Lecture 13: Bayesian Graphical Models

University of BRISTOL

EMAT31530/March 2018/Raul Santos-Rodriguez

Have a look at ...

... Russell and Norvig (Ch. 14)

... The introduction to the book Graphical Models; Foundations of Neural Computing (ed. M. Jordan)

... David barber's Bayesian reasoning and Machine Learning:
http://www.cs.ucl.ac.uk/staff/d.barber/brml/

... Kevin Murphy's Toolbox:
http://www.cs.ubc.ca/~murphyk/Software/

The road so far

- Machine learning

  Binary classification:

  $x \rightarrow \boxed{?} \rightarrow y \in \{-1, +1\}$, single action

- Search
- Bayesian networks
- Markov decision process
- Game theory

The road so far

- Machine learning

Binary classification:

$x \rightarrow \boxed{?} \rightarrow y \in \{-1, +1\}$, single action

- Search

Search problem:

$x \rightarrow \boxed{?} \rightarrow$ action sequence $(a_1, a_2, a_3, a_4, \ldots)$

- Bayesian networks
- Markov decision process
- Game theory

The road so far

- Machine learning

  Binary classification:

  $x \rightarrow \boxed{?} \rightarrow y \in \{-1, +1\}$, single action

- Search

  Search problem:

  $x \rightarrow \boxed{?} \rightarrow$ action sequence $(a_1, a_2, a_3, a_4, \dots)$

- Bayesian networks

  Bayesian Networks:

  Model uncertanty!

- Markov decision process

- Game theory

Outline

This lecture introduces probabilistic graphical models. These are very powerful representations that enable us to scale probabilistic models to large real problems. The objective is to learn the following topics:

- Bayes theorem
- Definition of a DAG (or Bayesian network)
- Domains of application of Bayesian nets
- Conditional independence in DAGs

Bayes theorem: The Horrible Disease?

You are about to be tested for a rare disease. How worried should you be if the test result is positive? The doctor has bad news and good news...

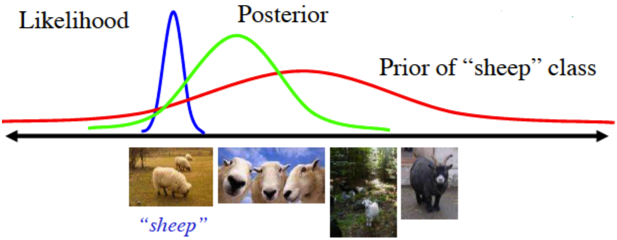
Bad You tested positive for a serious disease, and that the test is 99% accurate *(i.e., the probability of testing positive given that you have the disease is 0.99, as is the probability of testing negative given that you don't have the disease)*.

Good This is a rare disease, striking only 1 in 10,000 people.

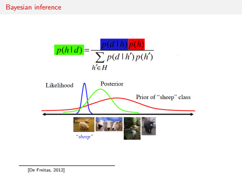
What are the chances that you actually have the disease?

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B)}$$

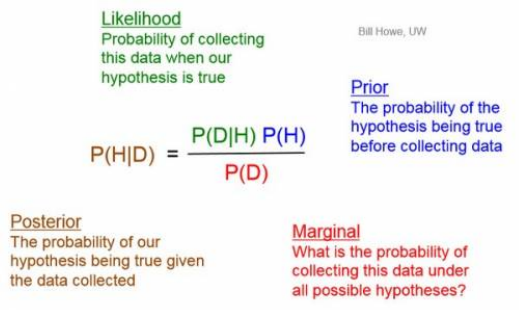
$$p(h \mid d) = \frac{p(d \mid h)\, p(h)}{\sum_{h' \in H} p(d \mid h')\, p(h')}$$



[De Freitas, 2012]



Bayesian inference derives the posterior probability as a consequence of two antecedents, a prior probability and a "likelihood function" derived from a probability model for the data to be observed. Bayesian inference computes the posterior probability according to Bayes' theorem.



Bayes theorem: example

The test is 99% accurate:

$$P(T = 1|D = 1) = 0.99$$

$$P(T = 0|D = 0) = 0.99$$

where T denotes test and D denotes disease.

The disease affects 1 in 10000:

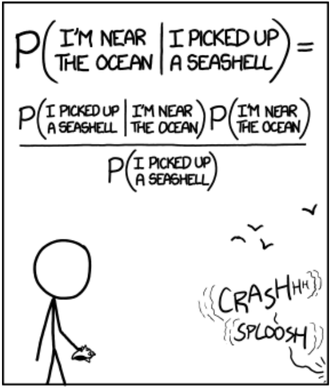
$$P(D = 1) = 0.0001$$

$$P(D = 1|T = 1) = \frac{P(T = 1|D = 1)P(D = 1)}{P(T = 1|D = 0)P(D = 0) + P(T = 1|D = 1)P(D = 1)}$$

$$P(D = 1|T = 1) = 0.0098$$

[https://www.intelligentinvestor.com.au/the-theory-that-cracked-the-enigma-code-1811436]

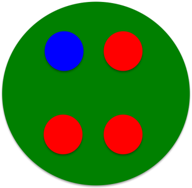
Bayes theorem



STATISTICALLY SPEAKING, IF YOU PICK UP A SEASHELL AND *DON'T* HOLD IT TO YOUR EAR, YOU CAN PROBABLY HEAR THE OCEAN.

A simple graphical model: Example

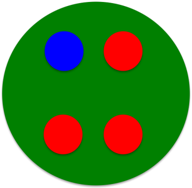
Problem: drawing balls from the set {r,r,r,b}



What are $P(B_1)$, $P(B_2|B_1)$ and $P(B_1, B_2)$?

A simple graphical model: Example

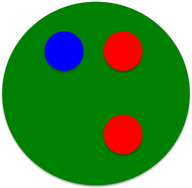
Problem: drawing balls from the set {r,r,r,b}



$$P(B_1) = \begin{array}{|c|c|} \hline B_1 & ? \\ \hline B_1 & ? \\ \hline \end{array}$$

A simple graphical model: Example

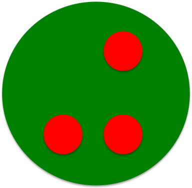
Problem: drawing balls from the set {r,r,r,b}



$$P(B_1) = \begin{array}{|c|c|}\hline B_1 & 3/4 \\ \hline B_1 & ? \\ \hline \end{array}$$

A simple graphical model: Example

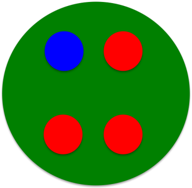
Problem: drawing balls from the set {r,r,r,b}



$P(B_1) =$

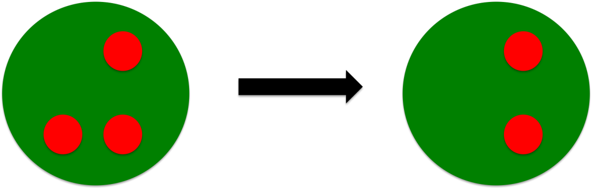
B_1	3/4
B_1	1/4

Problem: drawing balls from the set {r,r,r,b}



$$P(B_2|B_1) = \begin{array}{|c|c|c|} \hline & B_2 & B_2 \\ \hline B_1 & ? & ? \\ \hline B_1 & ? & ? \\ \hline \end{array}$$

Problem: drawing balls from the set {r,r,r,b}

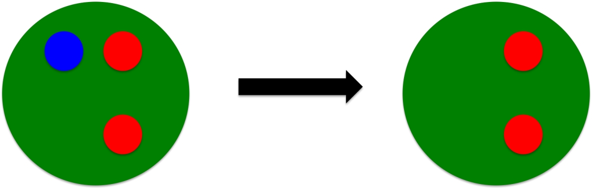


$P(B_2|B_1) =$

	B_2	B_2
B_1	?	?
B_1	1	0

A simple graphical model: Example

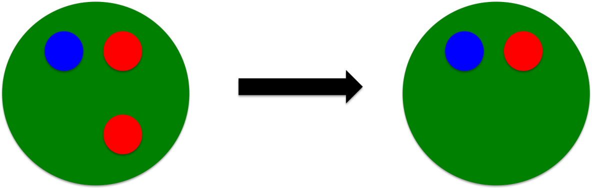
Problem: drawing balls from the set {r,r,r,b}



$P(B_2|B_1) =$

	B_2 (red)	B_2 (blue)
B_1 (red)	?	1/3
B_1 (blue)	1	0

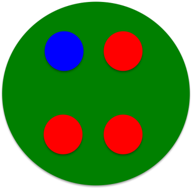
Problem: drawing balls from the set {r,r,r,b}



$P(B_2|B_1) =$

	B_2	B_2
B_1	2/3	1/3
B_1	1	0

Problem: drawing balls from the set {r,r,r,b}



$P(B_1, B_2) =$

	B_2 (r)	B_2 (b)
B_1 (r)	1/2	1/4
B_1 (b)	1/4	0

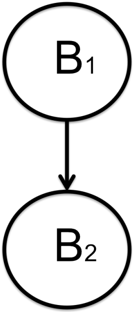
A simple graphical model: Example

Problem: drawing balls from the set {r,r,r,b}

$$P(B_1) = \begin{array}{|c|c|} \hline B_1 & 3/4 \\ \hline B_1 & 1/4 \\ \hline \end{array}$$

$$P(B_2|B_1) = \begin{array}{c|c|c|} \cline{2-3} & B_2 & B_2 \\ \hline \multicolumn{1}{|c|}{B_1} & 2/3 & 1/3 \\ \hline \multicolumn{1}{|c|}{B_1} & 1 & 0 \\ \hline \end{array}$$

$$P(B_1, B_2) = \begin{array}{c|c|c|} \cline{2-3} & B_2 & B_2 \\ \hline \multicolumn{1}{|c|}{B_1} & 1/2 & 1/4 \\ \hline \multicolumn{1}{|c|}{B_1} & 1/4 & 0 \\ \hline \end{array}$$



A simple graphical model: Example

Problem: drawing balls from the set {r,r,r,b}

$P(B_1) =$

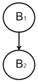
B_1	3/4
B_1	1/4

$P(B_2|B_1) =$

	B_2	B_2
B_1	2/3	1/3
B_1	1	0

$P(B_1, B_2) =$

	B_2	B_2
B_1	1/2	1/4
B_1	1/4	0



A Bayesian network represents the dependencies among variables. Bayesian networks can represent essentially any full joint probability distribution and in many cases can do so very concisely.

Directed probabilistic graphical models

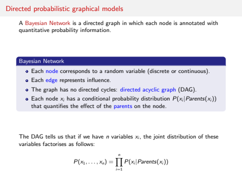
A Bayesian Network is a directed graph in which each node is annotated with quantitative probability information.

Bayesian Network

- Each node corresponds to a random variable (discrete or continuous).
- Each edge represents influence.
- The graph has no directed cycles: directed acyclic graph (DAG).
- 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 DAG tells us that if we have n variables x_i, the joint distribution of these variables factorises as follows:

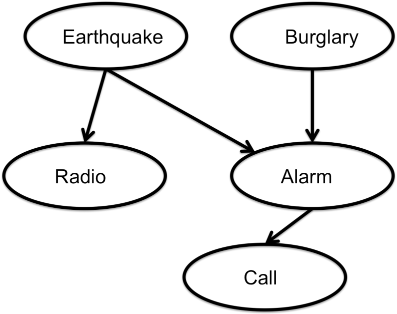
$$P(x_1, \dots, x_n) = \prod_{i=1}^{n} P(x_i|Parents(x_i))$$



- Each node corresponds to a random variable (discrete or continuous).
- A set of directed links or arrows connects pairs of nodes. If there is an arrow from node X to node Y , X is said to be a parent of Y. The graph has no directed cycles (and hence is a directed acyclic graph, or DAG).
- 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.

Directed probabilistic graphical models: example

A Bayesian Network is a directed graph in which each node is annotated with quantitative probability information.



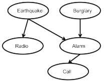
$$P(A, R, E, B) = P(A|R, E, B)P(R|E, B)P(E|B)P(B)$$

Can be reduced to:

$$P(A, R, E, B) = P(A|E, B)P(R|E)P(E)P(B)$$

Directed probabilistic graphical models: example

A Bayesian Network is a directed graph in which each node is annotated with quantitative probability information.



$$P(A, R, E, B) = P(A|R, E, B)P(R|E, B)P(E|B)P(B)$$

Can be reduced to:

$$P(A, R, E, B) = P(A|E, B)P(R|E)P(E)P(B)$$

We can look at Bayesian Networks as a "story". They offer the example of a story containing five random variables: "Burglary", "Earthquake", "Alarm", "Neighbour Call", and "Radio Announcement". In such a story, "Burglary" and "Earthquake" are independent, and "Burglary" and "Radio Announcement" are independent given "Earthquake." This is to say that there is no event which effects both burglaries and earthquakes. As well, "Burglary" and "Radio Announcements" are independent given "Earthquake" - meaning that while a radio announcement might result from an earthquake, it will not result as a repercussion from a burglary.

Because of the independence among these variables, the probability of $P(A, R, E, B)$ (the joint probability of an alarm, radio announcement, earthquake and burglary) can be reduced from:

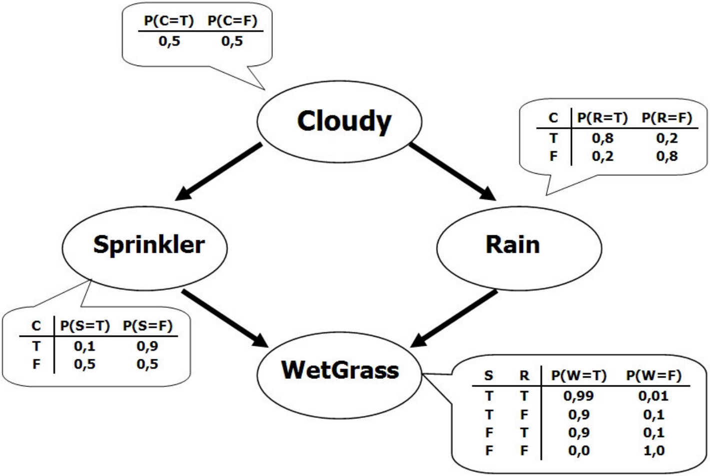
$$P(A, R, E, B) = P(A|R, E, B)P(R|E, B)P(E|B)P(B)$$

to:

$$P(A, R, E, B) = P(A|E, B)P(R|E)P(E)P(B)$$

This significantly reduce the number of probabilities involved.

$$p(C,S,R,W) = p(C)p(S|C)p(R|C)p(W|S,R)$$



P(C=T)	P(C=F)
0,5	0,5

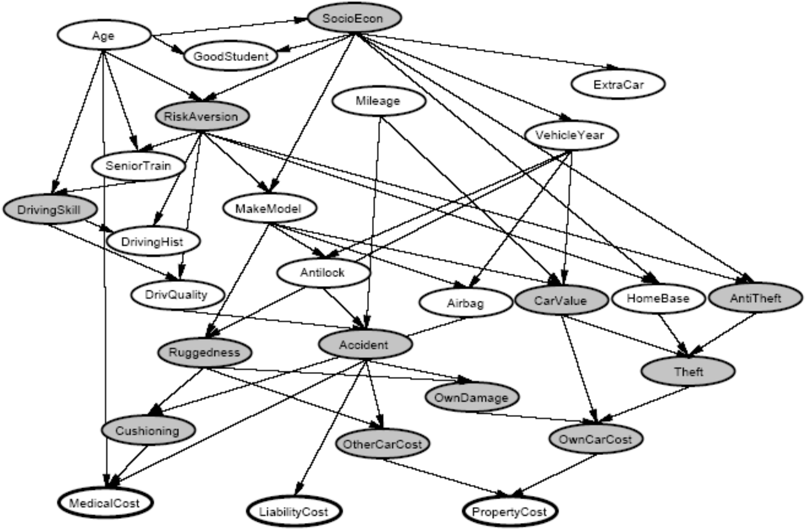
C	P(R=T)	P(R=F)
T	0,8	0,2
F	0,2	0,8

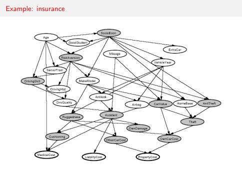
C	P(S=T)	P(S=F)
T	0,1	0,9
F	0,5	0,5

S	R	P(W=T)	P(W=F)
T	T	0,99	0,01
T	F	0,9	0,1
F	T	0,9	0,1
F	F	0,0	1,0

Joint distribution → 15 values vs Factorised joint distribution → 9 values

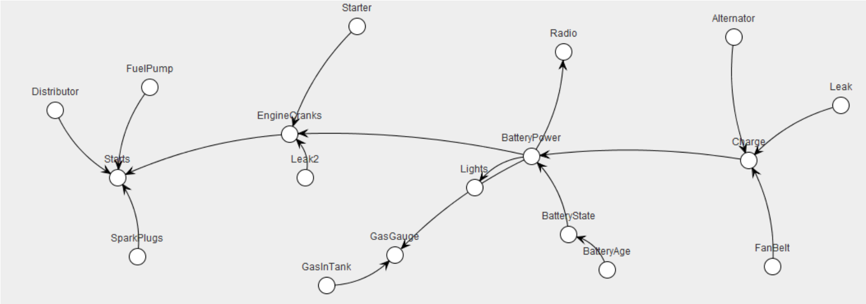
Example: insurance





The next Bayes net attempts to estimate the risk posed by an individual seeking car insurance. In other words, the network attempts to relate variables like the age and driving history of the individual to the cost and likelihood of property damage or bodily injury. The 12 shaded variables are considered hidden or unobservable, while the other 15 are observable. The network has over 1400 parameters. An insurance company would be interested in predicting the bottom three "cost" variables, given all or a subset of the other observable variables.

Example: car engine



[David Heckerman and John S. Breese and Koos Rommelse]

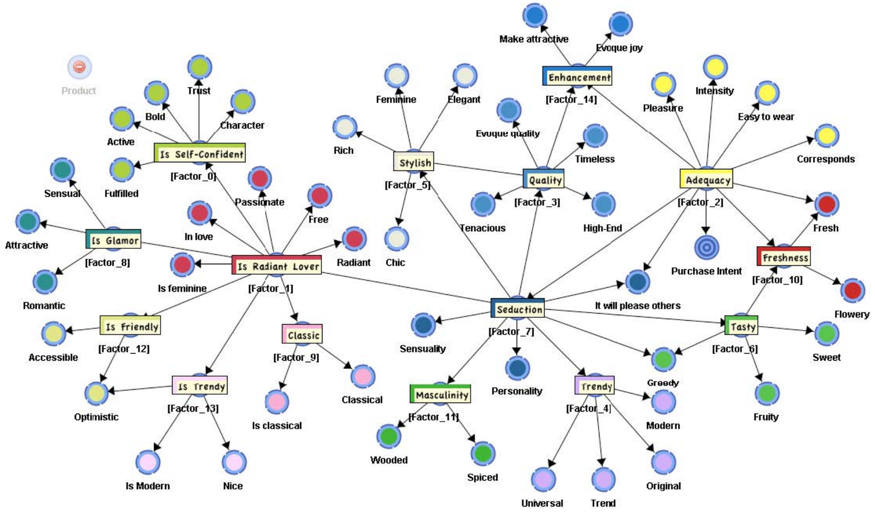
2018-03-08

Example: car engine

[David Heckerman and John S. Breese and Koos Rommelse]

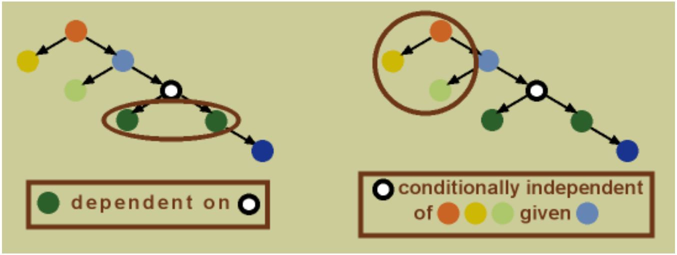
This is a Bayesian Network used to predict if a car will start.

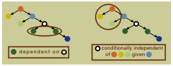
Example: perfume market analysis



[http://www.bayesia.com/en/applications/marketing.php]

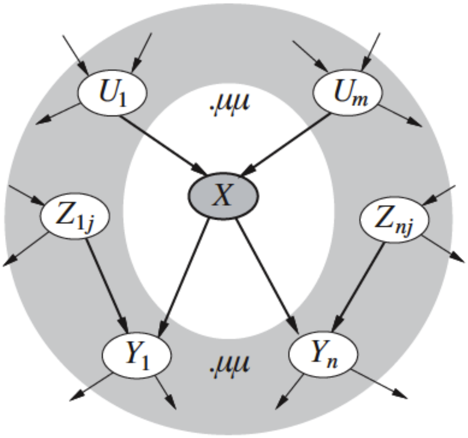
Conditional independence





In a Bayesian network, dependence is indicated by directed edges. A child node is dependent on its parent node. For example, in the graph above, the dark green nodes are both dependent on the black and white node. Additionally, any node is conditionally independent of its non-descendants, given its parent. For example, the black and white node is conditionally independent of the light green, orange, and yellow nodes, given the light blue node (its descends are the dark green and dark blue nodes, thus its non-descendants are all the other nodes).

Given its Markov Blanket, X is Independent of All Other Nodes



$$MB(X) = Parents(X) \cup Children(X) \cup Parents(Children(X))$$

Next lecture

We will talk about inference and discuss the influence of time on probabilistic models.