ML-2

Homework Assignment - 4

Shantanu Chandra
(12048461)

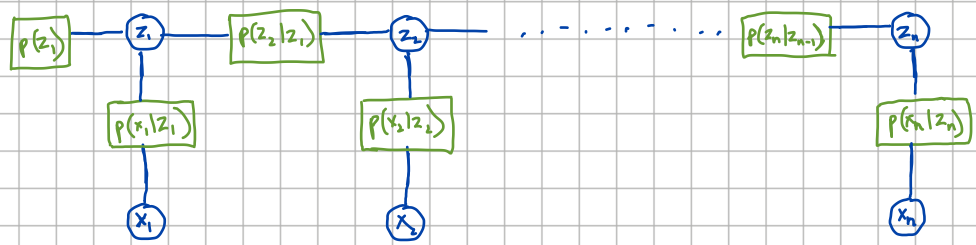
Q1.

1. Using Bishop 8.5, we can factorize the given graph as :

$$p(x,Z) = p(z_1)\,p(x_1|z_1)\prod_{i=2}^{N} p(z_i|z_{i-1})\,p(x_i|z_i)$$

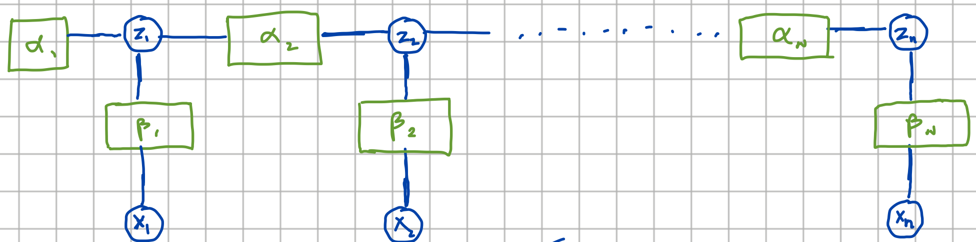
2. To convert a directed graph to a factor graph :

a) Create variable nodes in F.G. corresponding to the ones in DAG.
b) Create factor nodes corresponding to the original conditional distributions.
c) Add appropriate links.



NOTE : There are many factor graphs that can be formed that all represent the same DAG. Here is just one of such graphs.

d) To represent local conditional probabilities, we express them using $f_s(x_s)$:



// Ans.

3. Using Bishop 8.59 : $p(x) = \prod_s f_s(x_s)$ where, 's' is a subset of variables.
Here, we have 2 subsets, namely α & β.

Joint probability using the factors for these subsets is :

$$p(z,x) = f_{\alpha_1}(z_1)\prod_{i=2}^{N} f_{\alpha_i}(z_i, z_{i-1}) \prod_{j=1}^{N} f_{\beta_j}(x_j, z_j)$$

// Ans.

4

$$X = \{x_1 \dots x_N\}$$

$$P(z_n|X) = \frac{p(x_1 \dots x_N | z_n)\, p(z_n)}{P(x_1 \dots x_N)} \qquad \text{(Using Baye's Rule)}$$

$$= P(\underbrace{x_1 \dots x_n}, \underbrace{x_{n+1} \dots x_N} | z_n) \frac{p(z_n)}{P(x_1 \dots x_N)}$$

d-separated by z_n.
Hence independent $\Rightarrow P(A,B|C) = P(A|C)P(B|C)$

$$= \frac{p(x_1 \dots x_n|z_n)\, p(z_n) \cdot p(x_{n+1} \dots x_N|z_n)}{P(x_1 \dots x_N)}$$

$$= \frac{p(x_1 \dots x_n, z_n)\, p(x_{n+1} \dots x_N|z_n)}{P(X)}$$

Thus, we have:

$$\alpha(z_n) = p(x_1 \dots x_n, z_n)$$
$$= \sum_{z_{n-1}} p(x_1 \dots x_n, z_n | z_{n-1})\, p(z_{n-1})$$

Here, $x_1 \cdots x_{n-1}$ is d-separated from x_n by z_n, hence independent :

$$= \sum_{z_{n-1}} \underline{p(x_1 \dots x_{n-1}|z_{n-1})}\, p(x_n, z_n|z_{n-1})\, \underline{p(z_{n-1})}$$

$$= \sum \alpha(z_{n-1})\, p(x_n, z_n|z_{n-1}) \qquad \text{(following from the definition of } \alpha(z_n)\text{)}.$$

// Ans.

And we have :

$$\beta(z_n) = p(x_{n+1} \dots x_N|z_n)$$
$$= \sum_{z_{n+1}} p(x_{n+1} \dots x_N|z_n, z_{n+1})\, p(z_{n+1})$$

$$= \sum_{z_{n+1}} p\frac{(x_{n+1}, x_{n+2} \dots x_N, z_n|z_{n+1})}{P(z_n)}\, p(z_{n+1}) \qquad \left(\text{using } p(A|B) = \frac{P(A,B)}{P(B)}\right)$$

Here, $x_{n+2} \dots x_N$ is d-separated from x_{n+1}, z_n by z_{n+1} and hence independent :

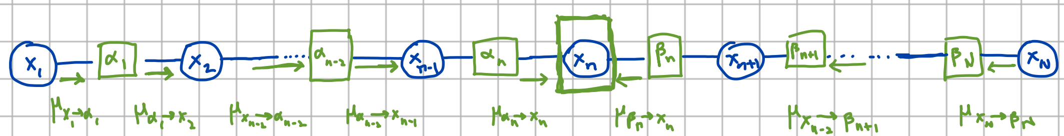
$$= \sum_{z_{n+1}} \underbrace{p(x_{n+2} \dots x_N|z_{n+1})}_{=\beta(z_{n+1})} \frac{p(x_{n+1}|z_n, z_{n+1})\, p(z_{n+1})\, \cancel{p(z_n)}}{\cancel{P(z_n)}}$$

$$= \sum_{z_{n+1}} \beta(z_{n+1})\; p(x_{n+1}, z_{n+1}|z_n) \qquad \text{(using } P(A|B,C) \cdot P(B) = p(A,B|C)\text{)}$$

// Ans.

Q2.

1



The given graph can be shown as a factor graph viewed above. From Bishop 8.4.4 we have :

<u>Factor → variable messages</u> : $\mu_{\alpha \to x}(x) = \sum_{x_1} \dots \sum_{x_m} f_\alpha(x, x_1, \dots x_m) \prod_{j \in \alpha \setminus i} \mu_{j \to \alpha}(x)$

<u>Variable → factor messages</u> : $\mu_{x \to \alpha}(x) = \prod_{\beta \in ne(j) \setminus \alpha} \mu_{\beta \to x}(x)$

Applying this to current graph :

$$\left.\begin{aligned}\mu_{x_1\to\alpha_1}(x_1) &= 1\\ \mu_{x_N\to\beta_N}(x_N) &= 1\end{aligned}\right\}$$ leaf nodes, all are initialized with 1.

$$\mu_{\alpha_1\to x_2}(x_2) = \sum_{x_1} f_{\alpha_1}(x_1,x_2)\,\mu_{x_1\to\alpha_1}(x_1) = \sum_{x_1} f_{\alpha_1}(x_1,x_2)$$

$$\mu_{x_2\to\alpha_2}(x_2) = \mu_{\alpha_1\to x_2}(x_2) = \sum_{x_1} f_{\alpha_1}(x_1,x_2)$$

$$\mu_{\alpha_2\to x_3}(x_3) = \sum_{x_2} f_{\alpha_2}(x_2,x_3)\,\mu_{x_2\to\alpha_2}(x_2)$$

$$\vdots$$

$$\underline{\underline{\mu_{\alpha_{n-1}\to x_n}(x_n) = \sum_{x_{n-1}} f_{\alpha_{n-1}}(x_{n-1},x_n)\,\mu_{x_{n-1}\to\alpha_{n-1}}(x_{n-1})}}$$

Thus we see correspondence between MRFs & FGs. :

MRF	FG	
$\mu_\alpha(x_n)$	$\mu_{\alpha_{n-1}\to x_n}(x_n)$	
$\Psi_{n-1,n}(x_{n-1},x_n)$	$f_{\alpha_{n-1}}(x_{n-1},x_n)$	(max cliques $\longleftrightarrow$ factors)

Similarly for β we see that :

$$\mu_{\beta_N\to x_{N-1}}(x_{N-1}) = \sum_{x_{N-1}} f_{\beta_N}(x_{N-1},x_N)\,\mu_{x_N\to\beta_N}(x_N)$$

$$\vdots$$

$$\underline{\underline{\mu_{\beta_{n+1}\to x_n}(x_n) = \sum_{x_{n+1}} f_{\beta_{n+1}}(x_n,x_{n+1})\,\mu_{x_{n+1}\to\beta_{n+1}}(x_{n+1})}}$$

Thus we see correspondence between MRFs & FGs. :

MRF	FG
$\mu_\beta(x_n)$	$\mu_{\beta_{n+1}\to x_n}(x_n)$
$\Psi_{n+1,n}(x_{n+1},x_n)$	$f_{x_{n+1}\to\beta_N}(x_{n+1})$

Finally for the marginal $p(x_n)$, using Bishop 8.63 :

$$p(x_n) = \prod_{s\in ne(x)} \mu_{f_s\to x}(x)$$

$$= \underline{\underline{\mu_{\alpha_{n-1}\to x_n}(x_n)\cdot\mu_{\beta_{n+1}\to x_n}(x_n)}}$$

Thus we see correspondence between MRFs & FGs.

MRF	FG
$p(x_n) = \frac{1}{Z}\mu_\alpha(x_n)\mu_\beta(x_n)$	$p(x_n) = \mu_{\alpha_{n-1}\to x_n}(x_n)\cdot\mu_{\beta_{n+1}\to x_n}(x_n)$

NOTE : Here, $Z=1$ since these are normalized as factors represent (conditional) probabilities. In MRF, 'Ψ' can be any function thus require normalization.
(+ve)

②

Q3. Given a F.G, the marginal is given by (Bishop 8.59)

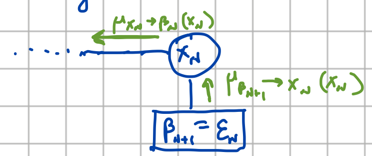
$$p(x) = \frac{1}{Z}\prod_{\alpha} f_{\alpha}(x_{\alpha})$$

To extend this to conditional, we introduce the evidence factor & renormalize :

$$p(x_n \mid x_N = \varepsilon_N) = \frac{1}{\tilde{Z}}\prod_{\beta} f_{\beta}(x_{\beta}) \cdot \delta_{x_N,\varepsilon_N} \qquad \text{where, } \delta_{x_N,\varepsilon_N} = \begin{cases} 0, & x_N \neq \varepsilon_N \\ 1, & x_N = \varepsilon_N \end{cases}$$

for the 'α' subset of variables : $p(x_n \mid x_N = \varepsilon) = \frac{1}{Z}\prod_{\alpha} f_{\alpha}(x_{\alpha})$

This means that we can add a new factor node to the existing F.G and perform the exact same algo on this new F.G. :



Now, following the same procedure as in Q2.1), we get :

for α :

$$\mu_{\alpha_{n-1} \to x_n}(x_n) = \sum_{x_{n-1}} f_{\alpha_{n-1}}(x_{n-1}, x_n)\, \mu_{x_{n-1} \to \alpha_{n-1}}(x_{n-1})$$

[this part remains unchanged from before]

for β :

$$\mu_{\beta_{N+1} \to x_N}(x_N) = \underbrace{f_{\beta_{N+1}}(x_N)}_{\text{(since leaf node)}} = \delta_{x_N,\varepsilon_N}$$

$$\mu_{x_N \to \beta_N}(x_N) = \mu_{\beta_{N+1} \to x_N}(x_N) = \delta_{x_N,\varepsilon_N}$$

$$\mu_{\beta_N \to x_{N-1}}(x_{N-1}) = \sum_{x_N} f_{\beta_N}(x_{N-1}, x_N)\, \mu_{x_N \to \beta_N}(x_N) = \sum_{x_N} f_{\beta_N}(x_{N-1}, x_N) \cdot \delta_{x_N,\varepsilon_N} = f_{\beta_N}(x_{N-1}, x_N)$$

Due to the δ, we don't sum over all possible values of x_N as its value is given $(=\varepsilon)$.

$\vdots$

$$\mu_{\beta_{n+1} \to x_n}(x_n) = \sum_{N-1} \cdots \sum_{n+1} f_{\beta_{n+1}}(x_n, x_{n+1})\, \mu_{x_{n+1} \to \beta_{n+1}}(x_{n+1})$$

$$= \sum_{N-1} \cdots \sum_{n+1} f_{\beta_{n+1}}(x_n, x_{n+1})\, \mu_{x_{n+1} \to \beta_{n+1}}(x_{n+1})$$

Thus, we see that sum-product algo. in this case :

1) Messages from $x_1 \cdots x_{n-1}$ (ie, α) remain unchanged.
2) Only messages coming from $x_N \cdots x_{n+1}$ (ie, β) change - they all have the term $\delta_{x_N,\varepsilon_N}$ now (which later cancels $\sum_N$).
3) This term in the end also makes the algorithm more efficient as we can use the exact same algo for conditional inference also (note, even after adding a new factor & introducing $\sum_N$ we still end up $\sum_{N-1}$ just as before).

Ans.

Q4.

To compute $p(x_s)$ we marginalize $p(x)$ over all the other variables :

$$p(x_s) = \sum_{x \setminus x_s} p(\mathbf{x})$$

Here, $p(\mathbf{x}) = \prod_s f_s(x_s)$ where x_s is a subset of variables (Bishop 8.59)

Combining this with the definition of $F_s(x_s, X)$ that it represents product of all factors in the group associated with factor f_s, we can write the marginal as :

$$p(x_s) = \sum_{x \setminus x_s} f_s(x_s) \prod_{i \in ne(f_s)} \prod_{j \in ne(x_i) \setminus f_s} F_j(x_i, X_{ij})$$

$$= f_s(x_s) \prod_{i \in ne(f_s)} \underbrace{\sum_{x \setminus x_s} \prod_{j \in ne(x_i) \setminus f_s} F_j(x_i, X_{ij})}$$

$$= f_s(x_s) \prod_{i \in ne(f_s)} \underbrace{\sum_{x \setminus x_s} G(x_i, X_{si})} \qquad \text{(Bishop 8.68)}$$

$$= f_s(x_s) \prod_{i \in ne(f_s)} \mu_{x_i \to f_s}(x_i) \qquad \text{(Bishop 8.67)}$$

// Hence proved