Weekly Exercise 10

Dr. Csaba Domokos
Technische Universität München, Computer Vision Group
July 10th, 2017

Mean field approximation (4 points)

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

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

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

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

b) Show that the entropy decomposes as:

$$H(\mathbf{y}) = \sum_{i \in \mathcal{V}} H_i(y_i) \ ,$$

in other words,

$$\sum_{y} p(\mathbf{y}) \log p(\mathbf{y}) = \sum_{i \in \mathcal{V}} \sum_{y_i \in \mathcal{Y}_i} -q_i(y_i) \log q_i(y_i) \ .$$

Solution. a) First remark that

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

Similarly, one can compute

$$\begin{aligned}
\mu_{F,y_F}(q) &= \sum_{\substack{\mathbf{y}'\in\mathcal{Y},\\ \mathbf{y}'_F=\mathbf{y}_F}} q(\mathbf{y}') \\
&= \sum_{\substack{\mathbf{y}'\in\mathcal{Y},\\ \mathbf{y}'_F=\mathbf{y}_F}} \prod_{i\in\mathcal{V}} q_i(y'_i) \\
&= \prod_{i\in N(F)} q_i(y_i) \underbrace{\sum_{\substack{y'_j\in\mathcal{Y}_j,\\ j\in\mathcal{V}\setminus N(F)}} \prod_{k\in\mathcal{V}\setminus N(F)} q_k(y'_k)}_{1} = \prod_{i\in N(F)} q_i(y_i) \; .
\end{aligned}$$

b) We show that given a set of independent random variables, the joint entropy can be calculated as the sum of the entropy of each variable. This property can readily be shown for two independent random variables x, y. That is

$$\begin{aligned}
H(xy) &= \sum_x \sum_y -p(xy)\log p(xy) \\
&= \sum_x \sum_y -p(x)p(y)\big(\log p(x) + \log p(y)\big) \\
&= \sum_x \sum_y -p(x)p(y)\log p(x) + \sum_x \sum_y -p(x)p(y)\log p(y) \\
&= \sum_x -p(x)\log p(x) \sum_y p(y) + \sum_x p(x) \sum_y -p(y)\log p(y) \\
&= \sum_x -p(x)\log p(x) + \sum_y -p(y)\log p(y) \\
&= H(x) + H(y) \; .
\end{aligned}$$

It is then obvious that this property can be generalized to a set of arbitrary number of independent random variables. Now, given the assumption (1), i.e. y_i are independent $\forall i \in \mathcal{V}$, hence

$$H(\mathbf{y}) = \sum_{i\in\mathcal{V}} H(y_i) = \sum_{i\in\mathcal{V}} \sum_{y_i\in\mathcal{Y}_i} -q_i(y_i)\log q_i(y_i) \; .$$

Programming (8 points)

Exercise 2 (**Semantic segmentation by applying a fully connected CRF model**, 8 points)**.** Let us consider the problem of *semantic image segmentation.* Assuming a label set $\mathcal{L}$, we define the *energy function* for $\mathbf{y} \in \mathcal{L}^{\mathcal{V}}$ as

$$E(\mathbf{y}) = \sum_{i\in\mathcal{V}} E_i(y_i) + \sum_{(i,j)\in\mathcal{E}} E_{ij}(y_i, y_j) \; , \tag{2}$$

on a fully connected CRF model, i.e. $\mathcal{E} = \{(i, j) \in \mathcal{V} \times \mathcal{V} \mid i < j\}$.

The goal of the exercise is to implement the **naïve mean field approximation** in order to obtain the results for test images shown in Figure 1. As you have seen in the previous exercises, these test images have been obtained from the MSRC image understanding dataset[1], which contains 21 classes, i.e. $\mathcal{L} = \{1, 2, \ldots, 21\}$. The meaning of the classes are given in the `21class.txt` file.

Figure 1: Test images for semantic image segmentation.

To define the **unary energy functions** E_i for all $i \in \mathcal{V}$, use the `*.c_unary` files provided in the supplementary material (`in2329-supplementary_material_10.zip`). Each test image has its own unary file, specified by the same filename. From each unary file, you can read out a $K \times H \times W$ array of float numbers. The H and W are the image height and width, and $K = 21$ is the number of classes. This array contains the 21-class probability distribution for each pixel. The unary energy functions E_i are then defined as the *negative log-likelihood*:

$$E_i(y_i = l) = -\log(p_l) \ .$$

To define the **pairwise energy functions** E_{ij} for $(i, j) \in \mathcal{E}$, use the *contrast-sensitive Potts model*:

$$E_{ij}(y_i, y_j) = [\![y_i \neq y_j]\!] \Bigg(w_1 \exp\Big(-\frac{|p_i - p_j|^2}{2\theta_\alpha^2} - \frac{|I_i - I_j|^2}{2\theta_\beta^2} \Big) + w_2 \exp\Big(-\frac{|p_i - p_j|^2}{2\theta_\gamma^2} \Big) \Bigg),$$

where p_i stands for the location of the pixel i and I_i denotes its intensity vector, moreover the parameters are chosen as

$$w_1 = 10,\ w_2 = 3,\ \theta_\alpha = 80,\ \theta_\beta = 13,\ \text{and}\ \theta_\gamma = 3\ .$$

[1] https://www.microsoft.com/en-us/research/project/image-understanding/