Natural Language Models and Interfaces

BSc Artificial Intelligence

Lecturer: Wilker Aziz
Institute for Logic, Language, and Computation

2018, week 3b — PGMs

Quick intro to PGMs

Check the lecture notes on PGMs

Tabular representation

Suppose A, B, and C are binary rvs

1. How do we represent $P_{A,B,C}$ (without making assumptions)?

Tabular representation

Suppose A, B, and C are binary rvs

1. How do we represent $P_{A,B,C}$ (without making assumptions)?

Joint assignments			Probability values
A	B	C	$P_{A,B,C}$
0	0	0	$P_{A,B,C}(0,0,0)$
0	0	1	$P_{A,B,C}(0,0,1)$
0	1	0	$P_{A,B,C}(0,1,0)$
1	0	0	$P_{A,B,C}(1,0,0)$
0	1	1	$P_{A,B,C}(0,1,1)$
1	1	0	$P_{A,B,C}(1,1,0)$
1	0	1	$P_{A,C,C}(1,0,1)$
1	1	1	$P_{B,B,C}(1,1,1)$

Table : Tabular joint distribution over 3 binary rvs

Tabular representation

Suppose A, B, and C are binary rvs

1. How do we represent $P_{A,B,C}$ (without making assumptions)?

Joint assignments			Probability values
A	B	C	$P_{A,B,C}$
0	0	0	$P_{A,B,C}(0,0,0)$
0	0	1	$P_{A,B,C}(0,0,1)$
0	1	0	$P_{A,B,C}(0,1,0)$
1	0	0	$P_{A,B,C}(1,0,0)$
0	1	1	$P_{A,B,C}(0,1,1)$
1	1	0	$P_{A,B,C}(1,1,0)$
1	0	1	$P_{A,C,C}(1,0,1)$
1	1	1	$P_{B,B,C}(1,1,1)$

Table : Tabular joint distribution over 3 binary rvs

2. How many probability values does it take in general for n variables with t outcomes each?

Tabular representation

Suppose A, B, and C are binary rvs

1. How do we represent $P_{A,B,C}$ (without making assumptions)?

Joint assignments			Probability values
A	B	C	$P_{A,B,C}$
0	0	0	$P_{A,B,C}(0,0,0)$
0	0	1	$P_{A,B,C}(0,0,1)$
0	1	0	$P_{A,B,C}(0,1,0)$
1	0	0	$P_{A,B,C}(1,0,0)$
0	1	1	$P_{A,B,C}(0,1,1)$
1	1	0	$P_{A,B,C}(1,1,0)$
1	0	1	$P_{A,C,C}(1,0,1)$
1	1	1	$P_{B,B,C}(1,1,1)$

Table : Tabular joint distribution over 3 binary rvs

2. How many probability values does it take in general for n variables with t outcomes each? n^t

Directed graphical models or Bayesian networks

A directed acyclic graph (DAG)

- nodes represent rvs
- edges represent direct influence
- a set of conditional independence statements
  - an rv is conditionally independent of its **non-descendants** given its **parents**

Conditional independence in BNs

Consider A, B, and C, due to chain rule we can write

$$P_{A,B,C}(a,b,c) = P_A(a) P_{B|A}(b|a) P_{C|AB}(c|a,b) \qquad (1)$$

Conditional independence in BNs

Consider A, B, and C, due to chain rule we can write

$$P_{A,B,C}(a, b, c) = P_A(a)P_{B|A}(b|a)P_{C|AB}(c|a, b) \tag{1}$$

But if we are given a particular set of assumptions

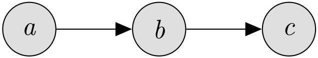

Figure : Examples of BN

then we can simplify it

$$P_{A,B,C}(a, b, c) = P_A(a)P_{B|A}(b|a)P_{C|AB}(c|a, b) \tag{2}$$

$$= P_A(a)P_{B|A}(b|a)P_{C|B}(c|b) \tag{3}$$

C is independent of non-descendants $\{A\}$ given its parents $\{B\}$

Chain rule for Bayesian networks

Chain rule (in general)

$$P_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \prod_{i=1}^{n} P_{X|X_{<i}}(x_i|x_{<i}) \tag{4}$$

Chain rule for Bayesian networks

$$P_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \prod_{i=1}^{n} P_{X|\mathsf{Pa}_X}(x|\mathsf{pa}_x) \tag{5}$$

where

- Pa_X set of rvs parents of X
- pa_X assignments of parents of X

Representing BNs

Each variable (given its parents) gets a tabular CPD

Thus for

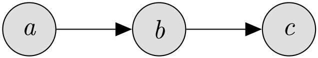

Figure : Examples of BN

A	P_A
0	$P_A(0)$
1	$P_A(1)$

A	B	$P_{B\|A}$
0	0	$P_{B\|A}(0\|0)$
0	1	$P_{B\|A}(1\|0)$
1	0	$P_{B\|A}(0\|1)$
1	1	$P_{B\|A}(0\|1)$

B	C	$P_{C\|B}$
0	0	$P_{C\|B}(0\|0)$
0	1	$P_{C\|B}(1\|0)$
1	0	$P_{C\|B}(0\|1)$
1	1	$P_{C\|B}(0\|1)$

Representation cost

- from $O(\prod_{i=1}^{n} |\text{supp}(X_i)|)$
- to $O(\sum_{i=1}^{n} |\text{supp}(X_i)| \times |\text{supp}(\text{Pa}_{X_i})|)$

Exercises

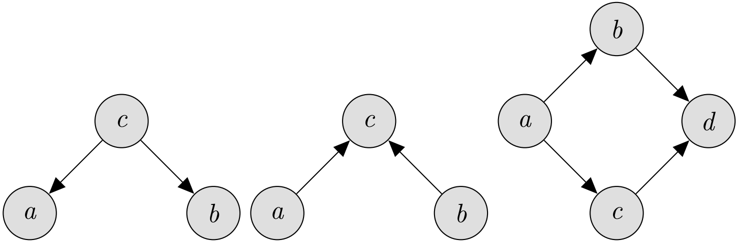

Figure : Write down the factorisation

Quiz

Inferences

So the BN shows us what are the CPDs in the problem

- but what if we want to reason about something that's not a CPD?

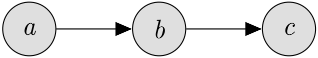

Figure : Examples of BN

Here we have CPDs P_A, $P_{B|A}$, and $P_{C|B}$

- how do we reason about $P_{B|C}$ or $P_{A|B}$?
- or P_B or P_C?
- or $P_{BC|A}$?

Inferences

So the BN shows us what are the CPDs in the problem

- but what if we want to reason about something that's not a CPD?

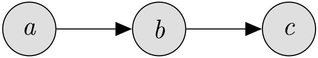

Figure : Examples of BN

Here we have CPDs P_A, $P_{B|A}$, and $P_{C|B}$

- how do we reason about $P_{B|C}$ or $P_{A|B}$?
- or P_B or P_C?
- or $P_{BC|A}$?

For whatever combination, we have rules of probability!

Conditional probability and marginalisation

If we have CPDs P_A, $P_{B|A}$, and $P_{C|B}$, infer $P_{B|C}$

Conditional probability and marginalisation

If we have CPDs P_A, $P_{B|A}$, and $P_{C|B}$, infer $P_{B|C}$

- start from the definition of conditional probability

$$P_{B|C}(b|c) = \frac{P_{BC}(b, c)}{P_C(c)}$$

Conditional probability and marginalisation

If we have CPDs P_A, $P_{B|A}$, and $P_{C|B}$, infer $P_{B|C}$

- start from the definition of conditional probability

$$P_{B|C}(b|c) = \frac{P_{BC}(b,c)}{P_C(c)}$$

- marginalise A in the numerator

$$P_{B|C}(b|c) = \frac{\sum_a P_{ABC}(a,b,c)}{P_C(c)}$$

Conditional probability and marginalisation

If we have CPDs P_A, $P_{B|A}$, and $P_{C|B}$, infer $P_{B|C}$

- start from the definition of conditional probability

$$P_{B|C}(b|c) = \frac{P_{BC}(b, c)}{P_C(c)}$$

- marginalise A in the numerator

$$P_{B|C}(b|c) = \frac{\sum_a P_{ABC}(a, b, c)}{P_C(c)}$$

- factorise the joint distribution to introduce the CPDs

$$P_{B|C}(b|c) = \frac{\sum_a P_A(a) P_{B|A}(b|a) P_{C|B}(c|b)}{P_C(c)}$$

Conditional probability and marginalisation

If we have CPDs P_A, $P_{B|A}$, and $P_{C|B}$, infer $P_{B|C}$

- start from the definition of conditional probability

$$P_{B|C}(b|c) = \frac{P_{BC}(b,c)}{P_C(c)}$$

- marginalise A in the numerator

$$P_{B|C}(b|c) = \frac{\sum_a P_{ABC}(a,b,c)}{P_C(c)}$$

- factorise the joint distribution to introduce the CPDs

$$P_{B|C}(b|c) = \frac{\sum_a P_A(a) P_{B|A}(b|a) P_{C|B}(c|b)}{P_C(c)}$$

- rearrange the terms for convenience

$$P_{B|C}(b|c) = \frac{P_{C|B}(c|b) \sum_a P_A(a) P_{B|A}(b|a)}{P_C(c)}$$

Conditional probability and marginalisation

If we have CPDs P_A, $P_{B|A}$, and $P_{C|B}$, infer $P_{B|C}$

- start from the definition of conditional probability

$$P_{B|C}(b|c) = \frac{P_{BC}(b, c)}{P_C(c)}$$

- marginalise A in the numerator

$$P_{B|C}(b|c) = \frac{\sum_a P_{ABC}(a, b, c)}{P_C(c)}$$

- factorise the joint distribution to introduce the CPDs

$$P_{B|C}(b|c) = \frac{\sum_a P_A(a) P_{B|A}(b|a) P_{C|B}(c|b)}{P_C(c)}$$

- rearrange the terms for convenience

$$P_{B|C}(b|c) = \frac{P_{C|B}(c|b) \sum_a P_A(a) P_{B|A}(b|a)}{P_C(c)}$$

- note that the last sum is the (inferred) marginal $P_B(b)$

Continuation

- we are here

$$P_{B|C}(b|c) = \frac{P_{C|B}(c|b) \sum_a P_A(a) P_{B|A}(b|a)}{P_C(c)}$$

Continuation

- we are here

$$P_{B|C}(b|c) = \frac{P_{C|B}(c|b) \sum_a P_A(a) P_{B|A}(b|a)}{P_C(c)}$$

- now obtain the marginal in the denominator as a function of tabular CPDs

$$\begin{aligned} P_C(c) &= \sum_a \sum_b P_{ABC}(a, b, c) \\ &= \sum_a \sum_b P_A(a) P_{B|A}(b|a) P_{C|B}(c|b) \\ &= \sum_a P_A(a) \sum_b P_{B|A}(b|a) P_{C|B}(c|b) \end{aligned}$$

Continuation

- we are here

$$P_{B|C}(b|c) = \frac{P_{C|B}(c|b) \sum_a P_A(a) P_{B|A}(b|a)}{P_C(c)}$$

- now obtain the marginal in the denominator as a function of tabular CPDs

$$\begin{aligned} P_C(c) &= \sum_a \sum_b P_{ABC}(a, b, c) \\ &= \sum_a \sum_b P_A(a) P_{B|A}(b|a) P_{C|B}(c|b) \\ &= \sum_a P_A(a) \sum_b P_{B|A}(b|a) P_{C|B}(c|b) \end{aligned}$$

- get back to the conditional

$$P_{B|C}(b|c) = \frac{P_{C|B}(c|b) \sum_a P_A(a) P_{B|A}(b|a)}{\sum_a P_A(a) \sum_b P_{B|A}(b|a) P_{C|B}(c|b)}$$

Quiz

References I