Lecture 13 - M3-nets

Tuesday, February 19, 2019 14:29

today: M3-net : M vs. L polytopes

<u>marginal polytope (M) formulation</u>:

$$\langle \alpha, s \rangle = \sum_{\tilde{y} \in \mathcal{Y}} \alpha(\tilde{y}) \left(\sum_{c} s_c(\tilde{y}_c) \right)$$

$$= \sum_{c \in \mathcal{C}} \sum_{\tilde{y}_c \in \mathcal{Y}_c} s_c(\tilde{y}_c) \underbrace{\left[\sum_{\substack{y' \in \mathcal{Y} \\ y'_c = \tilde{y}_c}} \alpha(y') \right]}_{\triangleq\, \mu_c(\tilde{y}_c)}$$

"clique marginal variable"

$$\mu = (\mu_c)_{c \in \mathcal{C}} \qquad \mu_c \in \Delta_{|\mathcal{Y}_c|}$$

$$\mu = A_M \alpha$$

A_M marginalization linear operator

$$\max_{\alpha \in \Delta_{|\mathcal{Y}|}} \langle \alpha, s \rangle = \max_{\mu \in A_M \Delta_{|\mathcal{Y}|}} \sum_{c \in \mathcal{C}} \underbrace{\sum_{\tilde{y}_c \in \mathcal{Y}_c} s_c(\tilde{y}_c)\, \mu_c(\tilde{y}_c)}_{\langle \mu_c, s_c \rangle}$$

$$M \triangleq A_M \Delta$$

$$M \triangleq \{ (\mu_c)_{c \in \mathcal{C}} : \exists \alpha \in \Delta_{|\mathcal{Y}|} \text{ s.t. } \mu_c(\tilde{y}_c) = \sum \alpha(y') \ \forall c \}$$

$\mathcal{M} \triangleq A_{\mathcal{M}} \Delta$

"marginal polytope"

$$\mathcal{M} \triangleq \Big\{ (\mu_c)_{c\in\mathcal{C}} : \exists \alpha \in \Delta_{|\mathcal{Y}|} \text{ s.t. } \mu_c(\tilde{y}_c) = \sum_{\substack{y'\in\mathcal{Y} \\ y'_c = \tilde{y}_c}} \alpha(y') \ \forall c \Big\}$$

$$\mathcal{M} \subseteq \bigtimes_{c\in\mathcal{C}} \Delta_{|\mathcal{Y}_c|}$$

<u>also</u>: $\mathcal{M} \subseteq \mathbb{L} \triangleq \{ (\mu_c)_{c\in\mathcal{C}} : \mu_c \in \Delta_{|\mathcal{Y}_c|}$ and μ_c's are "locally consistent" $\}$

"local consistency polytope"

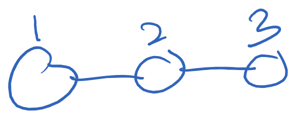
i.e. $\forall c, c' \in \mathcal{C}$

$\forall S \subseteq c \cap c'$

$\forall y_S$, $\displaystyle\sum_{\substack{y'_c\in\mathcal{Y}_c \\ \text{s.t. } y'_S = y_S}} \mu_c(y'_c) = \sum_{\substack{y''_{c'}\in\mathcal{Y}_{c'} \\ \text{s.t. } y''_S = y_S}} \mu_{c'}(y''_{c'})$

$=$ "$\mu_S(y_S)$"

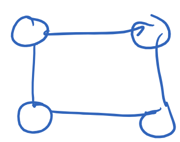
i.e.



$C_1 = \{1,2\}$ $C_2 = \{2,3\}$

(*) $\mathbb{L}$ can be seen as a <u>relaxation</u> of $\mathcal{M}$

if G is <u>not triangulated</u>, then we can have $\mathcal{M} \neq \mathbb{L}$



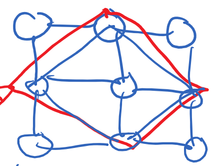
(*) key property: if G is <u>triangulated</u> then $\mathcal{M} = \mathbb{L}$?? (**)

implication here $\max \sum \langle s_c + l_c, \mu_c \rangle = \max \sum \langle s_c + l_c, \mu_c \rangle$ [assuming ...

implication here $\max_{\mu \in \mathcal{M}} \sum_c \langle s_c + \ell_c, \mu_c \rangle = \max_{\mu \in \mathcal{L}} \underbrace{\sum_c \langle s_c + \ell_c, \mu_c \rangle}_{g(w,z)}$ [assuming $\ell(y,y') = \sum_c \ell_c(y_c, y_c')$]

aside: triangulated graph review:

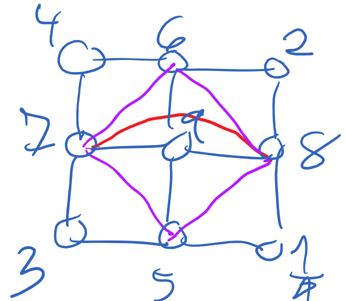
Δ-graph → iff it has no cycle of size 4 and more which is not chordal



not chordal

→ not triangulated

(*) you can triangulate a graph by running variable elimination



$\{6,7,8,9\}$ is biggest clique

→ treewidth ≤ 3

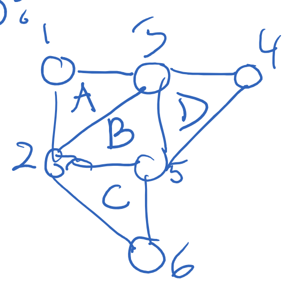
elimination ordering

(*) if G is triangulated → can construct a junction tree

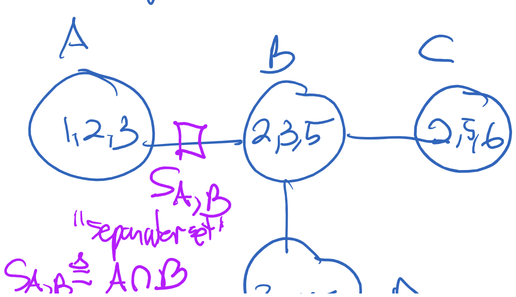
↳ clique tree & tree on the maximal cliques of G with the running intersection property, ie. every $C \in \mathcal{C}$ is a node of the J.T.

↳ if $v \in C_1 \cap C_2$ then v belongs to all cliques along the unique path from C_1 to C_2 in J.T.

G:



ex. of J.T. for G:



$S_{A,B}$

"separator set"

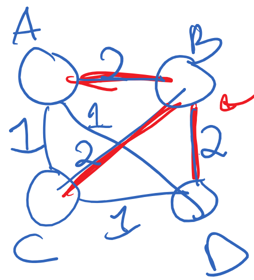
$S_{A,B} \triangleq A \cap B$

$\cup 6$

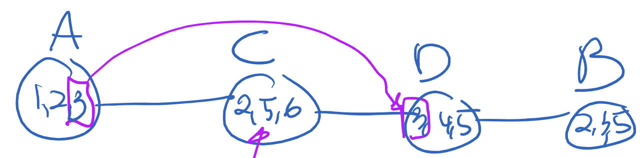
$S_{A,B} \triangleq A \cap B$
$= \{2,3\}$

$3,4,5$ D

* to build a J.T.: first build weighted complete graph on cliques where weight on edges is $|C_i \cap C_j|$



use maximal weight spanning tree to get J.T.



no 3 here ⇒ no good? ⇒ not a J.T.

15h28

$M = L$ for Δ-graphs

$M \subseteq L$

$\underline{L \subseteq M}$: $\displaystyle q(y) = \frac{\prod_{C \in \mathcal{C}} \mu_C(y_C)}{\prod_{S \in \mathcal{S}'} \mu_S(y_S)}$

maximal cliques

separator sets from a J.T.

Suppose $\mu \in L$; then we show μ is valid marginal from some joint q

we'll show that q is a valid joint

i.e. $\mu = A_M q$

when $\mathcal{C}$ are maximal cliques

and $\mathcal{S}'$ are separators from a J.T.

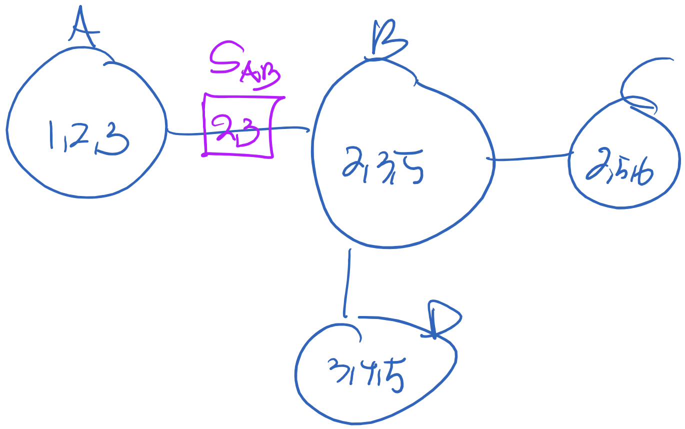
(this is generalization that for tree UGM $p(x) = \prod_{i,j} p(x_i, x_j) \prod_i p(x_i)$)

let's say want to show

(that for tree UGM $p(x) = \prod_{\{i,j\}} \frac{p(x_i, x_j)}{p(x_i)p(x_j)} \prod_{i \in V} p(x_i)$)

let's say want to show

$$\mu_B(y_B) = \sum_{y' : y'_B = y_B} \alpha(y')$$



$$\sum_{y' : y'_B = y_B} \left(\frac{\prod_C \mu_C(y_C)}{\prod_S \mu_S(y_S)} \right)$$

$$\cancel{\frac{\sum_{y_1} \mu_A(y_A)}{\mu_{S_{A,B}}(y_{S_{A,B}})}} \sum_{\text{rest}} \frac{\prod_{\substack{C \in \mathcal{C} \\ C \neq A}} \mu_C(y_C)}{\prod_{S \neq S_{AB}} \mu_S(y_S)}$$

there is no y_1 here, since $y_1 \notin S_{A,B}$ crucially using run. intersec. property

now, $\sum_{y_1} \mu_A(y_A) = \mu_{\{2,3\}}(y_{\{2,3\}})$ [by local consistency]

$= \mu_{S_{A,B}}(y_{S_{A,B}})$

keep plucking the leaves until you are left with $\mu_B(y_B)$ //

☆ why do we care to have shown $M = L$ for Δ-graph?

→ because describing L is simple i.e. Poly($\max_C |\mathcal{Y}_C|$) # of constraints

vs. M which needs exponential # " " in worst case

[<u>intuition</u>: triangulating a general graph yields big cliques C_{max} and thus $|\mathcal{Y}_{C_{max}}|$ is exponential in $|C_{max}|$]

one way to see source of exp. # of constraints

is Fourier-Motzkin elimination
(to project polytopes on lower dim. space)

$$a_i^T x \le z \quad \text{for } i \in I$$

$$a_j^T x \ge z \quad \text{for } j \in J$$

↳ for z feasible to exist

$$\Rightarrow a_i^T x \le a_j^T x \quad \forall (i,j) \in I \times J$$

<u>to get back to M^3net:</u>

when $M = L$ (i.e. Δ-graph)

$$\max_{y \in \mathcal{Y}} s(y) = \max_{\alpha \in \Delta_{|\mathcal{Y}|}} \langle \alpha, s \rangle = \max_{\mu \in \mathcal{M}} \sum_c \langle \mu_c, s_c \rangle \overset{?}{=} \max_{\mu \in \mathcal{L}} \sum_c \langle \mu_c, s_c \rangle$$

all the vertices have integer component (i.e. $\{0,1\}$)

i.e. $\mu^* = \arg\max_{\mu \in \mathcal{L}} \sum_c \langle \mu_c, s_c \rangle$

will have $\mu_c^*(\cdot) = \delta_{y_c^*}(\cdot)$

<u>note:</u> could use <u>junction algorithm</u> [max form] (generalization of Viterbi) to solve this LP (much more efficient than interior point method on $\mathcal{L}$)

overall, suppose that $\ell_i(y)$ decomposes also on G } $g(w, z)$

$$\max_{\tilde{y}\in\mathcal{Y}_i} \ell_i(\tilde{y}) - w^T\psi_i(\tilde{y}) = \max_{\mu\in L_i} a_i + (c_i + F_i^T w)^T \mu - w^T F_i \mu^{(i)} \leftarrow$$

(integer) marginal from ground truth $y^{(i)}$

last class: A) dualize this LP and plug it in structured SVM obj. to get a small "complicated" QP

use Mosek, Cplex, etc. to solve to 1e-16 accuracy using IPM e.g.

B) saddle pt. formulation $\min_{w} \max_{\mu_i\in L_i} f(w,\mu)$

extragradient → would require projection on L_i (local consistency polytope) which is more expensive than LP

[to check: can you use message passing for that?]

C) later: saddle pt. FW alg. which only requires LMOs "linear minimization oracles" on L_is (instead of projection)

next class: Lagrangian dual of SVMstruct