Homework 1
Inference and Representation

Aodong Li
NetID: al5350
al5350@nyu.edu

June 9, 2018

1. By using the mapping $x'_i = 2x_i - 1$, we can map from $\{0, 1\}$ to $\{-1, 1\}$. Under this mapping, the Ising model is

$$
\begin{aligned}
p(x) &= \frac{1}{Z}\exp(\sum_{(i,j)} w'_{ij}(2x_i - 1)(2x_j - 1) - \sum_i u'_i(2x_i - 1)) \\
&= \frac{1}{Z}\exp(\sum_{(i,j)} 4w'_{ij}x_ix_j - \sum_{(i,j)} 2w'_{ij}x_j - \sum_{(i,j)} 2w'_{ij}x_i + \sum_{(i,j)} w'_{ij} - \sum_i 2u'_ix_i - \sum_i u_i) \\
&= \frac{1}{Z * C}\exp(\sum_{(i,j)} 4w'_{ij}x_ix_j - \sum_i 2(u'_i + \sum_{(i,j)} w'_{ij})x_i)
\end{aligned}
$$

So the relationship is that $w_{ij} = 4w'_{ij}$ and $u_i = 2(u'_i + \sum_{(i,j)} w'_{ij})$.

2. We can transform each potential function in MRF G into a pairwise MRF G' by introducing an additional variable. So for each potential function $\psi(x_1, ..., x_m)$, we introduce an additional variable Y such that

$$\psi_i(x_1, ..., x_m) = \sum_y \psi_Y(y) \prod_i^m \psi_{iY}(x_i, y)$$

For each assignment of x for $\psi(x_1, ..., x_m)$, there is a corresponding potential value v.

Make the extra variable Y have the number of values that the original potential function have, and make edge potential as indicators. That is, $\psi_Y(y) = \psi(x_1, ..., x_m)$ for each assignment, and $\psi_{iY}(x_i, y) = 1$ if x_i occurs in the combination that y represents and 0 otherwise. And the new pairwise MRF indeed have a description which is polynomial in the size of the original MRF, $(1 + Val(X_1) + ... + Val(X_m)) \times Val(Y)$ as the number of parameters.

3. (a) i. Yes. The MVN can be rewritten as

$$P(x|\mu, I) = \frac{1}{\sqrt{(2\pi)^k}} \exp(-\frac{1}{2}x^\top x + \mu^\top x - \frac{1}{2}\mu^\top \mu).$$

We can take the following parameters

$$\begin{aligned} f(x) &= \begin{bmatrix} x^\top x \\ x \end{bmatrix} \\ \eta &= \begin{bmatrix} -\frac{1}{2} \\ \mu \end{bmatrix} \\ Z(\eta) &= \exp(\frac{1}{2}\mu^\top \mu) \\ h(x) &= \frac{1}{\sqrt{(2\pi)^k}} \end{aligned}$$

ii. Suppose the input is a valid one, and the Dirichlet distribution can be represented as

$$\begin{aligned} p(x; \alpha) &= \frac{1}{B(\alpha)} \prod_{k=1}^{K} x_k^{\alpha_k - 1} \\ &= \exp(\sum_{k=1}^{K} (\alpha_k - 1) \log x_k - \log B(\alpha)). \end{aligned}$$

So again, it is in exponential family parameterized by

$$\begin{aligned} h(x) &= 1 \\ \eta &= (\alpha_1 - 1, ..., \alpha_K - 1)^\top \\ f(x) &= (\log x_1, ..., \log x_K)^\top \\ Z(\eta) &= B(\alpha) \end{aligned}$$

iii. The density of log-Normal distribution is

$$
\begin{aligned}
p(y;\sigma) &= \frac{\mathrm{d}}{\mathrm{d}\ y}\int_{-\infty}^{\log y} \frac{1}{\sqrt{2\pi}\sigma}\exp(-\frac{x^2}{2\sigma^2})\mathrm{d}\ x \\
&= \frac{1}{\sqrt{2\pi}y}\exp\{-\frac{(\log y)^2}{2\sigma^2} - \log\sigma\}.
\end{aligned}
$$

So again, it is in exponential family parameterized by

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

iv. Yes, it is.

$$
p(x;\eta) = \frac{1}{Z(\eta)}\exp\{\frac{1}{2}x^\top Wx + u^\top x\}
$$

where $Z(\eta) = \sum_x \exp\{\frac{1}{2}x^\top Wx + u^\top x\}$ and W is a matrix and u is a vector.
So again, it is in exponential family parameterized by

$$
\begin{aligned}
h(x) &= 1 \\
\eta &= \begin{bmatrix} W \\ u \end{bmatrix} \\
f(x) &= \begin{bmatrix} xx^\top \\ x \end{bmatrix} \\
Z(\eta) &= \sum_x \exp\{\frac{1}{2}x^\top Wx + u^\top x\}
\end{aligned}
$$

(b)

$$
\begin{aligned}
p(y=1|x;\alpha) &= \exp\{-\log(1+\exp(-\alpha^\top x))\} \\
p(y=0|x;\alpha) &= \exp\{-\alpha^\top x - \log(1+\exp(-\alpha^\top x))\}.
\end{aligned}
$$

So it is in exponential family parameterized by

$$
\begin{aligned}
h(x, y) &= 1 \\
\eta &= -\alpha \\
f(x, y) &= \begin{cases} 0, & y = 1 \\ x, & y = 0 \end{cases} \\
Z(\eta) &= 1 + \exp(-\alpha^\top x)
\end{aligned}
$$

4. Since the tree structure dictates one node has at most one parent. So the moralization of a Bayesian network is the same as the structure of a markov random field without any additional edges.

   For each edge (x_i, x_j) (suppose $x_i \to x_j$), there is an conditional distribution $p(x_j|x_i)$ associated with it, so we can rewrite it as

   $$p(x_j|x_i) = \frac{p(x_i, x_j)}{p(x_i)p(x_j)} p(x_j).$$

   So including the marginal distribution of the root, we can represent the joint distribution as

   $$p(x) = \prod_{(i,j)\in E} \frac{p(x_i, x_j)}{p(x_i)p(x_j)} \prod_{i \in V} p(x_i).$$

   We then moralize the Bayesian network and express the marginal distribution and pairwise joint distribution by singleton and pairwise potential functions.

5. See the corresponding IPython Notebook in the same file folder.