

Lecture 13 - scribbles - M3-net (cont.)

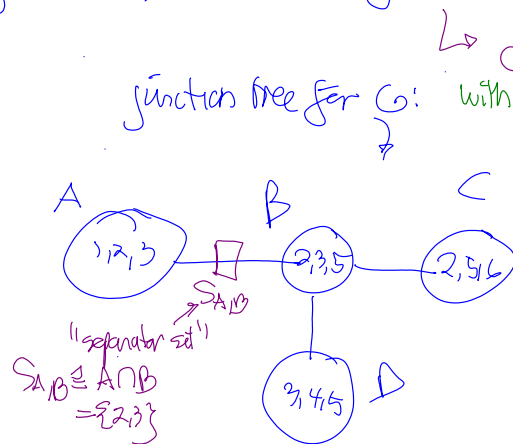
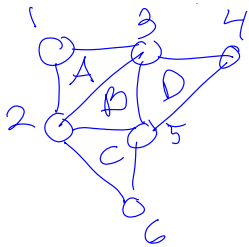
Tuesday, March 6, 2018
14:20

today: • finish Δ -graphs & M3-net
• dual SVMstruct

$M = L$ for Δ -graphs (continuation):

if G is triangulated \Rightarrow construct a junction tree

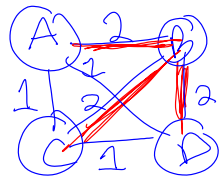
G :



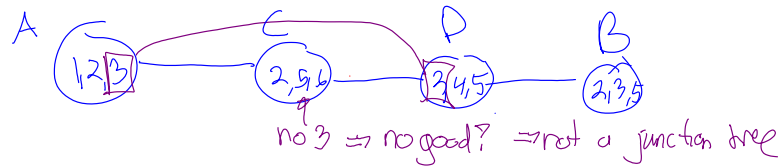
\hookrightarrow clique tree: every maximal clique of G is a node of the J.T.
with running intersection property:

\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 JT; first build weighted complete graph on cliques where weight on edges is $|C_i \cap C_j|$:



use the maximal weight spanning tree
to get JT



to show that $L \subseteq M$:

$$\alpha(y) = \bigwedge_{C \ni y} \mu_C(y_C)$$

this will be a valid joint $\alpha \in \mu = A_M \alpha$ when C are the maximal cliques
and S are separators
from a J.T.
to show that $\mu \in M$:

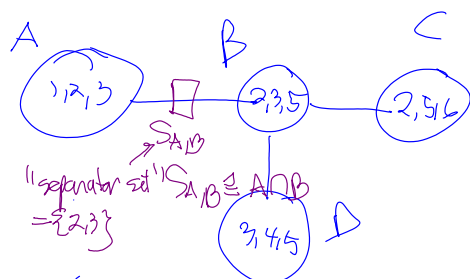
$$\frac{\prod_{C \in \mathcal{C}} \mu_S(y_S)}{\prod_{S \in \mathcal{S}} \mu_S(y_S)}$$

separator sets from S.T.

(note: this is generalization)

that for tree UGM

$$p(z) = \prod_{i: i \in E} \frac{p(x_i, z_i)}{p(x_i) p(z_i)} \prod_{i \in V} p(x_i)$$



(for correctness, here $\mu \in \Delta_A \times \Delta_B \times \Delta_C \times \Delta_D$)

we need to show that
 $\exists \alpha$ s.t. $\mu = \sum \alpha(y')$

let's say, want to show $\mu_B(y_B) = \sum_{y': y'_B = y_B} \alpha(y')$

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

$$\left(\sum_{y_1} \frac{\mu_A(y_A)}{\mu_{S_{AB}}(y_{S_{AB}})} \right) \times \left(\prod_{\substack{C \in \mathcal{C} \\ C \neq A}} \mu_C(y_C) \right) \times \left(\prod_{S \neq S_{AB}} \mu_S(y_S) \right)$$

there was no y_B here; crucially uses running intersection property

$$\sum_{y_1} \mu_A(y_A) = \mu_{\{2,3\}}(y_{\{2,3\}}) = \mu_{S_{AB}}(y_{S_{AB}})$$

thus $M=L$

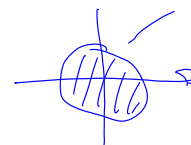
⊗ why do we care? because describing L is simple i.e. $\text{poly}(\max |S_C|)$ # of constraints

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

(intuition: triangulating a general graph yields big cliques C_{\max} and thus $|S_{C_{\max}}|$ is exponential)

polytope: convex hull of finite # of points

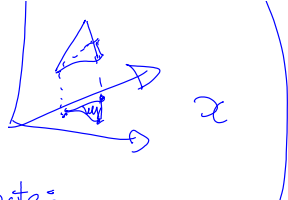
ℓ_2 ball = $\{x: \|x\|_2 \leq 1\}$
 is not a polytope



\mathbb{R}^n

Fourier-Motzkin elimination: $a_i^T x \leq z$ for $i \in I$
 $a_j^T x \geq z$ for $j \in J$

then use $a_i^T x \leq a_j^T x \quad \forall (i,j) \in I \times J$



$\max_{y \in S} s(y) = \max_{\alpha \in \Delta_{|S|}} \langle \alpha, S \rangle = \max_{\mu \in M} \sum_C \langle \mu_C, S_C \rangle \stackrel{\text{when } M=L \text{ i.e. } \Delta\text{-graph}}{=} \max_{\mu \in L} \sum_C \langle \mu_C, S_C \rangle$

note: could use junction tree algorithm (generalization of viterbi) to solve this LP

\rightarrow all the vertices have integer component (much more efficient than interior point method or simplex alg.)

i.e. $\mu^* = \arg \max_{\mu \in L} \sum_C \langle \mu_C, S_C \rangle$

will have $\mu_C^*(\cdot) = \delta(y_C^*)$

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

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

$\mu^{(i)}$ marginal from grand truth $y^{(i)}$

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

use Mosek, Cplex, Gurobi, etc.. to solve to $1e-16$ accuracy using interior point method e.g.

B) saddle point formulation $\min_w \max_{\mu \in L_i} \mathcal{L}(w, \mu)$

extragradient \rightarrow would require projecting on L_i (local consistency polytope)

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

and : there is a version of gradient algorithm which only requires LMOs "linear minimization oracles" on L_i 's instead of projection

"saddle pt. Frank-Wolfe algorithm" \rightarrow we'll come back later

interior point method:

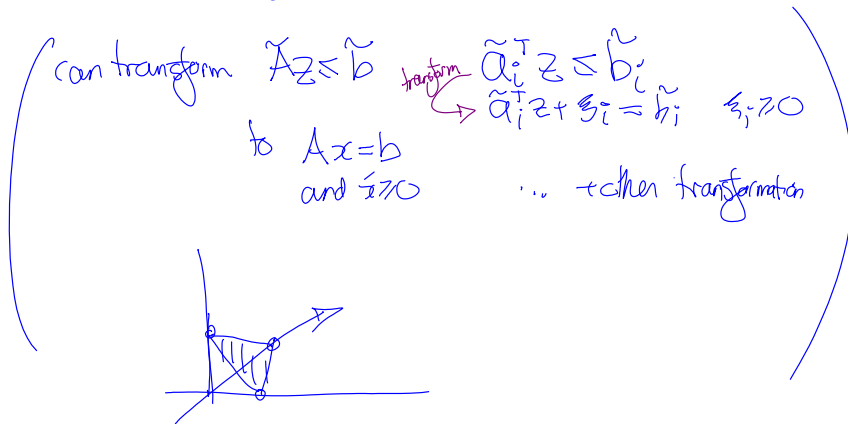
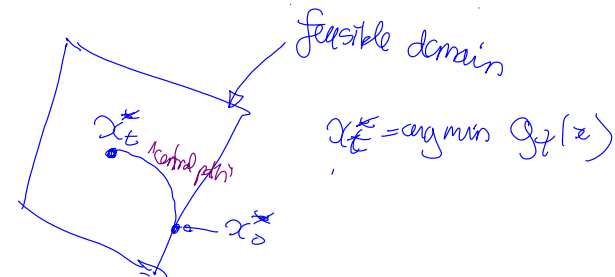
$$\begin{aligned} \min f(x) \\ \text{s.t. } Ax=b \\ x \geq 0 \end{aligned}$$

barrier method

$$\min f(x) - t \log x \triangleq g_t(x) \\ \text{s.t. } Ax=b$$

barrier function

$$\text{dom}(g_t(x)) = \{x; x > 0\}$$



interior point method : run Newton's method with warm start on $g_t(x)$ and decrease t with geometric schedule

convergence: Newton's method take $\log(\log \frac{1}{\epsilon}) + \text{cost}(f(x_0) - f^*)$ iterations to reach ϵ accuracy

iterations
IPM: $\log \frac{1}{\epsilon}$) • some constant overall convergence

"self-concordance analysis"

dimensionality is appearing; but no condition number of f (vs. gradient method)

cost per iteration is $\tilde{O}(nd^3)$

cost per iteration is $\sim O(d^3)$

(vs. ~~gradient~~ $\sim O(d^2)$)