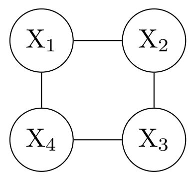
INFERENCE AND REPRESENTATION: HOMEWORK 2

STUDENT: MARGARITA BOYARSKAYA

1.



Let's examine the statement

$$P(X_1 = 0, X_2 = 0, X_3 = 1, X_4 = 0) = 0,$$

true for the given P. If we assume that P allows a factorization according to graph H (above), it follows that at least one of the potentials $\psi_1(X_1 = 0, X_2 = 0), \psi_1(X_2 = 0, X_3 = 1), \psi_1(X_3 = 1, X_4 = 0), \psi_1(X_4 = 0, X_1 = 0)$ must be equal to zero. However, it is easy to see that:
If $\psi_1(X_1 = 0, X_2 = 0) = 0$ or $\psi_1(X_4 = 0, X_1 = 0) = 0$, then $P(0, 0, 0, 0) = 0$;
If $\psi_1(X_2 = 0, X_3 = 1) = 0$, then $P(0, 0, 1, 1) = 0$;
If $\psi_1(X_1 = 0, X_2 = 0) = 0$, then $P(1, 1, 1, 0) = 0$,
each of which contradicts the formulation of the problem. $\square$

2. For the Boltzmann machine, the probability take form of:

$$P(x_1, ..., x_n) \propto exp(\sum_{(i,j)\in E} w_{i,j} x_i x_j - \sum_{i \in V} u_i x_i).$$

Let's denote $x_j^{'} = 2x_i - 1$, $x_i \in \{0, 1\}$. Denote also $w_{ij}^{'} = \frac{w_{i,j}}{4}$, $u_i^{'} = \frac{u_i}{2} - \sum_{j \in N} \frac{w_{i,j}}{4}$.
Then,

$$\begin{aligned}
P &\propto \sum_{i,j \in E} 4w_{i,j}^{'} \frac{x_i^{'}+1}{2} \frac{x_j^{'}+1}{2} - \sum_{i \in V} (2u_i^{'} + \sum_{j \in N} \frac{w_{i,j}^{'}}{2}) \frac{x_i^{'}+1}{2}, \\
&\propto -\sum_{i \in V} \Big[2(u_i^{'} + \sum_{j \in N} w_{i,j}^{'}) \frac{x_i^{'}+1}{2} - \sum_{j \in N} (2w_{i,j}^{'} \frac{x_i^{'}+1}{2} \frac{x_j^{'}+1}{2}) \Big] \\
&\propto -\sum_{i \in V} \Big[2u_i^{'} \frac{x_i^{'}+1}{2} - \sum_{j \in N} (2w_{i,j}^{'} \frac{x_i^{'}+1}{2} \frac{x_j^{'}+1}{2} - \frac{w_{i,j}^{'}}{2} \frac{x_i^{'}+1}{2}) \Big] \\
&\propto -\sum_{i \in V} u_i^{'} x_i^{'} - \sum_{j \in N} \frac{w_{i,j}^{'}}{2} x_i^{'} x_j^{'},
\end{aligned}$$

which is just a log probability expression for an undirected graph with the Ising model distribution:

$$P(x_1^{'}, ..., x_n^{'}) \propto \frac{1}{Z} exp\Big(\sum_{(i,j)\in E} w_{i,j}^{'} x_i^{'} x_j^{'} - \sum_{i \in V} u_i^{'} x_i^{'} \Big). \quad \square$$

3. To demonstrate that any discrete variables MRF can be converted into a pairwise Markov random field, we shall first find a pairwise form representation of the function $\psi_{123}(X_1, X_2, X_3)$. If variables X_i, X_2, X_3 take values in measurable spaces E_1, E_2, E_2, let us define a new variable Y: $\Omega \to E_1 \times E_2 \times E_3$ that takes values of the form (y_1, y_2, y_3), $y_i \in E_i$. We shall define functions

$$\psi_{Yi}(Y, X_i) := \begin{cases} (\psi_{123}(y_1, y_2, y_3))^{1/3} \text{ if } y_i = x_i \\ 0 \text{ otherwise} \end{cases},$$

for $i = 1, 2, 3$. This construction correctly reflects the relationship between X_1, X_2, and X_3, since:

$$\psi_{123}(X_1, X_2, X_3) = \sum_y \psi_{Y1}(Y, X_1)\psi_{Y2}(Y, X_2)\psi_{Y3}(Y, X_3).$$

Thus we learned to obtain the pairwise representation by introducing new variables for every triplewise formula. To generalize this procedure, we must convert every higher-order clause to a set of triplewise formulas. We can add auxiliary third node to each pair of nodes and find triple-clique potential functions, following logic similar to the above.

4. (a).i. Multivariate Gaussian. The probability density function is:

$$P(X|\mu) = \frac{1}{(2\pi)^{d/2}} exp\Big(-\frac{1}{2}x^T x + \mu^T x - \frac{1}{2}\mu^T \mu \Big),$$

thus the canonical parameters of the exponential family are:

$$h(x) = \frac{1}{(2\pi)^{d/2}} exp(-\frac{1}{2}x^T x); \quad T(x) = x; \quad \eta = \mu; \quad A(\eta) = \frac{1}{2}\eta^T \eta.$$

ii. Dirichlet. The density is:

$$P(\theta|\alpha) = \prod_i \frac{1}{\theta_i} exp\Big(\sum_i \alpha_i log\theta_i - log\frac{\prod \Gamma(\alpha_i)}{\Gamma \Sigma \alpha_i} \Big).$$

An exponential family is determined by parameters

$$h(\theta) = \prod_i \frac{1}{\theta_i}; \quad T(\theta) = [log\theta_1, ..., ; og\theta_K]^T; \quad \eta = \alpha; \quad A(\eta) = log\frac{\prod \Gamma(\eta_i)}{\Gamma \Sigma \eta_i}.$$

iii. Lognormal. The density is expressed as:

$$P(Y|\sigma) = \frac{1}{\sqrt{2\pi y}} exp\Big(-\frac{1}{2\sigma^2} log^2 y - \frac{1}{2} log(\sigma^2) \Big).$$

Thus, the exponential family is parametrized by:

$$h(y) = \frac{1}{\sqrt{2\pi y}}; \quad T(y) = log^2 y; \quad \eta = -\frac{1}{2\sigma^2}; \quad A(\eta) = -\frac{1}{2} log(-2\eta).$$

iv. Maxwell-Boltzmann. The probability distribution corresponding to parameter $a > 0$ has the probability density function:

$$P_a(x) = \sqrt{\frac{2}{\pi}} \frac{1}{a^3} exp\Big(- \frac{x^2}{2a^2}\Big).$$

The function $h(x)g(\theta)exp(\eta(\theta)T(x))$ with

$$h(x) = x^2; \quad g(\theta) = \sqrt{\frac{2}{\pi}}\theta^{-3}; \quad T(x) = x^2; \quad \eta(\theta) = -\frac{1}{2a^2}$$

is an exponential family.

(b)

$$P(Y = 1|x, \alpha) = \frac{1}{1 + exp\Big(- \alpha_0 - \sum \alpha_i \psi_i(x_1, ..., x_k)\Big)}.$$

A family of conditional densities for Y of the form above form will be in acontradiction with conditional densities for X given Y unless the conditional distributions for X are in an exponential family whose statistics are the link functions ψ_i.

5. Any node can be chosen and treated as a root, defining a directed graph with edges pointing away from this node. Let's consider eliminating variables bottom-up, up to node i. The resulting factor is denoted $m_{ji}(x_i)$. For an undirected graph, we define the sum-product equation for each message that propagates from x_j to x_i, where ψ_j is the node potential. Node j receives messages from its neighbors and sums over them to pass to the node i. Here the product signifies all "previous" information, the sum adds in more relevant information while eliminating x_j.

$$m_{ji}(x_i) = \sum_{x_j} \Big(\psi(x_j)\psi(x_i, x_j) \prod_{k\in(N(j)\setminus i)} m_{kj}(x_j)\Big)$$

A node passes a message after it receives messages from its neighbors.
It can be inductively proven (with induction in the number of nodes in the tree) that fixed-point messages can be used to compute the marginals in the form

$$p(x_i) \propto \psi(x_i) \prod_{k\in N(i)} m_{ki}(x_i).$$

For $n = 1$ the above is trivial. Let us now assume that the inductive claim holds for $n1$-th step. For an $n-$node tree reorder the nodes assigning them numbers in such way that the $n-$th node is a leaf that neighbors with the first node. The message sent from n-th to 1-st node is:

$$m_{n1}(x_i) = \sum_{x_n} \psi_n(x_n)\psi_{1n}(x_1, x_n).$$

Factorizing over potentials, we get

$$p(x_1, ..., x_{n-1}) = \sum_{x_n} p(x_1, ..., x_n) \propto m_{n1}(x_1) \prod_{i=1}^{n-1} \psi_i(x_i) \prod_{(k,l)\in E\setminus(1,n)} \psi_{kl}(x_k, x_l).$$

Note that all messages to be sent between all other nodes can be obtained via sum-product algorithm, and the potential on the first node will take the form:

$$\psi_1^{'}(x_1) = m_{n1}(x_1)\psi_1(x_1).$$

We must demonstrate that for node n it holds that $p(x_n) \propto \psi_n(x_n)m_{1n}(x_m)$. By inductive claim, for nodes $1, ..., n-1$ marginal probabilities are proportional to the product of incoming messages and the node potential. Thus, we have:

$$m_{1n} \propto \sum_{x_1} p(x_1)\frac{\psi_{1n}(x_1, x_n)}{m_{n1}(x_1, x_n)};$$

$$\psi_n(x_n)m_{1n}(x_n) \propto \psi_n(x_n)\sum_{x_1}\frac{p(x_1)\psi_{1n}(x_1, x_n)}{\sum_{x_n'}\psi_n(x_m')\psi_{1n}(x_1, x_n')}.$$

Noticing that $p(x_n|x_1) = \frac{p(x_1,x_n)}{p(x_1)} = \frac{\psi_n\psi_{1n}(x_1,x_n)}{\sum_{x_n'}\psi_n(x_n')\psi_{1n}(x_1,x_n')}$ we obtain, finally:

$$\psi_n(x_n)m_{1n}(x_n) \propto \sum_{x1} p(x_1)p(x_n|x_1) = p(x_n).$$