Probabilistic Inference & Learning

Exercise Sheet #10

Graphical Models

1. **The Sum-Product Algorithm:** Lecture 17 introduced the Sum-Product Algorithm for tree-structured factor graphs: Initialize all Leaf-Node-Messages from variables to factors, and from factors to messages, respectively, as $\mu_{x_i \to f_j}(x_i) = 1$ and $\mu_{f_j \to x_i}(x_i) = f_j(x_i)$, respectively, then iteratively move towards the root of the tree and compute variable-to-factor and factor-to-variable messages, respectively, as

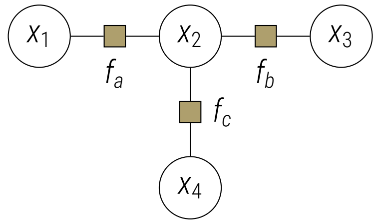
$$\mu_{f_j \to x_i}(x_i) = \sum_{x_1,\dots,x_M \in \mathrm{ne}(f_j)\setminus x_i} f_j(x_1,\dots,x_M,x_i) \prod_{m \in \mathrm{ne}(f_j)\setminus x_i} \mu_{x_m \to f_i}(x_m) \tag{1}$$

$$\mu_{x_m \to f_j}(x_m) = \prod_{\ell \in \mathrm{ne}(x_m)\setminus f_j} \mu_{f_\ell \to x_m}(x_m). \tag{2}$$

Then pass analogous messages from the root to the leaves, and compute local marginals by multiplying all incoming factor-to-variable nodes and normalizing the resulting local function.

Now consider the graph shown on the right, which encodes the (normalized) joint distribution

$$p(x_1,\dots,x_4) = f_a(x_1,x_2) \cdot f_b(x_2,x_3) \cdot f_c(x_2,x_4).$$



Let us designate x_3 as the root of the graph, and (x_1, x_2, x_4) as leaves. Explicitly write down all 12 messages (6 toward the root, 6 toward the leaves) constructed by the sum-product algorithm (many of them are of rather simple form). *30 points*

Use these messages to

(a) confirm that the product of incoming messages gives the correct marginal for x_2, i.e. that

$$\mu_{f_a \to x_2}(x_2) \cdot \mu_{f_b \to x_2}(x_2) \cdot \mu_{f_c \to x_2}(x_2) = \sum_{x_1,x_3,x_4} p(x_1,x_2,x_3,x_4) = p(x_2).$$

10 points

(b) confirm the analogous result for the root x_3 and one of the leaves, x_1. *10 points*

(c) show that $p(x_1,x_2) = f_a(x_1,x_2) \cdot \prod_{i=1,2} \mu_{x_i \to f_a}(x_i)$. *10 points*

2. Consider a tree-structured factor graph over discrete variables, and suppose we wish to evaluate the *joint* distribution $p(x_a, x_b)$ associated with two variables x_a and x_b that do not belong to a common factor. Define a procedure for using the sum-product algorithm to evaluate the joint, in which one of the variables is successively clamped to each of its allowed values (slide 8 in lecture 17 explains what that means). *20 points*

3. Consider two discrete variables x and y, each having three possible states (e.g. $x, y \in \{0; 1; 2\}$). Construct a joint distribution $p(x, y)$ (a 3×3 table of non-negative real numbers that sum to 1) over these variables having the property that the value $\hat{x}$ that maximizes the *marginal* $p(x)$, along with the value $\hat{y}$ that maximizes the marginal $p(y)$, together have probability zero under the joint distribution, i.e. $p(\hat{x}, \hat{y}) = 0$. Show that such a situation can *not* arise when x, y are binary variables (i.e. they can only take one of two possible values). *20 points*