Learning max-sum classifier by Perceptron

Vojtěch Franc

March 19, 2019

Classes of polynomialy solvable max-sum problems

Learning strictly trivial max-sum problems by Perceptron

XEP33SML – Structured Model Learning, Summer 2019

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
- $\boldsymbol{x} = (x_v \in \mathcal{X} \mid v \in \mathcal{V}) \in \mathcal{X}^{\mathcal{V}}$ inputs; $\boldsymbol{y} = (y_v \in \mathcal{Y} \mid v \in \mathcal{V}) \in \mathcal{Y}^{\mathcal{V}}$ labels
- $\boldsymbol{q} = (g_v \colon \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \mid v \in \mathcal{V})$
- $\boldsymbol{g} = (g_{vv'} \colon \mathcal{Y} \times \mathcal{Y} \to \mathbb{R} \mid \{v, v'\} \in \mathcal{E})$

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

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

Learning linear max-sum classifier from separable examples

Task:

- $q_v(x,y) = \langle \boldsymbol{w}, \boldsymbol{\phi}_v(x,y) \rangle$ and $g_{vv'}(y,y) = \langle \boldsymbol{w}, \boldsymbol{\phi}_{vv'}(y,y') \rangle$

- Given linearly separable training set

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

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

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

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

Tractable classification oracle

- The Perceptron algorithm requires a classification oracle

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

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

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

(2) $g_{vv'}\colon \mathcal{Y}\times\mathcal{Y}\to\mathbb{R}$, $\{v,v'\}\in\mathcal{E}$ 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) $\mathcal{P} = (\mathcal{V},\mathcal{E},\boldsymbol{q},\boldsymbol{g},\boldsymbol{x})$ has a strictly trivial equivalent:
The LP relaxation is tight and $\mathcal{P}$ has a unique solution.

LP relaxation of max-sum problem

- The max-sum problem

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

- The LP relaxation of the max-sum problem

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

subject to

$$\begin{aligned} \sum_{y'\in\mathcal{Y}} \mu_{vv'}(y,y') &= \mu_v(y), \quad \{v,v'\}\in\mathcal{E}, y\in\mathcal{Y}, \\ \sum_{y\in\mathcal{Y}} \mu_v(y) &= 1, \qquad v\in\mathcal{V} \\ \boldsymbol{\mu} &\geq \mathbf{0} \end{aligned}$$

Dual of LP relaxation

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

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

where $\boldsymbol{\varphi}$ comprises $\varphi_{vv'}\colon\mathcal{Y}\to\mathbb{R}$, $\varphi_{v'v}\colon\mathcal{Y}\to\mathbb{R}$, $\{v,v'\}\in\mathcal{E}$ and

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

- When is the LP relaxation tight, i.e. $E(\boldsymbol{x},\boldsymbol{q}^{\boldsymbol{\varphi}^*},\boldsymbol{g}^{\boldsymbol{\varphi}^*}) = f(\boldsymbol{x},\boldsymbol{y}^*,\boldsymbol{q},\boldsymbol{g})$?

- If the LP relaxation is tight how to get the labels $\boldsymbol{y}^*$?

Interpretation of the dual of LP relaxation

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

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

- $P = (\mathcal{V}, \mathcal{E}, \boldsymbol{q}, \boldsymbol{g}, \boldsymbol{x})$ and $P = (\mathcal{V}, \mathcal{E}, \boldsymbol{q}', \boldsymbol{g}', \boldsymbol{x})$ are equivalent, i.e. $f(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{q}, \boldsymbol{g}) = f(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{q}', \boldsymbol{g}')$, $\forall \boldsymbol{y} \in \mathcal{Y}^{\mathcal{V}}$, iff they are related by (R).

- The dual of LP relaxation finds in the class of equivalent problems $\{P^{\varphi} \mid \boldsymbol{\varphi} \in \mathbb{R}^{2|\mathcal{E}||\mathcal{Y}|}\}$ the one with the minimal energy $E(\boldsymbol{x}, \boldsymbol{q}^{\varphi}, \boldsymbol{g}^{\varphi})$.

Trivial and strictly trivial max-sum problems

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

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

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

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

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

Energy is tight $\iff$ trivial

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

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

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

PROOF:

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

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

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

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

■

Solving max-sum problems by the LP relaxation

(1) Solve the dual of LP relaxation

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

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

(A) Checking that $P^{\boldsymbol{\varphi}^*}$ 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 $(\boldsymbol{q}, \boldsymbol{g})$ taking only values $\{-\infty, 0\}$.

Learning strictly trivial max-sum classifiers

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

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

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

satisfy

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

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. $\boldsymbol{\varphi}^i \in \mathbb{R}^{2|\mathcal{E}||\mathcal{Y}|}$, $i \in \{1,\ldots,m\}$ and $\boldsymbol{w} \in \mathbb{R}^n$.

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

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

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

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

Strictly trivial cover acyclic and super-modular problems

Theorem 2. *Let* $P = (\mathcal{V}, \mathcal{E}, \boldsymbol{q}, \boldsymbol{g}, \boldsymbol{x})$ *be a max-sum problem and let* P *have a unique solution. If* $(\mathcal{V}, \mathcal{E})$ *is an acyclic graph or quality functions* $\boldsymbol{g}$ *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 $\{1, 2, \ldots, 9\}$.

Example: Sudoku solver

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

- $\mathcal{V} = \{(i,j) \in \mathbb{N}^2 \mid 1 \le i \le 9, 1 \le j \le 9\}$
- $\boldsymbol{x} = (x_v \in \{\square, 1, \ldots, 9\} \mid v \in \mathcal{V}) \in \mathcal{X}^{\mathcal{V}}$
- $\boldsymbol{y} = (y_v \in \{1, \ldots, 9\} \mid v \in \mathcal{V}) \in \mathcal{Y}^{\mathcal{V}}$
- $\mathcal{E} = \{\{(i,j),(i',j')\} \mid i = i' \vee j = j' \vee (\lceil i/3 \rceil = \lceil i'/3 \rceil \wedge \lceil j/3 \rceil = \lceil j'/3 \rceil)\}$
- $q\colon \{\square, 1, \ldots, 9\} \times \{1, \ldots, 9\} \to \{0, -\infty\}$ such that

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

- $g\colon \{1, \ldots, 9\}^2 \to \{0, -\infty\}$ 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