Recitation 3

Bayesian Networks

10-701: Introduction to Machine Learning

9/29/2023

1 Review: Bayesian Networks

A Bayesian network (a.k.a. directed acyclic graph (DAG) or directed graphical model) represents **a family of probability distributions over random variables** $X_1, \dots, X_n$ **which satisfy the conditional independence assumptions represented by its network structure**. Each node in the network corresponds to a random variable, and each directed edge represents the dependency between two variables.

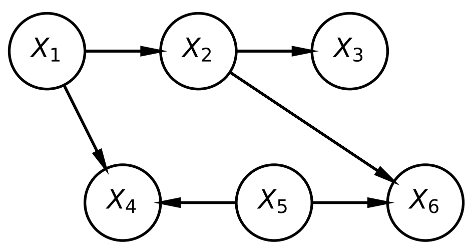


Figure 1: Example Bayesian network with 6 random variables $X_1, \dots, X_6$.

More specifically, we make the assumption that **a random variable is independent of any of its non-descendants given its parents** (local Markov assumption). As a consequence, the set of conditional independence assumptions lead to a compact factorization of the joint distribution. Using the chain rule and the conditional independence assumptions, we can represent the joint distribution as:

$$P(X_1, \dots, X_n) = P(X_n | X_{n-1}, \dots, X_1) \cdots P(X_2 | X_1) P(X_1) \quad (1)$$

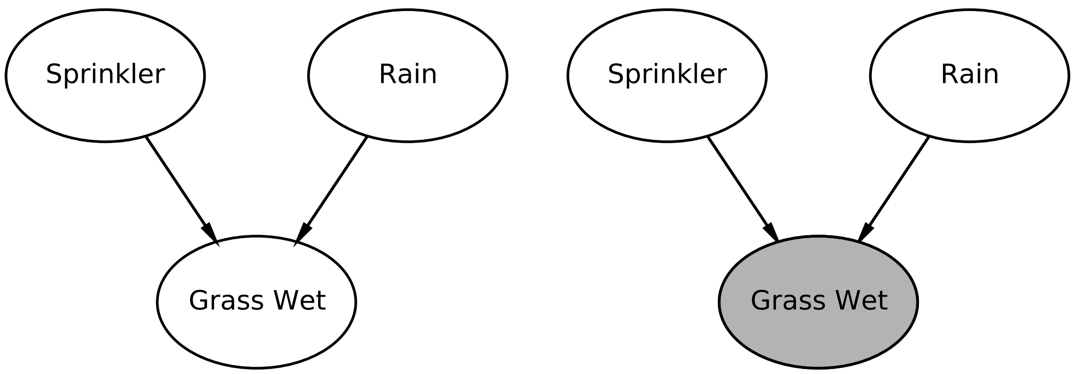
$$= \prod_{i=1}^{n} P(X_i | \text{Parents}(X_i)) \quad (2)$$

Thus, to fully characterize the probability distribution represented by a Bayesian network, all we need to do is learn the parameters of the conditional probabilities $P(X_i | \text{Parents}(X_i))$ $\forall i \in [n]$.

2 Conditional Independence

Given a Bayesian network, we are often interested in testing whether two random variables in the network are dependent under certain conditions/observations. In this section, we will go through a simple v-structure example and d-separation.

2.1 Simple Example



(a) Unobserved (b) Observed

Figure 2: A Bayesian network that encodes the dependencies between three random events/-variables: 1) the sprinkler was on; 2) it rained; 3) the grass is wet. The sprinkler being on and raining have the common effect of making the grass wet.

1. Based on the network in Fig. 2(a), how can we represent the joint probability of the three random events, i.e. $P(\text{Sprinkler}, \text{Grass Wet}, \text{Rain})$?

   $P(\text{Sprinkler}, \text{Grass Wet}, \text{Rain}) = P(\text{Grass Wet}|\text{Sprinkler}, \text{Rain}) \cdot P(\text{Sprinkler}) \cdot P(\text{Rain})$

2. As in Fig. 2(b), suppose we find that the grass is indeed wet. Given this observation, are the events that the sprinkler was on and that it rained independent of each other?

   No. Given that the grass is wet, the two events are *dependent*, i.e. the probability of one event is affected by what the other event is. More formally, we are saying

$$P(\text{Sprinkler}, \text{Rain}|\text{Grass Wet}) \neq P(\text{Sprinkler}|\text{Grass Wet}) \cdot P(\text{Rain}|\text{Grass Wet})$$

or identically,

$$\text{Sprinkler} \not\perp\!\!\!\perp \text{Rain}|\text{Grass Wet}$$

Here is one way to see why this is true: Given that the grass is wet, let us assume that it didn't rain. In this case, since we know that the grass is wet, the probability that the sprinkler was on becomes higher.

2.2 d-separation

d-separation is a criterion that we can use to easily test whether two random variables in a Bayesian network are conditionally independent. Using d-separation, all we need to do is check whether there exists an "unblocked" path between the two nodes that we are testing. If there exists such a path, then they are d-connected, or conditionally *dependent*. If all paths are "blocked," they are d-separated, or conditionally *independent*.

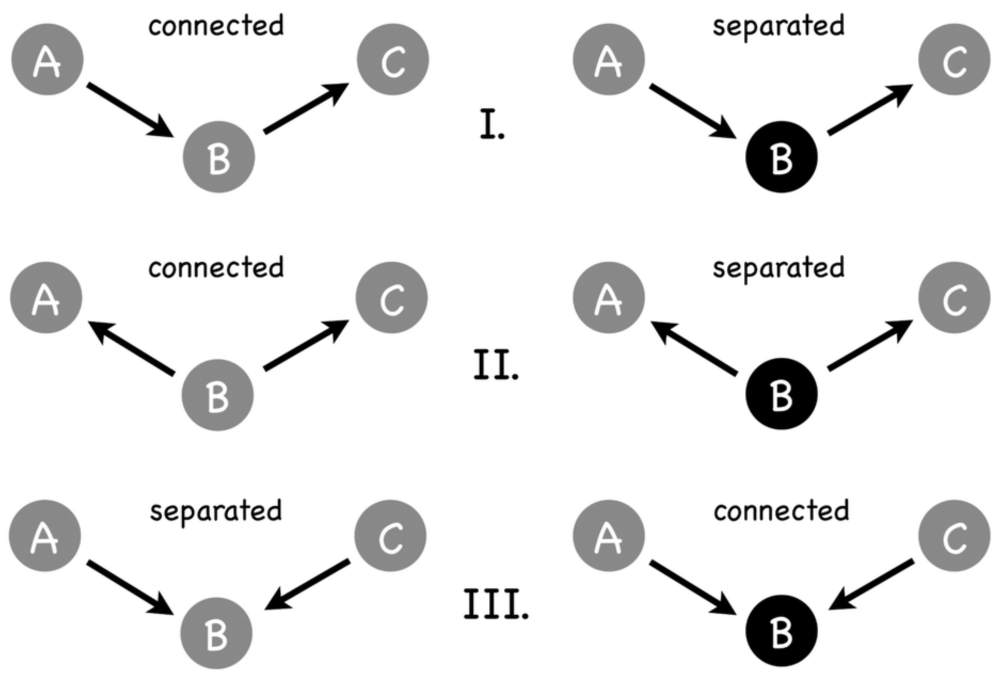
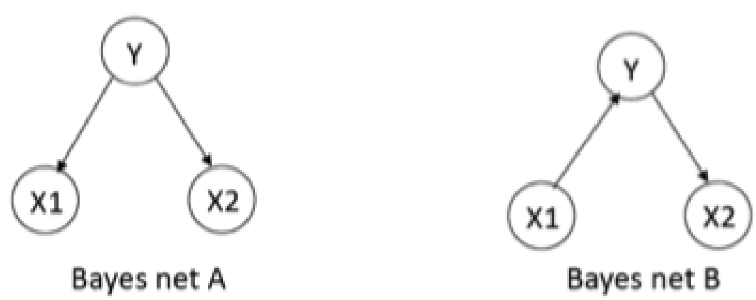


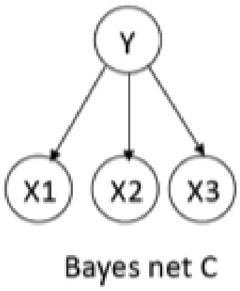
Figure 3: (I) Cascade, (II) Common Cause, (III) Common Effect (v-structure)

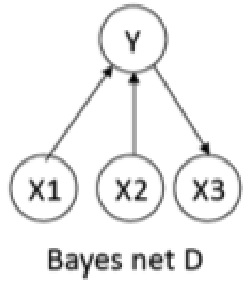
1. Do Bayesian networks A and B below encode the *same* set of conditional independence assumptions? If not, write down an independence assumption that holds for one but not for the other.



Yes, i.e. $X_1 \perp\!\!\!\perp X_2 | Y$

2. Do Bayesian networks C and D below encode the *same* set of conditional independence assumptions? If not, write down an independence assumption that holds for one but not for the other.





No. $X_1 \perp\!\!\!\perp X_2 | Y$ holds for C but not for D.

3. In the Bayesian network below, what are relationships (independent or not) between the following random variables, assuming X_3 is observed?

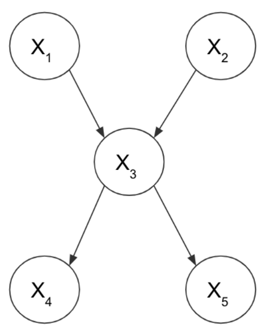


Figure 4

- X_1______$X_4 \mid X_3$
- X_1______$X_2 \mid X_3$
- X_4______$X_5 \mid X_3$

  $\perp\!\!\!\perp$, $\not\perp\!\!\!\perp$, $\perp\!\!\!\perp$

3 Additional Practice Problems

3.1 Problem 1

Consider the Bayesian network below to solve the following questions. We use $(X \perp\!\!\!\perp Y|Z)$ to denote the fact that X and Y are independent given Z.

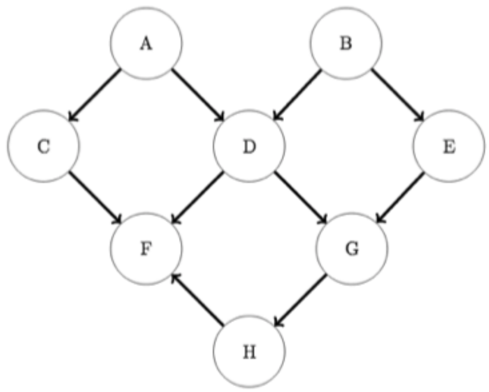


Figure 5

1. Write down the factorization of the joint distribution for the given Bayesian network.

$$P(A,B,C,D,E,F,G,H) = P(A)P(B)P(C|A)P(D|A,B)P(E|B)P(F|C,D,H) \cdot P(G|D,E)P(H|G)$$

2. Are there any pairs of points that are independent? If your answer is yes, please list out all such pairs.

   Answer: Yes. (C, E), (C, B), (A, E), (A, B).

3. Does $(B \perp\!\!\!\perp C|A, D)$ hold? Briefly explain.

Answer: It holds. We can see this by checking all possible paths between B and C and using d-separation.

4. Does $(B \perp\!\!\!\perp F | A, D)$ hold? Briefly explain.

Answer: It does not hold. The path B, E, G, H, F is still unblocked.

5. Assuming that there are $d = 10$ values that each of these variables can take (say 1 to 10), how many parameters do we need to model the full joint distribution without using the knowledge encoded in the graph (i.e. no independence/conditional independence assumptions)? How many parameters do we need for the Bayesian network for such setting? (you do not need to provide the exact number, a close approximation or a tight upper/lower bound will do).

Answer: Naive Approach: 10^8. Bayesian Network: about $10 \times 2 + 10^2 \times 3 + 10^3 \times 2 + 10^4 \times 1 = 12320$.

3.2 Problem 2

Consider the Bayesian network below to solve the following questions.

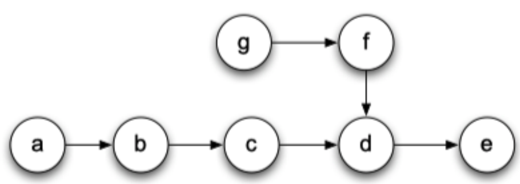


Figure 6

1. Write down the factorization of the joint distribution for the given Bayesian network.

$$p(a, b, c, d, e, f, g) = p(a)p(b|a)p(c|b)p(d|c, f)p(e|d)p(f|g)p(g)$$

2. Let $X = \{c\}, Y = \{b, d\}, Z = \{a, e, f, g\}$. Is $X \perp\!\!\!\perp Z|Y$? If yes, explain why. If not, show a path from X to Z that is not blocked.

   No. $c \rightarrow d \rightarrow f$ is not blocked

3. Suppose you are allowed to choose a set W such that $W \subset Z$. Define $Z^* = Z \setminus W$ and $Y^* = Y \cup W$. What is the smallest set W such that $X \perp\!\!\!\perp Z^*|Y^*$ is true?

   $W = \{f\}$ is the smallest subset. Y^* is then a Markov Blanket of node c.

4. From the graph, we can see that $a \perp\!\!\!\perp c, d|b$. Prove that this implies $a \perp\!\!\!\perp c|b$.

   Note that

$$a \perp\!\!\!\perp c, d|b \Rightarrow p(a, c, d|b) = p(a|b)p(c, d|b)$$

$$\begin{aligned} p(a, c|b) &= \sum_d p(a, c, d|b) \\ &= \sum_d p(a|b)p(c, d|b) \\ &= p(a|b) \sum_d p(c, d|b) \\ &= p(a|b)p(c|b) \end{aligned}$$