Lecture 13 - M3-nets

Monday, February 27, 2024

M3-net trick: M vs L prob

II) M3-net example (CRF score)

\[ \text{graph } G = (V, E) \quad y_p, p \in V \quad y_c = (y_p)_{c \in C} \]

set of cliques \( \mathcal{C} \)

MRF model for score

\[ p(y | x; w) = \frac{1}{Z(x; w)} \prod_{c \in \mathcal{C}} \exp \left( \sum_{c \in \mathcal{C}} \log Y_c(y_c; x; w) \right) \]

\[ Z(x; w) = \sum_{c \in \mathcal{C}} \exp \left( \sum_{c \in \mathcal{C}} \log Y_c(y_c; x; w) \right) \]

Example: OCR

chain graph on \( y \)

Inference:

\[ \max_{y_c \in \mathcal{C}} \exp \left( \sum_{c \in \mathcal{C}} \log Y_c(y_c; x; w) \right) \]

tractable (could use \( V_t \) for chain graph)

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

\[ \forall x, y, w; \text{ let } s \text{ be a vector of size } 2^{|x|} \text{ of scores} \]

\[ \max_{y \in \mathcal{C}(x)} s(x, y; w) = \max_k s_k = \max_k \langle e_k, s \rangle \]

\[ e_k = \begin{pmatrix} g_k \end{pmatrix} \text{ in position } k \]

Trick?

(Convex relaxation)

\[ \text{Trick?} \]

Note: any combinatorial opt. problem can be written (trivially) as a convex opt. (LP) with \( \exp \# \mathcal{C} \) dimensions

Insight of M3-net paper → use MRF structure to transform

\[ \max_k \langle e_k, s \rangle = \max_{c \in \mathcal{C}} \exp \left( \sum_{c \in \mathcal{C}} \log Y_c(y_c; x; w) \right) \]

Marginal polyhedral (M) formulation

\[ s(y) \]

\[ \langle a, s \rangle = \sum_{y \in \mathcal{C}} a(y) \left( \sum_{c \in \mathcal{C}} \log Y_c(y_c; x; w) \right) \]

\[ = \sum_{y \in \mathcal{C}} \log Y_c(y_c; x; w) \]

\[ = \sum_{y \in \mathcal{C}} \log Y_c(y_c; x; w) \]
\[ \mu \triangleq (\mu_c \circ \epsilon_\mathcal{E}) \quad \mu_c \in \Delta | \mathcal{E} | \]

\[ \mu \triangleq \text{max} \quad \text{i.e.} \quad \mu_c(y_c) = \sum_{y \in \mathcal{E}} \alpha(y) = (\mu_c)_c \text{ (c, } y_c) \quad \alpha \]

\[ \text{matrix of size } (| \mathcal{E} | \times | \mathcal{E} |) \]

which ensures the clique marginalization operation of \( \alpha \)

\[ \text{dim}(\alpha) = | \mathcal{E} | \]

\[ \text{dim}(\mu) = \sum | \mathcal{E} | \]

\[ \max \quad \langle \alpha, \nu \rangle = \max_{\mu \in \mathcal{M}} \sum_{\mathcal{E}} \sum_{y \in \mathcal{E}} \alpha(y) \mu_c(y_c) \]

\[ \mathcal{M} = \{ (\mu_c)_{c \in \mathcal{E}} : \forall \alpha \in \Delta | \mathcal{E} | \text{ s.t. } \mu_c(y_c) = \sum y \text{ (c, y_c) } \} \]

\[ \forall c, y_c, \quad y_c = y_c \]

\[ \mathcal{L} \subseteq \mathcal{E} \]

\[ \text{ie. } 1 \quad 2 \quad 3 \]

\[ C_1 = \{ 1, 2 \} \quad C_2 = \{ 2, 3 \} \]

\[ M_{12} \quad M_{23} \]

\[ \alpha_{12}(y_1, y_2) \quad \mu_{12}(y_1, y_2) \]

\[ \alpha_{11}(y_1) = \sum y_2 \quad \mu_{12}(y_1, y_2) \]

\[ \alpha_{11}(y_1) = \sum y_2 \quad \mu_{12}(y_1, y_2) \]

\[ \therefore \mathcal{L} \text{ can be seen as a relaxation of } \mathcal{M} \]

\[ \text{if } \mathcal{G} \text{ is not triangulated, then we can have that } \mathcal{M} \not\subseteq \mathcal{L} \]

\[ \therefore \text{key property } \mathcal{G} \text{ is triangulated, then } \mathcal{M} = \mathcal{L} \quad \therefore \quad \text{and so } \mathcal{L} \text{ has only} \]
key property: if \( G \) is triangulated, then \( M = L \) and \( \mathcal{D} \) (and so \( L \) has only integer vertices)

 implication here: \( \max_{\mu \in M} \langle \delta_c + \delta_L \mu \rangle = \max_{\mu \in L} 0(\mu, \beta) \) 

 assuming \( 0(\mu, \beta) = \sum \mu_c \delta_c(\mu, \delta_c') \)

**Aside: Triangulated Graphs**

A graph \( \Delta \) is triangulated if it has no cycle of size 4 or more that are not chordal.

\( \Rightarrow \) if you can triangulate a graph by running variable elimination

\( \Rightarrow \) \( 6, 7, 8, 13 \) is highest clique

\( \Rightarrow \) tree width \( \leq 3 \)

If \( G \) is triangulated \( \Rightarrow \) can construct a junction tree

**Junction Tree**: a tree with nodes being the maximal cliques of \( G \)

- \( v \in C_1 \cap C_2 \), then \( v \) belongs to all cliques along the unique path from \( C_1 \) to \( C_2 \) in \( J.T. \)

To build \( J.T. \); first build a weighted complete graph on cliques where weight on edge \( C_1 \) \( C_2 \) is \( C_1 \cap C_2 \)

Use maximum weight spanning tree (one this weighted complete graph)
Let's prove \( M = L \) for \( \Delta \)-graph

\[ x(y) = \frac{1}{\gamma(y)} \sum_{e \in E} \mu_e(y_c) \]

\[ \gamma(y) = \prod_{s \in S} \mu_s(y_s) \]

- \( M \subseteq L \)
- \( L \subseteq M \)

\( \alpha(y) = \frac{1}{\gamma(y)} \sum_{e \in E} \mu_e(y_c) \)

\( \gamma(y) = \prod_{s \in S} \mu_s(y_s) \)

Let's say we want to show

\( \mu_N(y_0) = \sum_{y_0 \neq y} \alpha(y) \)

\( \text{Suppose } \mu \in L \)

\( \text{We want } \exists r \text{ s.t. } \mu \text{ is weak } r \text{ w.r.t. } L \)

\( \text{i.e. } \mu \subseteq A \mu \theta \)

\( \text{(This is a generalization of } \mu(x) = \frac{1}{\gamma(x)} \prod_{e \in E} \mu_e(x) \text{ for free UGM)} \)

- \( \text{Keep plucking the leaves until you are left with } \mu_N(y_0) \)

\( \text{To get back to } M^2 \text{-net:} \)

\( \max_{y \in S} s(y) = \max_{\mu \in \Delta \mu} \langle \alpha, s \rangle = \max_{\mu \in \Delta \mu} \sum_{y_0 \neq y} \alpha(y) \)

\( \text{Note: could use } \mu \subseteq A \mu \theta \)
Overall, suppose $l_i(y)$ decomposes also on $L$

\[
\max_{y \in \mathcal{Y}} l_i(y) = \max_{y \in \mathcal{Y}} \alpha_i + (c_i + F_t^T w) \mu_i - w^T \psi_i \mu_i
\]

\[
\text{(integer) maximum ground truth } u_i
\]

Last class: A) dualize the LP and plug it in SVMstruct alg.

(to get a "small comapred LP")

use Mosek, Gurobi etc. to solve 1e-16 accuracy using I.P.M. etc.

B) saddle pt: minimax $\min_{\omega} \max_{i \in \mathcal{I}} S(\omega, \mu_i)$

extragradient - would require projection on $L_i$

which is more expensive than LP on $L_i$

C) later saddle pt: EW alg. which only requires I.M.O.

"linear min. oracle"

on $L_i$ instead of proj.

next class: Lagrangian dual for SVMstruct