Lecture 13 - M3-nets

Thursday, February 25, 2021 13:30

today: M^3net trick: M vs. L polytope

II) M^3-net example (CRF score)

* graph $G=(V,E)$ $\qquad y_p,\ p\in V \qquad y_C \triangleq (y_p)_{p\in C}$

set of cliques $\mathcal{C}$

MRF model for score

$$p(y|x;w) = \frac{1}{Z(x;w)} \prod_{C\in\mathcal{C}} \psi_C(y_C; x, w)$$

$$= \frac{1}{Z} \exp\Big(\underbrace{\sum_{C\in\mathcal{C}} \underbrace{\log \psi_C(y_C; x,w)}_{\hookrightarrow\ w^T\varphi_C(y_C;x)}}_{S(x,y;w)}\Big)$$

$$S(x,y;w) = \sum_{C\in\mathcal{C}} S_C(y_C; x,w)$$

example: OCR

chain graph on y

inference:

$$\max_{\tilde y\in\mathcal{Y}} \sum_{C\in\mathcal{C}} w^T\varphi_C(y_C;x)$$

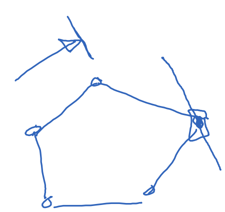
goal: rewrite as a LP

* fix x & w; let s be a vector of size $|\mathcal{Y}(x)|$ of scores

$$\max_{\tilde y\in\mathcal{Y}(x)} s(x,\tilde y;w) = \max_k s_k = \max_{\{e_k\}} \langle e_k, s\rangle \qquad e_k = \begin{pmatrix} 0\\0\\0\\1\\0\\\vdots\end{pmatrix} \leftarrow k^{th} \text{ position}$$

trick! (tight convex opt. relaxation)

$$= \max_{\alpha\in \underbrace{\mathrm{conv}(\{e_k\})}_{\triangleq \Delta_{|\mathcal{Y}|}}} \langle \alpha, s\rangle$$



note: any comb. opt. problem can be written (trivially) as a convex opt. (LP) [but with exp # of dimensions]

✪ insight of M^3-net paper → use MRF structure to transform $\displaystyle\max_{\alpha\in\Delta_{|\mathcal{Y}|}} \langle\alpha,s\rangle = \max_{\underset{\text{tractable}}{\mu\in L}} \sum_C \langle \underset{\text{reasonable size}}{\mu_C}, s_C\rangle$

marginal polytope (M) formulation

$$\langle\alpha,s\rangle = \sum_{\tilde y\in\mathcal{Y}} \alpha(\tilde y) \underbrace{\Big(\sum_{C\in\mathcal{C}} s_C(\tilde y_C)\Big)}_{s(\tilde y)} = \sum_C \sum_{\tilde y\in\mathcal{Y}} \alpha(\tilde y) s_C(\tilde y_C)$$

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

"clique marginal variable"

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

$$\mu = A_M \alpha \qquad \text{ie.} \rightarrow \mu_c(\tilde{y}_c) = \sum_{\substack{y' \in \mathcal{Y} \\ y'_c = \tilde{y}_c}} \alpha(y') = (A_M)_{c,\tilde{y}_c} \mu$$

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

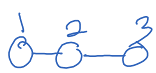
$$\boxed{M \triangleq A_M \Delta_{|\mathcal{Y}|} \qquad M \triangleq \left\{ (\mu_c)_{c \in \mathcal{C}} : \exists\, \alpha \in \Delta_{|\mathcal{Y}|} \text{ s.t. } \mu_c(\tilde{y}_c) = \sum_{\substack{y' \in \mathcal{Y} \\ y'_c = \tilde{y}_c}} \alpha(y') \ \forall c \ \forall \tilde{y}_c \right\}}$$

marginal polytope

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

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

"local consistency polytope"

ie. 

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

$\mu_{\{1,2\}}$ $\mu_{\{2,3\}}$

$\mu_{12}(y_1, y_2)$ $\mu_{2,3}(y_2, y_3)$

$$\text{"}\mu_2\text{"}(y_2) = \sum_{y_1} \mu_{12}(y_1, y_2)$$

$$\text{"}\mu_2\text{"}(y_2) = \sum_{y_3} \mu_{2,3}(y_2, y_3)$$

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

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

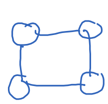
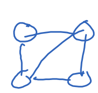
$\forall \tilde{y}_S$

$$\sum_{\substack{y'_c \in \mathcal{Y}_c \\ y'_S = \tilde{y}_S}} \mu_c(y'_c) = \sum_{\substack{y''_{c'} \in \mathcal{Y}_{c'} \\ y''_S = \tilde{y}_S}} \mu_{c'}(y''_{c'})$$

$$= \text{"}\mu_S(\tilde{y}_S)\text{"}$$

(*) L can be seen as a <u>relaxation</u> of M

if G is <u>not triangulated</u>, then we can have that $M \subsetneq L$

 e.g. → 

(*) key property: if G is triangulated, then $M = L$?? (*)(*)

implication here $\max_{\mu \in M} \underbrace{\sum_c \langle s_c + \ell_c, \mu_c \rangle}_{g(w;z)} = \max_{\mu \in L}$ ———

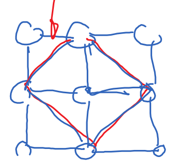
(assuming $\ell(y, y') = \sum_c \ell_c(y_c, y'_c)$]

14h26

aside: triangulated graph review:

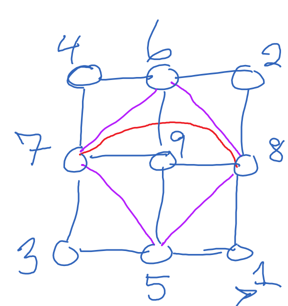
Δ-graph → "iff it has no cycle of size 4 or more which are not chordal"

not chordal



⇒ no triangulated

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



$\{6,7,8,9\}$ is biggest clique ⇒ treewidth ≤ 3

elim ordering

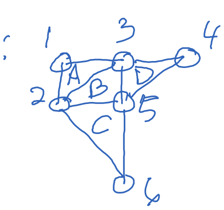
(*) if G is triangulated ⇒ can construct a junction tree

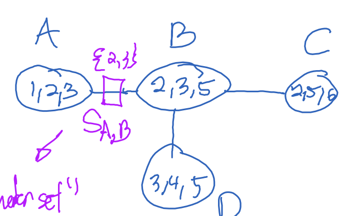
↳ clique tree with the running intersection property : tree with nodes being the maximal cliques of G

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

ex. of J.T. for G

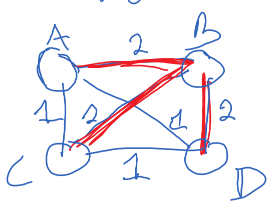
G:



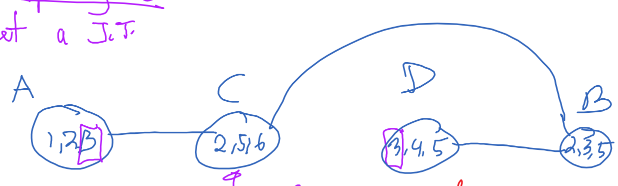


"separator set" $S_{A,B} \triangleq A \cap B$

(*) to build a J.T.: just build weighted complete graph on cliques where weight on edge $\{i,j\}$ is $|C_i \cap C_j|$



use maximum weight spanning tree to get a J.T.



no 3 ⇒ no good ⇒ no

no 3 => no good => no a J.T.

let's prove

$M = L$ for Δ-graphs

- $M \subseteq L$
- $L \subseteq M$:

$$\alpha(y) = \frac{\prod_{c \in \mathcal{C}} \mu_c(y_c)}{\prod_{S \in \mathcal{S}'} \mu_S(y_S)}$$

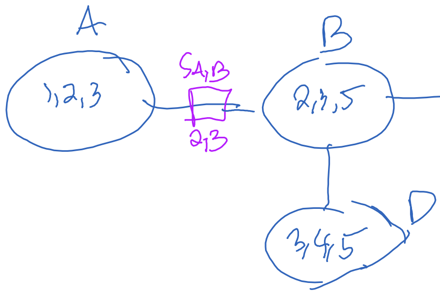
separator set from J.T.

suppose $\mu \in L$;

we want $\exists \alpha$ s.t. μ is marg. of α

i.e. $\mu = A_M \alpha$

(this is a gen. for the tree UGM $\quad p(x) = \prod_{\{i,j\} \in E} \frac{p(x_i, x_j)}{p(x_i)p(x_j)} \prod_{i \in V} p(x_i)$)



let's say want to show

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

$$\sum_{\substack{y' : \\ y'_B = y_B}} \left(\frac{\prod_c \mu_c(y'_c)}{\prod_S \mu_S(y'_S)} \right)$$

$$\sum_{y_{rest}} \left(\frac{\cancel{\boxed{\sum_{y_1} \mu_A(y_A)}}}{\cancel{\mu_{S_{A,B}}(y_{S_{A,B}})}} \right) (\text{rest})$$

by local consistency property

there is no y_1 here since $y_1 \notin S_{A,B}$ crucially using run. intersection 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 $\mu_B(y_B)$

to get back to M^3-net:

when $M = L$ (i.e. Δ-graph)

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

all vertices have integer

note: could use junction tree alg [max form] (generalization of Viterbi)

$(1, 1, 0)$

$$c = \{1,2\} \sim \begin{array}{c} 11 \\ 10 \\ 01 \\ 00 \end{array} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

$(1,1,0) \longrightarrow \begin{array}{cc} 10 & \\ 01 & \\ 00 & \end{array}\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$

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

i.e. $\mu^* = \underset{\mu \in L}{\text{argmax}} \sum_c \langle \mu_c, s_c \rangle$

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

[max form]

(generalization of Viterbi) to solve this LP much more efficiently (than interior point method)

Overall, suppose that $l_i(\tilde{y})$ decomposes also on G

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

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

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

use Mosek, Cplex, etc. to solve $1e^{-16}$ accuracy using I.P.M. e.g. ↳ (see next class)

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

extragradient → would require <u>projection</u> on L_i which is more expensive than $\underbrace{\text{LP on } L_i}_{\text{can be done using J.T. alg.?}}$

C) later: saddle pt. <u>FW</u> alg. which only requires LMO "linear min. oracle" on L_i's instead of projection

<u>next class</u>: Lagrangian dual for SVM struct