# Homework 4 - Introduction to Probabilistic Graphical Models

kipngeno koech - bkoech April 22, 2025

## 1 Structure Learning

#### 1.1 Tree-Selection and the Chow-Liu Algorithm

Use the Chow-Liu Algorithm to learn the model (tree and parameters) that generated the following data.

| Data |   |   |   |   |  |
|------|---|---|---|---|--|
|      | 1 | 0 | 0 | 1 |  |
|      | 0 | 1 | 0 | 0 |  |
|      | 1 | 1 | 0 | 1 |  |
|      | 1 | 1 | 1 | 1 |  |
|      | 0 | 1 | 0 | 1 |  |
|      | 1 | 0 | 0 | 1 |  |
|      | 1 | 1 | 0 | 1 |  |
|      | 0 | 0 | 1 | 0 |  |
|      | 0 | 1 | 0 | 1 |  |

Figure 1: The tree structure of the data.

### 1.2 Scoring Function

Using the BIC scoring metrics below, compute the model (graph and potentials) that generated the data below. BIC:

$$S(G, \theta; D) = LL(\theta; D) - \phi(|D|) ||G||$$
 where  $\phi(t) = \frac{\log(t)}{2}$ 

Maximizing the score:

$$S_{\max}(G, D) = \max_{\theta} \left( S(G, \theta; D) \right)$$

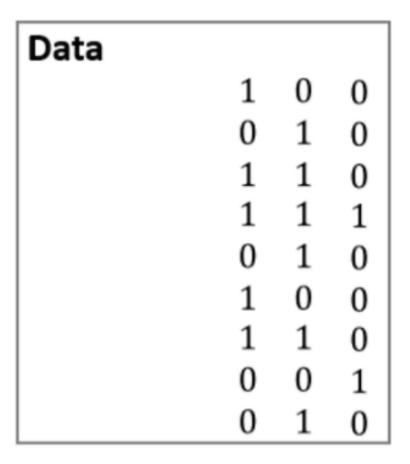


Figure 2: The data used to compute the BIC score.

#### 2 Variational Inference

#### 2.1 Mean-Field Approximation for Multivariate Gaussians

In this question, we'll explore how accurate a Mean-Field approximation can be for an underlying multivariate Gaussian distribution. Assume we have observed data  $X \in \mathbb{R}^{2 \times n}$  where each column  $X_{\cdot,i} \triangleq x^{(i)} \in \mathbb{R}^2$  is a sample that was drawn from a 2-dimensional Gaussian distribution  $x^{(i)} \sim p(\cdot; \mu, \Lambda^{-1})$ .

$$p(x; \mu, \Lambda) = \mathcal{N}\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}; \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix}^{-1}\right)$$
(1)

Note here that we're using the precision matrix  $\Lambda = \Sigma^{-1}$ . An additional property of the precision matrix is that it is symmetric, so  $\Lambda_{12} = \Lambda_{21}$ . (This is a convenient simplifying assumption.) We will approximate this 2-dimensional Gaussian with a mean field approximation,  $q(x) = q(x_1)q(x_2)$ , the product of two 1-dimensional distributions  $q(x_1)$  and  $q(x_2)$ . For now, we won't assume any form for these distributions.

- 1. Short Answer: Write down the equation for  $\log p(X)$ . (For this question, you can leave all of the parameters in terms of vectors and matrices, not their subcomponents.)
- 2. Short Answer: Group together everything that involves  $X_1$  and remove anything involving  $X_2$ . We claim that there exists some distribution  $q^*(X) = q^*(X_1)q^*(X_2)$  that minimizes the KL divergence  $q^* = \arg\min_q \mathrm{KL}(q||p)$ . Furthermore, said distribution will have a component  $q^*(X_1)$  that will be proportional to the quantity you find below. Write that term that is proportional to  $q^*(X_1)$ .

It can be shown that this implies that  $q(X_1)$  (and therefore  $q(X_2)$ ) is a Gaussian distribution:

$$q(x_1) = \mathcal{N}(x_1; m_1, \Lambda_{11}^{-1})$$

where

$$m_1 = \mu_1 - \Lambda_{11}^{-1} \Lambda_{12} (E[x_2] - \mu_2)$$

Using these facts, we'd like to explore how well our approximation can model the underlying distribution.

3. Suppose the parameters of the true distribution are

$$\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$
 and  $\Lambda = \begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{4} \end{bmatrix}$ .

- (a) **Numerical Answer:** What is the value of the mean of the Gaussian for  $q^*(X_1)$ ?
- (b) (2 points) Numerical Answer: What is the value of the variance of the Gaussian for  $q^*(X_1)$ ?
- (c) (2 points) Numerical Answer: What is the value of the mean of the Gaussian for  $q^*(X_2)$ ?
- (d) (2 points) Numerical Answer: What is the value of the variance of the Gaussian for  $q^*(X_2)$ ?
- (e) (5 points) **Plot:** Provide a computer-generated contour plot to show the result of our approximation  $q^*(X)$  and the true underlying Gaussian  $p(X; \mu, \Lambda)$  for the parameters given above.