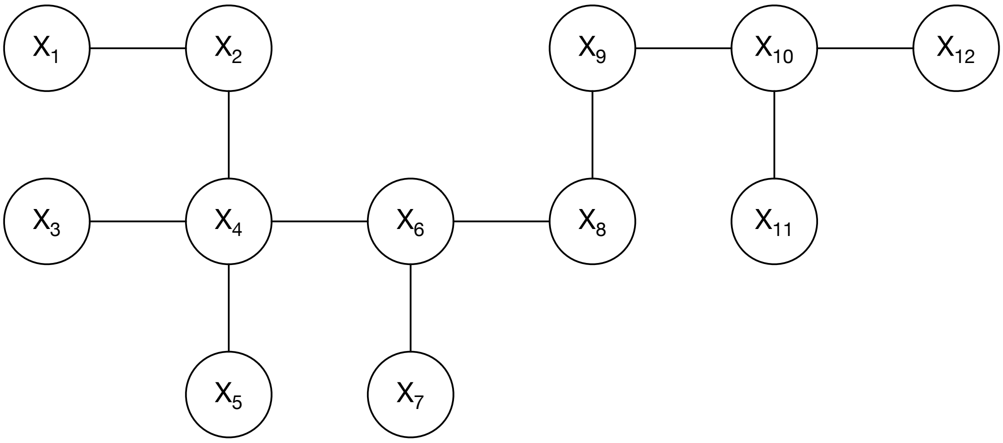
CS5340: Uncertainty Modeling in AI

Tutorial 5

Released: Feb. 23, 2023

Problem 1. (MRT Inference)
We have previously considered inference for this MRF which models the activity (low or high) at 12 MRT stations. This week, we will repeat the activity but using the *sum product algorithm.*



Recall that each node represents a random variable indicating whether the activity at a particular station is low (0) or high (1) and assume the following factorization:

$$p(x_1, x_2, \ldots, x_{12}) = \frac{1}{Z} \prod_{i \in V} \psi(x_i) \prod_{(i,j) \in E} \psi(x_i, x_j) \tag{1}$$

where V is the set of nodes, E is the set of edges, and that the unary and pairwise factors are given by:

x_i	$\psi(x_i)$
0	10
1	2

Figure 1: Unary Factors

x_i	x_j	$\psi(x_i, x_j)$
0	0	20
0	1	5
1	0	5
1	1	20

Figure 2: Pairwise Factors

Note that the factors are the same across the nodes. Your task is to compute the following conditional probabilities using *the sum-product algorithm.*

Problem 1.a. [2 points] Compute $p(x_{12} = 1 | x_1 = 0, x_7 = 0, x_9 = 1, x_{10} = 0)$.

Problem 1.b. [2 points] Compute $p(x_1 = 1 | x_3 = 0, x_4 = 1, x_6 = 0)$.

Problem 1.c. [2 points] Compute $p(x_{10} = 1 | x_9 = 1, x_{12} = 1, x_2 = 0)$.

Problem 1.d. [2 points] Compute $p(x_6 = 0|x_4 = 1, x_8 = 1, x_{10} = 0)$.

Problem 1.e. [2 points] Compute $p(x_8 = 1|x_1 = 0, x_6 = 0, x_9 = 1, x_{12} = 1)$.

Problem 1.f. [2 points] Compute $p(x_2 = 0|x_1 = 0, x_3 = 1, x_4 = 1, x_7 = 1, x_{11} = 0)$.

Problem 2. (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 2.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.*

Problem 2.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.*

Problem 2.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:

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

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

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?

Problem 2.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 $\mathbf{\Sigma}_j$. The conditional for the remaining nodes are also multivariate Gaussian:

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

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

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

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

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 are multivariate Gaussian.