

Learning max-sum classifier by Perceptron

Vojtěch Franc

February 24, 2020

Classes of polynomially solvable max-sum problems

Learning strictly trivial max-sum problems by Perceptron

XEP33SML – Structured Model Learning, Summer 2020

Max-Sum (Markov-Network) classifier

Setting:

- ◆ $(\mathcal{V}, \mathcal{E})$ is undirected graph; \mathcal{V} are parts and $\mathcal{E} \subseteq \binom{|\mathcal{V}|}{2}$ are related parts
- ◆ $\mathbf{x} = (x_v \in \mathcal{X} \mid v \in \mathcal{V}) \in \mathcal{X}^{\mathcal{V}}$ inputs; $\mathbf{y} = (y_v \in \mathcal{Y} \mid v \in \mathcal{V}) \in \mathcal{Y}^{\mathcal{V}}$ labels
- ◆ $\mathbf{q} = (q_v: \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R} \mid v \in \mathcal{V})$
- ◆ $\mathbf{g} = (g_{vv'}: \mathcal{Y} \times \mathcal{Y} \rightarrow \mathbb{R} \mid \{v, v'\} \in \mathcal{E})$

Max-sum classifier: $h: \mathcal{X}^{\mathcal{V}} \rightarrow \mathcal{Y}^{\mathcal{V}}$ returns labeling

$$h(\mathbf{x}; \mathbf{q}, \mathbf{g}) = \underset{\mathbf{y} \in \mathcal{Y}^{\mathcal{V}}}{\text{Argmax}} f(\mathbf{x}, \mathbf{y}; \mathbf{q}, \mathbf{g}) := \sum_{v \in \mathcal{V}} q_v(x_v, y_v) + \sum_{(v, v') \in \mathcal{E}} g_{vv'}(y_v, y_{v'})$$

Task:

◆ $q_v(x, y) = \langle \mathbf{w}, \phi_v(x, y) \rangle$ and $g_{vv'}(y, y') = \langle \mathbf{w}, \phi_{vv'}(y, y') \rangle$

◆ Given linearly separable training set

$$\mathcal{T}^m = \{(\mathbf{x}^i, \mathbf{y}^i) \in \mathcal{X}^{\mathcal{V}} \times \mathcal{Y}^{\mathcal{V}} \mid i = 1, \dots, m\}$$

find quality functions \mathbf{q}, \mathbf{g} (or \mathbf{w}) such that

$$\mathbf{y}^i = h(\mathbf{x}^i; \mathbf{q}, \mathbf{g}) = \operatorname{argmax}_{\mathbf{y} \in \mathcal{Y}^{\mathcal{V}}} f(\mathbf{x}^i, \mathbf{y}; \mathbf{q}, \mathbf{g}) = \operatorname{argmax}_{\mathbf{y} \in \mathcal{Y}^{\mathcal{V}}} \langle \mathbf{w}, \phi(\mathbf{x}^i, \mathbf{y}) \rangle$$

holds for all $i = 1, \dots, m$.

Tractable classification oracle

- ◆ The Perceptron algorithm requires a classification oracle

$$\mathbf{y}^* = h(\mathbf{x}; \mathbf{q}, \mathbf{g}) = \operatorname{argmax}_{\mathbf{y} \in \mathcal{Y}^{\mathcal{V}}} \left[\sum_{v \in \mathcal{V}} q_v(x_v, y_v) + \sum_{\{v, v'\} \in \mathcal{E}} g_{vv'}(y_v, y_{v'}) \right]$$

- ◆ The max-sum problem $\mathcal{P} = (\mathcal{E}, \mathcal{V}, \mathbf{q}, \mathbf{g}, \mathbf{x})$ is tractable if:

Tractable classification oracle

- ◆ The Perceptron algorithm requires a classification oracle

$$\mathbf{y}^* = h(\mathbf{x}; \mathbf{q}, \mathbf{g}) = \operatorname{argmax}_{\mathbf{y} \in \mathcal{Y}^{\mathcal{V}}} \left[\sum_{v \in \mathcal{V}} q_v(x_v, y_v) + \sum_{\{v, v'\} \in \mathcal{E}} g_{vv'}(y_v, y_{v'}) \right]$$

- ◆ The max-sum problem $\mathcal{P} = (\mathcal{E}, \mathcal{V}, \mathbf{q}, \mathbf{g}, \mathbf{x})$ is tractable if:

(1) $(\mathcal{V}, \mathcal{E})$ is acyclic graph

Tractable classification oracle

- ◆ The Perceptron algorithm requires a classification oracle

$$\mathbf{y}^* = h(\mathbf{x}; \mathbf{q}, \mathbf{g}) = \operatorname{argmax}_{\mathbf{y} \in \mathcal{Y}^{\mathcal{V}}} \left[\sum_{v \in \mathcal{V}} q_v(x_v, y_v) + \sum_{\{v, v'\} \in \mathcal{E}} g_{vv'}(y_v, y_{v'}) \right]$$

- ◆ The max-sum problem $\mathcal{P} = (\mathcal{E}, \mathcal{V}, \mathbf{q}, \mathbf{g}, \mathbf{x})$ is tractable if:

(1) $(\mathcal{V}, \mathcal{E})$ is **acyclic graph**

(2) $g_{vv'}: \mathcal{Y} \times \mathcal{Y} \rightarrow \mathbb{R}$, $\{v, v'\} \in \mathcal{E}$ are **super-modular**:

$$g_{vv'}(y_v, y_{v'}) + g_{vv'}(y'_v, y'_{v'}) \geq g_{vv'}(y_v, y'_{v'}) + g_{vv'}(y'_v, y_{v'})$$

holds for every $y_v > y'_v$ and $y_{v'} > y'_{v'}$.

Tractable classification oracle

- ◆ The Perceptron algorithm requires a classification oracle

$$\mathbf{y}^* = h(\mathbf{x}; \mathbf{q}, \mathbf{g}) = \operatorname{argmax}_{\mathbf{y} \in \mathcal{Y}^{\mathcal{V}}} \left[\sum_{v \in \mathcal{V}} q_v(x_v, y_v) + \sum_{\{v, v'\} \in \mathcal{E}} g_{vv'}(y_v, y_{v'}) \right]$$

- ◆ The max-sum problem $\mathcal{P} = (\mathcal{E}, \mathcal{V}, \mathbf{q}, \mathbf{g}, \mathbf{x})$ is tractable if:

(1) $(\mathcal{V}, \mathcal{E})$ is **acyclic graph**

(2) $g_{vv'}: \mathcal{Y} \times \mathcal{Y} \rightarrow \mathbb{R}$, $\{v, v'\} \in \mathcal{E}$ are **super-modular**:

$$g_{vv'}(y_v, y_{v'}) + g_{vv'}(y'_v, y'_{v'}) \geq g_{vv'}(y_v, y'_{v'}) + g_{vv'}(y'_v, y_{v'})$$

holds for every $y_v > y'_v$ and $y_{v'} > y'_{v'}$.

(3) $\mathcal{P} = (\mathcal{V}, \mathcal{E}, \mathbf{q}, \mathbf{g}, \mathbf{x})$ has a **strictly trivial equivalent**:

The LP relaxation is tight and optimal solution can be recovered easily.

LP relaxation of max-sum problem

- ◆ The max-sum problem

$$\mathbf{y}^* \in \operatorname{argmax}_{\mathbf{y} \in \mathcal{Y}^{\mathcal{V}}} f(\mathbf{x}, \mathbf{y}; \mathbf{q}, \mathbf{g}) := \sum_{v \in \mathcal{V}} q_v(x_v, y_v) + \sum_{\{v, v'\} \in \mathcal{E}} g_{v, v'}(y_v, y_{v'})$$

- ◆ The LP relaxation of the max-sum problem

$$\boldsymbol{\mu}^* = \operatorname{argmax}_{\boldsymbol{\mu}} \left[\sum_{v \in \mathcal{V}} \sum_{y \in \mathcal{Y}} \mu_v(y) q_v(x_v, y) + \sum_{\{v, v'\} \in \mathcal{E}} \sum_{(y, y') \in \mathcal{Y}^2} \mu_{vv'}(y, y') g_{vv'}(y, y') \right]$$

subject to

$$\sum_{y' \in \mathcal{Y}} \mu_{vv'}(y, y') = \mu_v(y), \quad \{v, v'\} \in \mathcal{E}, y \in \mathcal{Y},$$

$$\sum_{y \in \mathcal{Y}} \mu_v(y) = 1, \quad v \in \mathcal{V}$$

$$\boldsymbol{\mu} \geq \mathbf{0}$$

Dual of LP relaxation

- ◆ The dual of the LP relaxation leads to an unconstrained problem

$$\varphi^* = \operatorname{argmin}_{\varphi \in \mathbb{R}^{2|\mathcal{E}||\mathcal{Y}|}} E(\mathbf{x}, \mathbf{q}^\varphi, \mathbf{g}^\varphi) := \sum_{v \in \mathcal{V}} \max_{y \in \mathcal{Y}} q_v^\varphi(x_v, y) + \sum_{\{v, v'\} \in \mathcal{E}} \max_{\substack{y \in \mathcal{Y} \\ y' \in \mathcal{Y}}} g_{vv'}^\varphi(y, y')$$

where φ comprises $\varphi_{vv'}: \mathcal{Y} \rightarrow \mathbb{R}$, $\varphi_{v'v}: \mathcal{Y} \rightarrow \mathbb{R}$, $\{v, v'\} \in \mathcal{E}$ and

$$\begin{aligned} g_{vv'}^\varphi(y, y') &= g_{vv'}(y, y') + \varphi_{vv'}(y) + \varphi_{v'v}(y'), & \{v, v'\} \in \mathcal{E}, y, y' \in \mathcal{Y} \\ q_v^\varphi(y) &= q_v(y) - \sum_{v' \in \mathcal{N}(v)} \varphi_{vv'}(y), & v \in \mathcal{V}, y \in \mathcal{Y} \end{aligned}$$

- ◆ When is the LP relaxation tight, i.e. $E(\mathbf{x}, \mathbf{q}^{\varphi^*}, \mathbf{g}^{\varphi^*}) = f(\mathbf{x}, \mathbf{y}^*, \mathbf{q}, \mathbf{g})$?
- ◆ If the LP relaxation is tight how to get the labels \mathbf{y}^* ?

Interpretation of the dual of LP relaxation

- ◆ Let $P^\varphi = (\mathcal{V}, \mathcal{E}, \mathbf{q}^\varphi, \mathbf{g}^\varphi, \mathbf{x})$ be the max-sum problem constructed from $P = (\mathcal{V}, \mathcal{E}, \mathbf{q}, \mathbf{g}, \mathbf{x})$ by the **re-reparametrization**

$$\begin{aligned}
 g_{vv'}^\varphi(y, y') &= g_{vv'}(y, y') + \varphi_{vv'}(y) + \varphi_{v'v}(y'), & \{v, v'\} \in \mathcal{E}, y, y' \in \mathcal{Y} \\
 q_v^\varphi(y) &= q_v(y) - \sum_{v' \in \mathcal{N}(v)} \varphi_{vv'}(y), & v \in \mathcal{V}, y \in \mathcal{Y}
 \end{aligned}$$

(R)

- ◆ $P = (\mathcal{V}, \mathcal{E}, \mathbf{q}, \mathbf{g}, \mathbf{x})$ and $P' = (\mathcal{V}, \mathcal{E}, \mathbf{q}', \mathbf{g}', \mathbf{x})$ are **equivalent**, i.e. $f(\mathbf{x}, \mathbf{y}, \mathbf{q}, \mathbf{g}) = f(\mathbf{x}, \mathbf{y}, \mathbf{q}', \mathbf{g}')$, $\forall \mathbf{y} \in \mathcal{Y}^\mathcal{V}$, iff they are related by (R).
- ◆ The dual of LP relaxation finds in the class of equivalent problems $\{P^\varphi \mid \varphi \in \mathbb{R}^{2|\mathcal{E}||\mathcal{Y}|}\}$ the one with the minimal energy $E(\mathbf{x}, \mathbf{q}^\varphi, \mathbf{g}^\varphi)$.

Trivial and strictly trivial max-sum problems

Let us define a set $\mathcal{C}_P \subseteq \mathcal{Y}^{\mathcal{V}}$ which contains labelings $\mathbf{y} \in \mathcal{C}_P$ such that

$$\begin{aligned}
 q_v(x_v, y_v) &\geq \max_{y \in \mathcal{Y} \setminus \{y_v\}} q_v(x_v, y), & v \in \mathcal{V} \\
 g_{vv'}(y_v, y_{v'}) &\geq \max_{(y, y') \in \mathcal{Y}^2 \setminus \{y_v, y_{v'}\}} g_{vv'}(y, y'), & \{v, v'\} \in \mathcal{E}
 \end{aligned}
 \tag{Triv}$$

Definition 1. The max-sum problem $P = (\mathcal{V}, \mathcal{E}, \mathbf{q}, \mathbf{g}, \mathbf{x})$ is called *trivial* if $\mathcal{C}_P \neq \emptyset$.

Definition 2. The max-sum problem $P = (\mathcal{V}, \mathcal{E}, \mathbf{q}, \mathbf{g}, \mathbf{x})$ is called *strictly trivial* if it is trivial and all the inequalities (Triv) are satisfied strictly.

Definition 3. The max-sum problem $P = (\mathcal{V}, \mathcal{E}, \mathbf{q}, \mathbf{g}, \mathbf{x})$ has a (strictly) trivial equivalent iff there exist φ such $P = (\mathcal{V}, \mathcal{E}, \mathbf{q}^\varphi, \mathbf{g}^\varphi, \mathbf{x})$ is (strictly) trivial.

Energy is tight \iff trivial maxsum

Theorem 1. For any max-sum problem $P = (\mathcal{V}, \mathcal{E}, \mathbf{q}, \mathbf{g}, \mathbf{x})$ the inequality

$$E(\mathbf{x}, \mathbf{q}, \mathbf{g}) \geq \max_{\mathbf{y} \in \mathcal{Y}^{\mathcal{V}}} f(\mathbf{x}, \mathbf{y}, \mathbf{q}, \mathbf{g})$$

holds true. The bound is tight if and only if P is trivial.

Energy is tight \iff trivial maxsum

Theorem 1. For any max-sum problem $P = (\mathcal{V}, \mathcal{E}, \mathbf{q}, \mathbf{g}, \mathbf{x})$ the inequality

$$E(\mathbf{x}, \mathbf{q}, \mathbf{g}) \geq \max_{\mathbf{y} \in \mathcal{Y}^{\mathcal{V}}} f(\mathbf{x}, \mathbf{y}, \mathbf{q}, \mathbf{g})$$

holds true. The bound is tight if and only if P is trivial.

PROOF:

$$\sum_v \max_{y \in \mathcal{Y}} q_v(x_v, y) + \sum_{v, v'} \max_{\substack{y \in \mathcal{Y} \\ y' \in \mathcal{Y}}} g_{vv'}(y, y') \geq \max_{\mathbf{y} \in \mathcal{Y}^{\mathcal{V}}} \left(\sum_{v \in \mathcal{V}} q_v(x_v, y) + \sum_{v, v'} g_{vv'}(y, y') \right)$$

$$\mathbf{y}^* = \operatorname{argmax}_{\mathbf{y} \in \mathcal{Y}^{\mathcal{V}}} \left(\sum_{v \in \mathcal{V}} q_v(x_v, y) + \sum_{v, v'} g_{vv'}(y, y') \right)$$

$$\max_y q_v(x_v, y) = q_v(x_v, y_v^*), \quad \forall v \in \mathcal{V}$$

$$\max_{y, y'} g_{vv'}(y, y') = g_{vv'}(y_v^*, y_{v'}^*), \quad \forall \{v, v'\} \in \mathcal{E}$$

Solving max-sum problems by the LP relaxation

(1) Solve the dual of LP relaxation

$$\varphi^* = \underset{\varphi}{\operatorname{argmin}} E(\mathbf{x}, \mathbf{q}^\varphi, \mathbf{g}^\varphi)$$

(2) Check the tightness of the LP relaxation by try to find $\mathbf{y} \in \mathcal{C}_P$:

(A) Checking that P^{φ^*} is strictly trivial requires $\mathcal{O}(|\mathcal{V}||\mathcal{Y}| + |\mathcal{E}||\mathcal{Y}|^2)$ operations.

(B) Finding the consistent labeling can be expresses as a **constraint satisfaction problem** which is NP-complete in general.

CSP is an instance of max-sum problem with quality functions (\mathbf{q}, \mathbf{g}) taking only values $\{-\infty, 0\}$.

Learning strictly trivial max-sum classifiers

Task: Given $\{(\mathbf{x}^1, \mathbf{y}^1), \dots, (\mathbf{x}^m, \mathbf{y}^m)\} \in (\mathcal{X}^{\mathcal{V}} \times \mathcal{Y}^{\mathcal{V}})^m$ find (\mathbf{q}, \mathbf{g}) such that $\mathbf{y}^i = h(\mathbf{x}^i; \mathbf{q}, \mathbf{g})$, $i \in \{1, \dots, m\}$, and $P^i = (\mathcal{V}, \mathcal{E}, \mathbf{q}, \mathbf{g}, \mathbf{x}^i)$, $i \in \{1, \dots, m\}$, have a strictly trivial equivalent.

Learning strictly trivial max-sum classifiers

Task: Given $\{(\mathbf{x}^1, \mathbf{y}^1), \dots, (\mathbf{x}^m, \mathbf{y}^m)\} \in (\mathcal{X}^{\mathcal{V}} \times \mathcal{Y}^{\mathcal{V}})^m$ find (\mathbf{q}, \mathbf{g}) such that $\mathbf{y}^i = h(\mathbf{x}^i; \mathbf{q}, \mathbf{g})$, $i \in \{1, \dots, m\}$, and $P^i = (\mathcal{V}, \mathcal{E}, \mathbf{q}, \mathbf{g}, \mathbf{x}^i)$, $i \in \{1, \dots, m\}$, have a strictly trivial equivalent.

If $P = (\mathcal{V}, \mathcal{E}, \mathbf{q}, \mathbf{g}, \mathbf{x})$ has a strictly trivial equivalent and optimal solution is \mathbf{y}^* then there must exist φ such that the re-parametrized quality functions

$$q_v^{\varphi}(y) = q_v(y) - \sum_{v' \in \mathcal{N}(v)} \varphi_{vv'}(y), \quad v \in \mathcal{V}, y \in \mathcal{Y}$$

$$g_{vv'}^{\varphi}(y, y') = g_{vv'}(y, y') + \varphi_{vv'}(y) + \varphi_{v'v}(y'), \quad \{v, v'\} \in \mathcal{E}, y, y' \in \mathcal{Y}$$

satisfy

$$q_v^{\varphi}(x_v, y_v^*) > \max_{y \in \mathcal{Y} \setminus \{y_v^*\}} q_v^{\varphi}(x_v, y), \quad v \in \mathcal{V}$$

$$g_{vv'}^{\varphi}(y_v^*, y_{v'}^*) > \max_{(y, y') \in \mathcal{Y}^2 \setminus \{(y_v^*, y_{v'}^*)\}} g_{vv'}^{\varphi}(y, y'), \quad \{v, v'\} \in \mathcal{E}$$

Learning strictly trivial max-sum classifiers

Learning the max-sum problem with STE is equivalent to solving a set of $m(|\mathcal{V}|(|\mathcal{Y}| - 1) + |\mathcal{E}|(|\mathcal{Y}|^2 - 1))$ strict linear inequalities w.r.t. $\varphi^i \in \mathbb{R}^{2|\mathcal{E}||\mathcal{Y}|}$, $i \in \{1, \dots, m\}$ and $\mathbf{w} \in \mathbb{R}^n$.

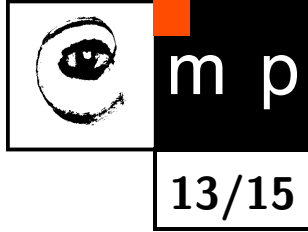
A) For all $i \in \{1, \dots, m\}$, $v \in \mathcal{V}$, $y \in \mathcal{Y} \setminus \{y_v^i\}$ satisfy

$$\langle \mathbf{w}, \phi_v(x_v^i, y_v^i) \rangle - \sum_{v' \in \mathcal{N}(v)} \varphi_{vv'}^i(y_v^i) > \langle \mathbf{w}, \phi_v(x_v^i, y) \rangle - \sum_{v' \in \mathcal{N}(v)} \varphi_{vv'}^i(y)$$

B) For all $i \in \{1, \dots, m\}$, $\{v, v'\} \in \mathcal{E}$, $(y, y') \in \mathcal{Y}^2 \setminus \{(y_v^i, y_{v'}^i)\}$ satisfy

$$\langle \mathbf{w}, \phi_{vv'}(y_v^i, y_{v'}^i) \rangle + \varphi_{vv'}^i(y_v^i) + \varphi_{v'v}^i(y_{v'}^i) > \langle \mathbf{w}, \phi_{vv'}(y, y') \rangle + \varphi_{vv'}^i(y) + \varphi_{v'v}^i(y')$$

Strictly trivial cover acyclic and super-modular problems



Theorem 2. *Let $P = (\mathcal{V}, \mathcal{E}, q, g, x)$ be a max-sum problem and let P have a unique solution. If $(\mathcal{V}, \mathcal{E})$ is an acyclic graph or quality functions g are super-modular then P is equivalent to some strictly trivial problem.*

Example: Sudoku solver

puzzle assignment

					8			
	1	9	5	6		2		
2	5			1		3	6	
9					2		8	1
	8	2	6		9			
5	7		1					2
	2	1		9			4	3
		5		7	6	8		
8	9		3					

solution

7	6	3	4	2	8	1	9	5
4	1	9	5	6	3	2	7	8
2	5	8	9	1	7	3	6	4
9	3	4	7	5	2	6	8	1
1	8	2	6	3	9	4	5	7
5	7	6	1	8	4	9	3	2
6	2	1	8	9	5	7	4	3
3	4	5	2	7	6	8	1	9
8	9	7	3	4	1	5	2	6

The task of Sudoku game is to fill empty fields such that each row, each column and each 3×3 field contains numbers $\{1, 2, \dots, 9\}$.

Example: Sudoku solver

$$\mathbf{y}^* = \operatorname{argmax}_{\mathbf{y} \in \mathcal{Y}^{\mathcal{V}}} \left(\underbrace{\sum_{v \in \mathcal{V}} q(x_v, y_v)}_{\text{copy given fields}} + \underbrace{\sum_{\{v, v'\} \in \mathcal{E}} g(y_v, y_{v'})}_{\text{neighbors must be different}} \right)$$

- ◆ $\mathcal{V} = \{(i, j) \in \mathbb{N}^2 \mid 1 \leq i \leq 9, 1 \leq j \leq 9\}$
- ◆ $\mathbf{x} = (x_v \in \{\square, 1, \dots, 9\} \mid v \in \mathcal{V}) \in \mathcal{X}^{\mathcal{V}}$
- ◆ $\mathbf{y} = (y_v \in \{1, \dots, 9\} \mid v \in \mathcal{V}) \in \mathcal{Y}^{\mathcal{V}}$
- ◆ $\mathcal{E} = \{\{(i, j), (i', j')\} \mid i = i' \vee j = j' \vee (\lceil i/3 \rceil = \lceil i'/3 \rceil \wedge \lceil j/3 \rceil = \lceil j'/3 \rceil)\}$
- ◆ $q: \{\square, 1, \dots, 9\} \times \{1, \dots, 9\} \rightarrow \{0, -\infty\}$ such that

$$q(x, y) = \begin{cases} -\infty & \text{if } x \neq \square \wedge y \neq x \\ 0 & \text{otherwise} \end{cases}$$

- ◆ $g: \{1, \dots, 9\}^2 \rightarrow \{0, -\infty\}$ such that $g(y, y') = \begin{cases} 0 & \text{if } y \neq y' \\ -\infty & \text{if } y = y' \end{cases}$

					8			
	1	9	5	6		2		
2	5			1		3	6	
9					2		8	1
	8	2	6		9			
5	7		1					2
	2	1		9			4	3
		5		7	6	8		
8	9		3					

7	6	3	4	2	8	1	9	5
4	1	9	5	6	3	2	7	8
2	5	8	9	1	7	3	6	4
9	3	4	7	5	2	6	8	1
1	8	2	6	3	9	4	5	7
5	7	6	1	8	4	9	3	2
6	2	1	8	9	5	7	4	3
3	4	5	2	7	6	8	1	9
8	9	7	3	4	1	5	2	6