Massachusetts Institute of Technology
Department of Electrical Engineering and Computer Science
6.438 Algorithms for Inference
Fall 2018

Problem Set 5

Issued: Thursday, Nov. 01, 2018 **Due:** Tuesday, Nov. 08, 2018

Problem 5.1

In this problem, we show that locally consistent marginals need not be globally consistent - that is, they need not be the marginals of any probability distribution. Consider the following factor graphical model over $\{0,1\}^3$:

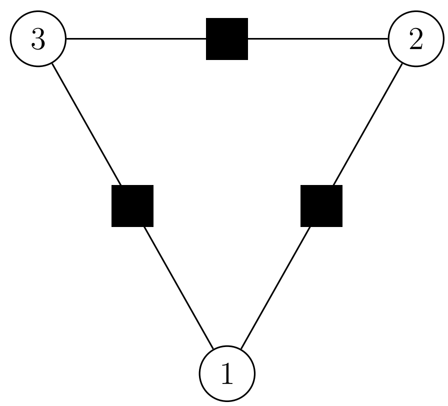


Figure 1: Factor Graph (G)

We consider the following local marginals - b_i ($i = 1,2,3$) for node i and b_{ij} $i \neq j$ for the factor between nodes i and j. Notation: $b_1 = \begin{bmatrix} a \\ b \end{bmatrix}$ means $b_1(0) = a$ and $b_1(1) = b$. $b_{12} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ means $b_{12}(0,0) = a$, $b_{12}(0,1) = b$, $b_{12}(1,0) = c$ and $b_{12}(1,1) = d$,

Consider the 'marginals'/'beliefs': $b_1 = b_2 = b_3 = b_4 = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$

$$b_{12} = b_{32} = \begin{bmatrix} 0.49 & 0.01 \\ 0.01 & 0.49 \end{bmatrix}$$

$$b_{31} = \begin{bmatrix} 0.01 & 0.49 \\ 0.49 & 0.01 \end{bmatrix}$$

(a) Show that the given 'marginals'/'beliefs' are locally consistent.

(b) Show that the locally consistent marginals are not globally consistent : that is, they are the not the marginals of any distribution $p(x_1, x_2, x_3)$ on $\{0, 1\}^3$.

Problem 5.2

In this exercise, you will construct an undirected graphical model for the problem of segmenting foreground and background in an image, and use loopy belief propagation to solve it.

Load the image `flower.bmp` into MATLAB using `imread`. (The command `imshow` may also come in handy in MATLAB, you may alternatively use your favorite language.) Partial labeling of the foreground and background pixels are given in the mask images `foreground.bmp` and `background.bmp`, respectively. In each mask, the white pixels indicate positions of representative samples of foreground or background pixels in the image.

Let $\mathbf{y} = \{y_i\}$ be an observed color image, so each y_i is a 3-vector (of RGB values between 0 and 1) representing the pixel indexed by i. Let $\mathbf{x} = \{x_i\}$, where $x_i \in \{0, 1\}$ is a foreground(1)/background(0) labeling of the image at pixel i. Let us say the probabilistic model for $\mathbf{x}$ and $\mathbf{y}$ given by their joint distribution can be factored in the form

$$p(\mathbf{x}, \mathbf{y}) = \frac{1}{Z} \prod_i \psi(x_i, y_i) \prod_{(j,k) \in E} \psi(x_j, x_k)$$

where E is the set of all pairs of adjacent pixels in the same row or column.

Suppose that we choose

$$\psi(x_j, x_k) = \begin{cases} 0.9, & \text{if } x_j = x_k \\ 0.1, & \text{if } x_j \neq x_k \end{cases}$$

This encourages neighboring pixels to have the same label — a reasonable assumption.

Suppose further that we use a simple model for the conditional distribution $p(y_i | x_i)$:

$$p(y_i | x_i = \alpha) \propto \frac{1}{(2\pi)^{3/2} (\det \mathbf{\Lambda}_\alpha)^{1/2}} \exp\left[-\frac{1}{2}(y_i - \mu_\alpha)^T \mathbf{\Lambda}_\alpha^{-1} (y_i - \mu_\alpha)\right] + \epsilon$$

for $y_i \in [0,1]^3$. That is, the distribution of color pixel values over the same type of image region is a modified Gaussian distribution, where ϵ accounts for outliers. Set $\epsilon = 0.01$ in this problem.

(a) Sketch and label the undirected graph with respect to which the distribution $p(\mathbf{x}, \mathbf{y})$ is Markov. What are the potential functions $\psi(x_i, y_i)$?

(b) Compute $\mu_\alpha, \mathbf{\Lambda}_\alpha$ from the labeled masks by finding the sample mean and covariance of the RGB values of those pixels for which the label $x_i = \alpha$ is known.

(c) We want to run the sum-product algorithm on the graph iteratively to find (approximately) the marginal distribution $p(x_i|\mathbf{y})$ at every i.

Write the local message update rule for passing the message $m_{j \to k}(x_k)$ from x_j to x_k, in terms of the messages from the other neighbors of x_j, the potential functions, and some arbitrary scaling constant.

Then write the final belief update rule on x_i, that is, the marginal computation in terms of the messages from all neighbors of x_i, the potential functions, and some arbitrary scaling constant.

(d) Implement the sum-product algorithm for this problem. There are four directional messages: down, up, left, and right, coming into and out of each x_i (except at the boundaries). Use a parallel update schedule, so all messages at all x_i are updated at once. Run for 30 iterations (or you can state and use some other reasonable termination criterion). You should renormalize the messages after each iteration to avoid overflow/underflow problems.

After the marginal distributions at the pixels are estimated, visualize their expectation (that would be the BLS estimates of x_i based on $\mathbf{y}$ if the inference were exact). Where are the beliefs "weak"? Where did the loopy belief propagation converge first and last?

(e) **(Practice)** Convert the graph in this problem to a factor graph and write the belief propagation messages passed between variable nodes and factor nodes.

Problem 5.3

In this problem, we consider a variational approximation to a multivariate Gaussian distribution using scalar Gaussian factors. Consider a Gaussian distribution $p_{\mathsf{x}}(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\mu, J^{-1})$ for which we identify the two components of the vector $\mathbf{x} \in \mathbb{R}^2$:

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \qquad \mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \qquad J = \begin{pmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{pmatrix}$$

where $J_{12} = J_{21}$ and each element of J is a scalar. The distribution is also assumed non-degenerate, i.e. J is positive definite.

We want to approximate the distribution p with a factorized distribution $q_{\mathsf{x}}(\mathbf{x}) = b_1(x_1)b_2(x_2)$, for some b_1 and b_2 which we do not a priori constrain to be of a particular class of distributions.

Analogously to the discrete case, if we drop reference to any graphical structure in the distribution (i.e. consider the graph to be fully connected) and consider the two-factor case, we have

$$b_1(x_1) = K' \exp\left\{E_{b_2}\left[\ln p_{\mathsf{x}}(\mathsf{x})\right]\right\} \tag{1}$$

where E_{b_2} denotes the expectation with respect to the variational distribution over node 2, i.e.

$$E_{b_2}\left[f(\mathsf{x}_2)\right] = \int f(x_2)b_2(x_2)\ dx_2$$

Note that we have replaced the unnormalized potential ψ with the joint probability distribution p_{x}, which only affects the normalization constant K'. The expression for the optimal update to b_2 is analogous to Eq.(1) with the factor b_1 switched in to the expression.

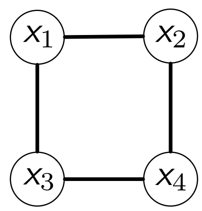
(a) Using the formula in Eq.(1), derive the optimal mean field updates for factors b_1 and b_2. You should show that the updated factors are themselves Gaussian densities, and use that observation and the technique of completing the square to find the normalization constant K' by comparison to the standard Gaussian form.

(b) Show that the update equations are satisfied when the variational mean is equal to the true mean. Discuss why the optimal variational approximation makes intuitive sense.

Problem 5.4 (practice)

(a) In our discussion of variational inference, we minimized $D(b_{\mathsf{x}} \| p_{\mathsf{x}})$ with respect to b_{x}. In this part, we consider the other direction. Suppose that calculating the marginals of $p_{\mathsf{x}}(\mathbf{x})$ is computationally intractable. Could it be tractable to find a fully factorized approximating distribution $b(\mathbf{x}) = \prod_{i=1}^{N} b_i(x_i)$ by minimizing $D(p_{\mathsf{x}} \| b_{\mathsf{x}})$ with respect to $b_{\mathsf{x}}(\mathbf{x})$?

For the remainder of this question, we return to the usual variational objective function $D(b_{\mathsf{x}} \| p_{\mathsf{x}})$. We restrict our attention to the following undirected graphical model:



Associated with each node in the graph is a random variable x_i, and the joint distribution for these variables is of the form

$$p_{\mathsf{x}_1,\mathsf{x}_2,\mathsf{x}_3,\mathsf{x}_4}(x_1, x_2, x_3, x_4) = \frac{1}{Z} \prod_{(i,j)\in\mathcal{E}} \psi_{i,j}(x_i, x_j) \tag{2}$$

where Z is the partition function, $\psi_{i,j}(x_i, x_j)$ is a clique potential for edge (i, j), and the set of edges is $\mathcal{E} = \{(1, 2), (1, 3), (2, 4), (3, 4)\}$.

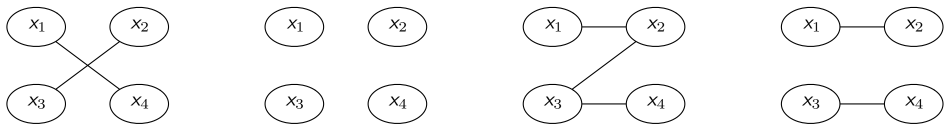
(b) Determine whether the following statement is TRUE or FALSE. If you answer TRUE, be sure to provide a proof; if you answer FALSE, be sure to provide a counterexample.

STATEMENT: Consider the family of approximating distributions

$$b_{\mathsf{x}}(\mathbf{x}) = \prod_{(i,j)\in\mathcal{E}_b} b_{i,j}(x_i, x_j)$$

for some alternative edge set $\mathcal{E}_b$. Then the optimal variational approximation in this family has the property that $b_{i,j}(x_i, x_j) \propto \psi_{i,j}(x_i, x_j)$ for all $(i, j) \in \mathcal{E}_b \cap \mathcal{E}$.

(c) Consider families of approximating distributions $b_{\mathsf{x}}(\mathbf{x})$ described by the four graphical models shown below.



I II III IV

Suppose that we have a black-box procedure that can obtain the best variational approximations within each of these families. Rank the graphs in descending order of the quality of their resulting approximations as obtained from the black-box procedure, and explicitly specify when there is a tie. Be sure to fully justify your reasoning.

Hint: If graph $\mathcal{G}_1$ is contained in $\mathcal{G}_2$, try deriving the variational inference update rule for the graph $\mathcal{G}_2$. Argue that if the updates are contained in the set of approximating distributions given by $\mathcal{G}_1$, then there is a tie.