Machine Learning Course - CS-433

Graphical Models – Factor Graphs

Dec 20, 2016

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Outline

In the last lecture we looked at Bayes nets, which are a graphical representation of a probability distribution via DAGs (directed acyclic graphs). Today we will see another useful graphical representation, called *factor graphs*. One of the main advantages of factor graphs is that they are the natural setting to discuss the sum-product algorithm, an algorithm that computes the marginal of a function, assuming that the factor graph of this function is a tree.

Basic Definition

As before, let $x_1, \cdots, x_D$ be the set of variables. These could be random variables as before, or just the variables of a multi-dimensional function. In particular, the function we consider neither has to be normalized nor positive. Assume that the function has the following factorized form

$$f(x_1, \cdots, x_D) = \prod_{a \in A} f_a(x_a).$$

By convention, the variables are indexed by integers and the factors are indexed by lower case Roman letters. There are $|A|$ factors and each factor involves a subset of the variables. This subset is denoted by x_a.
A factor graph represents such a factorization in terms of a bipartite graph. There is a node for each variable x_i (so D such nodes in total) and there is one node for each factor (so $|A|$ such nodes in total). Further, there is an edge from variable i to factor a if the variable x_i participates in the factor a.

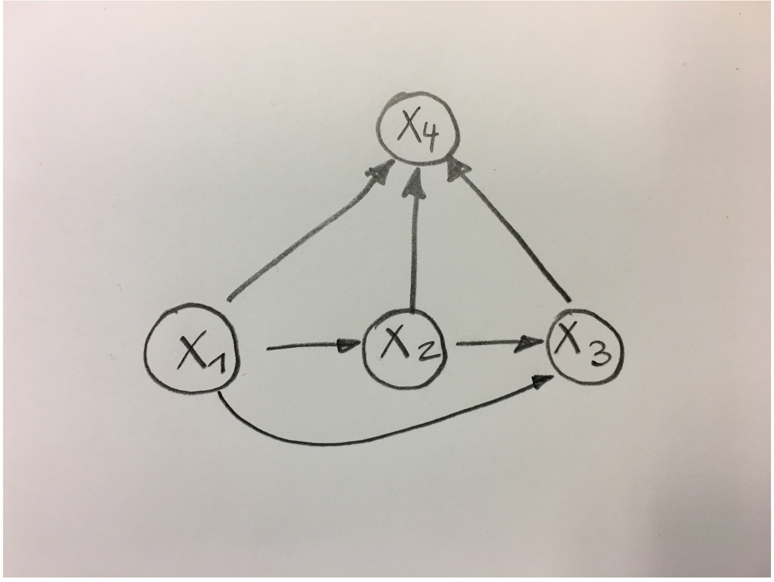

Figure 1: A Bayes net corresponding to the factorization written in (1).

Let us go back to the first example we have seen in the context of Bayes nets and let us look at the generic factorization of a probability distribution with $D = 4$. We have

$$p(X_1,X_2,X_3,X_4) = p(X_1)p(X_2|X_1)p(X_3|X_1,X_2)p(X_4|X_1,X_2,X_3). \quad (1)$$

To facilitate the comparison with factor graphs, let us reprint the Bayes net for this case. It is shown in Figure 1. Figure 2 shows the factor graph corresponding to (1).

A particularly important case is when the factor graph is a *tree*, i.e., the graph has no cycles.

Note that there is not a unique factor graph associated to a given function f. We have a considerable degree of freedom. E.g., we can join all factors and draw a "trivial" factor graph where we have only a single factor that is connected to all variables. Such a graph is of course a tree.

But we will soon see that the complexity of computing marginals is exponential in $\max_{a \in A} |x_a|$, the size of the largest subset

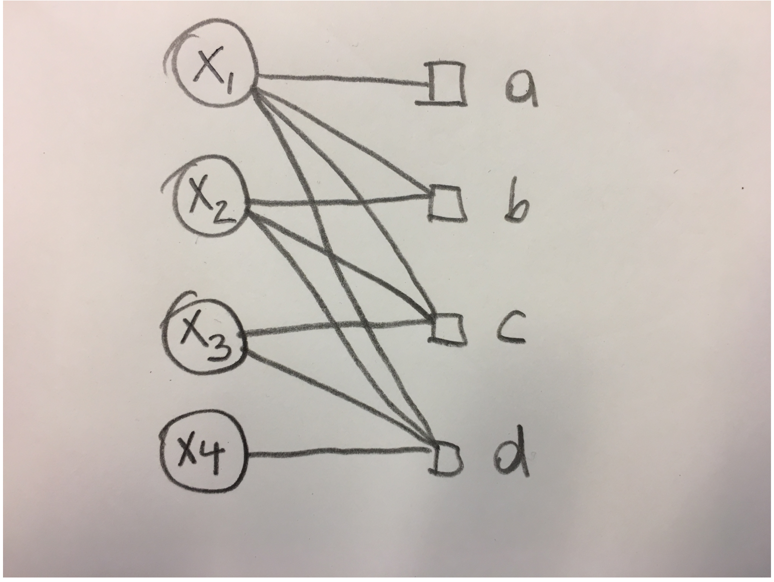

Figure 2: A factor graph corresponding to the factorization written in (1).

that is involved in any factor. Therefore, ideally we would like to have as many factor nodes as possible (so that each node has a small degree) but keep the factor graph a tree.

From Bayes Nets to Factor Graphs

It is pretty straighforward to go from a Bayes Net to a Factor Graph. In the simplest case each node in the Bayes net corresponds to a variable. So each such node will give rise to a node in the factor graph. (In the general case such a node could correspond to a whole set of variables and so such node might give rise to several nodes in the factor graph.) But each node in the Bayes net also corresponds to a factor and so gives rise to one factor in the factor graph. In the factor graph this node is connected to the parents of the corresponding node in the Bayes net.

The following claim is not very difficult to check. If the

Bayes net is a tree when seen as an undirected graph, then the factor graph is also a tree. The converse is not true. It is easy to construct a Bayes net that is not a tree but where the corresponding factor graph is a tree.

To see the first claim, let us show that if the factor graph has a cycle then so has the Bayes net. Let $f_{a_1}, x_1, \cdots, f_{a_k}, x_k$ denote a cycle in the factor graph, where $x_i \in a_i$, $i = 1, \cdots, a_k$, and $x_k \in a_1$. Then $x_1, \cdots, x_k$ is a cycle in the undirected Bayes net since x_i must be a parent or child of x_{i+1} for each $i = 1, \cdots, k-1$, and x_1 and x_k must be related in this sense as well.

To see that the converse is not true, consider the factorization

$$p(x_1)p(x_2|x_1)p(x_3|x_2)p(x_4|x_1, x_3).$$

This gives one cycle of length 4 in the undirected Bayes Net but no cycle in the factor graph.

Sum-Product Algorithm

See additional slides.

Loopy Belief Propagation