Probabilistic Graphical Models, Spring 2013

Problem Set 3: Undirected graphical models
Due: Thursday, February 21, 2013 at 5pm

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

2. Both of:
   (a) Exercise 4.2 from Koller & Friedman (reparameterization leaves distribution unchanged. See page 124).
   (b) Exercise 4.12 from Koller & Friedman (converting Boltzmann machine to Ising model. See page 126).

3. Give a procedure to convert any Markov network into a pairwise Markov random field. In particular, given a distribution \( p(X) \), specify a new distribution \( p(X, Y) \) which is a pairwise MRF, such that \( p(x) = \sum_y p(x, y) \), where \( Y \) 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: consider the factor graph representation of the Markov network, and introduce one new variable for each non-pairwise factor.

4. Exponential families (see Chap. 8.1-8.3). Probability distributions in the exponential family have the form:

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

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

   (a) Determine which of the following distributions are in the exponential family, exhibiting the \( f(x) \), \( Z(\eta) \), and \( h(x) \) functions for those that are.
      i. \( N(\mu, I) \)—multivariate Gaussian with mean vector \( \mu \) and identity covariance matrix.
      ii. \( \text{Dir}(\alpha) \)—Dirichlet with parameter vector \( \alpha = (\alpha_1, \alpha_2, \ldots, \alpha_K) \).
      iii. \( \text{log-Normal distribution} \)—the distribution of \( Y = \exp(X) \), where \( X \sim N(0, \sigma^2) \).
      iv. \( \text{Boltzmann distribution} \)—an undirected graphical model \( G = (V, E) \) involving a binary random vector \( X \) taking values in \( \{0,1\}^n \) with distribution \( p(x) \propto \exp\{\sum_i u_i x_i + \sum_{(i,j) \in E} w_{i,j} x_i x_j\} \).

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

   \[
   p(y | x; \eta) = h(x, y) \exp\{\eta \cdot f(x, y) - \ln Z(\eta, x)\}.
   \]
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.

5. **Conjugacy and Bayesian prediction.**

   (a) Let $\theta \sim \text{Dir}(\alpha)$. Consider multinomial random variables $(X_1, X_2, \ldots, X_N)$, where $X_i \sim \text{Mult}(\theta)$ for each $i$ (thus the $X_i$ are conditionally independent of one another given $\theta$).\(^1\) Show that the posterior $p(\theta \mid x_1, \ldots, x_N, \alpha)$ is given by $\text{Dir}(\alpha')$, where

   $$ \alpha'_k = \alpha_k + \sum_{i=1}^{N} 1[x_i = k]. $$

   This property, that the posterior distribution $p(\theta \mid \mathbf{x})$ is in the same family as the prior distribution $p(\theta)$, is called **conjugacy**. The Dirichlet distribution is the **conjugate prior** for the Multinomial distribution. Every distribution in the exponential family has a conjugate prior. For example, the conjugate prior for the mean of a Gaussian distribution can be shown to be another Gaussian distribution.

   (b) Now consider a random variable $X_{\text{new}} \sim \text{Mult}(\theta)$ that is assumed conditionally independent of $(X_1, X_2, \ldots, X_N)$ given $\theta$. Compute:

   $$ p(x_{\text{new}} \mid x_1, x_2, \ldots, x_N, \alpha) $$

   by integrating over $\theta$.

   *Hint:* Your result should take the form of a ratio of gamma functions.

   This is called **Bayesian prediction** because we put a prior distribution over the parameters $\theta$ (in this case, a Dirichlet) and are thus able to take into consideration our initial uncertainty over (and prior knowledge of) the parameters together with the evidence we observed (samples $x_1, \ldots, x_N$) when giving our predictions for $x_{\text{new}}$.

---

\(^1\)We are using the book’s definition of the Multinomial distribution, given on page 20, which Wikipedia calls the Categorical distribution.