

Lecture 20-21

Instructor: *Profs Peter J. Hass and Jie Xiong*SI Worksheet: *Juelin Liu*

1 Introduction

These lectures covered the Bayesian Networks.

2 Intuition

Suppose you want to represent the joint distribution of some set of random variables X_1, X_2, \dots, X_n . In the simplest case where all the random variables have only two possible outcomes, we need to store $2^n - 1$ numbers for the joint distribution. This is too large to store in memory even for small n .

The key insight is we can use the independent properties of the random variables to represent and store the joint distribution compactly. Bayesian Networks provide a natural representation of the conditional independence properties in a finer-grained way.

3 Independence Recap

Let's recap independent random variables first. Suppose you are tossing two different dices with possible outcomes $\{1, 2, 3, 4, 5, 6\}$. Let's also assume neither of the dice is fair. The face values can be represented with two discrete random variables X_1 and X_2 . If we store the joint distribution in memory, we can store the probability of each pair of values, which would consume $6^2 = 36$ memory space. Alternatively, you can store the first 35 values and compute the last one on the fly (using 1 minus the sum of the previous 35 probabilities). This approach takes 35 memory spaces.

However, X_1 and X_2 are two independent random variables. That is the face value of the first dice will not impact the face value of the second dice. As a result we can store the distribution for X_1 and X_2 respectively and compute $P(X_1 \cap X_2)$ on the fly using $P(X_1 \cap X_2) = P(X_1)P(X_2)$. Now we only need to store $(6 - 1) + (6 - 1) = 10$ entries in the memory.

The observation is that we can use the independence properties to reduce the memory required to store the distribution. In this case, the two random variables are marginally independent.

4 Conditional Independence Recap

Now, suppose you are tossing three dice, and their face values are characterized by three discrete random variables X_1, X_2, X_3 .

If we know that $P(X_1, X_2 | X_3) = P(X_1 | X_3)P(X_2 | X_3)$, how many entries do we need to store the distribution?

In this case,

$$P(X_1 \cap X_2 \cap X_3) = P(X_1, X_2|X_3)P(X_3) = P(X_1|X_3)P(X_2|X_3)P(X_3)$$

The joint distribution can be written as the product of three random variables $X_1|X_3$, $X_2|X_3$, and X_3 . Like the previous example, we can store the distribution for each random variable and compute the joint distribution on the fly.

For random variable X_3 , we need $6 - 1 = 5$ entries in memory. For the random variables $X_1|X_3$ and $X_2|X_3$, we need $(6 - 1) \times 6 = 30$ entries for each of them. Thus, the total entries required are $30 + 30 + 5 = 65$

Unlike the previous example, none of the variables in this joint probability are (marginally) independent. If the three random variables are (marginally) independent, we can store the distribution using $(6 - 1) + (6 - 1) + (6 - 1) = 15$ memory space.

5 Bayesian Networks

A Bayesian Network represents the conditional independent properties of the distribution in a directed acyclic graph (DAG). $G(V, E)$, whose nodes V are the random variables and the directed edges E correspond to direct influence of one node on another. Intuitively, for a given node, if we know the values of its direct parents, then the distribution of the node is independent of other nodes in the graph.

For example, to draw the Bayesian Network for the above example, we will have three nodes (X_1 , X_2 , and X_3) and two edges (X_3 to X_2 , X_3 to X_1) in the graph given that $P(X_1, X_2|X_3) = P(X_1|X_3)P(X_2|X_3)$.

6 Questions

Suppose that most mornings you set your alarm for 8 am so that you can catch the 8:30 am bus to school which usually arrives just in time for your class at 9 am. Unfortunately, sometimes you forget to set your alarm which might make you miss your bus and this increases the probability that you are late for class. Even if you catch the 8:30 am bus, there's a chance you'll be late for class. To help analyze the situation we introduce some random variables.

Let $A = 1$ if you remember to set your alarm and $A = 0$ otherwise.

Let $B = 1$ if you catch the 8:30 am bus and $B = 0$ otherwise.

Let $L = 1$ if you are late for class and $L = 0$ otherwise.

Suppose that $P(A = 1) = \frac{1}{2}$, $P(B = 1|A = 1) = \frac{4}{5}$, $P(B = 1|A = 0) = \frac{2}{5}$, $P(L = 1|B = 1) = \frac{1}{4}$, and $P(L = 1|B = 0) = \frac{2}{3}$.

1) What are the values of: $P(A = 0)$, $P(B = 0|A = 1)$, $P(B = 0|A = 0)$, $P(L = 0|B = 1)$, and $P(L = 0|B = 0)$?

2) Draw the Bayesian network for the random variables A , B , and L .

3) What is the value of $P(A = 1, B = 1, L = 0)$?

4) What is the value of $P(B = 1)$?

5) What is the value of $P(L = 0)$?

6) What is the value of $P(L = 1, A = 0)$?

7) What's the value of $P(L = 1|A = 0)$?

7 Answer

1. $P(A = 0) = 1 - P(A = 1) = 1/2$
 $P(B = 0|A = 1) = 1 - P(B = 1|A = 1) = 1 - \frac{4}{5} = \frac{1}{5}$
 $P(B = 0|A = 0) = 1 - 2/5 = 3/5$
 $P(L = 0|B = 1) = 1 - 1/4 = 3/4$
 $P(L = 0|B = 0) = 1 - 2/3 = 1/3$
2. A -i B -i L
3. $P(A = 1, B = 1, L = 0) = P(A = 1)P(B = 1|A = 1)P(L = 0|B = 1) = \frac{1}{2} \times \frac{4}{5} \times \frac{3}{4} = \frac{3}{10}$
4. $P(B = 1) = P(B = 1|A = 0)P(A = 0) + P(B = 1|A = 1)P(A = 1) = \frac{3}{5}$
5. $P(L = 0) = P(L = 0|B = 0)P(B = 0) + P(L = 0|B = 1)P(B = 1) = \frac{7}{12}$
6. $P(L = 1, A = 0) = P(L = 1, A = 0, B = 0) + P(L = 1, A = 0, B = 1) = P(A = 0)P(B = 0|A = 0)P(L = 1|B = 0) + P(A = 0)P(B = 1|A = 0)P(L = 1|B = 1) = \frac{1}{4}$
7. $P(L = 1|A = 0) = \frac{P(L=1, A=0)}{P(A=0)} = \frac{1}{2}$