

today: finish word-alignment example $f_i(w) = \max_{z_i \in Z_i} g(w, z_i)$
• M3-net

word alignment (cted.):

inference: $\max_{A_i \leq b_i, z \geq 0} w^T F_i z$

$$\begin{cases} z \in M_i \\ i \in S, z_{ij} \leq 1 & L \times L_F \\ \sum_j z_{ij} \leq 1 & L_F \\ \sum_i z_{ij} \leq 1 & L_F \end{cases}$$

what about loss?

example: Hamming loss

$$\begin{aligned} l(y, \tilde{y}) &= \sum_{k \in I} \mathbb{1}\{y_{ke} \neq \tilde{y}_{ke}\} \\ &= \sum_{k \in I} (y_{ke} - \tilde{y}_{ke})^2 \quad [\text{because } y_{ke} \in \{0, 1\}] \\ &= y_{ke}^2 - 2y_{ke}\tilde{y}_{ke} + \tilde{y}_{ke}^2 \end{aligned}$$

cool trick $y^2 = y$
when $y \in \{0, 1\}$

$$l_i(\tilde{y}) = a_i + \underbrace{(1 - 2y^{(i)})^T}_{c_i} \tilde{y}$$

loss augmented inference becomes

$$\begin{aligned} \max_{\substack{\tilde{y}_{ke} \in \{0, 1\} \\ \tilde{y} \in M_i}} & \underbrace{a_i + c_i^T \tilde{y}}_{l_i(\tilde{y})} - \underbrace{(w^T F_i y^{(i)} - w^T F_i \tilde{y})}_{w^T \psi(\tilde{y})} \\ &= a_i - w^T F_i y^{(i)} + \max_{\substack{\tilde{y}_{ke} \in \{0, 1\} \\ \tilde{y} \in M_i}} (F_i^T w + c_i)^T \tilde{y} \\ &= \max_{z \in M_i} (F_i^T w + c_i)^T z + a_i - w^T F_i y^{(i)} \end{aligned}$$

LP duality

$$\begin{aligned} \max_{\substack{A_i z \leq b_i \\ z \geq 0}} \tilde{C}_i^T z &= \min_{\substack{A_i v \geq \tilde{C}_i \\ v \geq 0}} b_i^T v \\ A_i &\in \frac{(2L+L^2)}{2} \times L^2 \quad \tilde{C}_i \in F_i^T w + c_i \quad (\text{note: here } b_i = 1) \end{aligned}$$

$$\max_{z \in M_i} g_i(w, z) = \min_{v \in V_i(w)} \tilde{C}_i(w, v)$$

$$V_i(w) = \sum_{v \geq 0} v_i \{ A_i^T v \geq F_i^T w + c_i \}$$

SVM dual objective becomes

$$\min_w \min_{\substack{\{v_i\}_{i=1}^n \\ A_i^T v_i \geq F_i^T w + c_i \\ v_i \geq 0}} \frac{\lambda \|w\|^2}{2} + \frac{1}{n} \sum_{i=1}^n [a_i - w^T F_i y^{(i)} + b_i^T v_i]$$

"small complicated QP"

$$\tilde{A}(w) \leq \tilde{b}$$

compare with saddle point formulation

$$\min_w \max_{\substack{\{z_i\} \\ z_i \in M_i}} \frac{\lambda \|w\|^2}{2} + \frac{1}{n} \sum_{i=1}^n [a_i - w^T F_i y^{(i)}] + [(F_i^T w + c_i)^T z_i]$$

simple constraints

* note: primal error vs. dual error not same!

II) M³-net example (CRF score)

suppose have a graph $G = (V, E)$ $y_p, p \in V$ $y_c \triangleq (y_p)_{p \in C}$

set of cliques \mathcal{C}

Markov random field (MRF)
undirected graphical model

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

$$= \frac{1}{Z(x;w)} \exp \left(\sum_c \underbrace{\log \psi_c(y_c; x, w)}_{w^T \phi_c(y_c; x)} \right) \rightarrow S(x, y; w)$$

inference: $\max_{\tilde{y} \in \mathcal{Y}} \underbrace{S(x, \tilde{y}; w)}_{\sum_{c \in \mathcal{C}} w^T \phi_c(y_c; x)}$

example: OCR
→ chain graph on y

goal: let's rewrite as a LP

fix x & w : let s be a vector of size $|V \cup E|$

k^{th} basis vector $e_k = \begin{pmatrix} 0 \\ \vdots \\ 1 \end{pmatrix}$

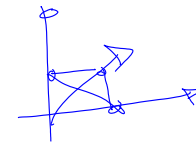
- y = ...

0 0 0 0 0 0

convex (S)

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

$$\begin{aligned} & \text{(tight convex dts. relaxation)} = \max_{\alpha \in \text{conv}(\xi_{\mathcal{K}})} \langle \alpha, s \rangle \\ & \quad \parallel \\ & \quad \Delta_{|\mathcal{K}|} \end{aligned}$$



$$\begin{aligned} \alpha \in \Delta_{|\mathcal{K}|} \\ \Leftrightarrow \alpha(\tilde{y}) \geq 0 \quad \forall \tilde{y} \in \mathcal{Y} \\ \sum_{\tilde{y} \in \mathcal{Y}} \alpha(\tilde{y}) = 1 \end{aligned}$$

note that any combinatorial optimization problem can be written as a convex LP (but with exp. # of dimensions)

⊛ insight of M²net paper ~ use MRF structure to transform $\max_{\alpha \in \Delta_{|\mathcal{K}|}} \langle \alpha, s \rangle$ to smaller problem

$$\begin{aligned} \langle \alpha, s \rangle &= \sum_{\tilde{y} \in \mathcal{Y}} \alpha(\tilde{y}) \left(\sum_c s_c(\tilde{y}_c) \right) \\ &= \sum_{c \in \mathcal{C}} \sum_{\tilde{y}_c \in \mathcal{X}_c} s_c(\tilde{y}_c) \left[\sum_{\substack{\tilde{y} \in \mathcal{Y} \\ \tilde{y}_c = \tilde{y}_c}} \alpha(\tilde{y}) \right] \\ & \quad \underbrace{\qquad\qquad\qquad}_{\triangleq \mu_c(\tilde{y}_c)} \quad \text{clique marginal variable} \end{aligned}$$

$$\begin{aligned} \mu &= (\mu_c)_{c \in \mathcal{C}} \quad \mu_c \in \Delta_{|\mathcal{X}_c|} \\ \mu &= A_M \alpha \end{aligned}$$

$$\max_{\alpha \in \Delta} \langle \alpha, s \rangle = \max_{\mu \in A_M \Delta} \sum_c \sum_{\tilde{y}_c \in \mathcal{X}_c} s_c(\tilde{y}_c) \mu_c(\tilde{y}_c) = \langle \mu, s \rangle$$

$$M \triangleq A_M \Delta$$

"marginal polytope"

$$M = \{ \xi_{\mu_c} \}_{c \in \mathcal{C}} : \exists \alpha \in \Delta_{|\mathcal{K}|} \text{ s.t. } \mu_c(\tilde{y}_c) = \sum \alpha(\tilde{y}) \quad \forall c \in \mathcal{C} \}$$

remark: went from exp. # of variables

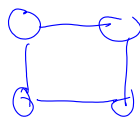
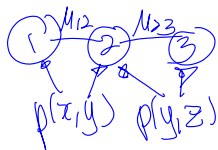
to $\sum_C |x_C|$ # of variables; but over more complicated set M

can we simplify M ?

consider the relaxation $M \subseteq L \triangleq \left\{ \{ \mu_C \}_{C \in \mathcal{C}} : \begin{array}{l} \mu_C \in \Delta(|x_C|) \text{ and} \\ \mu_C \text{ are "locally consistent"} \end{array} \right\}$

$$\Rightarrow \forall C, C'; \forall S \subseteq C \cap C', \forall y_S \sum_{\substack{y_C \in \mathcal{Y}_C \\ s.t. y_S = y_S}} \mu_C(y_C) = \sum_{\substack{y_{C'} \in \mathcal{Y}_{C'} \\ s.t. y_S = y_S}} \mu_{C'}(y_{C'}) = M_S(y_S)$$

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



$$\sum_z p(z,y) = p(y) = \sum_z p(y,z) \quad (\text{consistency at } y)$$

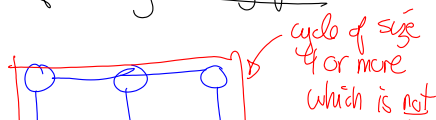
⊛ key property: if G is triangulated then $M = L$ ⊛ ⊛

less-augmented version:

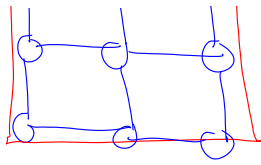
$$\text{implication here } \max_{\mu \in M} \sum_C \langle \Sigma_C + l_C, \mu_C \rangle = \max_{\substack{\mu \in L \\ p \\ z \in \mathcal{Z}}} \sum_C \langle \Sigma_C + l_C, \mu_C \rangle$$

$$(\text{assuming that } l(y,\tilde{y}) = \sum_C l(y,\tilde{y}_C))$$

review of triangulated graphs

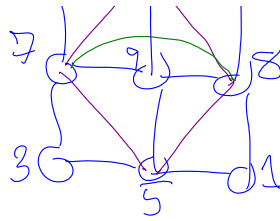


$\{6,7,8,9\}$ is biggest clique here



chordal

not triangulated



variable elimination ordering

next time: we show that LEM