CSC 446 Notes: Lecture 9

Typed by Hao Luo

February 27, 2012

1 Notations

We first formalize some notations. Conditional probability of X_i is denoted as:

$$P(X_i = x_i | X_1, X_2, \ldots, X_{i-1}, X_{i+1}, \ldots, X_N) = \frac{1}{Z} \sum P(X_1, X_2, \ldots, X_{i-1}, X_i = x_i, X_{i+1}, \ldots, X_N)$$

Conditional probability of X_i with respect to X_j and X_k is denoted as:

$$P(X_i | X_j, X_k) = \frac{1}{Z} \sum_{X_{t, t \neq i,j,k}} P(X_i, X_j, X_k) \tag{1}$$

Conditional independence is denoted as:

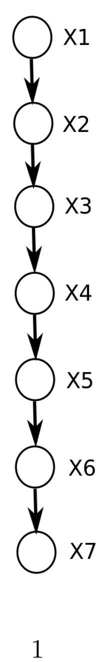
$$X_i \perp\!\!\!\perp X_a | X_j, X_k \Leftrightarrow P(X_i | X_a, X_j, X_k) = P(X_i | X_j, X_k)$$

In (1), note there are 2^{N-3} terms in the summation. Because of conditional independence, the probability can be simplified as

$$P(X_i | X_1, X_2, \ldots, X_{i-1}, X_{i+1}, \ldots, X_N) = P(X_i | Parent(X_i))$$

Only 2^M terms are involved, where M is the max number of parents.

2 Example



To compute $P(X_7|X_2)$, we have

$$P(x_7|x_2) = \frac{1}{Z}\sum_{X_3}\sum_{X_4}\sum_{X_5}\sum_{X_6} P(X_3|x_2)P(X_4|X_3)P(X_5|X_4)P(X_6|X_5)P(x_7|X_6)$$

Suppose every variable X_i is binary, then the summation has $2^4 = 16$ terms. On the other hand, we can use the same trick in dynamic programming by recording every probabilities we have computed for reuse. For example, in above example, if we define

$$f_5(x_5) = \sum_{X_6} P(X_6|X_5 = x_5)P(x_7|X_6) \tag{2}$$

$$f_4(x_4) = \sum_{X_5} P(X_5|X_4 = x_4)f_5(X_5) \tag{3}$$

$$f_3(x_3) = \sum_{X_4} P(X_4|X_3 = x_3)f_4(X_4) \tag{4}$$

$$f_2(x_2) = \sum_{X_3} P(X_3|X_2 = x_2)f_3(X_3) \tag{5}$$

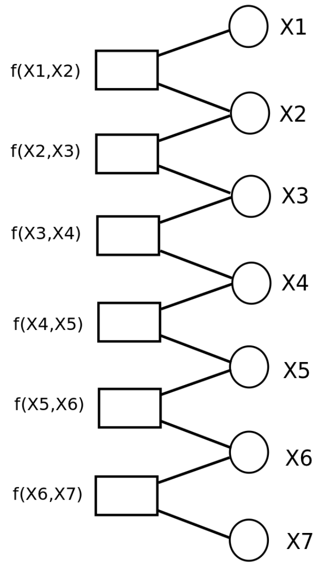
Then the probability above can be computed as

$$\begin{aligned}
P(X_7 = x_7|X_2 = x_2) &= \frac{1}{Z}\sum_{X_3}\sum_{X_4}\sum_{X_5}\sum_{X_6} P(X_3|X_2 = x_2)P(X_4|X_3)P(X_5|X_4)P(X_6|X_5)P(X_7 = x_7|X_6) && (6)\\
&= \frac{1}{Z}\sum_{X_3}\sum_{X_4}\sum_{X_5} P(X_3|x_2)P(X_4|X_3)P(X_5|X_4)f_5(X_5) && (7)\\
&= \frac{1}{Z}\sum_{X_3}\sum_{X_4} P(X_3|x_2)P(X_4|X_3)f_4(X_4) && (8)\\
&= \frac{1}{Z}\sum_{X_3} P(X_3|x_2)f_3(X_3) && (9)\\
&= \frac{1}{Z}f_2(x_2) && (10)
\end{aligned}$$

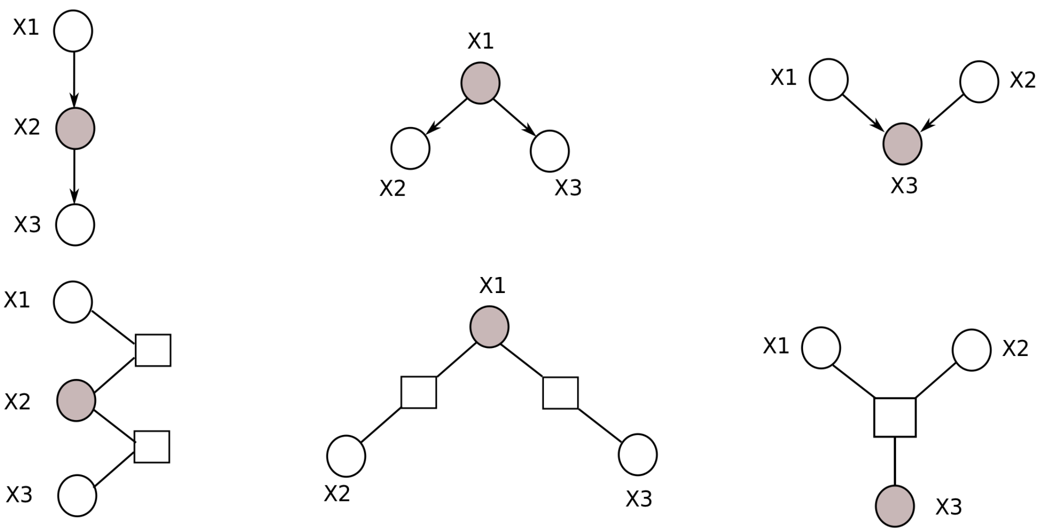
There are 4 sums and each sum needs to compute 2x2 probabilities, so a total of 16 steps.

3 Factor Graph

Factor graph is an undirected bipartite graph. There are two types of vertex in a factor graph, factor vertices and variable vertices. Factor vertices correspond to the function f_m in the above example, and each distinct variable vertex corresponds to a distinct variable. If factor function f_m is a function of X_i, its factor vertex is connected to X_i. So the factor graph for above example is,



For more examples,



Note that in the figures above, factor graphs illustrate that the shadowed variable nodes block the information flow from one variable node to another except the last one. In the last example, the two parent nodes are independent, although this cannot be seen from the graph structure. However, the blockage can be read from the table of the factor node in the center. Also note that the last two graphs have same undirected shape, but their factor graphs are different.

4 Message Passing (Belief Propagation)

We assume that the factor graph is a tree here. For each variable vertex n and its neighboring factor vertex f_m, the information propagated from n to f_m is,

$$q_{n \to m}(X_n) = \prod_{m' \in M(n)/m} r_{m' \to n}(X_n)$$

where $M(n)$ is the set of factors touching X_n. This table contains the information propagated from variable n to its neighboring factor vertex f_m. For each factor vertex f_m and its neighboring variable n, the information propagated from f_m to n is,

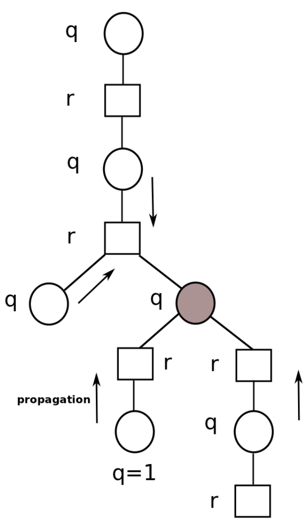
$$r_{m\to n}(X_n) = \sum_{\overrightarrow{X_m}\setminus X_n} f_m(\overrightarrow{X_m}) \prod_{n'\in N(m)\setminus\{n\}} q_{n'\to m}(X_{n'})$$

where $N(m)$ is the set of variables touching f_m. $\sum_{\overrightarrow{X_m}\setminus X_n}$ is the sum is over all variables connected to f_m except X_n. This table contains the information propagated from factor f_m to its neighbor variable n. Note that if variable vertex n is a leaf, $q_{n\to m} = 1$, and if factor vertex m is a leaf, $r_{m\to n} = f_m(X_n)$.

The procedure of message passing or belief propagation is first to propagate the information from leaf vertices to the center (i.e., from leaves to internal nodes) by filling in the tables for each message. Once all the messages variable x_n have been computed, the marginal probability of x_n is computed by combining the incoming messages:

$$P(X_n) = \frac{1}{Z}\prod_{m\in M(n)} r_{m\to n}(X_n)$$

To compute marginal probabilities for all variables, the information is propagated from center back to leaves.



5 Running Time

Suppose in a factor graph, there are N variable vertices and M factor vertices. For every variable vertex n, $|M(n)| < k$ and for every factor vertex f_m, $|N(m)| < l$, the running time is,

$$\mathcal{O}((N+M)(k+l)2^{l-1})$$