Lecture 13 - M3-nets

Tuesday, March 7, 2023 2:27 PM

today: M^3-net trick: $\mathcal{M}$ vs. $\mathcal{L}$ polytope

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

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

set of cliques $\mathcal{C}$
MRF model for score

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

$$= \frac{1}{Z} \exp\Big(\underbrace{\sum_{C\in\mathcal{C}} \log \psi_C(y_C; x; w)}\Big)$$

$\hookrightarrow w^T \varphi_C(y_C; x)$

$$s(x,y;w) \overset{?}{=} \sum_{C\in\mathcal{C}} s_C(y_C; x, w)$$

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 compact LP

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

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

trick ?!
(tight convex cts. 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

$$\max_{\alpha\in\Delta_{|\mathcal{Y}|}} \langle\alpha, s\rangle = \max_{\mu\in\mathcal{L}} \sum_C \langle \mu_C, s_C\rangle$$

$\mathcal{L}$: tractable; μ_C: reasonable size

marginal polytope ($\mathcal{M}$) formulation

$$\langle\alpha, s\rangle = \alpha(\tilde{y})\Big(\underbrace{\sum s_C(\tilde{y}_C)}_{s(\tilde{y})}\Big) = \sum \sum \alpha(\tilde{y}) s_C(\tilde{y}_C)$$

marginal polytope ($\mathcal{M}$) formulation

$$\langle \alpha, s\rangle = \sum_{\tilde y\in\mathcal{Y}}\alpha(\tilde y)\Big(\overbrace{\sum_{c\in\mathcal{C}} s_c(\tilde y_c)}^{s(\tilde y)}\Big) = \sum_{c\in\mathcal{C}}\sum_{\tilde y\in\mathcal{Y}}\alpha(\tilde y)\, s_c(\tilde y_c)$$

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

'clique marginal variable'

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

$\mu = A_M\,\alpha$ $\quad$ i.e. $\mu_c(\tilde y_c)=\sum_{\substack{y'\in\mathcal{Y}\\ y'_c=y_c}}\alpha(y') = (A_M)_{(c,\tilde y_c),:}\,\alpha$

A_M: $\left(\sum_c |\mathcal{Y}_c|\right)\times|\mathcal{Y}|$, $\quad |\mathcal{Y}_c|$ e.g. $2^{|c|}$

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

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

marginal polytope

$$\mathcal{M}\subseteq \mathop{\times}_{c\in\mathcal{C}}\Delta_{|\mathcal{Y}_c|}$$

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

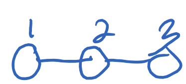
"local consistency polytope"

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

$\forall s\subseteq c\cap c'$

$\forall\tilde y_s$

$$\sum_{\substack{y'_c\in\mathcal{Y}_c\\ y'_s=\tilde y_s}}\mu_c(y'_c) = \text{"}\mu_s(\tilde y_s)\text{"} = \sum_{\substack{y''_{c'}\in\mathcal{Y}_{c'}\\ y''_s=\tilde y_s}}\mu_{c'}(y''_{c'})$$

i.e. 

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

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

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

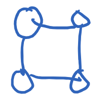
$$||$$

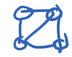
$$\text{"}\mu_2\text{"}(y_2)=\sum_{y_3}\mu_{23}(y_2,y_3)$$

(*) $\mathcal{L}$ can be seen as a relaxation of $\mathcal{M}$

⊛ L can be seen as a relaxation of M

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



e.g. → 

⊛ key property: if G is triangulated, then $M = L$?? ⊛⊛

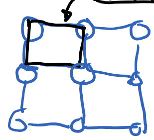
implication here $\max\limits_{\mu \in M} \underbrace{\sum_C \langle s_C + l_C, \mu_C \rangle}_{g(w;\, z, \mu)} = \max\limits_{\mu \in L}$ ———

[assuming $\ell(y, y') = \sum_C \ell_C(y_C, y'_C)$]

15h25

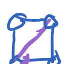
aside: triangulated graph review

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



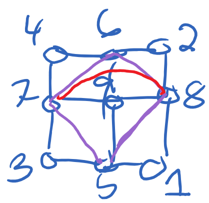
not chordal

⇒ no triangulated



chordal cycle

⊛ you can triangulate a graph by running variable elimination



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

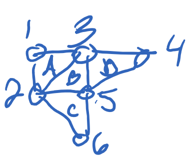
elim. ordering

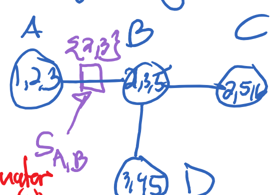
⊛ if G is triangulated ⇒ can construct a junction tree

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

↳ 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.

ex. of J.T. for G

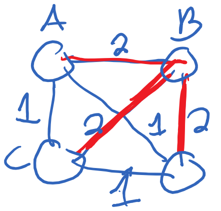




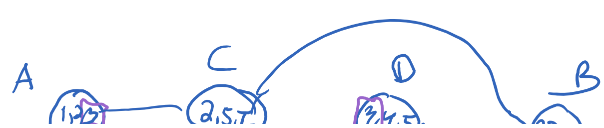
$S_{A,B}$ "separator set"

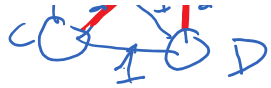
$S_{A,B} \triangleq A \cap B$

⊛ to build a J.T.: first build weighted complete graph on cliques where weights on edge $\{i, j\}$ is $|C_i \cap C_j|$



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





$\Rightarrow$ no good $\Rightarrow$ not a J.T.

let's prove that $M = L$ for Δ-graphs

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

$$\boxed{\alpha(y) = \frac{\prod_{C \in \mathcal{G}} \mu_C(y_C)}{\prod_{S \in S'} \mu_S(y_S)}}$$

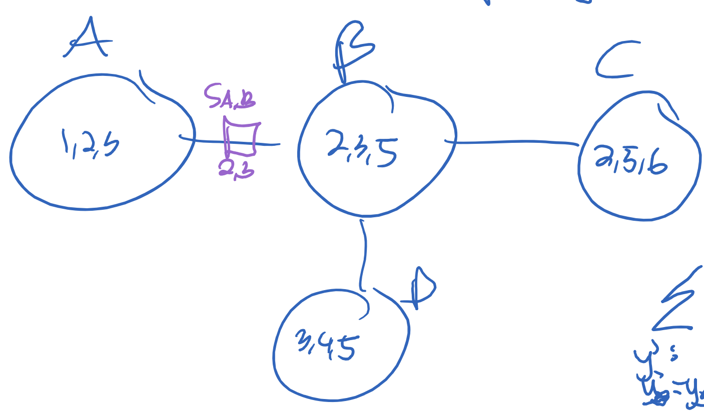
S': separator set from J.T.

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

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

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

there is no $\underline{y}$ here since $y_1 \notin S_{A,B}$ and then using run. intersection property

by local consistency property

$$\sum_{y_1} \mu_A(y_A) = \mu_{\{2,3\}}(y_{\{2,3\}}) \quad \text{(by local consistency)}$$
$$= \mu_{S_{A,B}}(y_{S_{A,B}})$$

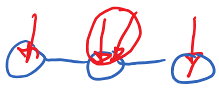
keep plucking the leaves until you are left with $\mu_B(y_B)$ //

to get back to M^3-net:

when $M = L$ (i.e. Δ-graph)

$$\max_{\tilde{y} \in \mathcal{Y}} s(\tilde{y}) = \max \quad \langle \alpha, s \rangle = \max \sum \langle \mu_C, s_C \rangle = \max \sum \langle \mu_C, s_C \rangle$$

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



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

ie $\mu^* = \arg\max_{\mu\in L} \sum_c \langle \mu_c, s_c\rangle$

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

note: could use junction tree alg. [max form] (generalization of Viterbi) to solve this LP more efficiently (than interior point method)

overall, suppose that $\ell_i(\tilde{y})$ decomposes also on

$$\max_{\tilde{y}\in\mathcal{Y}_i} \ell_i(\tilde{y}) - w^T\psi_i(\tilde{y}) = \max_{\mu\in L_i} c_{0i} + (c_i + F_i^T w)^T\mu - w^T F_i \mu^{(i)}$$

(integer) marginal ground truth $y(i)$

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

use Mosek, Cplex, etc. to solve to 1e-16 accuracy using IPM eg ↳(see next class)

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

extragradient ⟶ would require projection on L_i which is more expensive than LP on L_i

can be done via using J.T. alg.?

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

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