4/8/14

MRF MAP INFERENCE

① Goal:

→ Given MRF $G = (V, E)$

→ Given a set of factors $F = \{ \psi_c(x_c) \}$

Find $\vec{x}_{MAP} = \arg\max_{\vec{x}} \frac{1}{Z} \prod_c \psi_c(x_c)$

→ Notation: Log-Potentials/Factors or Local "Scores"

$$\theta_c(x_c) \equiv \log \psi_c(x_c)$$

Note 1: Notice that $\arg\max_{\vec{x}} \frac{1}{Z} \prod_c \psi_c(x_c)$

$$= \arg\max_{\vec{x}} \log\left[\frac{1}{Z} \prod_c \psi_c(x_c)\right]$$ [Why? ∵ log is an increasing fn]

$$= \arg\max_{\vec{x}} \left[\sum_c \theta_c(x_c) - \log Z\right]$$

$$= \arg\max_{\vec{x}} \sum_c \theta_c(x_c)$$

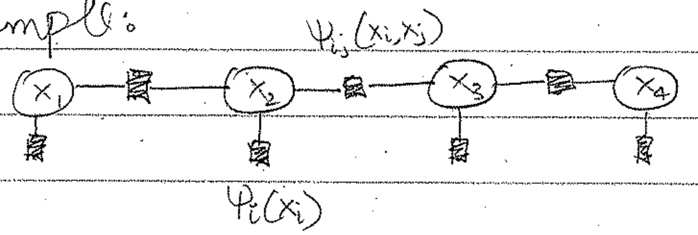
$$\equiv \arg\max_{\vec{x}} \left[\sum_i \theta_i(x_i) + \sum_{(i,j) \in E} \theta_{ij}(x_i, x_j)\right]$$ For pairwise MRFs.

② MAP Inference with message-passing / DP / BP

→ In short: all technique developed in previous lectures for marginal inference (or sum-product inference) generalize to max-product as well.

→ sum-product ⇒ max-product
(or max-sum in log-space)
(or min-sum in "energy" space)

Example:



MAP $\max_{x_1} \max_{x_2} \max_{x_3} \max_{x_4} \; \theta_1(x_1) + \theta_2(x_2) + \theta_3(x_3) + \theta_4(x_4) + \theta_{12}(x_1, x_2) + \theta_{23}(x_2, x_3) + \theta_{34}(x_3, x_4)$

Only terms that depend on x_4. Similar to old arguments, we can "push $\max_{x_4}$ in"

$$\lambda_{4\to3}(x_3) = \max_{x_4} \left[\theta_4(x_4) + \theta_{34}(x_3, x_4) \right]$$

$$\lambda_{3\to2}(x_2) = \max_{x_3} \left[\theta_3(x_3) + \theta_{23}(x_2, x_3) + \lambda_{4\to3}(x_3) \right]$$

$$\lambda_{2\to1}(x_1) = \max_{x_2} \left[\theta_2(x_2) + \theta_{12}(x_1, x_2) + \lambda_{2\to1}(x_1) \right]$$

MAP value (in log-space) $= \max_{x_1} \left[\theta_1(x_1) + \lambda_{2\to1}(x_1) \right]$

Note 1: For argmax backtrack

Note 2: max-marginal $\mu_i(x_i) = \theta_1(x_1) + \lambda_{2\to1}(x_1)$

$$= \left\{ \begin{array}{l} \max\limits_{\vec{X}} \sum_c \theta_c(x_c) \\ \text{s.t.} \quad X_i = x_i \end{array} \right\}$$

Interesting interpretation! of max-marginals. Force a variable to take a state & maximize over everything else.

Why should you care? Useful to prove things about

→ decoding MAP configuration without backtracking

→ 2nd best MAP search

→ etc.

(3) Marginal to MAP change in BP equations

Cluster Graph	Sum-Product	Max-Product	Max-Sum
General $\boxed{C_i} \overset{S_{ij}}{—} \boxed{C_j}$	$\delta_{i\to j}(S_{ij}) = \sum_{C_i \setminus S_{ij}} \psi_i(C_i) \prod_{k\in N(i)\setminus j} \delta_{k\to i}$	$\lambda_{i\to j}(S_{ij}) = \max_{C_i\setminus S_{ij}} [\quad]$	$\lambda_{i\to j}(S_{ij}) = \max_{C_i\setminus S_{ij}} \left[\theta_i(C_i) + \sum_{k\in N(i)\setminus j} \lambda_{k\to i}\right.$
Bethe $\boxed{X_c} \xrightarrow{X_i} \boxed{X_i}$	$\delta_{c\to i}(X_i) = \sum_{X_c\setminus X_i} \psi_c(X_c) \prod_{k\in c\setminus i} \delta_{k\to c}(x_k)$	$\lambda_{c\to i}(X_i) = \max_{X_c\setminus X_i} [\quad]$	$\lambda_{c\to i}(X_i)$ $\max_{X_c\setminus X_i} \left[\theta_c(X_c) + \sum_{k\in c\setminus i} \lambda_{k\to c}(x_k)\right.$
Loopy BP in Pairwise $(x_i) — (x_j)$	$\delta_{i\to j}(X_j) = \sum_{x_j} \psi_i(x_i) \psi_{ij}(x_i,x_j) \prod_{k\in N(i)\setminus j} \delta_{k\to i}(x_i)$	$\lambda_{i\to j}(x_j) = \max_{x_j} [\quad]$	$\lambda_{i\to j}(x_j)$ $\max_{x_j} \theta_i(x_i) + \theta_{ij}(x_i,x_j) + \sum_{k\in N(i)\setminus j} \lambda_{k\to i}(x_i)$

④ MAP as Linear Integer program

→ Trick: Convert $x_i = \underset{s}{\text{state}} \overset{\text{into}}{\Longrightarrow}$

$$\underbrace{\begin{bmatrix} \\ \mathbb{1}[\![x_i = s]\!] \\ \\ \\ \end{bmatrix}_{K\times 1}}_{\substack{\text{indicator vector} \\ \vec{\mu}_i}}$$

Do same for factors

$$\overset{\vec{X}_c}{(X_{c_1}, X_{c_2} \dots)} = \overset{\text{states}}{(S_1, S_2, \dots, S_{|c|})} \overset{\text{into}}{\Longrightarrow} \underbrace{\begin{bmatrix} \\ \mathbb{1}[\![X_{c_1} = s_1,\ X_{c_2} = s_2 \ \dots\ ]\!] \\ \\ \\ \end{bmatrix}}_{\vec{\mu}_c}$$

$|X_{c_1}|$ times
$|X_{c_2}|$ times
$\vdots$

Now $\text{Score}(\vec{x}) = \sum_c \vec{\theta}_c^{\,T} \vec{\mu}_c$

$= \vec{\theta}^{\,T} \vec{\mu}$

Now MAP-IP: $\max_{\vec{\mu}} \ \theta^T \mu$

$\mu_i(x_i) \in \{0,1\} \quad \mu_c(x_c) \in \{0,1\}$

s.t. $\left\{ \begin{array}{l} \sum_{x_i} \mu_i(x_i) = 1 \\ \\ \sum_{x_c} \mu_c(x_c) = 1 \end{array} \right\}$ unique labelling

$\sum_{x_c \setminus x_i} \mu_c(x_c) = \mu_i(x_i) \Big\}$ consistent labelling