# Inference and Representations
# Problem Set 2

Zhuoru Lin
zlin@nyu.edu

## 1 Exercise 4.1 from Koller & Friedman

Proof by contradiction:
If $P$ factorize over $H$, we must have:

$$p(0,0,0,0) = \phi_{1,2}(0,0)\phi_{2,3}(0,1)\phi_{3,4}(0,0)\phi_{4,1}(0,0) = \frac{1}{8}$$
$$p(0,0,1,0) = \phi_{1,2}(0,0)\phi_{2,3}(0,1)\phi_{3,4}(1,0)\phi_{4,1}(0,0) = 0$$
$$p(0,0,1,1) = \phi_{1,2}(0,0)\phi_{2,3}(0,1)\phi_{3,4}(1,1)\phi_{4,1}(1,0) = 0,$$

This shows that $\phi_{3,4}(1,0)$ is zero. However:

$$p(1,1,1,0) = \phi_{1,2}(0,0)\phi_{2,3}(0,1)\phi_{3,4}(1,0)\phi_{4,1}(0,0) = 0$$

shows that $\phi_{3,4}(1,0)$ is non-zero. This is a contradiction.

## 2 Ising Model

Let $H$ be the Hamiltonian of the distribution defined by Boltzmann Machine, i,e, $H = \sum_{(i,j)\in E} -w_{ij}x_i x_j + \sum_{i\in V} u_i x_i$. Let $f(x') = \frac{x'+1}{2}$, $f$ is a map from $X' \to X$ where $X'_i \in \{-1, 1\}$. Then we have:

$$-H = \sum_{(i,j)\in E} w_{ij}\frac{x'_i + 1}{2}\frac{x'_j + 1}{2} - \sum_{i\in V} u_i \frac{x'_i + 1}{2} \tag{1}$$

$$= \sum_{(i,j)\in E} \frac{1}{4}w_{ij}x'_i x'_j + \sum_{(i,j)\in E} \frac{1}{4}w_{ij}x'_i + \sum_{(i,j)\in E} \frac{1}{4}w_{ij}x'_j + \sum_{(i,j)\in E} \frac{1}{4}w_{ij} - \sum_{i\in V} \frac{1}{2}u_i x'_i - \sum_{i\in V} \frac{1}{2}u_i. \tag{2}$$

Define $nb(i) = \{j | j \text{ is a neightbour of } i\}$, we have: $\sum_{(i,j)\in E} \frac{1}{4}w_{ij}x'_i + \sum_{(i,j)\in E} \frac{1}{4}w_{ij}x'_j = \sum_{i\in V}\sum_{j\in nb(i)} \frac{1}{4}w_{ij}x'_j$.
Plug this into (2) we have:

$$-H = \sum_{(i,j)\in E} \frac{1}{4}w_{ij}x'_i x'_j + \sum_{i\in V}\left(\sum_{j\in nb(i)} w_{ij}x'_j - \frac{1}{2}u_i\right) + \sum_{(i,j)\in E} \frac{1}{4}w_{ij} - \sum_{i\in V} \frac{1}{2}u_i. \tag{3}$$

Let $w'_{ij} = \frac{1}{4}w_j$ , $h'_i = \sum_{j \in nb(i)} w_{ij} x'_j - \frac{1}{2} u_i$ and $-H' = \sum_{(i,j) \in E} w'_{ij} x'_i x'_j - \sum_{i \in V} u'_i x'_i$, (3) becomes:

$$-H = -H' + C \tag{4}$$

where $C$ is some constant independent from $X'$. We now have:

$$p(x) = \frac{1}{Z} exp(-H) = \frac{1}{Z'} exp(-H'), \tag{5}$$

where $Z$ and $Z'$ are the partition functions for $-H$ and $-H'$ respectively.

# 3   Convert any Markov network into pairwise Markov random field

Let's define a new random variable $\mathbf{Y}$ such that $\mathbf{Y}$ have the same number of states as $\mathbf{X}$. Let each $y \in \mathbf{Y}$ bijectively map to a specific state $\mathbf{x} \in \mathbf{X}$ by $\mathbf{x} = f(y)$. Define $p(y) = p(f(y))$ and $p(X_i|y) = 0$ if $X_i \notin f(y)$. With this definition we automatically have $\sum_y p(x, y) = p(x)$. In this case, we have $X_1, X_2, ..., X_n$ are trivially conditional independent given $Y$. Therefore:

$$p(\mathbf{x}, y) = p(y)p(\mathbf{x}|y)$$
$$= p(y) \prod_{i=1}^{n} p(x_i|y)$$
$$\propto \phi(y) \prod_{i=1}^{n} \phi(x_i, y),$$

which is a pairwise Markov random field.

# 4   Exponential Families

### (a-i) Multivariate Normal

$N(\mu, I)$ is in the Exponential Families, because for $X = N(\mu, I)$,

$$p(x) = \frac{1}{\sqrt{2\pi}} \exp[-\frac{1}{2}(x - \mu)^T (x - \mu)] \tag{6}$$

$$= \frac{1}{\sqrt{2\pi}} \exp[-\frac{1}{2}(x^T x - 2x^T \mu + \mu^T \mu)] \tag{7}$$

$$= \frac{1}{\sqrt{2\pi} \exp(\frac{1}{2}x^T x)} \exp(\mu^T x - \frac{1}{2}\mu^T \mu) \tag{8}$$

$$= h(x) \exp[\eta.f(x) - ln(Z(\eta))] \tag{9}$$

where $h(x) = \frac{1}{\sqrt{2\pi} \exp(\frac{1}{2}x^T x)}$, $\eta = \mu$ and $Z(\eta) = \exp(\frac{1}{2}\eta^T \eta)$.

## (a-ii) Dirichlet

$Dir(\alpha)$ is in the Exponential Families, because for $X = Dir(\alpha)$:

$$
\begin{aligned}
p(x) &= \frac{1}{B(\alpha)} \prod_{i=1}^{K} x_i^{\alpha_i - 1} \\
&= \frac{1}{B(\alpha)} \exp\left[\sum_{i=1}^{K} (\alpha_i - 1) ln(x_i)\right] \\
&= \frac{1}{B(\alpha)} \exp\left[\sum_{i=1}^{K} \alpha_i ln(x_i) - \sum_{i=1}^{K} ln(x_i)\right] \\
&= \frac{1}{\exp[\sum_{i=1}^{K} ln(x_i)]} \exp\left[\sum_{i=1}^{K} \alpha_i ln(x_i) - ln(B(\alpha))\right] \\
&= \frac{1}{\prod_{i=1}^{K} x_i} \exp\left[\sum_{i=1}^{K} \alpha_i ln(x_i) - ln(B(\alpha))\right] \\
&= h(x) \exp[\eta . f(x) - ln(Z(\eta))],
\end{aligned}
$$

where $h(x) = \frac{1}{\prod_{i=1}^{K} x_i}$, $f(x) = ln(x)$ , $\eta = \alpha$ and $Z(\eta) = B(\eta)$.

## (a-iii) Log-Normal

Let $Pr(x)$ denote the cumulative probability density function (Note that we already use p(x) to represent the probability density function). The probability of X<x is equivalent to ln(X)<ln(X). Therefore:

$$
\begin{aligned}
p(x) &= \frac{d\Phi(ln(x))}{d(x)} \\
&= \frac{1}{x} \phi(ln(x)) \\
&= \frac{1}{x} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{ln(x)}{2\sigma^2}\right] \\
&= \frac{1}{x\sqrt{2\pi}} \exp\left[-\frac{ln(x)}{2\sigma^2} - ln(\sigma)\right] \\
&= h(x) \exp[\eta . f(x) - ln(Z(\eta))],
\end{aligned}
$$

where $h(x) = \frac{1}{x\sqrt{2\pi}}$, $f(x) = ln(x)$, $\eta = \frac{-1}{2\sigma^2}$ and $Z(\eta) = \sqrt{\frac{-1}{2z}} = \sigma$

## (a-iv) Boltzmann

Define new vector $w' \in \mathbb{R}^{ij}$ so that $w'_{i \times j} = \frac{1}{2} w_{ij}$ if $(i, j) \in E$ and $w_{i \times j} = 0$ elsewhere. Let $x' \in \{0, 1\}^{ij}$ so that $x'_{i \times j} = x_i x_j$. Let $(w'; u)]$ be the concatenation of $w'$ and $u$ and $(x', x)$ be the concatenation of $x'$ and $x$. Then the Boltzmann distribution can be defined by:

$$p(x) = h(x) \exp[\eta.f(x) - ln(Z(\eta))], \tag{10}$$

where $h(x) = 1$, $\eta = (w'; u)$, $f(x) = (x'; x)$ and $Z(\eta) =$The partition function.

Therefore Boltzmann distribution is in exponential family.

## 4.1 (b)

By the form of $p(Y = 1|x)$, we understand that we make logistic function being in exponential family by making: $Z(\eta, x) = 1 + \exp[-\alpha_0 + \sum_{i=1}^{n} x_i]$, $h(x, 1) \exp[\eta.f(x, 1)] = 1$ and $h(x, 0) \exp[\eta.f(x, 0)] = \exp[-\alpha_0 - \sum_{i=1}^{n} x_i]$. We can achieve this by defining:

$$h(x, y) = 1$$
$$\eta = -\alpha$$
$$f(x, y) = ((1 - y), (1 - y)x_1, (1 - y)x_2, ..., (1 - y)x_n)$$
$$Z(\eta, x) = 1 + \exp[-\alpha_0 - \sum_{i=1}^{n} \alpha_i x_i]$$

# 5 Tree factorization

It can be shown by moralization that the tree structure represented by Bayesian network is equivalent to the Markov random field with the same edge. The conditional independence relations represented by Bayesian Networks also hold in Markov random field in tree. In Bayesian Network, it follows the factorization of:

$$p(\mathbf{x}) = p(x_r) \prod_{x_i \in \mathbf{x} - \{x_r\}} p(x_i \mid \pi(x_i)) \tag{11}$$

$$= p(x_r) \prod_{x_i \in \mathbf{x} - \{x_r\}} \frac{p(x, \pi(x))}{p(x)p(\pi(x))} p(x), \tag{12}$$

where $\pi(x_i)$ represents the parent of $x_i$ in Bayesian Network and $x_r$ is the root of the tree. One can always translate (11) into Markov random field by:

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

since each node have only one parent in tree and connected by edge in Markov random field.