Machine Learning 2 — Homework 4

Maurice Frank
11650656
maurice.frank@posteo.de

September 30, 2019

Note: I discussed solutions with Karan Malhotra.
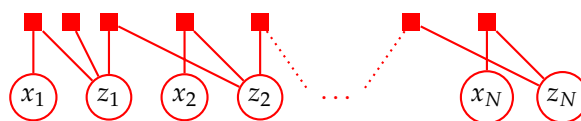
Problem 1.

We have $X = \{x_1, \ldots, x_N\}$ and $Z = \{z_1, \ldots, z_N\}$.

1.

$$p(Z, X) = p(z_1) \cdot \left(\prod_{i=2}^{N} p(z_i | z_{i-1}) \right) \cdot \left(\prod_{i=1}^{N} p(x_i | z_i) \right)$$

2.

We present the factor graph for the Markov chain:

3.

$$p(X) = f_1(z_1) \cdot \left(\prod_{i=2}^{N} f_i(z_i, z_{i-1}) \right) \cdot \left(\prod_{i=1}^{N} f_{N+1}(z_i, x_i) \right)$$

4.

$$p(z_n|\boldsymbol{X}) = \frac{p(\boldsymbol{X}|z_n)p(z_n)}{p(\boldsymbol{X})}$$

$$= \frac{p(x_1,\ldots,x_N|z_n)p(z_n)}{p(\boldsymbol{X})}$$

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

$$= \frac{\alpha(z_n)\beta(z_n)}{p(\boldsymbol{X})}$$

$$\implies$$

$$\alpha(z_n) := p(x_1,\ldots,x_n|z_n)$$

$$\beta(z_n) := p(x_{n+1},\ldots,x_N|z_n)p(z_n)$$

$$= p(x_{n+1},\ldots,x_N,z_n)$$

Step (1) is possible as z_n d-seperates the Markov chain at x_n, x_{n+1}.

$$\alpha(z_n) = p(x_1,\ldots,x_n|z_n)$$

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

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

$$= \sum_{z_{n-1}} \frac{p(x_1,\ldots,x_{n-1}|z_{n-1})p(x_n,z_n|z_{n-1})}{p(z_n)}p(z_{n-1}) \quad (2)$$

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

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

Again Step (2) is possible as z_{n-1} d-seperates $\{x_1,\ldots,x_n,z_n\}$ into $\{x_1,\ldots,x_{n-1}\}$ and $\{x_n,z_n\}$.

$$\beta(z_n) = p(x_{n+1}, \ldots, x_N, z_n)$$

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

$$= \sum_{z_{n+1}} p(x_{n+2}, \ldots, x_N | z_{n+1}) p(x_{n+1}, z_n | z_{n+1}) p(z_{n+1}) \qquad (3)$$

$$= \sum_{z_{n+1}} p(x_{n+2}, \ldots, x_N, z_{n+1}) p(x_{n+1}, z_n | z_{n+1})$$

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

Again Step (3) is possible as z_{n+11} d-seperates $\{x_{n+1}, \ldots, x_N, z_n\}$ into $\{x_{n+2}, \ldots, x_N\}$ and $\{x_{n+1}, z_n\}$.

Problem 2.

First we build the implicit factor graph for the chain with the implicit factors $\{f_{1,2}, \ldots, f_{N-1,N}\}$:
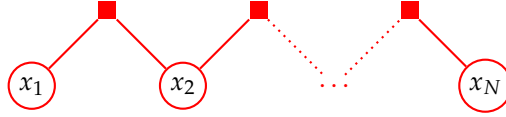
1.

From the factor graph we start the message passing of the sum-product algorithm from the left end to the root x_n:

Left side:

$$\mu_{x_1 \to \psi_{1,2}} = 1$$

$$\mu_{\psi_{1,2} \to x_2} = \sum_{x_1} \psi_{1,2}(x_1, x_2)$$

$$\mu_{x_i \to \psi_{i,i+1}} = \mu_{\psi_{i-1,i} \to x_i} \quad \forall 1 < i < n-1$$

$$\mu_{\psi_{i,i+1} \to x_{i+1}} = \sum_{x_i} \psi_{i,i+1}(x_i, x_{i+1}) \mu_{x_i \to \psi_{i,i+1}} \quad \forall 1 < i < n$$

With that we define μ_α:

$$\mu_\alpha(x_n) := \mu_{\psi_{n-1,n} \to x_n}$$
$$= \sum_{x_{n-1}} \psi_{n-1,n}(x_{n-1}, x_n) \mu_{x_{n-1} \to \psi_{n-1,n}}$$
$$= \sum_{x_{n-1}} \psi_{n-1,n}(x_{n-1}, x_n) \mu_{\psi_{n-2,n-1} \to x_{n-1}}$$
$$= \sum_{x_{n-1}} \psi_{n-1,n}(x_{n-1}, x_n) \mu_\alpha(x_{n-1})$$

Same we do from the right side:

$$\mu_{x_N \to \psi_{N-1,N}} = 1$$
$$\mu_{\psi_{N-1,N} \to x_{N-1}} = \sum_{x_N} \psi_{N-1,N}(x_{N-1}, x_N)$$
$$\mu_{x_i \to \psi_{i-1,i}} = \mu_{\psi_{i+1,i} \to x_i} \quad \forall n + 1 < i < N$$
$$\mu_{\psi_{i-1,i} \to x_{i-1}} = \sum_{x_i} \psi_{i-1,i}(x_{i-1}, x_i) \mu_{x_i \to \psi_{i-1,i}} \quad \forall n < i < N$$

With that we define μ_β:

$$\mu_\beta(x_n) := \mu_{\psi_{n,n+1} \to x_n}$$
$$= \sum_{x_{n+1}} \psi_{n,n+1}(x_n, x_{n+1}) \mu_{x_{n+1} \to \psi_{n,n+1}}$$
$$= \sum_{x_{n+1}} \psi_{n,n+1}(x_n, x_{n+1}) \mu_{\psi_{n+2,n+1} \to x_{n+1}}$$
$$= \sum_{x_{n+1}} \psi_{n,n+1}(x_n, x_{n+1}) \mu_\beta(x_{n+1})$$

Given the two neighbours of x_n we can write (Bishop 8.63):

$$p(x_n) = \frac{1}{Z} \cdot \mu_{\psi_{n-1,n} \to x_n} \cdot \mu_{\psi_{n,n+1} \to x_n}$$
$$= \frac{1}{Z} \cdot \mu_\alpha(x_n) \cdot \mu_\beta(x_n)$$

The normalizing constant Z can be computed from the complete joint distribution.

Problem 3.

When x_N is observed this will only affect the last potential $\psi_{N-1,N}(x_{N-1}, x_N)$. Thus we can extend this to only pass the observed value. Assuming we have

a observed value of $x_N = \chi$ we can introduce a indicator function $\mathbb{1}[x_N = \chi]$ which one-hot encode the observed value:

$$\mathbb{1}[x_N = \chi] = \begin{cases} 1 & \text{if } x_N = \chi \\ 0 & \text{else} \end{cases}$$

Given that the last potential becomes:

$$\psi_{N-1,N}(x_{N-1}, x_N) \leftarrow \psi_{N-1,N}(x_{N-1}, x_N)\mathbb{1}[x_N = \chi]$$

and we can keep all message passes after that.

Problem 4.

We get the marginal distribution by summing over the join distribution of x without x_s and get:

$$\begin{aligned} p(x_s) &= \sum_{x\backslash x_s} p(x) \\ &= \sum_{x\backslash x_s} f_s(x_s) \prod_{i\in\text{ne}(f_s)} \prod_{j\in\text{ne}(x_i)\backslash f_s} F_j(x_i, X_j) \\ &= f_s(x_s) \prod_{i\in\text{ne}(f_s)} \prod_{j\in\text{ne}(x_i)\backslash f_s} \left(\sum_{X_j} F_j(x_i, X_j) \right) \\ &= f_s(x_s) \prod_{i\in\text{ne}(f_s)} \prod_{j\in\text{ne}(x_i)\backslash f_s} \mu_{f_j\to x_i}(x_i) \\ &= f_s(x_s) \prod_{i\in\text{ne}(f_s)} \mu_{x_i\to f_s}(x_i) \end{aligned}$$

With X_j being the nodes in the sub-graph of x_j without x_i and $F_j(x_i, X_j)$ being the product of this sub-tree.