Inference and Representation, Fall 2018

Problem Set 2: Undirected graphical models & Modeling exercise

Due: Monday, September 25, 2018 at 11:59pm (as a PDF file uploaded to NYU Classes)

Important: *See problem set policy on the course web site.*

---

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

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

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 $\Pr(\mathbf{X})$ (assuming the state space $\{0, 1\}$ is mapped to $\{-1, +1\}$).

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

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

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.

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

      iii. log-Normal distribution—the distribution of $Y = \exp(X)$, where $X \sim N(0, \sigma^2)$.

      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 $\Pr(\mathbf{x}) \propto \exp\big\{ \sum_i u_i x_i + \sum_{(i,j)\in E} w_{i,j} x_i x_j \big\}$.

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

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

   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,

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

   Show that this conditional distribution is in the exponential family.

3. **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.

4. *Hammersley-Clifford and Gaussian models:* Consider a zero-mean Gaussian random vector $(X_1, \ldots, X_N)$ with a strictly positive definite $N \times N$ covariance matrix $\Sigma \succ 0$. For a given undirected graph $G = (V, E)$ with N vertices, suppose that $(X_1, \ldots, X_N)$ obeys all the basic conditional independence properties of the graph G (i.e., one for each vertex cut set).

   (a) Show the sparsity pattern of the inverse covariance $\Theta = (\Sigma)^{-1}$ must respect the graph structure (i.e., $\Theta_{ij} = 0$ for all indices i, j such that $(i, j) \notin E$.)

   (b) Interpret this sparsity relation in terms of cut sets and conditional independence.

5. *Undirected trees and marginals:* Let $G = (V, E)$ be an undirected graph. For each vertex $i \in V$, let μ_i be a strictly positive function such that $\sum_{x_i} \mu_i(x_i) = 1$. For each edge, let μ_{ij} be a strictly positive function such that $\sum_{x_i} \mu_{ij}(x_i, x_j) = \mu_j(x_j)$ for all x_j, and $\sum_{x_j} \mu_{ij}(x_i, x_j) = \mu_i(x_i)$ for all x_i. Suppose moreover that $\mu_{ij}(x_i, x_j) \neq \mu_i(x_i)\mu_j(x_j)$ for at least one configuration (x_i, x_j). Given integers $m_1, \ldots, m_n$, consider the function

$$r(x_1, \ldots, x_n) \quad = \quad \prod_{i=1}^{n} \left[\mu_i(x_i)\right]^{m_i} \prod_{(i,j)\in E} \mu_{ij}(x_i, x_j).$$

   Supposing that G is a tree, can you give choices of integers $m_1, \ldots, m_n$ for which r is a valid probability distribution? If so, prove the validity. (*Hint:* It may be easiest to first think about a Markov chain.)