Machine Learning I

Bjoern Andres

Machine Learning for Computer Vision
TU Dresden

Contents. This part of the course introduces algorithms for supervised structured inference with conditional graphical models.

The **inference problem** w.r.t. a **conditional graphical model** has the form of an unconstrained binary optimization problem:

$$\underset{y \in \{0,1\}^S}{\operatorname{argmin}} \; H_\theta(x, y) \tag{1}$$

It is NP-hard. (This can be shown, e.g., by reduction of binary integer programming, which is one of Karp's 21 problems).

We consider transformations that change one decision at a time:

Definition. For any $s \in S$, let $\mathrm{flip}_s \colon \{0,1\}^S \to \{0,1\}^S$ such that for any $y \in \{0,1\}^S$ and any $t \in S$:

$$\mathrm{flip}_s[y](t) = \begin{cases} 1 - y_t & \text{if } t = s \\ y_t & \text{otherwise} \end{cases} . \tag{2}$$

The greedy local search algorithm w.r.t these transformations is known as **Iterated Conditional Modes**, or ICM (Besag 1986).

$y' = \mathsf{icm}(y)$
choose $s \in \operatorname*{argmin}_{s' \in S} \; H_\theta(x, \mathrm{flip}_{s'}[y]) - H_\theta(x, y)$
if $H_\theta(x, \mathrm{flip}_s[y]) < H_\theta(x, y)$
$y' := \mathsf{icm}(\mathrm{flip}_s[y])$
else
$y' := y$

- The **inference problem** consists in computing the minimum of a sum of functions:

$$\begin{aligned} &\underset{y \in \{0,1\}^S}{\operatorname{argmin}} \; H_\theta(x, y) \\ = &\underset{y \in \{0,1\}^S}{\operatorname{argmin}} \; \sum_{f \in F} h_{f\theta}(x_f, y_{S(f)}) \end{aligned} \tag{3}$$

- This problem is analogous to that of computing the sum of a product of functions (from the previous lecture) in that both $(\mathbb{R}, \min, +)$ and $(\mathbb{R}, +, \cdot)$ are commutative semi-rings.
- This analogy is sufficient to transfer the idea of **message passing**, albeit with messages adapted to the $(\mathbb{R}, \min, +)$ semi-ring:

Definition. (Kschischang 2001) For any variable node $s \in S$ and any factor node $f \in F$, the functions

$$\mu_{s \to f}, \mu_{f \to s} \colon \{0,1\} \to \mathbb{R} \ , \tag{4}$$

called **messages**, are defined such that for all $y_s \in \{0,1\}$:

$$\mu_{s \to f}(y_s) = \sum_{f' \in F(s) \setminus \{f\}} \mu_{f' \to s}(y_s) \tag{5}$$

$$\mu_{f \to s}(y_s) = \min_{y_{S(f) \setminus \{s\}}} \psi_{f\theta}(x_f, y_{S(f)}) + \sum_{s' \in S(f) \setminus \{s\}} \mu_{s' \to f}(y_{s'}) \tag{6}$$

Lemma. If the factor graph is acyclic, messages are defined recursively by (5) and (6), beginning with the messages from leaves. Moreover, for any $s \in S$:

$$\begin{aligned} &\underset{y \in \{0,1\}^S}{\operatorname{argmin}} \; H_\theta(x, y) \\ = &\min_{y \in \{0,1\}^S} \sum_{f \in F} h_{f\theta}(x_f, y_{S(f)}) \\ = &\min_{y_s \in \{0,1\}} \sum_{f' \in F(s)} \mu_{f' \to s}(y_s) \end{aligned} \qquad (7)$$

Proof. Analogous to that of Lemma 18 in the lecture notes.

Summary

- For conditional graphical models whose factor graph is **acylic**, the inference problem can be solved efficiently by means of **min-sum message passing**.
- For conditional graphical models whose factor graph is **cyclic**, one local search algorithm for the inference problem is known as **Iterated Conditional Modes (ICM)**.