

today: M3-net trick: M vs L polytope

II) M³-net example (CRF score)

* graph $G=(V,E)$

$y_p, p \in V \quad y_c \triangleq (y_p)_{p \in C}$

set of cliques \mathcal{C}

MRF model for score

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

$$= \frac{1}{Z} \exp\left(\sum_{C \in \mathcal{C}} \log \psi_C(y_C; x; \omega)\right)$$

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

example: OCR
chain graph on y

inference:

$$\max_{\tilde{y} \in \mathcal{Y}} \sum_{C \in \mathcal{C}} w^T \phi_C(y_C; \tilde{z}) \quad (\text{could use Viterbi for chain graph})$$

goal for today: rewrite this as a (small) LP

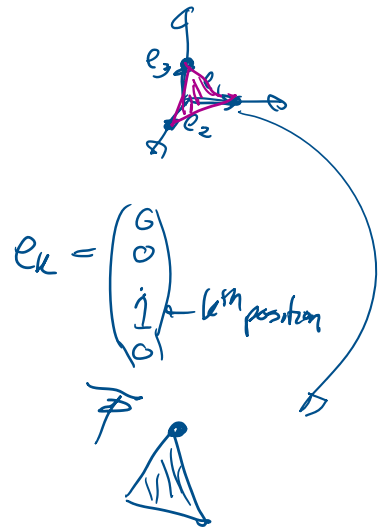
* for x & ω ; let s be a vector of size $|\mathcal{Y}(x)|$ of scores

$$\max_{\tilde{y} \in \mathcal{Y}(x)} s(x, \tilde{y}; \omega) = \max_k s_k = \max_{\{e_k\}} \langle e_k, s \rangle$$

$$= \max_{\alpha \in \text{conv}\{\{e_k\}\}} \langle \alpha, s \rangle$$

$$\triangleq \Delta_{|S|}$$

trick!
(tight convex relaxation)



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

⊛ insight of M³-net paper → use MRF structure

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

↑
tractable ↑
reasonable size

marginal polytope (M) formulation

$$\langle \alpha, s \rangle = \sum_{\tilde{y} \in \mathcal{Y}} \alpha(\tilde{y}) \left(\sum_{C \in \mathcal{C}} s_C(\tilde{y}_C) \right) = \sum_C \sum_{\tilde{y} \in \mathcal{Y}} \alpha(\tilde{y}) s_C(\tilde{y}_C)$$

$$= \sum_C \sum_{\tilde{y}_C} s_C(\tilde{y}_C) \left[\sum_{\tilde{y}} \alpha(\tilde{y}) \right]$$

$$= \sum_C \sum_{\tilde{y}_c \in \mathcal{Y}} s_c(\tilde{y}_c) \left[\sum_{\substack{y' \in \mathcal{Y} \\ y'_c = \tilde{y}_c}} \alpha(y') \right]$$

$$\mu \triangleq (\mu_c)_{c \in \mathcal{C}}$$

$$\mu_c \in \Delta_{|\mathcal{Y}_c|}$$

$\underbrace{y'_c = \tilde{y}_c}_{\triangleq \mu_c(\tilde{y}_c)}$ "clique marginal variable"

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

matrix of size $\dim(\mu) \times \dim(\alpha)$ which encodes the clique marginalization operation of α

$$\dim(\alpha) = |\mathcal{Y}|$$

$$\dim(\mu) = \sum_C |\mathcal{Y}_c|$$

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

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

marginal polytope

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

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

ie. $\overset{1}{\circ} - \overset{2}{\circ} - \overset{3}{\circ}$

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

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

$$\mu_{1,2}(y_1, y_2) \quad \mu_{2,3}(y_2, y_3)$$

consistency part

$$\mu_2(y_2) = \sum_{y_1} \mu_{1,2}(y_1, y_2)$$

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

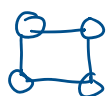
ie. $\forall C, C' \in \mathcal{C}$

$$\forall S \subseteq C \cap C'$$

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

\otimes L can be seen as a relaxation of M

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



(in particular, L can have fractional corners)

\star key property: if G is triangulated, then $M=L$ \iff \otimes \otimes (and so L has only

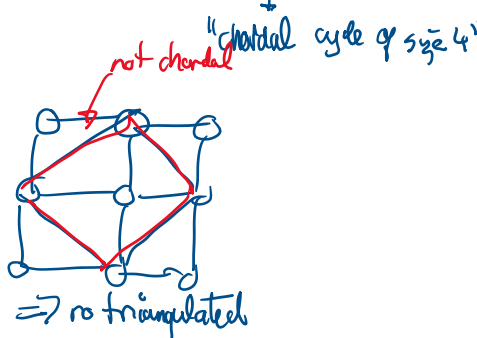
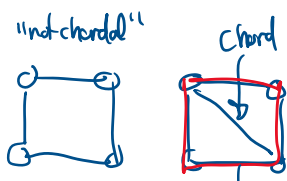
⊛ key property: if G is frangulated, then $M=L$ \Leftrightarrow ⊛ ⊛ (and so L has only integer vertices)

implication here $\max_{\mu \in M} \sum_C \langle s_C + l_C, \mu_C \rangle = \max_{\mu \in L} \dots$
 $g(w; \frac{z}{u})$ [assuming $l(y,y') = \sum_C l_C(y, y')$]

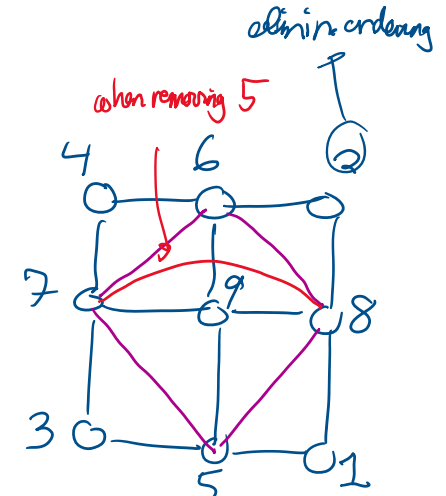
0h30

aside: frangulated review

Δ -graph \rightarrow iff it has no cycle of size 4 or more that are not chordal

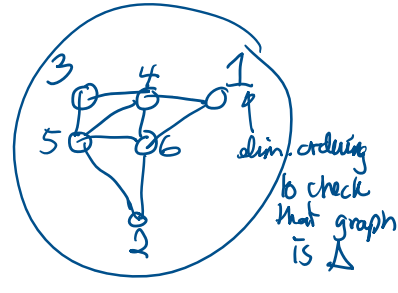


⊛ you can frangulate a graph by running variable elimination



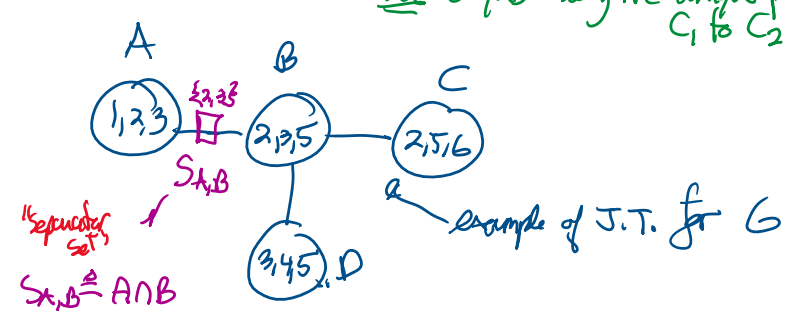
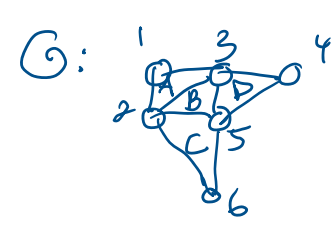
$\{6,7,8,9\}$ is biggest clique \Rightarrow tree width ≤ 3

⊛ if G is frangulated \Rightarrow can construct a junction tree

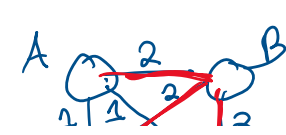


clique tree: tree with nodes being the maximal cliques of G with the running intersection property

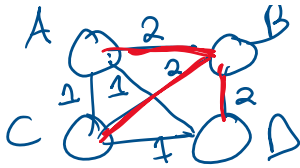
\hookrightarrow if $v \in C_i \cap C_j$, then v belongs to all cliques along the unique path from C_i to C_j in J.T.



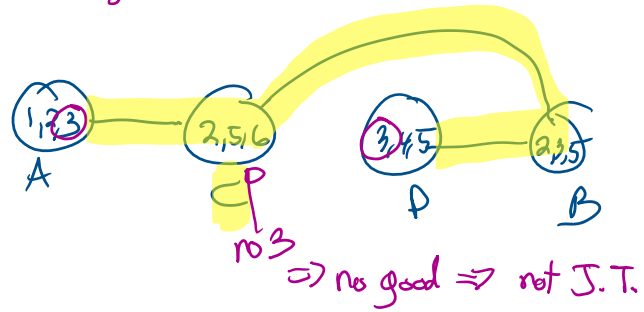
⊛ to build a J.T.; first build a weighted complete graph on cliques where weight on edge $\{i,j\}$ is $|C_i \cap C_j|$



use maximum weight spanning tree (on this weighted complete graph)



Use maximum weight spanning tree (on this weighted complete graph) to get a J.T.



Let's prove $M=L$ for Δ -graph

• $M \subseteq L$

• $L \subseteq M$:

$$\alpha(y) = \frac{\prod_{c \in \mathcal{C}} \mu_c(y_c)}{\prod_{s \in \mathcal{S}} \mu_s(y_s)}$$

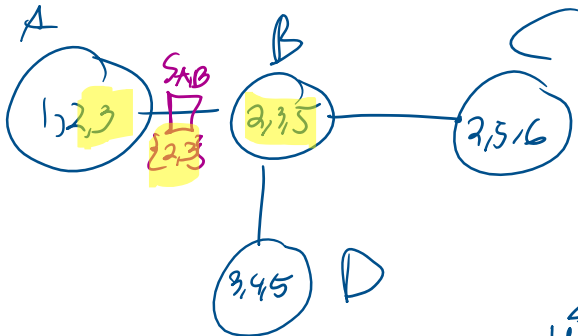
separator set from a J.T.

Suppose $\mu \in L$

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

i.e. $\mu = \sum \alpha \mu_\alpha$

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



Let's say want to show

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

$$\sum_{y'_B: y'_B = y_B} \left(\frac{\prod_c \mu_c(y_c)}{\prod_s \mu_s(y_s)} \right)$$

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

by local consistency prop

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

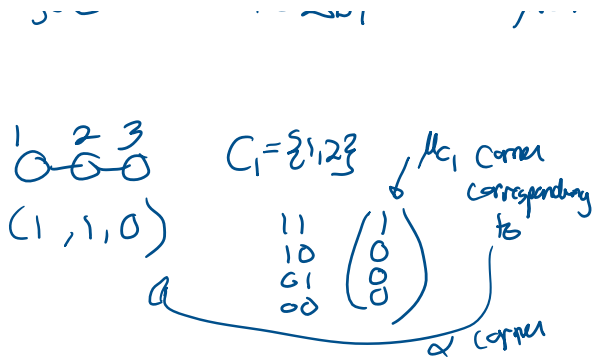
→ keep plucking the leaves until you are left with $\mu_B(y_B)$

to get back to M^3 -net:

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

Use $M=L$ (ie. Δ -graph)

note: could use ...



all vertices are integer

note: could use (max form) junction tree alg. (generaliz. of Viterbi) to solve this LP much more eff. than IPM

i.e. $\mu^* = \arg \max_{\mu \in L} \sum_C T(\mu, \epsilon)$

will have $\mu_C^*(\cdot) = \delta_{y_C^*}(\cdot)$

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

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

(integer) marginal ground truth y_i

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

use Mosek, CPLEX etc. to solve $1e-16$ accuracy using I.P.M. e.g. L_0 (see next class)

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

extragradient \rightarrow would require projection on L_i

which is more expensive than LP on L_i

C) later saddle pt. FW alg. which only requires LMO can be done using J.T. "linear min. oracle" on L_i 's instead of proj.

next class: Lagrangian dual for SVM struct