Chapter 19. Undirected graphical models (Markov random fields)

19.1.
By definition,

$$Z(\boldsymbol{\theta}) = \sum_{\mathbf{y}} \prod_{c \in \mathcal{C}} \psi_c(\mathbf{y}_c|\boldsymbol{\theta}_c)$$

Differentiating this gives:

$$\frac{\partial \log Z(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}_c} = \frac{1}{Z(\boldsymbol{\theta})} \sum_{\mathbf{y}} \frac{\partial}{\partial \boldsymbol{\theta}_c} \prod_{c' \in \mathcal{C}} \psi_{c'}(\mathbf{y}_{c'}|\boldsymbol{\theta}_{c'})$$

$$= \frac{1}{Z(\boldsymbol{\theta})} \sum_{\mathbf{y}} \prod_{c' \in \mathcal{C} \setminus \{c\}} \psi_{c'}(\mathbf{y}_{c'}|\boldsymbol{\theta}_{c'}) \frac{\partial}{\partial \boldsymbol{\theta}_c} e^{\boldsymbol{\theta}_c^T \boldsymbol{\phi}_c(\mathbf{y}_c)}$$

$$= \frac{1}{Z(\boldsymbol{\theta})} \sum_{\mathbf{y}} \boldsymbol{\phi}_c(\mathbf{y}_c) \prod_{c' \in \mathcal{C}} \psi_{c'}(\mathbf{y}_{c'}|\boldsymbol{\theta}_{c'})$$

$$= \sum_{\mathbf{y}} \boldsymbol{\phi}_c(\mathbf{y}_c) p(\mathbf{y}|\boldsymbol{\theta}) = \mathbb{E}[\boldsymbol{\phi}_c(\mathbf{y}_c)|\boldsymbol{\theta}].$$

19.2.
a. We have:

$$\boldsymbol{\Sigma}^{-1} = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}$$

There is no marginal independency. There is only one conditional independency: $X_1 \perp X_3 | X_2$.
b. We have:

$$\boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \frac{3}{4} & -\frac{1}{2} & \frac{1}{4} \\ -\frac{1}{2} & 1 & -\frac{1}{2} \\ \frac{1}{4} & -\frac{1}{2} & \frac{3}{4} \end{pmatrix}$$

There is no conditional independency, therefore, the model have to be a fully connected graph, canceling the marginal independency $X_1 \perp X_3$.
c. The exponential term of the joint distribution is:

$$-\frac{1}{2}e^{x_1^2+(x_2-x_1)^2+(x_3-x_2)^2}$$

Which gives $\boldsymbol{\mu} = (0,0,0)^T$ and the precision and covariance matrix as:

$$\mathbf{\Sigma}^{-1} = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{pmatrix}, \mathbf{\Sigma} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{pmatrix}$$

d. There is no marginal independency. The only conditional independency is $X_1 \perp X_3|X_2$. The corresponding undirected graphical model is $X_1 - X_2 - X_3$.

19.3.
We have:

$$\mathbf{A}^{-1} = \begin{pmatrix} \frac{1}{8} & -\frac{1}{24} & 0 \\ -\frac{1}{24} & \frac{5}{36} & -\frac{1}{24} \\ 0 & -\frac{1}{24} & \frac{1}{8} \end{pmatrix}, \mathbf{B}^{-1} = \begin{pmatrix} \frac{1}{7} & \frac{1}{21} & 0 \\ \frac{1}{21} & \frac{1}{7} & \frac{1}{21} \\ 0 & \frac{1}{21} & \frac{1}{7} \end{pmatrix},$$

$$\mathbf{C}^{-1} = \begin{pmatrix} \frac{8}{63} & -\frac{1}{21} & \frac{1}{63} \\ -\frac{1}{21} & \frac{1}{7} & -\frac{1}{21} \\ \frac{1}{63} & -\frac{1}{21} & \frac{8}{63} \end{pmatrix}, \mathbf{D}^{-1} = \begin{pmatrix} \frac{1}{8} & \frac{1}{24} & \frac{1}{72} \\ \frac{1}{24} & \frac{1}{8} & \frac{1}{24} \\ \frac{1}{72} & \frac{1}{24} & \frac{1}{8} \end{pmatrix}.$$

a. Since we have $X_1 \perp X_3|X_2$, only A and B can be the covariance matrix.
b. Only C and D can be the inverse covariance matrix.
c. Since we have $X_1 \perp X_3$, only C and D can be the covariance matrix.
d. Only A and B can be the inverse covariance matrix.
e. A is true by the properties of marginal Gaussian density function.
B is not true. We have:

$$\mathbf{\Omega}_{(1:2)} = \begin{pmatrix} \frac{4}{3} & -\frac{2}{3} \\ -\frac{2}{3} & \frac{4}{3} \end{pmatrix}.$$

19.4.
To train an MRF: For each iteration, we compute all the marginals for each data case, taking $O(Nc)$ time, and compute the expected feature, taking

$O(1)$ time, so the whole training time is $O(r(Nc+1))$.
To train an CRF: For each iteration, we compute all the marginals and perform inference to compute the partition function for each data case, taking $O(Nc+N)$ time, so the whole training time is $O(r(Nc+N))$.

19.5.

$$p(x_i = 1|\mathbf{x}_{-i}, \boldsymbol{\theta}) = \frac{p(x_i = 1, \mathbf{x}_{-i}|\boldsymbol{\theta})}{p(\mathbf{x}_{-i}|\boldsymbol{\theta})} = \frac{p(x_i = 1, \mathbf{x}_{-i}|\boldsymbol{\theta})}{p(x_i = 0, \mathbf{x}_{-i}|\boldsymbol{\theta}) + p(x_i = 1, \mathbf{x}_{-i}|\boldsymbol{\theta})}$$
$$= \frac{1}{1 + \frac{p(x_i=0,\mathbf{x}_{-i}|\boldsymbol{\theta})}{p(x_i=1,\mathbf{x}_{-i}|\boldsymbol{\theta})}} = \frac{1}{1 + \frac{1}{e^{h_i}\prod_{<i,j>} e^{J_{ij}x_j}}} = \sigma(h_i + \sum_{j\neq i} J_{ij}x_j).$$