Graphical models overview

Brooks Paige

COMP0171

Factorizations and dependency structures

From the **product rule**, we can always factorize a joint distribution over three random variables a, b, c into a product of conditionals, for example

$$p(a, b, c) = p(a|b, c)p(b|c)p(c)$$

Factorizations and dependency structures

From the **product rule**, we can always factorize a joint distribution over three random variables a, b, c into a product of conditionals, for example

$$p(a, b, c) = p(a|b, c)p(b|c)p(c)$$
$$p(a, b, c) = p(a|b, c)p(c|b)p(b)$$

Factorizations and dependency structures

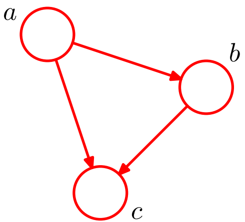
From the **product rule**, we can always factorize a joint distribution over three random variables a, b, c into a product of conditionals, for example

$$
\begin{aligned}
p(a, b, c) &= p(a|b, c)p(b|c)p(c) \\
p(a, b, c) &= p(a|b, c)p(c|b)p(b) \\
p(a, b, c) &= p(c|b, a)p(b|a)p(a) \\
&\vdots
\end{aligned}
$$

Factorizations and dependency structures

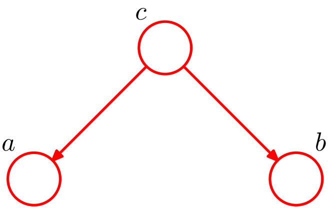
From the **product rule**, we can always factorize a joint distribution over three random variables a, b, c into a product of conditionals, for example

$$p(a, b, c) = p(a|b, c)p(b|c)p(c)$$
$$p(a, b, c) = p(a|b, c)p(c|b)p(b)$$
$$p(a, b, c) = p(c|b, a)p(b|a)p(a)$$
$$\vdots$$



Structured models and conditional independence

"Tail to tail"

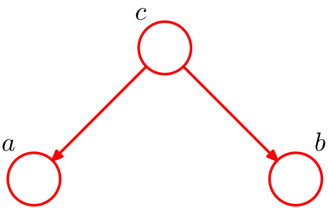


$$p(a, b, c) = p(a|c)p(b|c)p(c)$$

Figures: Bishop, Ch. 8

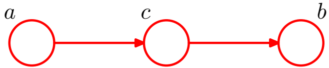
Structured models and conditional independence

"Tail to tail"



$$p(a, b, c) = p(a|c)p(b|c)p(c)$$

"Chain"

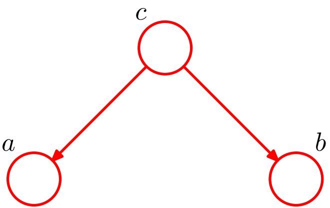


$$p(a, b, c) = p(a)p(c|a)p(b|c)$$

Figures: Bishop, Ch. 8

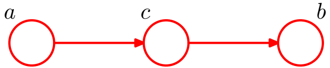
Structured models and conditional independence

"Tail to tail"



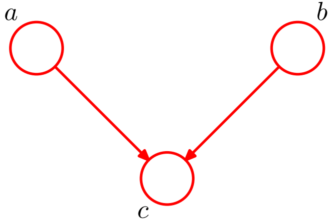
$$p(a, b, c) = p(a|c)p(b|c)p(c)$$

"Chain"



$$p(a, b, c) = p(a)p(c|a)p(b|c)$$

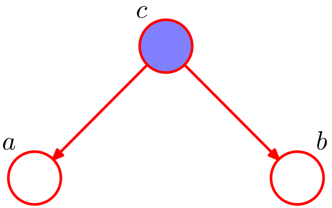
"Head to head"



$$p(a, b, c) = p(a)p(b)p(c|a, b)$$

Figures: Bishop, Ch. 8

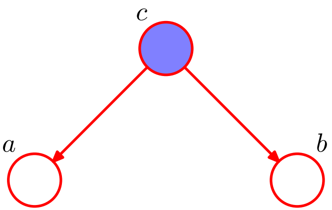
Conditioning on observed variables



$$p(a, b|c) = \frac{p(a|c)p(b|c)p(c)}{p(c)}$$
$$= p(a|c)p(b|c)$$

Figures: Bishop, Ch. 8

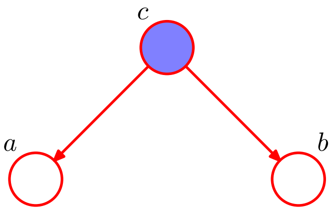
Conditioning on observed variables



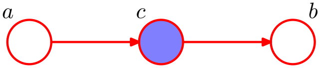
$$p(a, b|c) = \frac{p(a|c)p(b|c)p(c)}{p(c)}$$
$$= p(a|c)p(b|c)$$
$$\Rightarrow a \perp b | c$$

Figures: Bishop, Ch. 8

Conditioning on observed variables



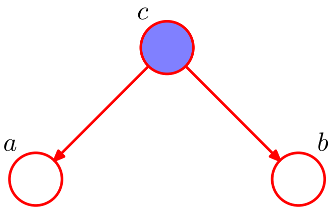
$$p(a,b|c) = \frac{p(a|c)p(b|c)p(c)}{p(c)}$$
$$= p(a|c)p(b|c)$$
$$\Rightarrow a \perp b|c$$



$$p(a,b|c) = \frac{p(a)p(c|a)p(b|c)}{p(c)}$$
$$= p(a|c)p(b|c)$$

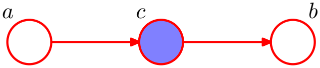
Figures: Bishop, Ch. 8

Conditioning on observed variables



$$p(a,b|c) = \frac{p(a|c)p(b|c)p(c)}{p(c)}$$
$$= p(a|c)p(b|c)$$

$\Rightarrow a \perp b | c$

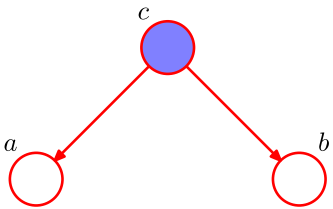


$$p(a,b|c) = \frac{p(a)p(c|a)p(b|c)}{p(c)}$$
$$= p(a|c)p(b|c)$$

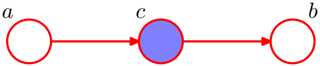
$\Rightarrow a \perp b | c$

Figures: Bishop, Ch. 8

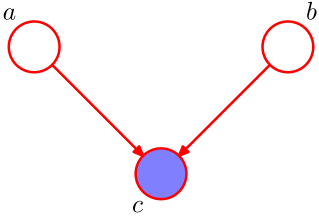
Conditioning on observed variables



$$p(a, b|c) = \frac{p(a|c)p(b|c)p(c)}{p(c)}$$
$$= p(a|c)p(b|c)$$
$$\Rightarrow a \perp b | c$$



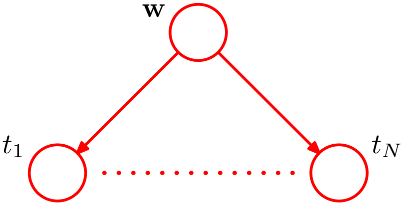
$$p(a, b|c) = \frac{p(a)p(c|a)p(b|c)}{p(c)}$$
$$= p(a|c)p(b|c)$$
$$\Rightarrow a \perp b | c$$

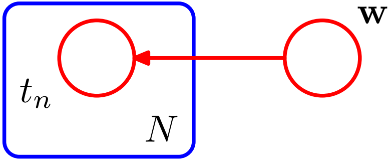


$$p(a, b|c) = \frac{p(a)p(b)p(c|a, b)}{p(c)}$$

Figures: Bishop, Ch. 8

Plate notation



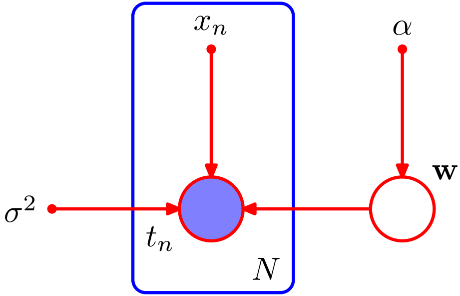


$$p(\mathbf{w}, t_1, \ldots, t_N) = p(\mathbf{w}) \prod_{n=1}^{N} p(t_n|\mathbf{w})$$

Figures: Bishop, Ch. 8

Including deterministic variables

Hyperparameters just get "dots", no circles.



$$p(\mathbf{w}, t_1, \ldots, t_N | \alpha, \sigma^2, x_{1:N}) = p(\mathbf{w}|\alpha) \prod_{n=1}^{N} p(t_n|\mathbf{w}, x_n, \sigma^2)$$

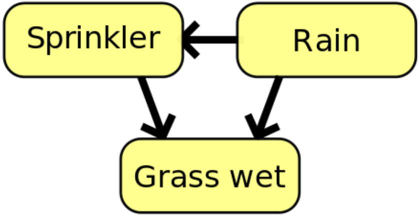
Figure: Bishop, Ch. 8

Why do this?

Example: building larger discrete distributions

Graphical models are a way of visually representing dependencies (and conditional dependencies) in models.

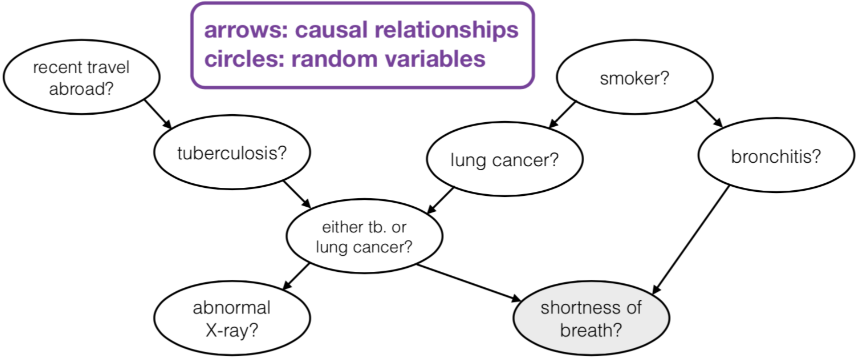
$$p(G, S, R) = p(G|S, R)p(G|R)p(R)$$



This is a "Bayes net" which describes the relationship between three random variables, each a discrete distribution. There exist fast algorithms for estimating the posterior over any particular random variable, conditioned on any set of other random variables.

Example: medical diagnostics

Easy way of encoding (probabilistic) domain knowledge and causal relationships.



These models have a long history in medical diagnosis systems. Each conditional probability in this graph can be estimated reliably from historical clinical data.

Lauritzen & Spiegelhalter, 1988

Estimating effects of COVID-19 interventions

This is a generative model for COVID-19 deaths (Flaxman et al., 2020).

- The observed data are dates of deaths and of different types of interventions, in a variety of countries.
- The goal is to estimate the time-varying reproduction number R.