Lecture 14: Graphical Models

Machine Learning
Queens College
Today

• Graphical Models
  – Representing joint probability distributions compactly and graphically by utilizing conditional dependence
Joint Probability Distribution

Representation: What is the joint probability distribution $P(B, E, R, A, N)$ over five binary variables?

• How many states in total?

2^5 = 32

• How many independent probabilities do we need?

31

• How many table entries should we sum over to get $P(R=r)$?

$\frac{16}{P(B, E, R, A, N)}$
Joint Probability Distribution

• Full joint distribution is sufficient to represent the complete domain and to do any type of probabilistic inferences

• Problems: $n$ - number of random variables, $d$ – number of values
  – Space complexity. A full joint distribution requires to remember $O(d^n)$ numbers
  – Inference complexity. Computing queries requires $O(d^n)$ steps
  – Acquisition problem. Who is going to define all the probabilities?
Probability Models

• What if the variables are independent?
• If x and y are independent:
  \[ p(x, y) = p(x)p(y) \]
• The original distribution can be factored if B, E, A, R, N are independent

\[ P(B, E, A, R, N) = \frac{P(B)P(E)P(A)P(R)P(N)}{1 + 1 + 1 + 1 + 1} = 5 \]
Conditional Independence

• Independence assumptions are convenient (Naïve Bayes), but rarely true.
• More often some groups of variables are dependent, but others are independent.
• Still others are **conditionally independent**.
Conditional Independence

• If two variables are conditionally independent.

\[ p(x, z | y) = p(x | y)p(z | y) \]

\[ p(x, z) \neq p(x)p(z) \]

• E.g. \( y = \text{flu?} \), \( x = \text{achiness?} \), \( z = \text{headache?} \)

\[ x \perp z | y \]
Factorization of a joint

• Assume $x \perp z \mid y$

• How do you factorize: $p(x, y, z)$

\[
p(x, y, z) = p(x, z \mid y) p(y) = p(x \mid y) p(z \mid y) p(y)
\]
Factorization of a joint

• What if there is no conditional independence?

• How do you factorize:  
  \[ p(x, y, z) \]

\[ p(x, y, z) = p(y)p(x|y)p(z|y) \]
Bayesian networks

• A Bayesian Network is a graph in which:
  – A set of random variables makes up the nodes in the network.
  – A set of directed links or arrows connects pairs of nodes.
  – Each node has a conditional probability table that quantifies the effects the parents have on the node.
  – Directed, acyclic graph (DAG), i.e. no directed cycles.

A Bayesian network modeling the joint probability distribution $P(B, E, R, A, N)$
Bayesian networks

Conditional probability tables (CPT):

- $P(B)$
- $P(E)$
- $P(N|A)$
- $P(R|E)$
- $P(A|B,E)$
Independences in BNs

• Three basic independence structures:
Indepedences in BNs

• **Indirect cause:** Burglary is independent of NeighborCall given Alarm
  
  \[ P(\text{N}|A, B) = P(\text{N}|A) \]
  
  \[ P(\text{N, B}|A) = P(\text{N}|A)P(\text{B}|A) \]
Independences in BNs

- **Common cause**: Alarm is independent of RadioAnnounce given Earthquake
  - $P(A|E, R) = P(A|E)$
  - $P(A, R|E) = P(A|E)P(R|E)$
Independences in BNs

- **Common effect**: Burglary and Earthquake are marginally independent, but become dependant given Alarm
  - $P(B, E) = P(B)P(E)$
  - $P(B, E|A) \neq P(B|A)P(E|A)$!!!
Bayes Ball Algorithm

\[ x_a \perp x_b \mid x_c \]

- Shade nodes \( x_c \)
- Place a “ball” at each node in \( x_a \)
- Bounce balls around the graph according to rules
- If no balls reach \( x_b \), then conditionally independent.
Ten rules of Bayes Ball Theorem
Independences in BNs

- Earthquake ⊥ Burglary | NeighborCall?
- Burglary ⊥ NeighborCall?
- Burglary ⊥ RadioAnnounce | Earthquake?
- Burglary ⊥ RadioAnnounce | NeighborCall?

\[ \text{false} \]
Markov Assumption

• Each variable is independent on its non-descendants, given its parents in Bayesian networks

• \( N \perp B, E, R \mid A \)
• \( R \perp B, A, N \mid E \)
• ...

\[ \text{Diagram showing relationships between variables} \]
Full Joint Distribution in BNs

Product Rule:

\[ P(A, B) = P(A \mid B)P(B) = P(B \mid A)P(A) \]

\[ P(B, E, A, R, N) = P(N \mid B, E, A, R)P(B, E, A, R) \]

\[ = P(N \mid A)P(R1 \mid E, B, A)P(E, B, A) \]

\[ = P(N \mid A)P(R1 \mid E) \]

\[ = \cdots = P(N \mid A)P(R1 \mid E)P(A \mid B, E, R)P(B)P(E) \]
Efficient inference

\[ P(R) = \sum_{B,E,A,N} P(B)P(E)P(A | B, E)P(R | E)P(N | A) \]

\[ P(R) = \sum_{B,E,A,N} P(B)P(E)P(A | B, E)P(R | E)P(N | A) \]

\[ = \sum_{B,E,A,N} P(B)P(E)P(A | B, E)P(R | E)P(N | A) \]

\[ = \sum_{B,E,A,N} P(B)P(E)P(A | B, E)P(R | E)P(N | A) \]

\[ = \sum_{B,E,A,N} P(B)P(E)P(A | B, E)P(R | E)P(N | A) \]

\[ = \sum_{B,E,A,N} P(B)P(E)P(A | B, E)P(R | E)P(N | A) \]

\[ = \sum_{B,E,A,N} P(B)P(E)P(A | B, E)P(R | E)P(N | A) \]

\[ = \sum_{B,E,A,N} P(B)P(E)P(A | B, E)P(R | E)P(N | A) \]

\[ = \sum_{B,E,A,N} P(B)P(E)P(A | B, E)P(R | E)P(N | A) \]
Undirected Graphs

• What if we allow undirected graphs?
• What do they correspond to?
• Not Cause/Effect, or Trigger/Response, but general dependence
• Example: Image pixels, each pixel is a bernouli
  – $P(x_{11}, ..., x_{1M}, ..., x_{M1}, ..., x_{MM})$
  – Bright pixels have bright neighbors
• No parents, just probabilities.
• Grid models are called Markov Random Fields
Undirected Graphs

• Undirected separability is easy.
• To check conditional independence of A and B given C, check the Graph reachability of A and B without going through nodes in C
Next Time

• Inference in Graphical Models