Inference and Representation, Fall 2017

Problem Set 2: Undirected graphical models & Modeling exercise

Selected solutions

---

1. Exercise 4.1 from Koller & Friedman (requirement of positivity in Hammersley-Clifford theorem; see page 116).

2. Recall that an Ising model is given by the distribution

$$p(x_1, \cdots, x_n) = \frac{1}{Z} \exp\Big(\sum_{(i,j)\in E} w_{i,j} x_i x_j - \sum_{i \in V} u_i x_i \Big), \tag{1}$$

   where the random variables $X_i \in \{-1, +1\}$. Related to the Ising model is the *Boltzmann machine*, which is parameterized the same way (i.e., using Eq. 1), but which has variables $X_i \in \{0, 1\}$. Here we get a non-zero contribution to the energy (i.e. the quantity in the parentheses in Eq. 1) from an edge (i, j) only when $X_i = X_j = 1$.

   Show that a Boltzmann machine distribution can be rewritten as an Ising model. More specifically, given parameters $\vec{w}, \vec{u}$ corresponding to a Boltzmann machine, specify new parameters $\vec{w}', \vec{u}'$ for an Ising model and prove that they give the same distribution $p(\mathbf{X})$ (assuming the state space $\{0, 1\}$ is mapped to $\{-1, +1\}$).

3. Give a procedure to convert any Markov network on discrete variables into a pairwise Markov random field. In particular, given a distribution $p(\mathbf{X})$, specify a new distribution $p'(\mathbf{X}, \mathbf{Y})$ which is a pairwise MRF, such that $p(\mathbf{x}) = \sum_{\mathbf{y}} p'(\mathbf{x}, \mathbf{y})$, where $\mathbf{Y}$ are any new variables added.

   Clarification: Assume that the input is specified as full tables specifying the value of the potential for every assignment to the variables for each potential. The new pairwise MRF must have a description which is polynomial in the size of the original MRF.

   Hint: First consider a simple case, such as a MRF on 3 binary variables with a single potential function for the 3 variables, i.e. $p(\mathbf{X}) \propto \psi_{123}(X_1, X_2, X_3)$. Introduce a new variable Y with $2^3 = 8$ states and let $p'(\mathbf{X}, Y) \propto \psi_Y(Y)\psi_{1Y}(X_1, Y)\psi_{2Y}(X_2, Y)\psi_{3Y}(X_3, Y)$. Figure out how to set the new potential functions $\psi_Y(Y), \psi_{1Y}(X_1, Y), \psi_{2Y}(X_2, Y)$ and $\psi_{3Y}(X_3, Y)$ so as to have $p(\mathbf{x}) = \sum_y p'(\mathbf{x}, y)$ for all assignments $\mathbf{x}$.

   Answer:

   Construct a new Markov random field defined on the original variables $\mathbf{X}$ in addition to a new variable Y_C for each potential ψ_C of the original Markov network. Denote ψ_C's scope as $C = \{X_{i_1}, \ldots, X_{i_{|C|}}\}$. Each state of Y_C corresponds to a choice of states for $X_{i_1}, ..., X_{i_{|C|}}$ (e.g., if each of the X variables takes k states, the number of states of Y_C would be $k^{|C|}$). For each variable Y_C and each variable $X_i \in C$, we have one edge (X_i, Y_C).

   We define single node potentials for each variable Y_C to be

$$\psi_{Y_C}(y_C) = \psi_{Y_C}((x'_{i_1}, ..., x'_{i_{|C|}})) = \psi_C(x'_{i_1}, ..., x'_{i_{|C|}}),$$

   where $y_C = (x'_{i_1}, ..., x'_{i_{|C|}})$ is a state of Y_C.

We define the pairwise potentials $\psi_{X_iY_C}$ for each edge to be

$$\psi_{X_iY_C}(x_i, y_C) = \psi_{X_iY_C}(x_i, (x'_{i_1}, ..., x'_{i_{|C|}})) = \begin{cases} 1 & x_i = x'_i \\ 0 & x_i \neq x'_i \end{cases}$$

The marginal distribution is then

$$\begin{aligned}
\sum_{\mathbf{y}} p'(\mathbf{x}, \mathbf{y}) &= \sum_{\mathbf{y}} \frac{1}{Z} \prod_C \psi_{Y_C}(y_C) \prod_{X_i \in C} \psi_{X_iY_C}(x_i, y_C) \\
&= \sum_{\mathbf{y}} \frac{1}{Z} \prod_C \psi_{Y_C}((x_{i_1}(y_C), ..., x_{i_{|C|}}(y_C))) \prod_{X_i \in C} \psi_{X_iY_C}(x_i, (x_{i_1}(y_C), ..., x_{i_{|C|}}(y_C))) \\
&= \frac{1}{Z} \prod_C \psi_{Y_C}((x_{i_1}, ..., x_{i_{|C|}})) \\
&= p(\mathbf{x}),
\end{aligned}$$

where the simplification from the second to third lines is because the product is zero (due to the edge potentials) for any assignment $\mathbf{y}$ which is inconsistent with $\mathbf{x}$.

4. **Exponential families**. Probability distributions in the exponential family have the form:

$$p(\mathbf{x}; \eta) = h(\mathbf{x}) \exp\{\eta \cdot \mathbf{f}(\mathbf{x}) - \ln Z(\eta)\}$$

for some scalar function $h(\mathbf{x})$, vector of functions $\mathbf{f}(\mathbf{x}) = (f_1(\mathbf{x}), \ldots, f_d(\mathbf{x}))$, canonical parameter vector $\eta \in \mathbb{R}^d$ (often referred to as the *natural parameters*), and $Z(\eta)$ a constant (depending on η) chosen so that the distribution normalizes.

(a) Determine which of the following distributions are in the exponential family, exhibiting the $\mathbf{f}(\mathbf{x})$, $Z(\eta)$, and $h(\mathbf{x})$ functions for those that are.

i. $N(\mu, I)$—multivariate Gaussian with mean vector μ and identity covariance matrix.

Answer: The density for a d-dimensional Gaussian with mean μ and covariance matrix I is

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

so we have an exponential family with parameters

$$\begin{aligned}
h(x) &= \frac{1}{(2\pi)^{d/2}} \exp(-\frac{1}{2}x^T x), \\
f(x) &= x, \\
\eta &= \mu, \\
\ln Z(\eta) &= \frac{1}{2}\eta^T \eta.
\end{aligned}$$

ii. $\text{Dir}(\alpha)$—Dirichlet with parameter vector $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_K)$ (see Sec. 2.5.4).

Answer: The Dirichlet density for $\theta \in \mathbb{R}^K$ is

$$p(\theta|\alpha) = \frac{1}{B(\alpha)} \prod_{i=1}^K \theta_i^{\alpha_i - 1} = \prod_{i=1}^K \frac{1}{\theta_i} \exp\left(\sum_{i=1}^K \alpha_i \log \theta_i - \log B(\alpha)\right),$$

where $B(\alpha) = \frac{\prod_{i=1}^K \Gamma(\alpha_i)}{\Gamma(\sum_{i=1}^K \alpha_i)}$. Hence, we have an exponential family with parameters

$$\begin{aligned} h(\theta) &= \prod_{i=1}^K \frac{1}{\theta_i}, \\ f(\theta) &= [\log(\theta_1), \cdots, \log(\theta_K)]^T, \\ \eta &= \alpha, \\ \ln Z(\eta) &= \log B(\eta). \end{aligned}$$

iii. log-Normal distribution—the distribution of $Y = \exp(X)$, where $X \sim N(0, \sigma^2)$.
Answer: The log normal density has the form

$$p(y|\sigma) = \frac{1}{y\sigma\sqrt{2\pi}} \exp\left(-\frac{(\log y)^2}{2\sigma^2}\right) = \frac{1}{y\sqrt{2\pi}} \exp\left(\frac{-1}{2\sigma^2}(\log y)^2 - 0.5\log\left(\sigma^2\right)\right).$$

Hence, the log normal distributions over σ^2 are an exponential family with

$$\begin{aligned} h(y) &= \frac{1}{y\sqrt{2\pi}}, \\ f(y) &= (\log y)^2, \\ \eta &= -\frac{1}{2\sigma^2}, \\ \ln Z(\eta) &= -0.5\log(-2\eta). \end{aligned}$$

iv. Boltzmann distribution—an undirected graphical model $G = (V, E)$ involving a binary random vector $\mathbf{X}$ taking values in $\{0,1\}^n$ with distribution $p(\mathbf{x}) \propto \exp\left\{\sum_i u_i x_i + \sum_{(i,j)\in E} w_{i,j} x_i x_j\right\}$.
Answer: Let η be of dimension $|V| + |E|$, where $\eta_i = u_i \ \forall i$ and $\eta_{(i,j)} = w_{i,j} \ \forall (i,j) \in E$. Then,

$$Z(\eta) = \sum_{\mathbf{x}} \exp\left\{\sum_i \eta_i x_i + \sum_{(i,j)\in E} \eta_{(i,j)} x_i x_j\right\}.$$

We have an exponential family with parameters

$$\begin{aligned} h(\mathbf{x}) &= 1, \\ f(\mathbf{x})_i &= x_i, \forall i, \\ f(\mathbf{x})_{(i,j)} &= x_i x_j, \forall (i,j) \in E. \end{aligned}$$

(b) *Conditional models.* One can also talk about conditional distributions being in the exponential family, being of the form:

$$p(\mathbf{y} \mid \mathbf{x}; \eta) = h(\mathbf{x}, \mathbf{y}) \exp\{\eta \cdot \mathbf{f}(\mathbf{x}, \mathbf{y}) - \ln Z(\eta, \mathbf{x})\}.$$

The partition function Z now depends on $\mathbf{x}$, the variables that are conditioned on. Let Y be a binary variable whose conditional distribution is specified by the logistic function,

$$p(Y = 1 \mid \mathbf{x}; \alpha) = \frac{1}{1 + e^{-\alpha_0 - \sum_{i=1}^n \alpha_i x_i}}$$

Show that this conditional distribution is in the exponential family.

Answer:

Set

$$\begin{aligned}\eta &= \alpha \\ \mathbf{f}(\mathbf{x}, y) &= y \cdot (1, \mathbf{x}) = (y, x_1 y, ..., x_n y) \\ Z(\eta, \mathbf{x}) &= 1 + e^{\alpha_0 + \sum_{i=1}^n \alpha_i x_i} \\ h(\mathbf{x}, \mathbf{y}) &= 1\end{aligned}$$

Then we have that

$$p(y \mid \mathbf{x}; \eta) = \frac{e^{\alpha_0 y + \sum_{i=1}^n \alpha_i x_i y}}{1 + e^{\alpha_0 + \sum_{i=1}^n \alpha_i x_i}} = \begin{cases} \frac{e^{\alpha_0 + \sum_{i=1}^n \alpha_i x_i}}{1+e^{\alpha_0 + \sum_{i=1}^n \alpha_i x_i}} & y = 1 \\ \frac{1}{1+e^{\alpha_0 + \sum_{i=1}^n \alpha_i x_i}} & y = 0 \end{cases}$$

In the case of $y = 1$, multiplying the numerator and denominator by $e^{-\alpha_0 + \sum_{i=1}^n -\alpha_i x_i}$ results in the logistic function.

5. **Tree factorization.** Let T denote the edges of a tree-structured pairwise Markov random field with vertices V. For the special case of trees, prove that *any* distribution $p_T(\mathbf{x})$ corresponding to a Markov random field over T admits a factorization of the form:

$$p_T(\mathbf{x}) = \prod_{(i,j) \in T} \frac{p_T(x_i, x_j)}{p_T(x_i) p_T(x_j)} \prod_{j \in V} p_T(x_j), \tag{2}$$

where $p_T(x_i, x_j)$ and $p_T(x_i)$ denote pairwise and singleton marginals of the distribution p_T, respectively.

Hint: consider the Bayesian network where you choose an arbitrary node to be a root and direct all edges away from the root. Show that this is equivalent to the MRF. Then, looking at the BN's factorization, reshape it into the required form.

Answer: The corresponding distribution over the BN is given by

$$\begin{aligned}p(\mathbf{x}) &= \prod_{i \in V} p(x_i | x_{Pa(x_i)}) \\ &= p_T(x_1) \prod_{(i,j) \in E} \frac{p_T(x_i, x_j)}{p_T(x_j)}\end{aligned}$$

since each node is connected to its parent by one edge, and every edge is traversed since we have a tree property. But this is the distribution over a pairwise MRF. Lastly,

$$\begin{aligned}p_T(x_1) \prod_{(i,j) \in E} \frac{p_T(x_i, x_j)}{p_T(x_j)} &= p_T(x_1) \prod_{(i,j) \in E} \frac{p_T(x_i) p_T(x_i, x_j)}{p_T(x_i) p_T(x_j)} \\ &= \prod_{i \in V} p_T(x_i) \prod_{(i,j) \in E} \frac{p_T(x_i, x_j)}{p_T(x_i) p_T(x_j)}\end{aligned}$$