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

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

$$\min \left\{ H_\theta(x, y) \mid y \in \{0,1\}^S \right\} \tag{1}$$

$$= \min \left\{ \sum_{f \in F} h_{f\theta}(x_f, y_{S(f)}) \,\middle|\, y \in \{0,1\}^S \right\} \tag{2}$$

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

$$\operatorname{flip}_s[y](t) = \begin{cases} 1 - y_t & \text{if } t = s \\ y_t & \text{otherwise} \end{cases} . \tag{3}$$

Algorithm. Iterated conditional modes

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

Definition 2. (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 $y_s \in \{0,1\}$:

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

$$\mu_{f \to s}(y_s) = \min_{y_{S(f) \setminus \{s\}} \in \{0,1\}^{S(f) \setminus \{s\}}} h_{f\theta}(x_f, y_{S(f)}) + \sum_{s' \in S(f) \setminus \{s\}} \mu_{s' \to f}(y_{s'}) \tag{6}$$

Lemma 1. If the factor graph is acyclic, messages are defined recursively by (5) and (6), beginning with the messages from leaves. Moreover, for any $s \in S$:

$$\min_{y \in \{0,1\}^S} \ H_\theta(x,y) = \min_{y_s \in \{0,1\}} \sum_{f' \in F(s)} \mu_{f' \to s}(y_s) \tag{7}$$

Definition 3. For any finite set S and any $d \in \{0, \dots, |S|\}$, let

$$J_{Sd} := \bigcup_{m=0}^{d} \binom{S}{m} \qquad C_{Sd} := \mathbb{R}^{J_{Sd}} \tag{8}$$

and call any $c \in C_{Sd}$ an S-variate **multi-linear polynomial form** of degree at most d. If $d = |S|$, c is also called an S-variate **multi-linear polynomial form**.

Example. For $S = d = 2$, we have

$$\begin{aligned} J_{22} &= \bigcup_{m=0}^{2} \binom{\{0,1\}}{m} \\ &= \binom{\{0,1\}}{0} \cup \binom{\{0,1\}}{1} \cup \binom{\{0,1\}}{2} \\ &= \{\emptyset\} \cup \{\{0\}, \{1\}\} \cup \{\{0,1\}\} \\ &= \{\emptyset, \{0\}, \{1\}, \{0,1\}\} \end{aligned}$$

Definition 4. An **(S-variate) pseudo-boolean function (PBF)** is any $f\colon \{0,1\}^S \to \mathbb{R}$ with S a finite set.

Definition 5. For any finite set S, any $d \in \{0, \dots, |S|\}$ and any $c \in C_{Sd}$, the function f_c defined below is called the **PBF defined by** c.

$$f_c\colon \quad \{0,1\}^S \to \mathbb{R}\colon \quad x \mapsto \sum_{m=0}^{d} \sum_{J \in \binom{S}{m}} c_J \prod_{j \in J} x_j \tag{9}$$

Example. For any $c \in C_{22}$, f_c is such that for all $x \in \{0,1\}^2$:

$$f_c(x_0, x_1) = c_\varnothing + c_{\{0\}} x_0 + c_{\{1\}} x_1 + c_{\{0,1\}} x_0 x_1 \quad .$$

Lemma 2. Every PBF has a unique multi-linear polynomial form.

Proof. For any $J \subseteq S$, let $x^J \in \{0,1\}^S$ such that for all $j \in S$:

$$x^J_j = \begin{cases} 1 & \text{if } j \in J \\ 0 & \text{otherwise} \end{cases} .$$

Now,

$$\forall x \in \{0,1\}^S: \quad f(x) = \sum_{J \subseteq S} c_J \prod_{j \in J} x_j$$

is written equivalently as

$$f(x^\varnothing) = c_\varnothing$$
$$\forall J \neq \varnothing: \quad f(x^J) = c_J + \sum_{J' \subset J} c_{J'} \ .$$

Thus, c is defined uniquely (by induction over the cardinality of J). □

Example 1. For $S = d = 2$ and any $f : \{0,1\}^2 \to \mathbb{R}$, the existence of a $c \in C_{22}$ such that $f = f_c$ means

$$\forall x \in \{0,1\}^2 : \quad f(x_0, x_1) = c_\varnothing + c_{\{0\}} x_0 + c_{\{1\}} x_1 + c_{\{0,1\}} x_0 x_1 \ .$$

Explicitly,

$$\begin{aligned}
f(0,0) &= c_\varnothing \\
f(1,0) &= c_\varnothing + c_{\{0\}} \\
f(0,1) &= c_\varnothing \qquad\quad + c_{\{1\}} \\
f(1,1) &= c_\varnothing + c_{\{0\}} + c_{\{1\}} + c_{\{0,1\}} \ .
\end{aligned}$$

Definition 6. For any finite set S and any $d \in \{0, \ldots, |S|\}$, an S-**variate PBF of degree at most** d is any $f \in F_{Sd}$ where

$$F_{Sd} := \{f : \{0,1\}^S \to \mathbb{R} \mid \exists c \in C_{Sd} : \; f = f_c\} \; . \tag{10}$$

In particular, an S-**variate quadratic PBF (QPBF)** is any $f \in F_{S2}$.

Remark. For any finite set S, $F_{S|S|}$ is the set of all S-variate PBFs (by Lemma 2).

Definition 7. An instance of **pseudo-boolean optimization (PBO)** is a pair (S, c) with S a finite set and $c \in C_{S|S|}$. The objective function is f_c. The feasible set is $\{0,1\}^S$. The solutions are those $\hat{x} \in \{0,1\}^S$ with

$$f(\hat{x}) = \min \left\{ f_c(x) \mid x \in \{0,1\}^S \right\} \tag{11}$$

An instance of **quadratic pseudo-boolean optimization (PBO)** is a pair (S, c) with S a finite set and $c \in C_{S2}$. The objective function is f_c. The feasible set is $\{0,1\}^S$. The solutions are those $\hat{x} \in \{0,1\}^S$ with

$$f(\hat{x}) = \min \left\{ f_c(x) \mid x \in \{0,1\}^S \right\} \tag{12}$$

We show that PBO is polynomially reducible to QPBO.

Definition 8. For any finite set S and any $c \in C_{S|S|}$, define the **size** of c as

$$\mathrm{size}(c) := \sum_{J \subseteq S\colon\, c_J \neq 0} |J| \ . \tag{13}$$

Lemma 3. For any $x, y, z \in \{0, 1\}$:

$$z = xy \quad \Leftrightarrow \quad xy - 2xz - 2yz + 3z = 0 \ , \tag{14}$$

$$z \neq xy \quad \Leftrightarrow \quad xy - 2xz - 2yz + 3z > 0 \ . \tag{15}$$

Proof. By verifying equivalence for all eight cases. □

Algorithm. (Boros and Hammer 2001)

Input: $c \in C_{S|S|}$
$M := 1 + 2\sum_{J \subseteq S} |c_J|$
$t := 0$
$c^t := c$
while there exists a $J \subseteq S$ such that $|J| > 2$ and $c^t_J \neq 0$
  $c^{t+1} := c^t$
  $t := t + 1$
  Choose $j, k \in J$ such that $j \neq k$
  $S := S \,\dot\cup\, \{s_t\}$
  $c^t_{\{j,k\}} := c^t_{\{j,k\}} + M$
  $c^t_{\{j,s_t\}} := -2M$
  $c^t_{\{k,s_t\}} := -2M$
  $c^t_{\{s_t\}} := 3M$
  for all $\{j,k\} \subset J' \subseteq S$ such that $c^t_{J'} \neq 0$
    $c^t_{J' \setminus \{j,k\} \cup \{s_t\}} := c^t_{J'}$
    $c^t_{J'} := 0$
return c^t

Theorem 1.

- The algorithm terminates in polynomial time in $\mathrm{size}(c)$.
- $\mathrm{size}(c')$ is polynomially bounded by $\mathrm{size}(c)$.
- Upon termination in iteration t, c^t is a quadratic multi-linear polynomial form. Moreover, for all $x \in \{0,1\}^{S \cup \{s_1, \ldots s_t\}}$: x minimizes the QPBF f_{c^t} iff $x|_S$ minimizes the PBF c.

Proof. The algorithm replaces the occurrence of $x_j x_k$ by x_{s_t} and adds the form $M(x_j x_k - 2x_j x_{s_t} - 2x_k x_{s_t} + 3x_{s_t})$. By Lemma 3 and definition of M:

– If $x_{s_t} = x_j x_k$:

$$f_{c_t}(x) \; = \; f_c(x|_S) \; \leq \; \max_{x' \in \{0,1\}^S} f_c(x') \; < \; M/2 \; .$$

– Otherwise:

$$f_{c_t}(x) \geq M/2$$

Moreover, for every $t > 0$:

$$|\{J \subseteq S \colon |J| > 2 \wedge c^t_J \neq 0\}| \; < \; |\{J \subseteq S \colon |J| > 2 \wedge c^{t-1}_J \neq 0\}| \; ,$$

which proves the complexity claims. □

Definition 9. For any finite set S and any $d \in \{0, \dots, |S|\}$, let

$$K^+_{Sd} := \{(K^1, K^0) \mid K^1, K^0 \subseteq S \wedge K^1 \cap K^0 = \emptyset \wedge |K^1| + |K^0| = d\}$$

$$J^+_{Sd} := \bigcup_{m=0}^{d} K^+_{Sm}$$

$$C^+_{Sd} := \{c : J^+_{Sd} \to \mathbb{R} \mid \forall j \in J^+_{Sd} \setminus \{(\emptyset, \emptyset)\} : \ 0 \leq c_j\}$$

and call any $c \in C^+_{Sd}$ an S-variate **posiform** of degree at most d.

Example. For $S = d = 2$,

$$\begin{aligned} J^+_{S2} = \quad & \{ \ (\emptyset, \emptyset) \ \} \\ & \cup \{ \ (\{0\}, \emptyset), \ (\emptyset, \{0\}), \ (\{1\}, \emptyset), \ (\emptyset, \{1\}) \ \} \\ & \cup \{ \ (\{0,1\}, \emptyset), \ (\{0\}, \{1\}), \ (\{1\}, \{0\}), \ (\emptyset, \{0,1\}) \ \} \end{aligned}$$

Definition 10. For any finite set S, any $d \in \{0, \dots, |S|\}$ and any $c \in C^+_{Sd}$, $f_c : \{0,1\}^S \to \mathbb{R}$ such that

$$\forall x \in \{0,1\}^S \qquad f_c(x) := \sum_{(J^1, J^0) \in J^+_{Sd}} c_{J^1 J^0} \prod_{j \in J^1} x_j \prod_{j' \in J^0} (1 - x'_j) \qquad (16)$$

is called the **PBF defined by** c.

Example. For any $c \in C^+_{S2}$, $f_c : \{0,1\}^S \to \mathbb{R}$ is such that $\forall x \in \{0,1\}^2$:

$$\begin{aligned} f(x) = \quad & c_{\varnothing\varnothing} \\ & + c_{\{0\}\varnothing} x_0 + c_{\varnothing\{0\}}(1 - x_0) + c_{\{1\}\varnothing} x_1 + c_{\varnothing\{1\}}(1 - x_1) \\ & + c_{\{0,1\}\varnothing} x_0 x_1 + c_{\{0\}\{1\}} x_0 (1 - x_1) + c_{\{1\}\{0\}} (1 - x_0) x_1 \\ & \quad + c_{\varnothing\{0,1\}} (1 - x_0)(1 - x_1) \ . \end{aligned}$$

Definition 11. For any finite set S and any $f : \{0,1\}^S \to \mathbb{R}$, the posiform defined by

$$\forall x \in \{0,1\}^S: \quad K_x^1 := \{j \in S \mid x_j = 1\}$$
$$K_x^0 := \{j \in S \mid x_j = 0\}$$

and

$$J := \{(\varnothing, \varnothing)\} \cup \bigcup_{x \in \{0,1\}^S} \{(K_x^1, K_x^0)\}$$

and $c : J \to \mathbb{R}$ such that

$$c_{\varnothing\varnothing} := \min_{x \in \{0,1\}^S} f(x)$$
$$\forall x \in \{0,1\}^S \quad c_{K_x^1 K_x^0} := f(x) - c_{\varnothing\varnothing}$$

is called **min-term posiform** of f.

Lemma 4. For any finite set S and any $f : \{0,1\}^S \to \mathbb{R}$, the min-term posiform c of f is such that $f_c = f$.

Corollary 1. For any finite set S and any $f : \{0,1\}^S \to \mathbb{R}$, there exists a posiform $c \in C^+_{Sn}$ such that $f_c = f$.

Proof. Let S be any finite set, let $f : \{0,1\}^S \to \mathbb{R}$, and let $c : J \to \mathbb{R}$ be the min-term posiform of f.

c is a posiform (by definition).

Let $g : \{0,1\}^S \to \mathbb{R}$ be the PBF defined by this posiform.

Then, for any $x \in \{0,1\}^S$,

$$(J^1, J^0) \in \{(\emptyset, \emptyset), (K_x^1, K_x^0)\} \subseteq J$$

are the only elements of J for which

$$0 \neq \prod_{j \in J^1} x_j \prod_{j' \in J^0} (1 - x'_j) = 1 \ .$$

Thus,

$$\begin{aligned} \forall x \in \{0,1\}^S : \quad g(x) &= c_{\emptyset\emptyset} + c_{K_x^1 K_x^0} \\ &= c_{\emptyset\emptyset} + f(x) - c_{\emptyset\emptyset} \qquad \text{(by definition of } c\text{)} \\ &= f(x) \ . \end{aligned}$$

□

Remark. Unlike multi-linear polynomial forms, posiforms of PBFs need not be unique, e.g., $x_0 = x_0 x_1 + x_0(1 - x_1)$.

Definition 12. For any finite set S, any $f : \{0,1\}^S \to \mathbb{R}$ and any $d \in \{0, \ldots, |S|\}$, let

$$C^+_{Sd}(f) := \left\{ c \in C^+_{Sd} \mid f_c = f \right\} \quad . \tag{17}$$

Remark. For any finite set S and any $f : \{0,1\}^S \to \mathbb{R}$, $C^+_{Sn}(f)$ contains at least the min-term posiform of f.

Lemma 5.

$$\forall f : \{0,1\}^S \to \mathbb{R} \quad \forall c \in C^+_{nn}(f) \quad \forall x \in \{0,1\}^S : \quad c_{\emptyset\emptyset} \leq f(x) \ .$$

Proof. By definition, we have, for all $x \in \{0,1\}^S$,

$$\begin{aligned} f(x) &= \sum_{m=0}^{d} \sum_{(K^1,K^0)\in K^+_{Sm}} c_{K^1K^0} \prod_{j\in K^1} x_j \prod_{j'\in K^0} (1 - x'_j) \\ &= c_{\emptyset\emptyset} + \sum_{m=1}^{d} \sum_{(K^1,K^0)\in K^+_{Sm}} c_{K^1K^0} \prod_{j\in K^1} x_j \prod_{j'\in K^0} (1 - x'_j) \ , \end{aligned}$$

and all coefficients $c_{K^1K^0}$ in the second sum are non-negative.

Therefore, the second sum is non-negative.

Thus, for all $x \in \{0,1\}^S$:

$$c_{\emptyset\emptyset} \leq f(x) \ .$$

□

Definition 13. For any posiform $c : J \to \mathbb{R}$, a pair (T, y) such that $T \subseteq S$ and $y : S \to \{0, 1\}$ is called a **contractor** of c iff

$$\begin{aligned}\forall (J^1, J^0) \in J\colon \qquad & (J^1 \cap T = \varnothing \ \wedge \ J^0 \cap T = \varnothing) \\ & \vee \ (\exists j \in J^1 \cap T\colon y_j = 0) \\ & \vee \ (\exists j \in J^0 \cap T\colon y_j = 1) \ . \end{aligned} \tag{18}$$

Theorem 2 (partial optimality)**.** For any finite set S, any $f : \{0,1\}^S \to \mathbb{R}$, any posiform $c \in C^+_{S|S|}(f)$ and any contractor (T, y) of c, there exists a solution x^* to the problem $\min\{f(x) \mid x \in \{0,1\}^S\}$ such that

$$\forall j \in S\colon \quad x^*_j = y_j \quad . \tag{19}$$

Proof. Let $\sigma_{Ty} : \{0,1\}^S \to \{0,1\}^S$ such that $\forall x \in \{0,1\}^S\ \forall j \in S$:

$$\sigma_{Ty}(x)_j = \begin{cases} y_j & \text{if } j \in T \\ x_j & \text{otherwise} \end{cases} . \tag{20}$$

Let $J^{\bar{T}} := \{(J^1, J^0) \in J^+_{S|S|} \mid J^1 \cap T = J^0 \cap T = \varnothing\}$ and $J^T := J \setminus J^{\bar{T}}$.

Now, $\forall x \in \{0,1\}^S$:

$$f(x) = \underbrace{\sum_{(J^1,J^0)\in J^T} c_{J^1 J^0} \prod_{j\in J^1} x_j \prod_{j'\in J^0} (1 - x'_j)}_{=:f^T(x)} + \underbrace{\sum_{(J^1,J^0)\in J^{\bar{T}}} c_{J^1 J^0} \prod_{j\in J^1} x_j \prod_{j'\in J^0} (1 - x'_j)}_{=:f^{\bar{T}}(x)} .$$

Furthermore, $\forall x \in \{0,1\}^S$:

$$\begin{aligned} f^T(\sigma_{Ty}(x)) &= 0 && \text{(by definition)} \\ 0 &\le f^T(x) && \text{(because } (\varnothing,\varnothing) \notin J^T\text{)} \\ f^{\bar{T}}(\sigma_{Ty}(x)) &= f^{\bar{T}}(x) && \text{(by definition)} . \end{aligned}$$

Adding the lhs. and rhs. shows that σ_{Ty} is improving for the problem $\min\{f(x) \mid x \in \{0,1\}^S\}$. □

For any finite set S, consider S-variate **quadratic** forms, i.e.

– any **multi-linear polynomial form** $c \in C_{S2}$, and f_c, i.e. for all $x \in \{0,1\}^S$:

$$f_c(x) = c_\varnothing + \sum_{j \in S} c_{\{j\}} x_j + \sum_{\{j,k\} \in \binom{S}{2}} c_{\{j,k\}} x_j x_k$$

– any **posiform** $c' \in C^+_{S2}$, and f'_c, i.e. for all $x \in \{0,1\}^S$:

$$\begin{aligned} f'_{c'}(x) = c'_{\varnothing\varnothing} &+ \sum_{j \in S} \left(c'_{\{j\}\varnothing} x_j + c'_{\varnothing\{j\}} (1 - x_j) \right) \\ &+ \sum_{\{j,k\} \in \binom{S}{2}} \Big(c'_{\{j,k\}\varnothing} x_j x_k + c'_{\{j\}\{k\}} x_j (1 - x_k) \\ &\qquad + c'_{\{k\}\{j\}} x_k (1 - x_j) + c'_{\varnothing\{j,k\}} (1 - x_j)(1 - x_k) \Big) \end{aligned}$$

Lemma 6. For any finite set S, any QPBF $f : \{0,1\}^S \to \mathbb{R}$, the $c \in C_{S2}$ such that $f_c = f$ and any $c' \in C^+_{S2}(f)$:

$$c_\varnothing = c'_{\varnothing\varnothing} + \sum_{j \in S} c'_{\varnothing\{j\}} + \sum_{\{j,k\} \in \binom{S}{2}} c'_{\varnothing\{j,k\}}$$

$$\forall j \in S\colon \quad c_{\{j\}} = c'_{\{j\}\varnothing} - c'_{\varnothing\{j\}} + \sum_{k \in S \setminus \{j\}} \left(c'_{\{j\}\{k\}} - c'_{\varnothing\{j,k\}}\right)$$

$$\forall \{j,k\} \in \tbinom{S}{2}\colon \quad c_{\{j,k\}} = c'_{\{j,k\}\varnothing} + c'_{\varnothing\{j,k\}} - c'_{\{j\}\{k\}} - c'_{\{k\}\{j\}}$$

Proof. Expansion of the posiform c' yields a quadratic multi-linear polynomial form. Comparison with c yields the conditions stated in the Lemma. □

Definition 14 (Complementation)**.** For any finite set S and any QPBF $f : \{0,1\}^S \to \mathbb{R}$, the real number $\max\{c'_{\varnothing\varnothing} \mid c' \in C^+_{S2}(f)\}$ is called the **floor dual** of f.

Corollary 2 (of Lemma 6)**.** For any finite set S and any QPBF $f : \{0,1\}^S \to \mathbb{R}$, the floor dual is the solution to the linear optimization problem

$$\begin{aligned}
\max_{c' : J^+_{S2} \to \mathbb{R}} \quad & c_\varnothing - \sum_{j \in S} c'_{\varnothing\{j\}} - \sum_{\{j,k\} \in \binom{S}{2}} c'_{\varnothing\{j,k\}} \\
\text{subject to} \quad & \forall j \in S\colon \quad c_{\{j\}} = c'_{\{j\}\varnothing} - c'_{\varnothing\{j\}} + \sum_{k \in \{1,\ldots,n\} - \{j\}} \left(c'_{\{j\}\{k\}} - c'_{\varnothing\{j,k\}}\right) \\
& \forall \{j,k\} \in \tbinom{S}{2}\colon \quad c_{\{j,k\}} = c'_{\{j,k\}\varnothing} + c'_{\varnothing\{j,k\}} - c'_{\{j\}\{k\}} - c'_{\{k\}\{j\}} \\
& \forall J \in J^+_{S2} - \{(\varnothing,\varnothing)\}\colon \quad 0 \le c'_J \ .
\end{aligned}$$