CS 337, Fall 2023
Bayesian Networks (I)

Scribes*: Nandigama Nikhil, Premankur Chakraborty, Kanad Shende, Chaitanya Aggarwal, Aaryan Dangi
Hriswitha Ijjada, Khushang Singla, S.Harsha Varma , Yash Ruhatiya
Edited by: Vedang Asgaonkar

November 13, 2023

Disclaimer. Please note this document has not received the usual scrutiny that formal publications enjoy. This may be distributed outside this class only with the permission of the instructor.

1 Recall: The Concept of Independence

Two events are said to be independent when the occurrence of one event does not affect the probability of occurrence of the other. Similarly, two random variables are said to be independent when the realization of one does not change the probability distribution of the other.

A and B are said to be independent (denoted as A ⊥ B) if:

$$P(A \cap B) = P(A)P(B) \tag{1}$$

A and B are said to be **conditionally independent** given C (denoted as A ∩ B | C) if:

$$P(A, B|C) = P(A|C)P(B|C) \tag{2}$$

Lastly, a finite set of events is said to be **mutually independent** if every event is independent of any intersection of other events from the set, i.e. if for every $k \leq n$ and for every k indices $1 \leq i_1 < ... < i_k \leq n$, we have:

$$P(\bigcap_{j=1}^{k} A_{i_j}) = \prod_{j=1}^{k} P(A_{i_j}) \tag{3}$$

As we will see below, the above proves to be very useful because it allows us to use a small number of probabilities to estimate a full joint distribution across all random variables in the system.

2 Bayesian Networks

A Bayesian network, also called a **Belief Network** is a probabilistic model that represents random variables and the conditional dependencies between them via a **Directed Acyclic Graph (DAG)**. Bayesian networks are a form of causal notation that allows us to take an event that occurs and predict the likelihood that any of the possible causes is the contributing factor. For example, a Bayesian network can be used to find the probabilities of the presence of any disease given the symptoms.

Formally, a Bayesian Network (G) can be described with the following properties:

*Both teams contributed equally to the scribe.

- Every node of G (remember that G is a DAG) represents one random variable (either discrete or continuous)

- There are directed edges connecting different pairs of nodes. If there exists an edge from node X to node Y, then X is the parent of Y.

- Each node in G has an accompanying conditional probability distribution associated with its random variable.

For any node (random variable) X_i in G, the Conditional Probability Distribution is given by P(X_i|Parents(X_i)). The Bayesian Network model assumes that each random variable is independent of its non-descendants given its parents. Hence the topology of G, coupled with the conditional probability distributions of all nodes, suffices to describe the full joint distribution over all random variables:

$$P(X_1, X_2, ..., X_n) = \prod_{i=1}^{n} P(X_i|Parents(X_i)) \tag{4}$$

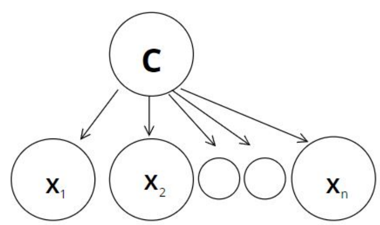
3 Naive Bayes Model

3.1 Assumptions

The naive bayes model is a simple model that assumes:

1. The features are all conditionally independent, given the cause of the instance

2. All the features contribute equally to the outcome



(5)

$$P(X_1, X_2...X_n|C) = \prod_{i=1}^{n} P(X_i|C)$$
$$P(C, X_1, X_2...X_n) = P(C)P(X_1, X_2...X_n|C)$$
$$P(C, X_1, X_2...X_n) = P(C)\prod_{i=1}^{n} P(X_i|C)$$

The number of parameters to train are linear in N.

3.2 An Example

Consider the following Bayesian Network, which shows the relation between a student's intelligence(I), course's difficulty(D), grade in that course(G), Student's SAT score(S) and student's chance of getting a LOR from their professor(L).

Here the joint probability distribution would be

$$P(D, I, G, S, L) = P(I)P(D)P(G|I, D)P(S|I)P(L|G)$$

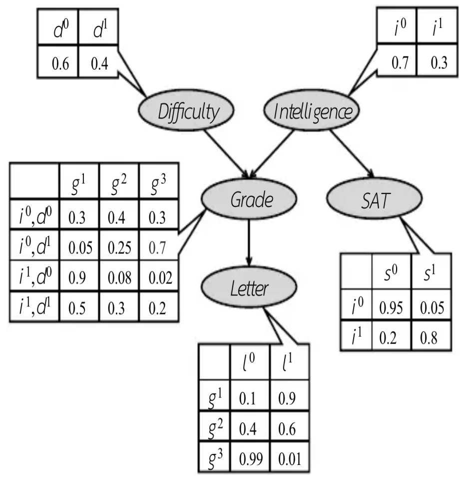


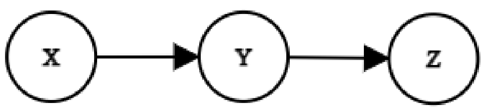
Figure 3.4 Student **Bayesian network** $\mathcal{B}^{student}$ **with CPDs**

4 Independence in a Bayesian Network

Question: Given two nodes, X and Y, are they independent, given an evidence node Z?
To solve this, consider a simple three-node network X, Y and a node Z

Case1: X and Y are directly connected

X and Y are correlated regardless of any evidence about any other variables



They are not independent because X is a parent of Y. So,

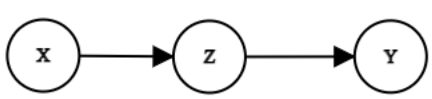
$$P(Y) = P(X)P(Y|X), \text{ where } P(Y|X) \text{ is the conditional probability distribution of } Y \tag{6}$$

Case2: X and Y are indirectly connected via Z

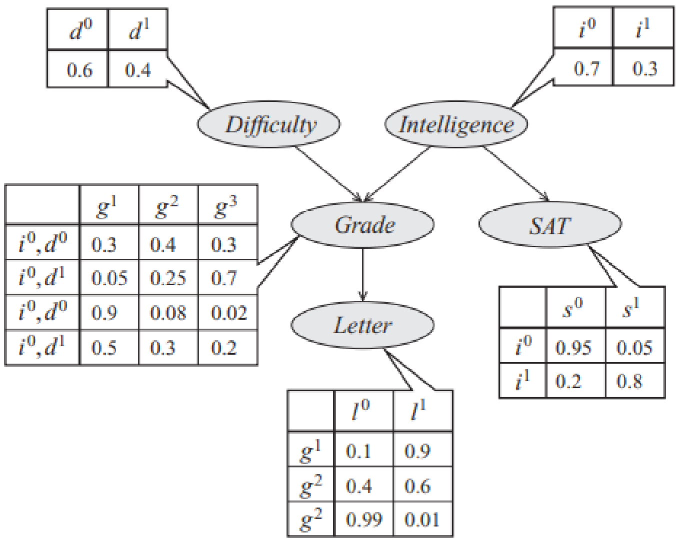
There are 4 cases

1. Indirect casual effect or Casual trail
2. Indirect evidential effect or Evidential trail
3. Common cause
4. Common effect

4.1 Casual trail



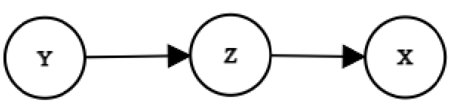
- Cause X cannot influence effect Y if Z is observed
- If Z is observed, it blocks the influence of X, as X can affect Y only via Z
- In the below example, intelligence does not matter for the letter if the grade is already known.



$$P(y|x,z) = \frac{P(x,y,z)}{P(x,z)} = \frac{P(x)P(z|x)P(y|z)}{P(x)P(z|x)} = P(y|z) \tag{7}$$

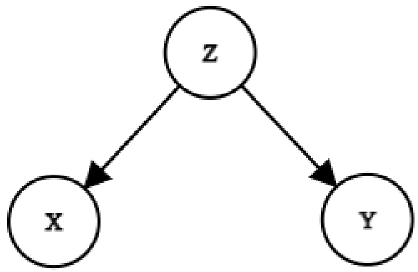
Therefore, X cannot influence Y if Z is observed

4.2 Evidential trail



- Effect X can influence Y via Z only if Z is unobserved
- If Z is observed, X cannot influence Y
- In the above example, if the grade is unobserved, the letter influences the assessment of Intelligence
- Dependency is the symmetric notion. If X can influence Y, then Y can influence X and vice versa
- So the condition for the dependency is the same for evidential trial and casual trial because dependency is symmetric

4.3 Common cause

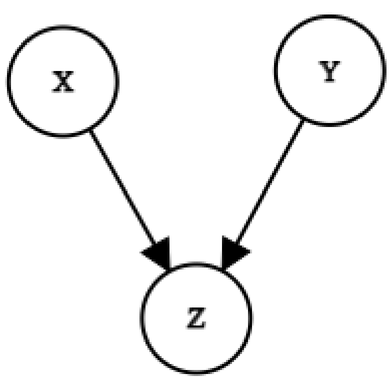


- X can influence Y if and only if Z is not observed
- If Z is observed, X cannot influence as it blocks the influence
- In the above example, the grade is correlated with the sat score, but if intelligence is observed, then the sat provides no additional information

$$P(y|x,z) = \frac{P(x,y,z)}{P(x,z)} = \frac{P(z)P(x|z)P(y|z)}{P(z)P(x|z)} = P(y|z) \tag{8}$$

- Therefore, X and Y are independent given Z
- Conclusion: Observing the cause blocks the influence between the effects

4.4 Common effect



- X can influence Y if and only if Z or a descendant of Z is observed
- In the above example, if the grade is unobserved, the difficulty and intelligence are independent
- If Grade is observed, Intelligence and difficulty are correlated; suppose the difficulty is more, and if the grade is observed to be more, then we can say they have more intelligence.

Active trail: The path between X and Y is said to be active, iff X can influence/dependent on Y

4.5 Summary

1. Casual trail: active if and only if Z is unobserved
2. Evidential trail: active if and only if Z is unobserved
3. Common Cause: active if and only if Z is unobserved
4. Common effect: active iff Z or descendant of Z is observed

5 D-Separation

Given sets X, Y and Z of random variables, X and Y are said to be d-separated given Z if and only if there is no active path from $x \in X$ to $y \in Y$ via Z. D-Separation for X, Y given Z implies that X and Y are conditionally independent given Z. So, two random variables are independent if all possible paths connecting them in the Bayesian Network are inactive.
Note that the paths here are all possible routes from X to Y in the Bayesian Network disregarding the directionality of the edges. This is because d-separation is concerned with the flow of information between variables, not the direction of causality.

5.1 An example

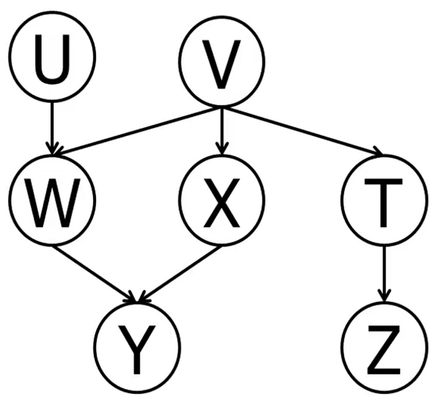


Figure 1: A Bayesian Network

Given the above bayesian network let us answer some questions:

Is U⊥V?
The answer is yes.
Solution: Let us consider all possible (undirected) paths from U to V:

1. U-W-V : Forms a v-structure. Inactive as the middle node V is unobserved.
2. U-W-Y-X-V : Consider a part of the path W-Y-X, this forms a v-structure as well and is inactive as Y is unobserved. Now, that we have a trail in the path which is inactive, the complete path is inactive.

Since, all possible paths are inactive, we can guarantee the independence.

Is W⊥X?
Can't guarantee.

Solution: Consider the path W-V-X, this forms a common cause and as discussed earlier this trail is active iff the middle node (V here) is unobserved which is the case here. Since we have an active path, we cannot guarantee the independence.

Is X$\perp$T | V?
Yes, X and T are conditionally independent given V.
Solution: Note that we are given V, i.e V is observed. Let us consider all possible (undirected) paths from X to T via V:

1. X-V-T : Common cause. Inactive as the middle node V is observed.
2. X-Y-W-V-T : Consider the trail X-Y-W. This forms a v-structure and is inactive as Y is unobserved. Now, that we have a trail in the path which is inactive, the complete path is inactive.

Since, all possible paths are inactive, we can guarantee the independence given V.