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

So far, we have considered

- **learning problems** wrt. **labeled data** (S, X, x, y) where, for every $s \in S$, a label $y_s \in \{0, 1\}$ is given
- **inference problems** wrt. **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
- 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 $\mathcal{Y}$ 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 a **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**.

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** has the form

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

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

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

$$\min_{y \in \mathcal{Y}} \quad \sum_{s \in S} L(f_{\hat{\theta}}(x_s), y_s) \tag{2}$$

Special cases of the learning and inference problem:

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

$$\exists s \in S \; \exists b \in \{0,1\} \; \forall y \in \mathcal{Y}\colon \quad y_s = b \tag{3}$$

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

$$\forall s \in S \; \forall b \in \{0,1\} \; \exists y \in \mathcal{Y}\colon \quad y_s = b \tag{4}$$

Remark. The inference problem

$$\min_{y \in \mathcal{Y}} \quad \sum_{s \in S} L(f_{\hat{\theta}}(x_s), y_s) \tag{5}$$

can be stated equivalently with a linear objective function:

$$\begin{aligned}
& \operatorname*{argmin}_{y \in \mathcal{Y}} \quad \sum_{s \in S} L(f_{\hat{\theta}}(x_s), y_s) && (6)\\
= \; & \operatorname*{argmin}_{y \in \mathcal{Y}} \quad \sum_{s \in S} y_s \, L(f_{\hat{\theta}}(x_s), 1) + (1 - y_s) \, L(f_{\hat{\theta}}(x_s), 0) && (7)\\
= \; & \operatorname*{argmin}_{y \in \mathcal{Y}} \quad \sum_{s \in S} y_s \, \underbrace{(L(f_{\hat{\theta}}(x_s), 1) - L(f_{\hat{\theta}}(x_s), 0)}_{=: \, c_s} && (8)\\
= \; & \operatorname*{argmin}_{y \in \mathcal{Y}} \quad \sum_{s \in S} c_s \, y_s \quad . && (9)
\end{aligned}$$