CS5340: Uncertainty Modeling in AI

Tutorial 5: Solutions

Released: Feb. 21, 2024

Problem 1. (Linear Gaussian)

For this tutorial problem, we will consider a specific DGM that is the basis for more sophisticated models such as Probabilistic PCA and Linear Dynamical Systems. This model is called the Linear-Gaussian Model. *Note:* for this problem, we will be denoting random variables with lower case letters, and bolded lowercase letters to represent vectors, and bolded uppercase letters to represent matrices.

Problem 1.a. We will build our way up towards this model. As a prelude, consider K *independent* univariate Gaussian random variables $x_1, x_2, \ldots, x_K$,

$$p(x_k) = \mathcal{N}(\mu_k, \sigma_k^2)$$

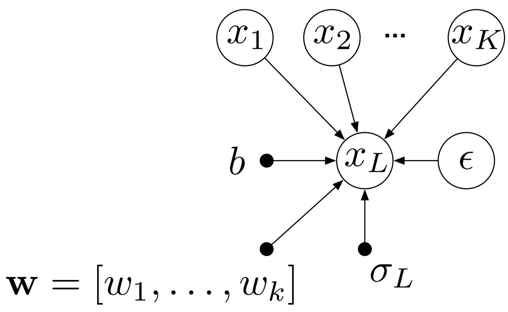
for $k = 1, 2, \ldots, K$. Define the random variable x_L,

$$x_L = b + \sigma_L \epsilon + \sum_{k=1}^{K} w_k x_k$$

where $\epsilon \sim \mathcal{N}(0, 1)$.

1. Draw out the DGM for the model described above.
2. Show that $p(x_L|x_1, \ldots, x_K) = \mathcal{N}\left(b + \sum_{k=1}^{K} w_k x_k, \sigma_L^2\right)$. In other words, x_L is Gaussian distributed with mean $b + \sum_{k=1}^{K} w_k x_k$ and variance σ_L^2.
3. Define the random variable $\mathbf{x} = (x_1, x_2, \ldots, x_K, x_L)$. Show that $\mathbf{x}$ is a *multivariate* Gaussian random variable. *Hint: Consider the definition of the multivariate Gaussian and the properties of Gaussians.*

Solution:



Solution: Since ϵ follows Gaussian distribution, then the linear transformation of ϵ plus a constant is still Gaussian. As such, we just need to compute mean and variance of the Gaussian distribution.

$$\mathbb{E}\left[x_L|x_1, \cdots, x_K\right] = \mathbb{E}\left[b + \sigma_L \epsilon + \sum_{k=1}^{K} w_k x_k\right] \quad (1)$$

$$= \mathbb{E}[b] + \sigma_L \mathbb{E}[\epsilon] + \sum_{k=1}^{K} w_k \mathbb{E}[x_k] \quad (2)$$

Since $x_1, \cdots, x_K$ are known, $\mathbb{E}[x_k] = x_k$, therefore $\sum_{k=1}^K w_k \mathbb{E}[x_k] = \sum_{k=1}^K w_k x_k$. Since b is a constant, $\mathbb{E}[b] = b$. Also, given $\mathbb{E}[\epsilon] = 0$. Thus,

$$\mathbb{E}\left[x_L | x_1, \cdots, x_K\right] = b + \sum_{k=1}^K w_k x_k \tag{3}$$

Similarly, for variance, we have

$$\begin{aligned} \mathrm{Var}\left[x_L | x_1, \cdots, x_K\right] &= \mathrm{Var}\left[b + \sigma_L \epsilon + \sum_{k=1}^K w_k x_k\right] &(4)\\ &= \mathrm{Var}[b] + \sigma_L^2 \mathrm{Var}[\epsilon] + \sum_{k=1}^K w_k^2 \mathrm{Var}[x_k] &(5)\\ &= \sigma_L^2 \mathrm{Var}[\epsilon] = \sigma_L^2 &(6) \end{aligned}$$

Solution: Since

$$p(\mathbf{x}_{\pi_i}) \propto \exp\left\{-\frac{1}{2}\left[\sum_{k \in \mathbf{x}_{\pi_i}} \frac{1}{\sigma_k^2}(x_k - \mu_k)^2\right]\right\} \tag{7}$$

$$p(x_L | \mathbf{x}_{\pi_i}) \propto \exp\left\{-\frac{1}{2}\left[\frac{1}{\sigma_L^2}(x_L - (b + \sum_{k \in \mathbf{x}_{\pi_i}} w_k x_k))^2\right]\right\} \tag{8}$$

$$p(\mathbf{x}_{\pi_i}) p(x_L | \mathbf{x}_{\pi_i}) \propto \exp\left\{-\frac{1}{2}\left[\sum_{k \in \mathbf{x}_{\pi_i}} \frac{1}{\sigma_k^2}(x_k - \mu_k)^2 + \frac{1}{\sigma_L^2}(x_L - (b + \sum_{k \in \mathbf{x}_{\pi_i}} w_k x_k))^2\right]\right\} \tag{9}$$

From the form of $p(\mathbf{x}_{\pi_i}) p(x_L | \mathbf{x}_{\pi_i})$ is of the form $\frac{1}{Z} exp(-g(\mathbf{x})$ where $g(\mathbf{x})$ is a quadratic function that is positive definite. Hence, $p(\mathbf{x}_{\pi_i}) p(x_L | \mathbf{x}_{\pi_i})$ is multivariate Gaussian.

We can obtain the mean and covariance matrix for this distribution. Denote $\mathbf{x}_{\pi_i} = (x_1, \cdots, x_K)$, $\mathbf{w} = [w_1, \cdots, w_K]^T$. We look at the terms in exponential,

$$\begin{aligned} &\sum_{k \in \mathbf{x}_{\pi_i}} \frac{1}{\sigma_k^2}(x_k - \mu_k)^2 + \frac{1}{\sigma_L^2}(x_L - (b + \sum_{k \in \mathbf{x}_{\pi_i}} w_k x_k))^2 &(10)\\ =&(\mathbf{x}_{\pi_i} - \boldsymbol{\mu}_{\pi_i})^T \boldsymbol{\Sigma}_{\pi_i}^{-1} (\mathbf{x}_{\pi_i} - \boldsymbol{\mu}_{\pi_i}) + \left(x_L - (b + \mathbf{w}^T \mathbf{x}_{\pi_i})\right)^T \boldsymbol{\Sigma}_L^{-1} \left(x_L - (b + \mathbf{w}^T \mathbf{x}_{\pi_i})\right) &(11)\\ =&(\mathbf{x}_{\pi_i} - \boldsymbol{\mu}_{\pi_i})^T \boldsymbol{\Sigma}_{\pi_i}^{-1} (\mathbf{x}_{\pi_i} - \boldsymbol{\mu}_{\pi_i}) &(12)\\ &+ \left(x_L - (b + \mathbf{w}^T \boldsymbol{\mu}_{\pi_i}) - \mathbf{w}^T (\mathbf{x}_{\pi_i} - \boldsymbol{\mu}_{\pi_i})\right)^T \boldsymbol{\Sigma}_L^{-1} \left(x_L - (b + \mathbf{w}^T \boldsymbol{\mu}_{\pi_i}) - \mathbf{w}^T (\mathbf{x}_{\pi_i} - \boldsymbol{\mu}_{\pi_i})\right) &(13)\\ =&(\mathbf{x}_{\pi_i} - \boldsymbol{\mu}_{\pi_i})^T \left[\boldsymbol{\Sigma}_{\pi_i}^{-1} + \boldsymbol{w} \boldsymbol{\Sigma}_L^{-1} \boldsymbol{w}^T\right] (\mathbf{x}_{\pi_i} - \boldsymbol{\mu}_{\pi_i}) &(14)\\ &+ \left(x_L - (b + \mathbf{w}^T \boldsymbol{\mu}_{\pi_i})\right)^T \boldsymbol{\Sigma}_L^{-1} \left(x_L - (b + \mathbf{w}^T \boldsymbol{\mu}_{\pi_i})\right) - 2(\mathbf{x}_{\pi_i} - \boldsymbol{\mu}_{\pi_i})^T \boldsymbol{w} \boldsymbol{\Sigma}_L^{-1} (x_L - \mu_L) &(15)\\ =&\begin{bmatrix} \mathbf{x}_{\pi_i} - \boldsymbol{\mu}_{\pi_i} \\ x_L - \mu_L \end{bmatrix}^T \begin{bmatrix} \boldsymbol{\Sigma}_{\pi_i}^{-1} + \boldsymbol{w} \boldsymbol{\Sigma}_L^{-1} \boldsymbol{w}^T & \boldsymbol{\Sigma}_L^{-1} \boldsymbol{w}^T \\ \boldsymbol{w} \boldsymbol{\Sigma}_L^{-1} & \boldsymbol{\Sigma}_L^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{x}_{\pi_i} - \boldsymbol{\mu}_{\pi_i} \\ x_L - \mu_L \end{bmatrix} &(16) \end{aligned}$$

where $\mu_L = b + \mathbf{w}^T \mathbf{x}_{\pi_i}$, $\boldsymbol{\Sigma}_L = [\sigma_L^2]$ and $\boldsymbol{\Sigma}_{\pi_i} = \text{diagonal}(\sigma_1^2, \cdots, \sigma_K^2)$.

Now we verify

$$\begin{bmatrix} \boldsymbol{\Sigma}_{\pi_i}^{-1} + \boldsymbol{w}\boldsymbol{\Sigma}_L^{-1}\boldsymbol{w}^T & \boldsymbol{\Sigma}_L^{-1}\boldsymbol{w}^T \\ \boldsymbol{w}\boldsymbol{\Sigma}_L^{-1} & \boldsymbol{\Sigma}_L^{-1} \end{bmatrix}$$

is positive definite. For any $\mathbf{u}$, we have

$$\mathbf{u}^T \begin{bmatrix} \boldsymbol{\Sigma}_{\pi_i}^{-1} + \boldsymbol{w}\boldsymbol{\Sigma}_L^{-1}\boldsymbol{w}^T & \boldsymbol{\Sigma}_L^{-1}\boldsymbol{w}^T \\ \boldsymbol{w}\boldsymbol{\Sigma}_L^{-1} & \boldsymbol{\Sigma}_L^{-1} \end{bmatrix} \mathbf{u} = \mathbf{u}^T \begin{bmatrix} 1 & 0 \\ -\boldsymbol{w}^T & 1 \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_L^{-1} & 0 \\ 0 & \boldsymbol{\Sigma}_{\pi_i}^{-1} \end{bmatrix} \begin{bmatrix} 1 & -\boldsymbol{w} \\ 0 & 1 \end{bmatrix} \mathbf{u} \tag{17}$$

$$= \left[\mathbf{u} \begin{bmatrix} 1 & -\boldsymbol{w} \\ 0 & 1 \end{bmatrix}\right]^T \begin{bmatrix} \boldsymbol{\Sigma}_L^{-1} & 0 \\ 0 & \boldsymbol{\Sigma}_{\pi_i}^{-1} \end{bmatrix} \left[\begin{bmatrix} 1 & -\boldsymbol{w} \\ 0 & 1 \end{bmatrix} \mathbf{u}\right] > 0 \tag{18}$$

The last inequality is given by that $\boldsymbol{\Sigma}_L$ and $\boldsymbol{\Sigma}_{\pi_i}$ are positive definite.

Therefore,

$$\begin{bmatrix} \boldsymbol{\Sigma}_{\pi_i}^{-1} + \boldsymbol{w}\boldsymbol{\Sigma}_L^{-1}\boldsymbol{w}^T & \boldsymbol{\Sigma}_L^{-1}\boldsymbol{w}^T \\ \boldsymbol{w}\boldsymbol{\Sigma}_L^{-1} & \boldsymbol{\Sigma}_L^{-1} \end{bmatrix}$$

is a covariance matrix.

Define

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_{\pi_i} \\ b + \mathbf{w}^T\mathbf{x}_{\pi_i} \end{bmatrix}$$

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{\pi_i}^{-1} + \boldsymbol{w}\boldsymbol{\Sigma}_L^{-1}\boldsymbol{w}^T & \boldsymbol{\Sigma}_L^{-1}\boldsymbol{w}^T \\ \boldsymbol{w}\boldsymbol{\Sigma}_L^{-1} & \boldsymbol{\Sigma}_L^{-1} \end{bmatrix}^{-1}$$

We have

$$p(x_L, \mathbf{x}_{\pi_i}) = p(\mathbf{x}_{\pi_i})p(x_L|\mathbf{x}_{\pi_i}) \propto \exp\left\{-\frac{1}{2}\left(\begin{bmatrix} \mathbf{x}_{\pi_i} \\ x_L \end{bmatrix} - \boldsymbol{\mu}\right)^T \boldsymbol{\Sigma}^{-1} \left(\begin{bmatrix} \mathbf{x}_{\pi_i} \\ x_L \end{bmatrix} - \boldsymbol{\mu}\right)\right\} \tag{19}$$

Therefore, $p(x_L, \mathbf{x}_{\pi_i})$ is Gaussian distributed.

Problem 1.b. Let's now move to the more complex case. Consider an *arbitrary* DGM G where each node j without any parents is Gaussian distributed with mean μ_j and variance σ_j^2. The remaining nodes are defined as

$$x_i = b_i + \left(\sum_{j \in x_{\pi_i}} w_{i,j}x_j\right) + \sigma_i\epsilon_i$$

where x_{π_i} denotes the set of node i's parents and ϵ_i is the standard normal random variable $\epsilon_i \sim \mathcal{N}(0, 1)$.

1. Show that each node x_i has the conditional distribution: $p(x_i|x_{\pi_i}) = \mathcal{N}\left(b_i + \sum_{j\in x_{\pi_i}} w_{i,j}x_j, \sigma_i^2\right)$
2. Define the random variable $\mathbf{x} = (x_1, x_2, \ldots, x_D)$. Show that $\mathbf{x}$ is a *multivariate Gaussian*.

Solution: We just apply the same derivation from problem 2.a.

Since ϵ_i follows Gaussian distribution, then the linear transformation of ϵ_i plus constant is still Gaussian distribution. Then we just need to compute mean and variance of the Gaussian distribution.

$$\mathbb{E}\left[x_i|x_{\pi_i}\right] = \mathbb{E}\left[b_i + \sigma_i\epsilon_i + \sum_{j\in x_{\pi_i}} w_jx_j\right] \tag{20}$$

$$= \mathbb{E}[b_i] + \sigma_i\mathbb{E}[\epsilon_i] + \sum_{j\in x_{\pi_i}} w_j\mathbb{E}[x_j] \tag{21}$$

$$\tag{22}$$

Since x_{π_i} are known, $\mathbb{E}[x_j] = x_j, (j \in x_{\pi_i})$, therefore $\sum_{j \in x_{\pi_i}} w_j \mathbb{E}[x_j] = \sum_{k \in x_{\pi_i}} w_j x_j$. Since b_i is a constant, $\mathbb{E}[b_i] = b_i$. Also, given $\mathbb{E}[\epsilon_i] = 0$. Thus,

$$\mathbb{E}\left[x_i | x_{\pi_i}\right] = b_i + \sum_{j \in x_{\pi_i}} w_j x_j \tag{23}$$

Similarly, for variance, we have

$$\begin{aligned}
\mathrm{Var}\left[x_i | x_{\pi_i}\right] &= \mathrm{Var}\left[b_i + \sigma_i \epsilon_i + \sum_{k \in x_{\pi_i}} w_j x_j\right] && (24)\\
&= \mathrm{Var}[b_i] + \sigma_i^2 \mathrm{Var}[\epsilon_i] + \sum_{k \in x_{\pi_i}} w_j^2 \mathrm{Var}[x_j] && (25)\\
&= \sigma_i^2 \mathrm{Var}[\epsilon_i] = \sigma_i^2 && (26)
\end{aligned}$$

Solution: We know that $p(x_i|\mathbf{x}_{\pi_i})$ is Gaussian distributed. Let's first assume $p(\mathbf{x}_{\pi_i})$ is also Gaussian. Next, we prove $p(x_i, \mathbf{x}_{\pi_i}) = p(x_i|\mathbf{x}_{\pi_i})p(\mathbf{x}_{\pi_i})$ is also Gaussian. We can perform induction from root node following the topological order in the graph to show that the joint distribution of all nodes is a multivariate Gaussian. Below, we just outline the key argument for the inductive step (a full proof would require starting from the base case, making the inductive hypothesis, followed by the inductive step)

Since

$$p(\mathbf{x}_{\pi_i}) \propto \exp\left\{-\frac{1}{2}\left[\sum_{k \in \mathbf{x}_{\pi_i}} \frac{1}{\sigma_k^2}(x_k - \mu_k)^2\right]\right\} \tag{27}$$

$$p(x_i|\mathbf{x}_{\pi_i}) \propto \exp\left\{-\frac{1}{2}\left[\frac{1}{\sigma_i^2}(x_i - (b_i + \sum_{k \in \mathbf{x}_{\pi_i}} w_k x_k))^2\right]\right\} \tag{28}$$

$$p(\mathbf{x}_{\pi_i})p(x_i|\mathbf{x}_{\pi_i}) \propto \exp\left\{-\frac{1}{2}\left[(\mathbf{x}_{\pi_i} - \boldsymbol{\mu}_{\pi_i})^\top \Sigma^{-1}{}_{\pi_i}(\mathbf{x}_{\pi_i} - \boldsymbol{\mu}_{\pi_i}) + \frac{1}{\sigma_i^2}(x_i - (b_i + \sum_{k \in \mathbf{x}_{\pi_i}} w_k x_k))^2\right]\right\} \tag{29}$$

As before, we see that the terms within the exponential form a positive definite quadratic function.

Therefore, $p(x_i, \mathbf{x}_{\pi_i})$ is Gaussian distributed. We start from the nodes without parents and analyze each node in topological order, and by induction, we can show that the joint distribution of all nodes is multivariate Gaussian.

Problem 1.c. We can determine the mean of $\mathbf{x}$ using a recursive method. Note that $\mathbb{E}[\mathbf{x}] = (\mathbb{E}[x_1], \ldots, \mathbb{E}[x_D])^\top$. Show that the expectation of each component $\mathbb{E}[x_i]$ is given by:

$$\mathbb{E}[x_i] = b_i + \sum_{j \in x_{\pi_i}} w_{i,j} \mathbb{E}[x_j]$$

Solution: Since $x_i = b_i + \left(\sum_{j \in x_{\pi_i}} w_{i,j} x_j\right) + \sigma_i \epsilon_i$ and $\epsilon_i \sim \mathcal{N}(0, 1)$

$$\mathbb{E}[x_i] = \mathbb{E}\left[b_i + \left(\sum_{j \in x_{\pi_i}} w_{i,j} x_j\right) + \sigma_i \epsilon_i\right] \tag{30}$$

$$= \mathbb{E}[b_i] + \sum_{j \in x_{\pi_i}} w_{i,j} \mathbb{E}[(x_j)] + \sigma_i \mathbb{E}[\epsilon_i] \tag{31}$$

$$= b_i + \sum_{j \in x_{\pi_i}} w_{i,j} \mathbb{E}[(x_j)] \tag{32}$$

Problem 1.d. Likewise, we can determine the covariance matrix of $\mathbf{x}$. Note that

$$\mathbf{\Sigma}_{ij} = \mathrm{Cov}[x_i, x_j] = \mathbb{E}[(x_i - \mathbb{E}[x_i])(x_j - \mathbb{E}[x_j])]$$

1. Show that $\mathrm{Cov}[x_i, x_j] = I_{ij} \sigma_j^2 + \sum_{k \in x_{\pi_j}} w_{j,k} \mathrm{Cov}[x_i, x_k]$
2. If the DGM G has no edges, is the covariance matrix $\mathbf{\Sigma}$ a spherical, diagonal, or general symmetric covariance matrix? How many parameters does it have?
3. If the DGM G is fully-connected, what kind of matrix is the covariance matrix $\mathbf{\Sigma}$? Is it spherical, diagonal, or a general symmetric covariance matrix? How many parameters does it have?

Solution: Since $x_j = b_j + \left(\sum_{k \in x_{\pi_j}} w_{j,k} x_k\right) + \sigma_j \epsilon_j$

$$\mathrm{Cov}[x_i, x_j] = \mathbb{E}\left[\left[x_i - \mathbb{E}[x_i]\right]\left[x_j - \mathbb{E}[x_j]\right]\right] \tag{33}$$

$$= \mathbb{E}\left[\left[x_i - \mathbb{E}[x_i]\right]\left[b_j + \left(\sum_{k \in x_{\pi_j}} w_{j,k} x_k\right) + \sigma_j \epsilon_j - \mathbb{E}[x_j]\right]\right] \tag{34}$$

$$= \mathbb{E}\left[\left[x_i - \mathbb{E}[x_i]\right]\left[b_j + \left(\sum_{k \in x_{\pi_j}} w_{j,k} x_k\right) + \sigma_j \epsilon_j - \left(b_j + \sum_{k \in x_{\pi_j}} w_{j,k} \mathbb{E}[x_k]\right)\right]\right] \tag{35}$$

$$= \mathbb{E}\left[\left[x_i - \mathbb{E}[x_i]\right]\left[\sum_{k \in x_{\pi_j}} w_{j,k} \left(x_k - \mathbb{E}[x_k]\right) + \sigma_j \epsilon_j\right]\right] \tag{36}$$

$$= \sum_{k \in x_{\pi_j}} w_{j,k} \mathbb{E}\left[\left[x_i - \mathbb{E}[x_i]\right]\left[x_k - \mathbb{E}[x_k]\right]\right] + \sigma_j \mathbb{E}\left[\left[x_i - \mathbb{E}[x_i]\right] \epsilon_j\right] \tag{37}$$

$$= \sum_{k \in x_{\pi_j}} w_{j,k} \mathrm{Cov}\left(x_i, x_k\right) + \sigma_j \mathbb{E}\left[\left[x_i - \mathbb{E}[x_i]\right] \epsilon_j\right] \tag{38}$$

Since $x_i = b_i + \left(\sum_{j \in x_{\pi_i}} w_{i,j} x_j\right) + \sigma_i \epsilon_i$

$$\sigma_j \mathbb{E}\left[\left[x_i - \mathbb{E}[x_i]\right] \epsilon_i\right] = \sigma_j \mathbb{E}\left[\left[\sum_{k \in x_{\pi_i}} w_{i,k} \left(x_k - \mathbb{E}[x_k]\right) + \sigma_i \epsilon_i\right] \epsilon_j\right] \tag{39}$$

$$\tag{40}$$

Since x_k and ϵ_j are independent, therefore,

$$\mathbb{E}\left[\left[\sum_{k\in x_{\pi_i}} w_{i,k}\left(x_k - \mathbb{E}[x_k]\right)\right]\epsilon_j\right] = \mathbb{E}\left[\sum_{k\in x_{\pi_i}} w_{i,k}\left(x_k - \mathbb{E}[x_k]\right)\right]\mathbb{E}\left[\epsilon_j\right] \tag{41}$$

$$= \mathbb{E}\left[\sum_{k\in x_{\pi_i}} w_{i,k}\left(x_k - \mathbb{E}[x_k]\right)\right]\cdot 0 \tag{42}$$

$$= 0 \tag{43}$$

Then, if $i = j$ we have

$$\sigma_i\mathbb{E}\left[\left[x_i - \mathbb{E}[x_i]\right]\epsilon_i\right] = \sigma_i\sigma_i\mathbb{E}\left[\epsilon_i\epsilon_i\right] \tag{44}$$

$$= \sigma_i^2\mathbb{E}\left[\epsilon_i\epsilon_i\right] \tag{45}$$

$$= \sigma_i^2\mathbb{E}\left[\left[\epsilon_i - \mathbb{E}[\epsilon_i]\right]\left[\epsilon_i - \mathbb{E}[\epsilon_i]\right]\right] \tag{46}$$

$$= \sigma_i^2\mathrm{Var}[\epsilon_i] = \sigma_i^2 \tag{47}$$

$$\tag{48}$$

If $i \neq j$, we have ϵ_i and ϵ_j independent.

$$\sigma_i\mathbb{E}\left[\left[x_i - \mathbb{E}[x_i]\right]\epsilon_i\right] = \sigma_i\sigma_j\mathbb{E}\left[\epsilon_i\epsilon_j\right] \tag{49}$$

$$= \sigma_i\sigma_j\mathbb{E}\left[\epsilon_i\right]\mathbb{E}\left[\epsilon_j\right] \tag{50}$$

$$= 0 \tag{51}$$

$$\tag{52}$$

Put all these together,

$$\mathrm{Cov}[x_i, x_j] = I_{ij}\sigma_j^2 + \sum_{k\in x_{\pi_j}} w_{j,k}\mathrm{Cov}[x_i, x_k] \tag{53}$$

where I_{ij} equals 1, if $i = j$; and equals 0, if $i \neq j$.

Solution: If G has no edges, the covariance matrix is a diagonal matrix and the number of parameters is D.

Solution: If G is fully-connected, the covariance matrix is a general symmetric matrix. The number of parameters is $\frac{D(D-1)}{2}$.

Problem 1.e. (Challenge) Consider now the situation where each node in G is a multivariate Gaussian random variable. More concretely, each node j without parents is multivariate Gaussian distributed with mean $\boldsymbol{\mu}_j$ and variance $\boldsymbol{\Sigma}_j$. The conditional for the remaining nodes are also multivariate Gaussian:

$$p(\mathbf{x}_i|\mathbf{x}_{\pi_i}) = \mathcal{N}\left(\mathbf{b}_i + \sum_{j\in x_{\pi_i}} \mathbf{W}_{ij}\mathbf{x}_j, \boldsymbol{\Sigma}_i\right)$$

Show that the joint distribution over *all* variables is multivariate Gaussian.

Solution: First assume $\mathbf{x}_{\pi_i} \sim \mathcal{N}(\mu_{\pi i}, \mathbf{\Sigma}_{\pi_i})$, and we show that $p(\mathbf{x}_i, \mathbf{x}_{\pi_i}) = p(\mathbf{x}_i|\mathbf{x}_{\pi_i})p(\mathbf{x}_{\pi_i})$ is Gaussian distributed. Then, we apply this property on the graph G. We start from root nodes and, by induction, we can prove the whole joint distribution is a multivariate Gaussian.

Since

$$p(\mathbf{x}_i|\mathbf{x}_{\pi_i}) \propto \exp\left\{-\frac{1}{2}\left(\mathbf{x}_i - (\mathbf{b}_i + \sum_{j\in x_{\pi_i}} \mathbf{W}_{ij}\mathbf{x}_j)\right)^T \mathbf{\Sigma}_i^{-1} \left(\mathbf{x}_i - (\mathbf{b}_i + \sum_{j\in x_{\pi_i}} \mathbf{W}_{ij}\mathbf{x}_j)\right)\right\} \tag{54}$$

$$p(\mathbf{x}_{\pi_i}) \propto \exp\left\{-\frac{1}{2}\left(\mathbf{x}_{\pi_i} - \mu_{\pi i}\right)^T \mathbf{\Sigma}_{\pi_i}^{-1} \left(\mathbf{x}_{\pi_i} - \mu_{\pi i}\right)\right\} \tag{55}$$

$$\begin{aligned} p(\mathbf{x}_i|\mathbf{x}_{\pi_i})p(\mathbf{x}_{\pi_i}) \propto \exp\{-\frac{1}{2}\left(\left(\mathbf{x}_{\pi_i} - \mu_{\pi i}\right)^T \mathbf{\Sigma}_{\pi_i}^{-1} \left(\mathbf{x}_{\pi_i} - \mu_{\pi i}\right)\right) && (56)\\ -\frac{1}{2}\left(\mathbf{x}_i - (\mathbf{b}_i + \sum_{j\in x_{\pi_i}} \mathbf{W}_{ij}\mathbf{x}_j)\right)^T \mathbf{\Sigma}_i^{-1} \left(\mathbf{x}_i - (\mathbf{b}_i + \sum_{j\in x_{\pi_i}} \mathbf{W}_{ij}\mathbf{x}_j)\right)\} && (57)\\ && (58)\end{aligned}$$

We analyze the terms in the exponential.

$$\begin{aligned} &\left(\mathbf{x}_{\pi_i} - \mu_{\pi i}\right)^T \mathbf{\Sigma}_{\pi_i}^{-1} \left(\mathbf{x}_{\pi_i} - \mu_{\pi i}\right) + \left(\mathbf{x}_i - (\mathbf{b}_i + \sum_{j\in x_{\pi_i}} \mathbf{W}_{ij}\mathbf{x}_j)\right)^T \mathbf{\Sigma}_i^{-1} \left(\mathbf{x}_i - (\mathbf{b}_i + \sum_{j\in x_{\pi_i}} \mathbf{W}_{ij}\mathbf{x}_j)\right)\} && (59)\\ =&\left(\mathbf{x}_{\pi_i} - \mu_{\pi i}\right)^T \mathbf{\Sigma}_{\pi_i}^{-1} \left(\mathbf{x}_{\pi_i} - \mu_{\pi i}\right) && (60)\\ &+\left(\mathbf{x}_i - \left((\mathbf{b}_i + \sum_{j\in x_{\pi_i}} \mathbf{W}_{ij}\boldsymbol{\mu}_j) + \sum_{j\in x_{\pi_i}} \mathbf{W}_{ij}(\mathbf{x}_j - \boldsymbol{\mu}_j)\right)\right)^T && (61)\\ &\mathbf{\Sigma}_i^{-1}\left(\mathbf{x}_i - \left((\mathbf{b}_i + \sum_{j\in x_{\pi_i}} \mathbf{W}_{ij}\boldsymbol{\mu}_j) + \sum_{j\in x_{\pi_i}} \mathbf{W}_{ij}(\mathbf{x}_j - \boldsymbol{\mu}_j)\right)\right) && (62)\\ =&\left[\begin{bmatrix}\mathbf{x}_i\\ \mathbf{x}_{\pi_i}\end{bmatrix} - \begin{bmatrix}\mathbf{b}_i + \sum_{j\in x_{\pi_i}} \mathbf{W}_{ij}\boldsymbol{\mu}_j\\ \boldsymbol{\mu}_{\pi_i}\end{bmatrix}\right]^T \begin{bmatrix}1 & 0\\ -\mathbf{W}_i^T & 1\end{bmatrix} && (63)\\ &\begin{bmatrix}\mathbf{\Sigma}_i^{-1} & 0\\ 0 & \mathbf{\Sigma}_{\pi_i}^{-1}\end{bmatrix}\begin{bmatrix}1 & -\mathbf{W}_i\\ 0 & 1\end{bmatrix}\left[\begin{bmatrix}\mathbf{x}_i\\ \mathbf{x}_{\pi_i}\end{bmatrix} - \begin{bmatrix}\mathbf{b}_i + \sum_{j\in x_{\pi_i}} \mathbf{W}_{ij}\boldsymbol{\mu}_j\\ \boldsymbol{\mu}_{\pi_i}\end{bmatrix}\right] && (64)\end{aligned}$$

where $\mathbf{W_i} = [\mathbf{W}_{i,1}, \cdots, \mathbf{W}_{i,K_{\pi_i}}]$

According to *Schur complement*[1],

$$\begin{bmatrix}1 & 0\\ -\mathbf{W}_i^T & 1\end{bmatrix}\begin{bmatrix}\mathbf{\Sigma}_i^{-1} & 0\\ 0 & \mathbf{\Sigma}_{\pi_i}^{-1}\end{bmatrix}\begin{bmatrix}1 & -\mathbf{W}_i\\ 0 & 1\end{bmatrix} = \begin{bmatrix}\mathbf{\Sigma}_i + \mathbf{W}_i^T\mathbf{\Sigma}_{\pi_i}\mathbf{W}_i & \mathbf{\Sigma}_{\pi_i}^T\mathbf{W}_i\\ \mathbf{W}_i^T\mathbf{\Sigma}_{\pi_i} & \mathbf{\Sigma}_{\pi_i}\end{bmatrix}^{-1} \tag{65}$$

Since $\mathbf{\Sigma}_{\pi_i}$ and $\mathbf{\Sigma}_i$ are positive definite, then

$$\mathbf{u}^T\begin{bmatrix}1 & 0\\ -\mathbf{W}_i^T & 1\end{bmatrix}\begin{bmatrix}\mathbf{\Sigma}_i^{-1} & 0\\ 0 & \mathbf{\Sigma}_{\pi_i}^{-1}\end{bmatrix}\begin{bmatrix}1 & -\mathbf{W}_i\\ 0 & 1\end{bmatrix}\mathbf{u} = \left[\begin{bmatrix}1 & -\mathbf{W}_i\\ 0 & 1\end{bmatrix}\mathbf{u}\right]^T\begin{bmatrix}\mathbf{\Sigma}_i^{-1} & 0\\ 0 & \mathbf{\Sigma}_{\pi_i}^{-1}\end{bmatrix}\left[\begin{bmatrix}1 & -\mathbf{W}_i\\ 0 & 1\end{bmatrix}\mathbf{u}\right] > 0 \tag{66}$$

[1]https://en.wikipedia.org/wiki/Schur_complement

hold for any $\mathbf{u}$, thus,

$$\begin{bmatrix} 1 & 0 \\ -\mathbf{W}_i^T & 1 \end{bmatrix} \begin{bmatrix} \mathbf{\Sigma}_i^{-1} & 0 \\ 0 & \mathbf{\Sigma}_{\pi_i}^{-1} \end{bmatrix} \begin{bmatrix} 1 & -\mathbf{W}_i \\ 0 & 1 \end{bmatrix}$$

is positive definite.

Therefore,

$$\left(\mathbf{x}_{\pi_i} - \mu_{\pi i}\right)^T \mathbf{\Sigma}_{\pi_i}^{-1} \left(\mathbf{x}_{\pi_i} - \mu_{\pi i}\right) + \left(\mathbf{x}_i - (\mathbf{b}_i + \sum_{j \in x_{\pi_i}} \mathbf{W}_{ij}\mathbf{x}_j)\right)^T \mathbf{\Sigma}_i^{-1} \left(\mathbf{x}_i - (\mathbf{b}_i + \sum_{j \in x_{\pi_i}} \mathbf{W}_{ij}\mathbf{x}_j)\right)\} \tag{67}$$

$$= \left[\begin{bmatrix} \mathbf{x}_i \\ \mathbf{x}_{\pi_i} \end{bmatrix} - \begin{bmatrix} \mathbf{b}_i + \sum_{j \in x_{\pi_i}} \mathbf{W}_{ij}\boldsymbol{\mu}_j \\ \boldsymbol{\mu}_{\pi_i} \end{bmatrix}\right]^T \begin{bmatrix} \mathbf{\Sigma}_i + \mathbf{W}_i^T \mathbf{\Sigma}_{\pi_i} \mathbf{W}_i & \mathbf{\Sigma}_{\pi_i}^T \mathbf{W}_i \\ \mathbf{W}_i^T \mathbf{\Sigma}_{\pi_i} & \mathbf{\Sigma}_{\pi_i} \end{bmatrix}^{-1} \left[\begin{bmatrix} \mathbf{x}_i \\ \mathbf{x}_{\pi_i} \end{bmatrix} - \begin{bmatrix} \mathbf{b}_i + \sum_{j \in x_{\pi_i}} \mathbf{W}_{ij}\boldsymbol{\mu}_j \\ \boldsymbol{\mu}_{\pi_i} \end{bmatrix}\right] \tag{68}$$

Denote

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_i \\ \mathbf{x}_{\pi_i} \end{bmatrix}$$

$$\boldsymbol{\mu} = \begin{bmatrix} \mathbf{b}_i + \sum_{j \in x_{\pi_i}} \mathbf{W}_{ij}\boldsymbol{\mu}_j \\ \boldsymbol{\mu}_{\pi_i} \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} \mathbf{\Sigma}_i + \mathbf{W}_i^T \mathbf{\Sigma}_{\pi_i} \mathbf{W}_i & \mathbf{\Sigma}_{\pi_i}^T \mathbf{W}_i \\ \mathbf{W}_i^T \mathbf{\Sigma}_{\pi_i} & \mathbf{\Sigma}_{\pi_i} \end{bmatrix}$$

Put everything together, we have

$$p(\mathbf{x}_i, \mathbf{x}_{\pi_i}) = p(\mathbf{x}_i | \mathbf{x}_{\pi_i}) p(\mathbf{x}_{\pi_i}) \propto \exp\left\{-\frac{1}{2} \left(\mathbf{x} - \boldsymbol{\mu}\right)^T \Sigma^{-1} \left(\mathbf{x} - \boldsymbol{\mu}\right)\right\} \tag{69}$$

Therefore,

$$(\mathbf{x}_i, \mathbf{x}_{\pi_i}) \sim \mathcal{N}\left(\begin{bmatrix} \mathbf{b}_i + \sum_{j \in x_{\pi_i}} \mathbf{W}_{ij}\boldsymbol{\mu}_j \\ \mathbf{x}_{\pi_i} \end{bmatrix}, \begin{bmatrix} \mathbf{\Sigma}_i + \mathbf{W}_i^T \mathbf{\Sigma}_{\pi_i} \mathbf{W}_i & \mathbf{\Sigma}_{\pi_i}^T \mathbf{W}_i \\ \mathbf{W}_i^T \mathbf{\Sigma}_{\pi_i} & \mathbf{\Sigma}_{\pi_i} \end{bmatrix}\right)$$

Given this property, it follows easily by induction that the joint distribution over the graph G is multivariate Gaussian.