

Special cases of the learning and inference problem:

- **Semi-supervised learning**: Some labels are fixed, i.e.:

$$\exists s \in S \ \exists b \in \{0,1\} \ \forall y \in \mathcal{Y}\colon \quad y_s = b \tag{5}$$

- **Unsupervised learning**: No label is fixed, i.e.:

$$\forall s \in S \ \forall b \in \{0,1\} \ \exists y \in \mathcal{Y}\colon \quad y_s = b \tag{6}$$

Remark. The inference problem

$$\min_{y \in \mathcal{Y}} \quad \sum_{s \in S} L(f_{\hat{\theta}}(x_s), y_s) \tag{7}$$

can be stated equivalently with a linear objective function:

$$\begin{aligned}
& \operatorname*{argmin}_{y \in \mathcal{Y}} \quad \sum_{s \in S} L(f_{\hat{\theta}}(x_s), y_s) && (8)\\
= \;& \operatorname*{argmin}_{y \in \mathcal{Y}} \quad \sum_{s \in S} y_s \, L(f_{\hat{\theta}}(x_s), 1) + (1 - y_s) \, L(f_{\hat{\theta}}(x_s), 0) && (9)\\
= \;& \operatorname*{argmin}_{y \in \mathcal{Y}} \quad \sum_{s \in S} y_s \, \underbrace{(L(f_{\hat{\theta}}(x_s), 1) - L(f_{\hat{\theta}}(x_s), 0)}_{=: \, c_s} && (10)\\
= \;& \operatorname*{argmin}_{y \in \mathcal{Y}} \quad \sum_{s \in S} c_s \, y_s \quad . && (11)
\end{aligned}$$

Summary.

- Semi-supervised and unsupervised learning are optimization problems.
- Feasible solutions to these optimization problems consist of both:
  - a labeling y of the samples
  - a parameter vector θ that defines a function f_θ
- Even if the parameter vector θ is learned in a supervised manner from labeled data, the inference problem can be non-trivial due to the constraint $y \in \mathcal{Y}$.