Independence and Bayesian Networks

Alice Gao

Lecture 11

Readings: RN 13.4 & 14.2. PM 8.2 & 8.3.

Outline

Learning Goals

By the end of the lecture, you should be able to

- Given a probabilistic model,
  determine if two variables are unconditionally independent, or conditionally independent given a third variable.
- Give examples of deriving a compact representation of a joint distribution by using independence assumptions.
- Describe components of a Bayesian network.
- Compute a joint probability given a Bayesian network.
- Explain the independence relationships in the three key structures.

(Unconditional) Independence

Definition ((unconditional) independence)

X and Y are (unconditionally) independent iff

$$P(X|Y) = P(X)$$

$$P(Y|X) = P(Y)$$

$$P(X \land Y) = P(X)P(Y)$$

Learning Y does NOT influence your belief about X.

Conditional Independence

Definition (conditional independence)

X and Y are conditionally independent given Z if

$$P(X|Y \wedge Z) = P(X|Z).$$

$$P(Y|X \wedge Z) = P(Y|Z).$$

$$P(Y \wedge X|Z) = P(Y|Z)P(X|Z).$$

Learning Y does NOT influence your belief about X if you already know Z.

CQ: Deriving a compact representation

CQ: Consider a model with three random variables, A, B, C.

1. What is the minimum number of probabilities required to specify the joint distribution?

2. Assume that A, B, and C are independent.
   What is the minimum number of probabilities required to specify the joint distribution?

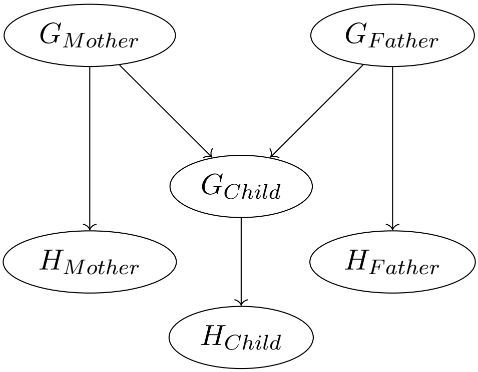
CQ: Deriving a compact representation

CQ: Consider a model with three random variables, A, B, C.

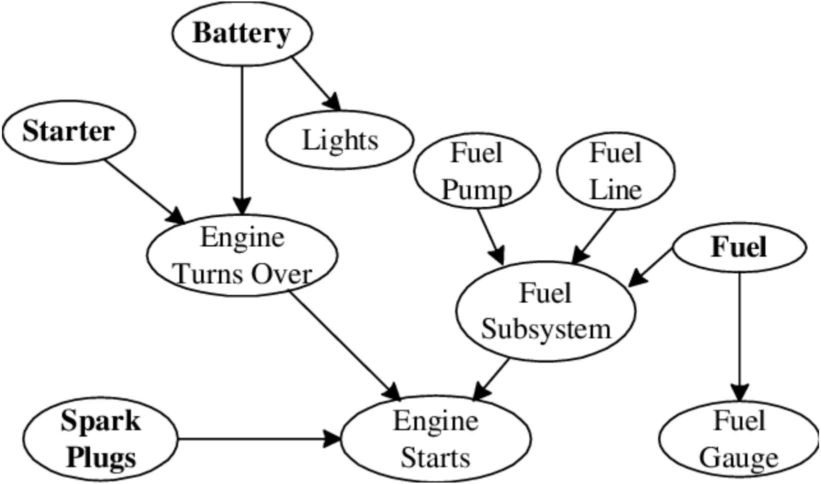
1. What is the minimum number of probabilities required to specify the joint distribution?

2. Assume that A and B are conditionally independent given C. What is the minimum number of probabilities required to specify the joint distribution?

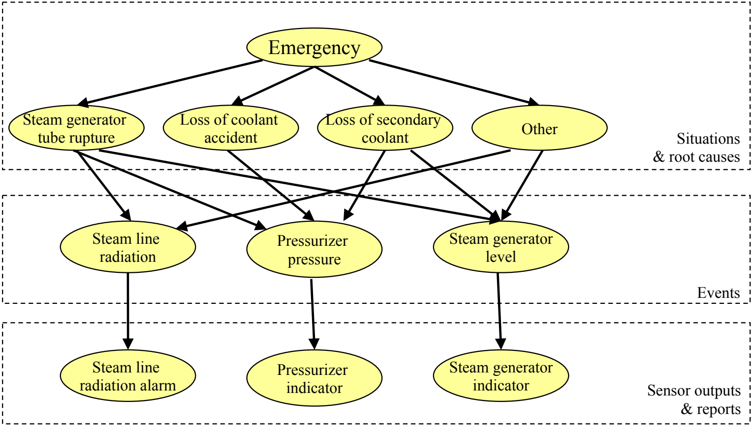
Inheritance of Handedness



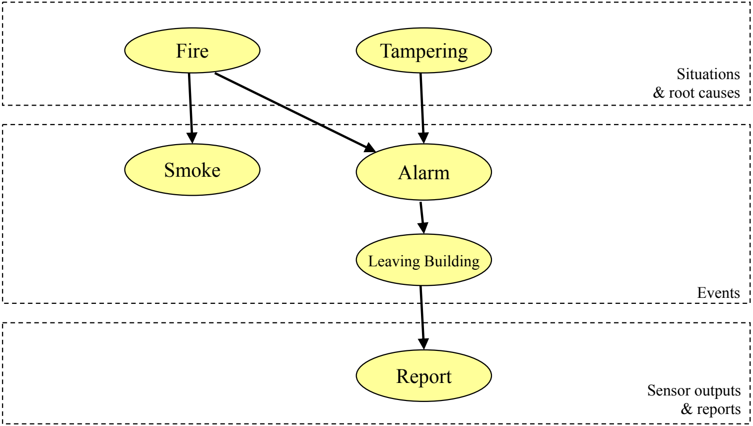
Car Diagnostic Network



Example: Nuclear power plant operations

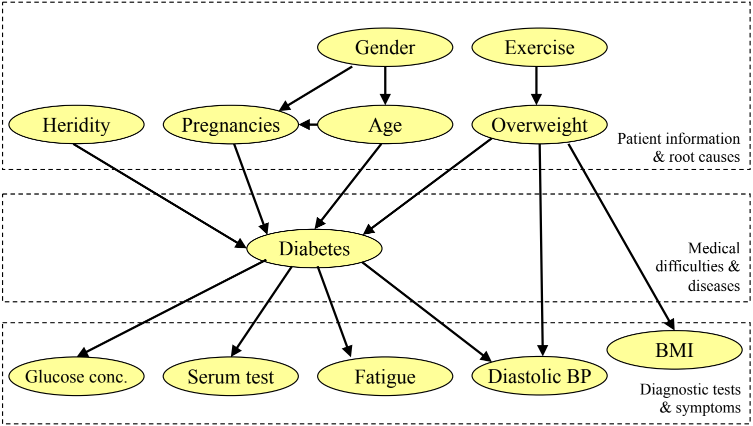


Example: Fire alarms



Report: “report of people leaving building because a fire alarm went off”

Example: Medical diagnosis of diabetes



Why Bayesian Networks?

A probabilistic model of the Holmes scenario:

- The random variables:
  Earthquake, Radio, Burglary, Alarm, Watson, and Gibbon.
- # of probabilities in the joint distribution: $2^6 = 64$.
- For example,
  $P(E \wedge R \wedge B \wedge A \wedge W \wedge G) = ?$
  $P(E \wedge R \wedge B \wedge A \wedge W \wedge \neg G) = ?$
  ... etc ...

We can compute any probability using the joint distribution, but

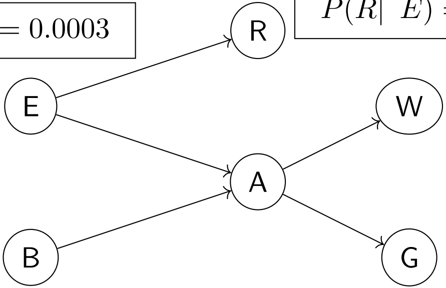
- Quickly become intractable as the number of variables grows.
- Unnatural and tedious to specify all the probabilities.

Why Bayesian Networks?

A Bayesian Network

- is a compact version of the joint distribution, and
- takes advantage of the unconditional and conditional independence among the variables.

A Bayesian Network for the Holmes Scenario



$$P(E) = 0.0003$$

$$P(B) = 0.0001$$

$$P(R|\neg E) = 0.0002$$
$$P(R|\ E) = 0.9$$

$$P(W|\neg A) = 0.4$$
$$P(W|\ A) = 0.8$$

$$P(G|\neg A) = 0.04$$
$$P(G|\ A) = 0.4$$

$$P(A|\neg B \wedge \neg E) = 0.01$$
$$P(A|\neg B \wedge\ E) = 0.2$$
$$P(A|\ B \wedge \neg E) = 0.95$$
$$P(A|\ B \wedge\ E) = 0.96$$

Bayesian Network

A Bayesian Network is a directed acyclic graph.

- Each node corresponds to a random variable.
- X is a parent of Y if there is an arrow from node X to node Y.
- Each node X_i has a conditional probability distribution $P(X_i|Parents(X_i))$ that quantifies the effect of the parents on the node.

The Semantics of Bayesian Networks

Two ways to understand Bayesian Networks:

- A representation of the joint probability distribution
- An encoding of the conditional independence assumptions

Representing the joint distribution

We can compute each joint probability using the following formula.

$$P(X_n \wedge \cdots \wedge X_1) = \prod_{i=1}^{n} P(X_i | Parents(X_i))$$

Representing the joint distribution

Example: What is the probability that

- The alarm has sounded,
- Neither a burglary nor an earthquake has occurred,
- Both Watson and Gibbon call and say they hear the alarm, and
- There is no radio report of an earthquake?

CQ: Calculating the joint probability

CQ: What is the probability that

- NEITHER a burglary NOR an earthquake has occurred,
- The alarm has NOT sounded,
- NEITHER of Watson and Gibbon is calling, and
- There is NO radio report of an earthquake?

(A) 0.5699

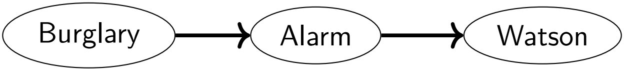
(B) 0.6699

(C) 0.7699

(D) 0.8699

(E) 0.9699

Burglary, Alarm and Watson



CQ Unconditional Independence

CQ: Are Burglary and Watson independent?

(A) Yes

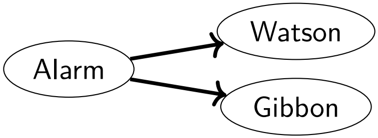
(B) No

CQ: Conditional Independence

CQ: Are Burglary and Watson conditionally independent given Alarm?

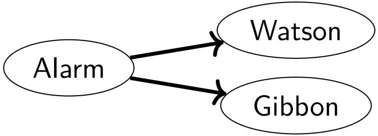
(A) Yes
(B) No

Alarm, Watson and Gibbon



CQ Unconditional Independence

CQ: Are Watson and Gibbon independent?

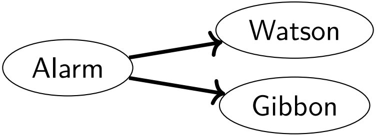


(A) Yes

(B) No

CQ Conditional Independence

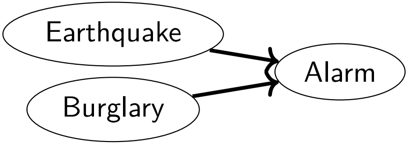
CQ: Are Watson and Gibbon conditionally independent given Alarm?



(A) Yes

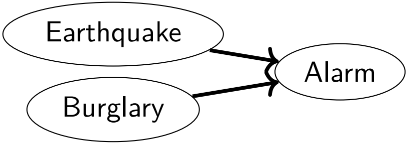
(B) No

Earthquake, Burglary, and Alarm



CQ Unconditional Independence

CQ: Are Earthquake and Burglary independent?

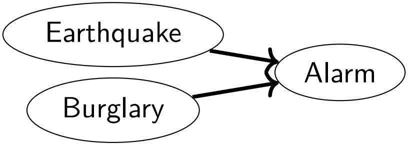


(A) Yes

(B) No

CQ: Conditional Independence

CQ: Are Earthquake and Burglary conditionally independent given Alarm?



(A) Yes
(B) No

Revisiting the Learning Goals

By the end of the lecture, you should be able to

- Given a probabilistic model,
  determine if two variables are unconditionally independent, or conditionally independent given a third variable.
- Give examples of deriving a compact representation of a joint distribution by using independence assumptions.
- Describe components of a Bayesian network.
- Compute a joint probability given a Bayesian network.
- Explain the independence relationships in the three key structures.