Machine Learning I

Bjoern Andres

Machine Learning for Computer Vision
TU Dresden

Contents. This part of the course is about structured data, the structured learning problem and the structured inference problem.

Motivation. Even the most general learning and inference problem w.r.t. constrained data $(S, X, x, \mathcal{Y})$ we have considered is too restrictive for certain applications:

- Attributes $x_s \in X$ are defined for single elements $s \in S$ only.
- Dependencies between decisions $y_s, y_{s'} \in \{0, 1\}$ for distinct $s, s' \in S$ are only due to hard constraints definded by the feasible set $\mathcal{Y} \subset \{0, 1\}^S$.

Example: Pixel classification: Given a digital image, we need to decide for every pixel $s \in S$, by the contents of the image around that pixel, whether the pixel is of interest $(y_s = 1)$ or not of interest $(y_s = 0)$.

Typically, decisions at neighboring pixels $s, s' \in S$ are more likely to be equal $(y_s = y_{s'})$ than unequal $(y_s \neq y_{s'})$, and we wish to learn how this increased probability depends on the contents of the image.

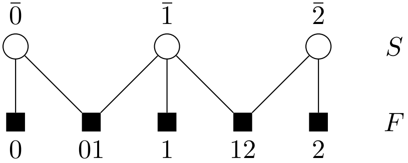
The mathematical abstractions of learning we have considered so far are insufficient to express these dependencies.

In order to lift this restriction, we will define the **supervised structured learning** problem and the **structured inference** problem in which

- attributes are associated with subsets of S
- decisions can be tied by probabilistic dependencies.

More specifically, we will

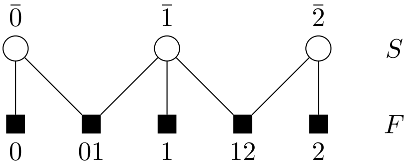
- introduce a family $H : \Theta \to \mathbb{R}^{X \times Y}$ of functions that quantify by $H_\theta(x, y)$ how incompatible attributes $x \in X$ are with a combination of decisions $y \in \{0, 1\}^S$
- define supervised structured learning as a problem of finding one function from this family
- define structured inference as the problem of finding a combination of decisions $y \in \{0, 1\}^S$ that minimizes $H_\theta(x, \cdot)$.



Definition. A triple (S, F, E) is called a **factor graph** with **variable nodes** S and **factor nodes** F iff $S \cap F = \emptyset$ and $(S \cup F, E)$ is a bipartite graph such that $\forall e \in E \, \exists s \in S \, \exists f \in F \colon e = \{s, f\}$.

- For any factor node $f \in F$, we denote by $S_f = \{s \in S \mid \{s, f\} \in E\}$ the set of those variable nodes that are neighbors of f.
- For any variable node $s \in S$, we denote by $F_s = \{f \in F \mid \{s, f\} \in E\}$ the set of those factor nodes that are neighbors of s.

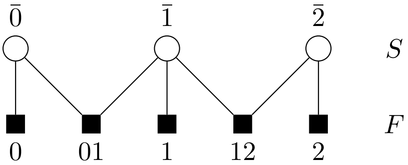
Supervised Structured Learning



Definition. A tuple $T = (S, F, E, \{X_f\}_{f \in F}, x)$ is called **unlabeled structured data** iff the following conditions hold:

- (S, F, E) is a factor graph
- Every set X_f is non-empty, called the **attribute space** of f
- $x \in \prod_{f \in F} X_f$, where the Cartesian product $\prod_{f \in F} X_f$ is called the **attribute space** of T.

A tuple $(S, F, E, \{X_f\}_{f \in F}, x, y)$ is called **labeled structured data** iff $(S, F, E, \{X_f\}_{f \in F}, x)$ is unlabeled structured data, and $y \in \{0, 1\}^S$.

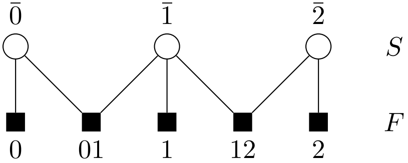


Definition. W.r.t. any labeled structured data $(S, F, E, \{X_f\}_{f \in F}, x, y)$,

- the attribute space $X = \prod_{f \in F} X_f$
- the set $Y = \{0, 1\}^S$
- any $\Theta \neq \emptyset$ and family of functions $H : \Theta \to \mathbb{R}^{X \times Y}$
- any $R : \Theta \to \mathbb{R}_0^+$, called a **regularizer**
- any $L : \mathbb{R}^Y \times Y \to \mathbb{R}_0^+$, called a **loss function**
- any $\lambda \in \mathbb{R}_0^+$, called a **regularization parameter**,

the instance of the **supervised structured learning problem** has the form

$$\inf_{\theta \in \Theta} \quad \lambda R(\theta) + L(H_\theta(x, \cdot), y) \tag{1}$$



Definition. With respect to

- any unlabeled structured data $T = (S, F, E, \{X_f\}_{f \in F}, x)$
- any $\hat{H} \colon X \times \{0,1\}^S \to \mathbb{R}$

the instance of the **structured inference problem** has the form

$$\min_{y \in \{0,1\}^S} \quad \hat{H}(x, y) \tag{2}$$

Summary.

- **Structured data** consists of a factor graph (S, F, E) and attributes $x_f \in X_f$ for every factor $f \in F$.
- The **structured learning problem** is an optimization problem whose feasible solutions θ define functions $H_\theta : X \times Y \to \mathbb{R}$ whose values $H_\theta(x, y)$ quantify an incompatibility of attributes $x \in X$ and combinations of decisions $y \in \{0, 1\}^S$.
- The **structured inference problem** consists in finding decisions $y \in \{0, 1\}^S$ compatible with given attributes $x \in X$, by minimizing a given incompatibility function $\hat{H}(x, \cdot)$.