Homework #1

04654: Intro to Probabilistic Graphical Model

Prof. Assane GUEYE

Early deadline: Thursday, Feb 06, 2025 at 11:59 PM CAT

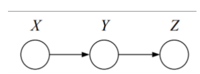
Due: Tuesday, Feb 11, 2025 at 11:59 PM CAT

Please remember to show your work for all problems and to write down the names of any students that you collaborate with. The full collaboration and grading policies are available on the course website: `https://labayifa.github.io/04654`. You are strongly encouraged (but not required) to use Latex to typeset your solutions.

Your solutions should be uploaded to Gradescope (`https://www.gradescope.com/`) in PDF format by the deadline. We will not accept hardcopies. If you choose to hand-write your solutions, please make sure the uploaded copies are legible. Gradescope will ask you to identify which page(s) contain your solutions to which problems, so make sure you leave enough time to finish this before the deadline. We will give you a 30-minute grace period to upload your solutions in case of technical problems.

1 Bayesian Networks *[45 points]*

1. (25 points) * Given the following Markov Chain graph, for which we are considering all variables to be binary:



   (a) Write a code (using you preferred programming language) that generates a distribution (not necessarily a valid BN one) over the 3 variables.

   (b) Write a code that verifies whether a distribution is a valid BN distribution.

   (c) Using your code, generate 10000 distributions and compute the fraction of distributions that are valid BN distributions.

   * Given the following Bayesian Network

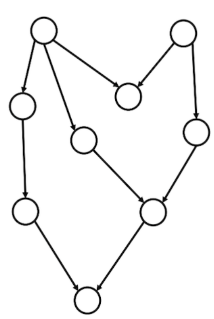


Figure 1: A Bayesian network.

(a) Propose a topological ordering for this graph.

(b) Let $\mathbf{X}$ be a random vector that is Markov with respect to the graph. We assume that the random variables X_i are binary.
Write all local conditional independence.

2. (20 points) State True or False, and briefly justify your answer within 3 lines. The statements are either direct consequences of theorems in Koller and Friedman (2009, Ch. 3), or have a short proof. In the follows, P is a distribution and $\mathcal{G}$ is a BN structure.

(a) If $A \perp B \mid C$ and $A \perp C \mid B$, then $A \perp B$ and $A \perp C$. (Suppose the joint distribution of A, B, C is positive.) (This is a general probability question not related to BNs.)

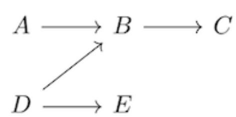


Figure 2: A Bayesian network.

(b) In Figure 2, $E \perp C \mid B$.

(c) In Figure 2, $A \perp E \mid C$

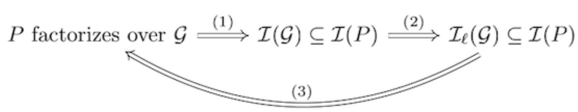


Figure 3: Some relations in Bayesian networks.

Recall the definitions of local and global independences of $\mathcal{G}$ and independences of P.

$$\mathcal{I}_l(\mathcal{G}) = \{(X \perp \text{NonDescendants}_{\mathcal{G}}(X) \mid \text{Parents}_{\mathcal{G}}(X))\} \tag{1}$$

$$\mathcal{I}(\mathcal{G}) = \{(X \perp Y \mid Z) : \text{d-separated}_{\mathcal{G}}(X, Y \mid Z)\} \tag{2}$$

$$\mathcal{I}(P) = \{(X \perp Y \mid Z) : P(X, Y \mid Z) = P(X \mid Z)P(Y \mid Z)\} \tag{3}$$

(d) In Figure 3, relation 1 is true.

(e) In Figure 3, relation 2 is true.

(f) In Figure 3, relation 3 is true.

(g) If $\mathcal{G}$ is an I-map for P, then P may have extra conditional independencies than $\mathcal{G}$.

(h) Two BN structures $\mathcal{G}_1$ and $\mathcal{G}_2$ are I-equivalent if they have the same skeleton and the same set of v-structures.

(i) If $\mathcal{G}_1$ is an I-map of distribution P, and $\mathcal{G}_1$ has fewer edges than $\mathcal{G}_2$, then $\mathcal{G}_2$ is not a minimal I-map of P.

(j) The P-map of a distribution, if it exists, is unique.

2 Markov Networks *[35 points]*

Let $\mathbf{X} = (X_1, \ldots, X_d)$ be a random vector (not necessarily Gaussian) with mean $\boldsymbol{\mu}$ and covariance matrix Σ. The partial correlation matrix R of $\mathbf{X}$ is a $d \times d$ matrix where each entry $R_{ij} = \rho(X_i, X_j | \mathbf{X}_{-ij})$ is the partial correlation between X_i and X_j given the $d-2$ remaining variables $\mathbf{X}_{-ij}$. Let $\Theta = \Sigma^{-1}$ be the inverse covariance matrix of $\mathbf{X}$.

We will prove the relation between R and Θ, and furthermore how Θ characterizes conditional independence in Gaussian graphical models.

1. (10 points) Show that

$$\begin{pmatrix} \Theta_{ii} & \Theta_{ij} \\ \Theta_{ji} & \Theta_{jj} \end{pmatrix} = \begin{pmatrix} \mathrm{Var}[e_i] & \mathrm{Cov}[e_i, e_j] \\ \mathrm{Cov}[e_i, e_j] & \mathrm{Var}[e_j] \end{pmatrix}^{-1} \tag{4}$$

   for any $i, j \in [d]$, $i \neq j$. Here e_i is the residual resulting from the linear regression of $\mathbf{X}_{-ij}$ to X_i, and similarly e_j is the residual resulting from the linear regression of $\mathbf{X}_{-ij}$ to X_j.

2. (10 points) Show that

$$R_{ij} = -\frac{\Theta_{ij}}{\sqrt{\Theta_{ii}}\sqrt{\Theta_{jj}}} \tag{5}$$

3. (15 points) From the above result and the relation between independence and correlation, we know $\Theta_{ij} = 0 \iff R_{ij} = 0 \impliedby X_i \perp X_j \mid \mathbf{X}_{-ij}$. Note the last implication only holds in one direction.

   Now suppose $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$ is jointly Gaussian. Show that $R_{ij} = 0 \implies X_i \perp X_j \mid \mathbf{X}_{-ij}$.

3 Exact Inference (Variable Elimination) *[25 points]*

Reference materials for this problem:

- Jordan textbook Ch. 3, available at

  `https://people.eecs.berkeley.edu/ jordan/prelims/chapter3.pdf`

- Koller and Friedman (2009, Ch. 9 and Ch. 10)

3.1 Variable elimination on a grid [10 points]

Consider the following Markov network:

$$\begin{array}{ccccc} A & \text{——} & B & \text{——} & C \\ | & & | & & | \\ D & \text{——} & E & \text{——} & F \\ | & & | & & | \\ G & \text{——} & H & \text{——} & I \end{array}$$

We are going to see how *tree-width*, a property of the graph, is related to the intrinsic complexity of variable elimination of a distribution.

1. (5 points) Write down largest clique(s) for the elimination order $E, D, H, F, B, A, G, I, C$.
2. (5 points) Write down largest clique(s) for the elimination order $A, G, I, C, D, H, F, B, E$.
3. (5 points) Which of the above ordering is preferable? Explain briefly.
4. (10 points) Using this intuition, give a reasonable ($\ll n^2$) upper bound on the tree-width of the $n \times n$ grid.