Combinatorial Optimization in Computer Vision
Lecture: F. R. Schmidt and C. Domokos
Exercises: T. Möllenhoff and T. Windheuser
Winter Semester 2015/2016

Computer Vision Group
Institut für Informatik
Technische Universität München

Weekly Exercises 6

Room: 02.09.023
Tuesday, 01.12.2015, 14:15-15:45
Submission deadline: Tuesday, 01.12.2015, 11:15 , Room 02.09.023

Mean Field Inference (10 Points)

Exercise 1 (4 Points)**.** Assume a graphical model $G = (\mathcal{V}, \mathcal{E})$ and consider a factorized distribution in the following form:

$$q(y) = \prod_{i \in \mathcal{V}} q_i(y_i). \tag{1}$$

a) Show that the marginal distribution of a factor F is given by:

$$\mu_{F,y_F}(q) = q_{N(F)}(y_F) = \prod_{i \in N(F)} q_i(y_i).$$

b) Show that the entropy decomposes as:

$$H(q) = \sum_{i \in \mathcal{V}} H_i(q_i),$$

where

$$H_i(q) = -\sum_{i \in \mathcal{V}} \sum_{y_i \in \mathcal{Y}_i} q_i(y_i) \log q_i(y_i).$$

Solution.

a) First note that:

$$\begin{aligned}
\sum_{y' \in \mathcal{Y}, y'_1 = y_1} q(y) &= \sum_{y' \in \mathcal{Y}, y'_1 = y_1} \prod_{i \in \mathcal{V}} q_i(y_i) && (2)\\
&= q_1(y_1) \sum_{y_2 \in \mathcal{Y}_2} \dots \sum_{y_k \in \mathcal{Y}_k} q_2(y_2) \dots q_k(y_k) && (3)\\
&= q_1(y_1) \underbrace{\sum_{y_2 \in \mathcal{Y}_2} q_2(y_2)}_{1} \dots \underbrace{\sum_{y_k \in \mathcal{Y}_k} \dots q_k(y_k)}_{1} = q_1(y_1). && (4)
\end{aligned}$$

Similarly, one computes:

$$\mu_{F,y_F}(q) = \sum_{y' \in \mathcal{Y}, y'_F = y_F} q(y') = \sum_{y' \in \mathcal{Y}, y'_F = y_F} \prod_{i \in \mathcal{V}} q_i(y'_i) \tag{5}$$

$$= \prod_{i \in N(F)} q_i(y_i) \underbrace{\sum_{y_j \in \mathcal{Y}_j, j \in \mathcal{V} \setminus N(F)} \prod_{k \in \mathcal{V} \setminus N(F)} q_k(y_k)}_{=1} = \prod_{i \in N(F)} q_i(y_i). \tag{6}$$

b)

$$H(q) = -\sum_{y \in \mathcal{Y}} q(y) \log q(y) = -\mathbb{E}[\log q(y)] \tag{7}$$

$$= -\mathbb{E}\left[\sum_{i \in \mathcal{V}} \log q_i(y_i)\right] = -\sum_{i \in \mathcal{V}} \mathbb{E}[\log q_i(y_i)] \tag{8}$$

$$= -\sum_{i \in \mathcal{V}} \sum_{y_i \in \mathcal{Y}_i} q_i(y_i) \log q_i(y_i) = \sum_{i \in \mathcal{V}} H_i(q_i) \tag{9}$$

Exercise 2 (6 Points)**.** In the lecture we introduced $D_{\mathrm{KL}}(q(y) \| p(y \mid x))$ for some "simple" distribution q as a tractable approximation of $p(y \mid x)$. We now assume that $q(y)$ follows the factorization (1).

a) Show that the reverse KL-divergence $D_{\mathrm{KL}}(p(y \mid x) \| q(y))$ can be written as

$$D_{\mathrm{KL}}(p(y \mid x) \| q(y)) = \sum_{y \in \mathcal{Y}} p(y \mid x) \left(\sum_{i \in \mathcal{V}} \log q_i(y_i)\right) + \text{const}, \tag{10}$$

where the constant is meant with respect to $q(y)$.

b) Using the technique of Lagrange multipliers, show that minimizing (10) with respect to $q_i(y_i)$ while holding all other variables fixed yields the corresponding marginal distribution of p:

$$q_i^*(y_i) = \operatorname*{argmin}_{q_i} D_{\mathrm{KL}}(p(y \mid x) \| q(y)) = p(y_i \mid x).$$

Solution.

a)

$$\begin{aligned} D_{\mathrm{KL}}(p(y \mid x) \| q(y)) &= \sum_{y \in \mathcal{Y}} p(y \mid x) \log \frac{p(y \mid x)}{q(y)} \\ &= -\sum_{y \in \mathcal{Y}} p(y \mid x) \log q(y) + \sum_{y \in \mathcal{Y}} p(y \mid x) \log p(y \mid x) \\ &= -\sum_{y \in \mathcal{Y}} p(y \mid x) \left(\sum_{i \in \mathcal{V}} \log q_i(y_i)\right) + \underbrace{\sum_{y \in \mathcal{Y}} p(y \mid x) \log p(y \mid x)}_{\text{const}}. \end{aligned}$$

b) Since we are optimizing over $q_j(y_j)$ we can absorb all the other terms q_i for $i \neq j$ into the constant term:

$$\begin{aligned}
D_{\mathrm{KL}}(p\|q) &= -\sum_{y\in\mathcal{Y}} p(y \mid x)\left[\sum_{i\in\mathcal{V}} \log q_i(y_i)\right] + \text{const} \\
&= -\sum_{y\in\mathcal{Y}} p(y \mid x) \log q_j(y_j) + \text{const} \\
&= -\sum_{y_j\in\mathcal{Y}_j} \log q_j(y_j)\left[\sum_{i\neq j, y_i\in\mathcal{Y}_i} p(y \mid x)\right] + \text{const.} \\
&= -\sum_{y_j\in\mathcal{Y}_j} \log q_j(y_j) p(y_j \mid x) + \text{const.}
\end{aligned}$$

Introducing a Lagrange multiplier λ to ensure that $q_j(y_j)$ sums up to one yields the Lagrangian

$$\mathcal{L}(q_j, \lambda) = \sum_{y_j\in\mathcal{Y}_j} p(y_j|x) \log q_j(y_j) + \lambda(\sum_{y_j\in\mathcal{Y}_j} q_j(y_j) - 1).$$

Setting the derivative of the Lagrangian $\mathcal{L}(q_j, \lambda)$ w.r.t to q_j to zero yields:

$$-\frac{p(y_j|x)}{q_j(y_j)} + \lambda = 0.$$

Hence $\lambda q_j(y_j) = p(y_j)$. Summing both sides over all possible y_j yields:

$$\lambda = \sum_{y_j\in\mathcal{Y}_j} p(y_j \mid x) = 1,$$

and thus

$$q_j(y_j) = p_j(y_j \mid x).$$

Programming (10 Points)

Exercise 3 (Parametric Max Flow, 10 Points)**.**
In this programming exercise we are interested in finding the breakpoints of the energy function

$$E(x, \lambda) = \sum_{i=1}^{n} f_i x_i + \lambda \sum_{i=1}^{n} \sum_{j \in \mathcal{N}(i)} c_{ij} x_i \overline{x}_j \tag{11}$$

from Exercise Sheet 3. The file `06_supp.zip` from the lecture website contains the code to find the minimizer $x^* = \arg\min_x E(X, \lambda)$ for a given $\lambda \geq 0$.

Extend the program such that it finds all breakpoints of energy function (11), given some image. Afterwards display them, sorted by increasing λ.