

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

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

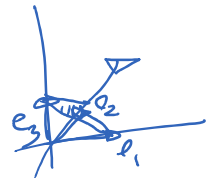
$$= \frac{1}{Z} \exp\left(\underbrace{\sum_{C \in \mathcal{C}} \log \psi_C(y_C; x, w)}_{s(x, y; w)}\right)$$

$w^T \psi_C(y_C; x)$

example: CRF
chain graph on y

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

goal: let's rewrite as a LP



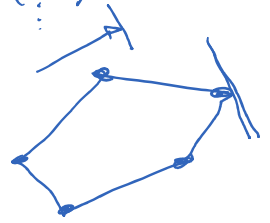
* for $x \& w$; let s be a vector of size $|\mathcal{Y}(x)|$ of scores

$$\max_{\tilde{y} \in \mathcal{Y}(x)} s(x, \tilde{y}; w) = \max_K s_K = \max_{\{e_k\}} \langle e_k, s \rangle$$

$$e_k = \begin{pmatrix} 0 \\ \dots \\ 0 \\ 1 \\ 0 \\ \dots \end{pmatrix}$$

$\leftarrow k^{\text{th}}$ position

trick! (tight convex pts. relaxation) $\rightarrow \max_{\alpha \in \text{conv}(\{e_k\})} \langle \alpha, s \rangle \triangleq \Delta_{\mathcal{Y}}$



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

insight of M³-net paper \rightarrow use MRF structure to transform $\max_{\alpha \in \Delta_{\mathcal{Y}}} \langle \alpha, s \rangle = \max_{\mu \in \mathcal{L} \subset \mathcal{C}} \langle \mu, s \rangle$
reasonable size
tractable

marginal polytope (M) formulation

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

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

$\triangleq \mu_c(\tilde{y}_c)$ "clique marginal variable"

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

$$\mu = A_M \alpha$$

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

$\langle \mu, s \rangle$

$$M \triangleq A_M \Delta_{|\mathcal{Y}|}$$

$$M \triangleq \left\{ (\mu_c)_{c \in \mathcal{C}} : \exists \alpha \in \Delta_{|\mathcal{Y}|} \text{ s.t. } \mu_c(\tilde{y}_c) = \sum_{\substack{\tilde{y} \in \mathcal{Y} \\ \tilde{y}_c = \tilde{y}_c}} \alpha(\tilde{y}) \quad \forall c \right\}$$

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

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"

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

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

$$\forall y_S$$

$$\sum_{\substack{y_{c'} \in \mathcal{Y}_{c'} \\ \text{s.t. } y_S = y_S}} \mu_c(y_{c'}) = \sum_{\substack{y_{c'} \in \mathcal{Y}_{c'} \\ \text{s.t. } y_S = y_S}} \mu_{c'}(y_{c'})$$

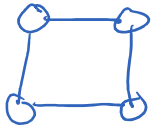
i.e. $\begin{matrix} 1 & 2 & 3 \\ \circ & \circ & \circ \\ C_1 = \{1,2\} & C_2 = \{2,3\} \end{matrix}$

$$= \|\mu_S(y_S)\|$$

⊗ L can be seen as a relaxation of M

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



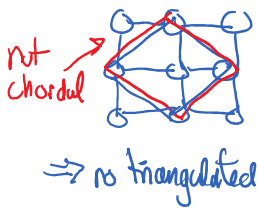


⊗ Key property: if G is triangulated then $M=L$!! ⊗ ⊗

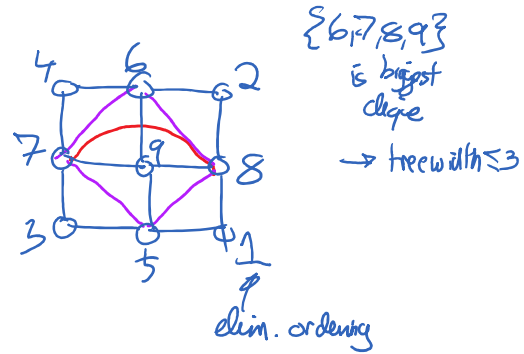
implication here $\max_{M \in M} \underbrace{\sum (c_e + l_c)_{M_e}}_{g(w,z)} = \max_{M \in L} \dots$ [assuming $l(y,y') = \sum l_c(y_c, y'_c)$]

aside: triangulated graph review:

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



⊗ you can triangulate a graph by running variable elimination

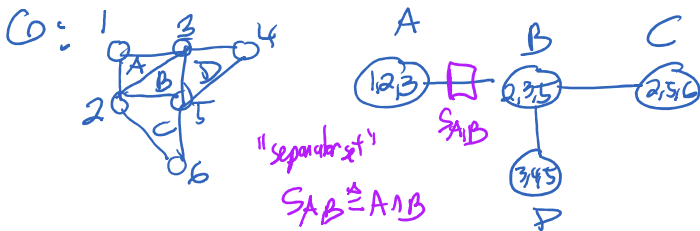


⊗ If G is triangulated \Rightarrow can construct a junction tree

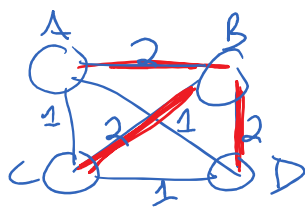
\hookrightarrow clique tree: tree with nodes being the maximal cliques of G with the running intersection property

\hookrightarrow 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



⊗ to build a J.T. first 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 J.T.





M=L for Δ-graphs

• $M \subseteq L$

• L ⊆ M:
$$\alpha(y) = \frac{\prod_{C \in \mathcal{C}} \mu(y_C)}{\prod_{S \in \mathcal{S}} \mu_S(y_S)}$$

↑
separator set
from J.T.

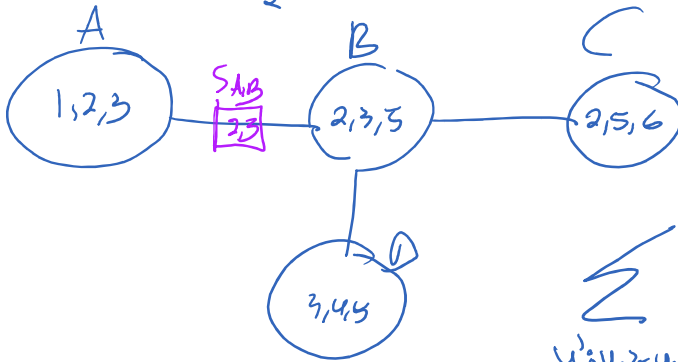
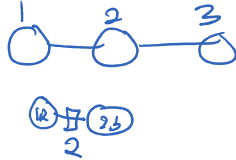
Suppose $\mu \in L$;

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

ie. $\mu = A \mu \alpha$

(this is gen. that for the free UGM

$$p(x) = \prod_{i,j \in E} \frac{p(x_i, x_j)}{p(x_i)p(x_j)} \prod_{i \in V} p(x_i)$$



Let's say want to show that
$$\mu_B(y_B) = \sum_{y': y_B' = y_B} \alpha(y')$$

$$\sum_{y' : y_B' = y_B} \left(\frac{\prod_{C \in \mathcal{C}} \mu(y_C)}{\prod_{S \in \mathcal{S}} \mu_S(y_S)} \right)$$

~~$\sum_{y' \in \mathcal{Y}} \left(\frac{\mu(y_A)}{\mu_{S_{A,B}}(y_{S_{A,B}})} \right) \left(\frac{\prod_{C \in \mathcal{C}} \mu(y_C)}{\prod_{S \in \mathcal{S}} \mu_S(y_S)} \right)$~~

There is no y_A here since $y_A \notin S_{A,B}$ crucially using marg. inter. property

now,
$$\sum_{y_A} \mu_A(y_A) = \mu_{\{2,3,5\}}(y_{\{2,3,5\}}) \quad \text{[by local consistency]}$$

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

$$y_1, \dots, y_n = M_{\text{stab}}(y_{\text{stab}})$$

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. $\text{Poly}(\max_C \{s_c\})$ # of constraints

vs. M which needs exponential # of constraints in worst case

to get back to MZnet:

$$\max_{\tilde{y} \in \mathcal{Y}} s(\tilde{y}) = \max_{\alpha \in \Delta_{\mathcal{R}^1}} \langle \alpha, s \rangle = \max_{\mu \in M_C} \sum \langle \mu_c, s_c \rangle = \max_{\mu \in L^c} \sum \langle \mu_c, s_c \rangle$$

when M=L (i.e. Δ graph)

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

ie. $\mu^* = \text{argmax}_{\mu \in L^c} \sum \langle \mu_c, s_c \rangle$
will have $\mu_C^*(\cdot) = \text{DU}_L(\cdot)$

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

overall, suppose that $l_i(y)$ decomposes also on G

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

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

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-6 accuracy use IPM eg

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

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

c) later: saddle pt. FW alg which only requires LMOs
"linear min. oracles"
on L^*
instead of projection

next class: Lagrangian dual of SVM struct