The Artificial Intelligence Toolbox Part II – CS26210

Elio Tuci
elt7@aber.ac.uk

Using Qwizdom QVR

- On any web-enabled device go to:
- http://qvr.qwizdom.com
- Select **I have a Session Key**
- Enter the code **Q5VN94**

If you aren't already using AU Eduroam wireless have a look at

http://www.inf.aber.ac.uk/advisory/faq/253

Program

Week 1

7/02 Set Theory, Fuzzy Logic (319)
8/02 Fuzzy Logic (B20) - Hand-out Assignment 1

Week 2

14/02 Fuzzy Logic - Further Exercises (319)
15/02 Theory of Probability (B20)

Week 3

21/02 Conditional Probability (319)
22/02 Probabilistic Methods (B20) - Hand-in Assignment 1 (Blackboard)

Week 4

28/02 Bayesian Networks (B20)
1/03 In Class Test (319) (Set Theory, Theory of Probability, Conditional Probability)

Week 5

7/03 Bayesian networks (319)
8/03 Discussion, further exercises (B20) - Hand-out Assignment 2
22/03 Hand-in Assignment 2 (Blackboard)

In-Class Test

Assessment 2

You work as an AI consultant, and have been called in by a small-but growing Insurance Company. In order to remain in business the company needs to decide to whom they should provide car insurance.

……

The company has asked you to write an assessment of both approaches (Probabilistic and Fuzzy Logic) and to make a recommendation for the approach that the company should take. You have until.

Background Reading – Briefly assess similar systems you found during a day or so of background reading, and their relevance to this project.

Exploring the Possibilities: From Theory to Application – Explore the arguments for and against using probabilistic and fuzzy logic/systems for the various aspects of the system.

A Summary of Your Recommendations – Highlight any problems or issues about which the company should be aware, and summarise your recommendations.

Thursday 28th, March 2013

- First-order Markov chain
- Bayes' Theorem
- Bayes' Theorem - Law of total probability
- Bayes' Theorem - Multiple Causes
- Bayes' Theorem - Multiple Effects

Bayes' Rules

$$P(\text{cause} \mid \text{effect}) = \frac{P(\text{cause})\, P(\text{effect} \mid \text{cause})}{P(\text{effect})}$$

Bayes ‘ Rules

Law of total probability

$$P(E) = P(E|C)P(C) + P(E|\neg C)P(\neg C)$$

$$P(C|E) = \frac{P(E|C)P(C)}{P(E)}$$

$$P(C|E) = \frac{P(E|C)P(C)}{P(E|C)P(C) + P(E|\neg C)P(\neg C)}$$

Bayes' Rules with multiple mutually exclusive causes

$E \wedge c_1$	$E \wedge c_2$	$E \wedge c_3$	$E \wedge c_4$
$E \wedge c_5$	$E \wedge c_6$	$E \wedge c_7$	$E \wedge c_8$
$E \wedge c_9$	$E \wedge c_{10}$	$E \wedge c_{11}$	$E \wedge c_{12}$

$$P(C_k \mid E) = \frac{P(E \mid C_k)P(C_k)}{\sum_{i=1}^{N} P(E \mid C_i)P(C_i)} \text{ with } k \in [1, N]$$

Bayes' Rules

multiple conditionally independent effects

One cause, multiple symptoms

$P(E_1 \wedge E_2|C) = P(E_1|c)\,P(E_2|c)$

Naive Bayes Classifier

$$P(C \mid E_1, E_2, E_3, \ldots., E_n) = \frac{\prod_{i=1}^{n} P(E_i \mid C)\,P(C)}{P(E_1, E_2, E_3, \ldots., E_n)}$$

Thursday 7th, March 2013

- Bayesian networks

Problem?

Making Inference with Full Joint Probability Distribution Tables is hard when there are too many variables, because the problem becomes intractably large.

Problem

The FJPD table can get ENORMOUS very quickly. n Boolean variables requires an input table of size $O(2^n)$.
With two variables that each have two values, there are four entries. However:

If there are two variables each with four values then there are (4*4) 16 entries

If there are four variables each with three values then there are (3*3*3*3) 81 entries

If there are 20 variables, each with three values then there are 3,486,784,401 entries

Things get out of control very fast.

So how can we model real-world systems probabilistically?

Solutions

1. Absolute Independence
2. Conditional Independence
3. Bayesian Networks

Absolute Independence

Independent Events and Unconditional Probability
then
there is no problem for inference

$P(a \land b \land c) = P(a)P(b)P(c)$

E.g., we roll four fair dice. The sample space is:

$S = \{(x_1, x_2, x_3, x_4): x_1=1,\dots,6; x_2=1,\dots,6; x_3=1,\dots,6; x_4=1,\dots,6;\}$

The probability of having four 6?

$X(\omega) = (x_1=6; x_2=6; x_3=6; x_4=6;)$

$P(X) = 1/6 * 1/6 * 1/6 * 1/6 = 1/1296$

Absolute Independence

P(Toothache, Catch, Cavity, Weather)
Independence of Weather from dental problems

P(Toothache,Catch,Cavity,Weather) = P(Weather)P(Toothache,Catch,Cavity)

	toothache		¬*toothache*	
	catch	¬catch	*catch*	¬catch
cavity	*.108*	*.012*	*.072*	*.008*
¬cavity	*.016*	*.064*	*.144*	*.576*

weather	
cloudy	0.6
rain	0.1
snow	0.29
sunny	0.01

From (4*2*2*2) 32 entries to 12 entries
P(toothache,catch,cavity,cloudy) = P(cloudy)P(toothache,catch,cavity)

Conditional Independence

First case

$P(A|B) = P(A)$

A and B Ind.

Second case

$P(A|B \cap C) = P(A|C)$

A and B Indep. given C

Third case

$P(A \cap B|C)=P(A|C)\ P(B|C)$

A and B Indep. given C

Bayesian Networks

Joint vs. Conditional Probability

What is the difference between **joint** and **conditional** probability?

Joint probability takes a big picture view of things and gets the probability of something without any conditions being known E.g., Holiday ={true, false}; Sunny={true, false}

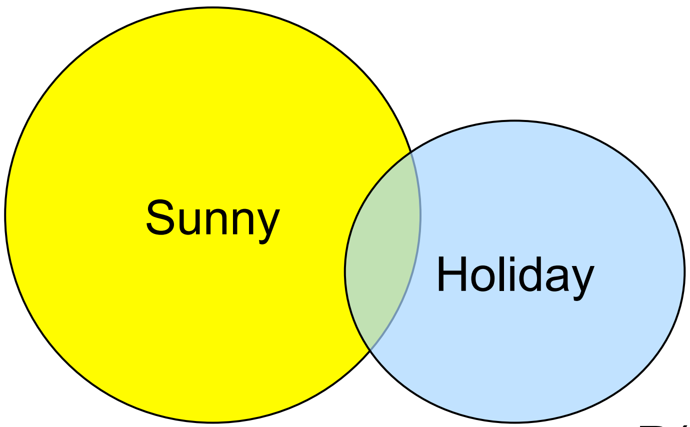
$P(holiday=true \wedge sunny=true)$

$P(holiday=true \wedge sunny=false)$

Etc.

Conditional probability "zooms in" on just one condition of things and gets the probability of something else given relative to that condition. E.g., $P(holiday=true \mid sunny=true)$

Full Joint Probability Distribution Table



Here, we know both sunny and holiday are true
but what is the probability of this happening?

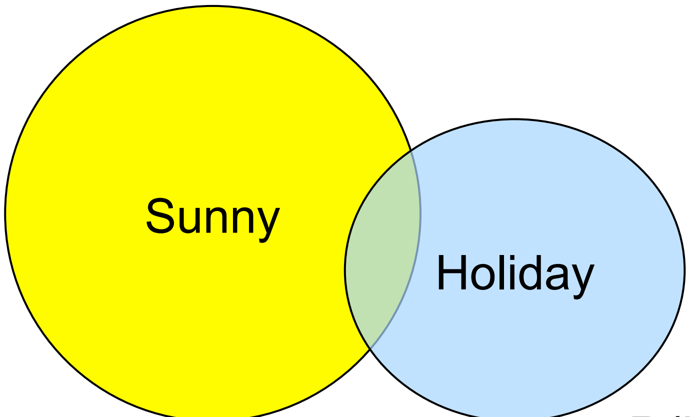
P(sunny=true) = 0.3 P(sunny=false) = 0.7 (sums to 1)
P(holiday=true) = 0.08 P(holiday=false)=0.92 (sums to 1)

The probabilities in the joint table are globally observed probabilities.
$P(A \wedge B)=P(A|B)*P(B)$
In this case they are the global probabilities for the various possible values of S and H

Joint	H=true	H=false
S=true	0.045	0.255
S=false	0.035	0.665

(Table sums to 1)

Conditional Probability Table



Here, we know sunny is true,
what is the probability that holiday
is also true?

P(sunny=true) = 0.3 P(sunny=false) = 0.7 (sums to 1)
P(holiday=true) = 0.08 P(holiday=false)=0.92 (sums to 1)

Conditional probabilities are the locally observed probabilities for a variable, given known values for other variables
$P(A|B) = P(A \wedge B)/P(B)$
In this case we want to know the probabilities for Holiday relative to known knowledge about Sunny

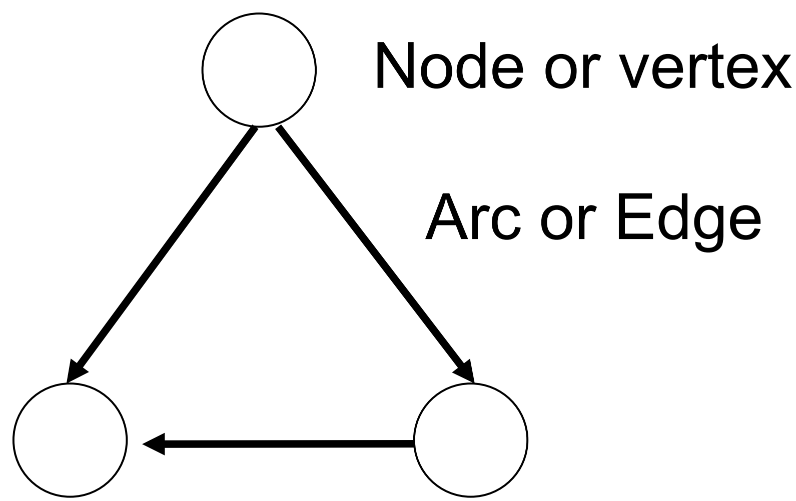
Conditional	H=true	H=false
|S=true	0.15	0.85
|S=false	0.05	0.95

(Rows sum to 1)

Bayesian Networks

A Bayesian network is **directed acyclic graph** which represents dependencies among variables.

A **directed acyclic graph** is formed by a collection of nodes or vertices and directed edges or arcs, each edge connecting one vertex to another, such that there is no way to start at some vertex v and follow a sequence of edges that eventually loops back to v again.



Bayesian Networks

Each node is a variable (continuous or discrete).

Connection from Y to X means that Y is the cause (parent) of X.

The Network Topology defines Dependencies among variables. Each variable is conditionally independent of its non-descendants, given its parents.

Each node as a prior P(Y) or conditional probability $P(X_i|Parents(X_i))$

Scenario with multiple variables Cavity, Toothache, Catch, Weather

It is sunny, but it could be rainy, cloudy or with snow

A person goes to the dentist with toothache or just for a check.

The probability to find a cavity, without toothache, is P(Cavity)

The probability to find a cavity, with toothache, is P(cavity | toothache)

The first thing the dentist does, is to check for cavities with a steel probe. There is a cavity if the probe catches in the person's teeth. P(Catches, Cavity)

Anatomy of Bayesian Networks

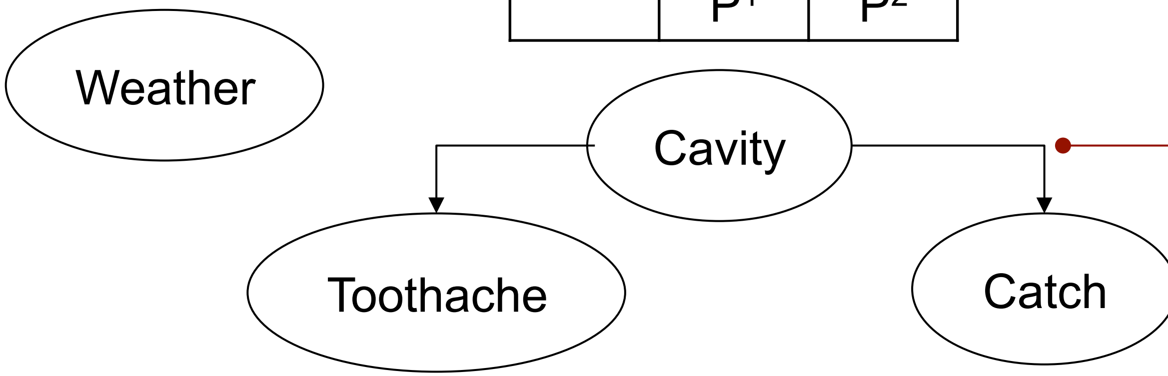
(E.g., Toothache, Catch, Cavity, Weather)

Prior probabilities table
(no condition)

	Cavity	
	T	F
	P^1	P^2

Arrows show direction
of influence

	Weather	
	T	F
	P^1	P^2



Each node
represent
a variable

|Cavity	Toothache	
	T	F
T	P^1	P^2
F	P^4	P^3

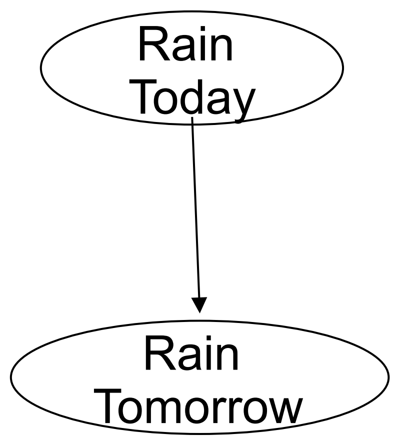
Conditional
probabilities
tables

|Cavity	Catch	
	T	F
T	P^1	P^2
F	P^4	P^3

Bayesian Networks
first (simple) example

Given a situation where it might rain today, and might rain tomorrow, and suppose that P(rain today) = 0.2, P(rain tomorrow given that it rains today) = 0.7, what is the probability it rains both days?

$$P(E^1) = 0.2,\ P(E^2|E^1) = 0.7,\ P(E^2 \wedge E^1)=?$$



E^1	true	false
	0.2	x

	E^1	
E^2	true	false
false	?	?
true	0.7	?

$$P(E^2 \wedge E^1)=P(E^2|E^1)P(E^1)$$

$$0.7*0.2 = 0.14$$

Bayesian Networks
An example from wikipedia

There are two events which could cause grass to be wet: either the sprinkler is on or it's raining. Also, suppose that the rain has a direct effect on the use of the sprinkler (namely that when it rains, the sprinkler is usually not turned on). All three variables (Sprinkler, Rain, Grass Wet) have two possible values T (for true) and F (for false).

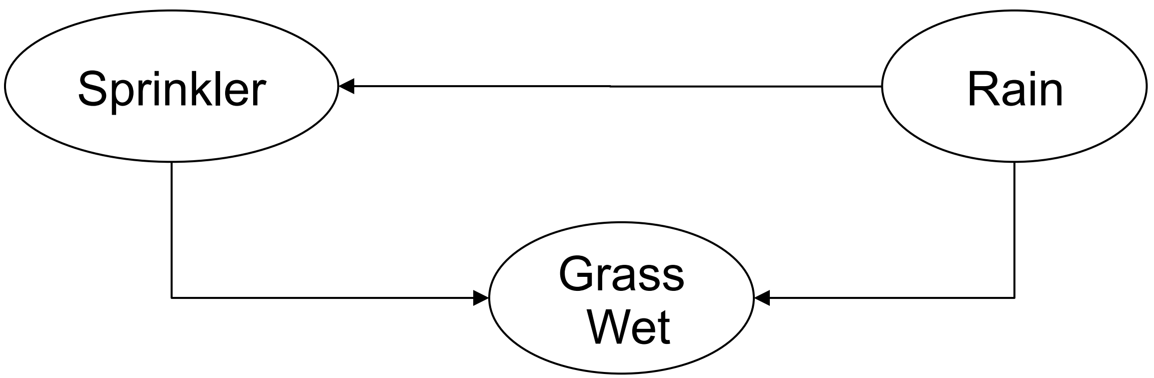
Bayesian Networks
An example from wikipedia

Three variables (Sprinkler, Rain, Grass Wet)

Sprinkler (true) or Rain (true) make Grass Wet (true)

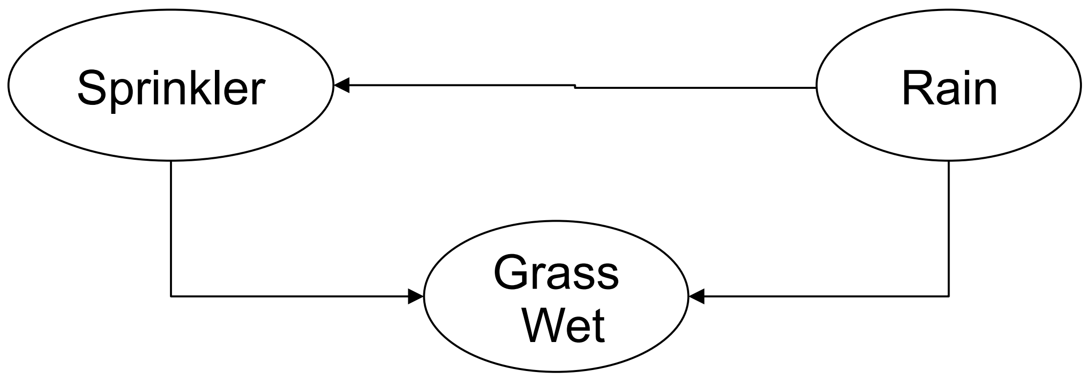
Rain (true) determine Sprinkler (false)

Conditional probability table



Bayesian Networks (from wikipedia)

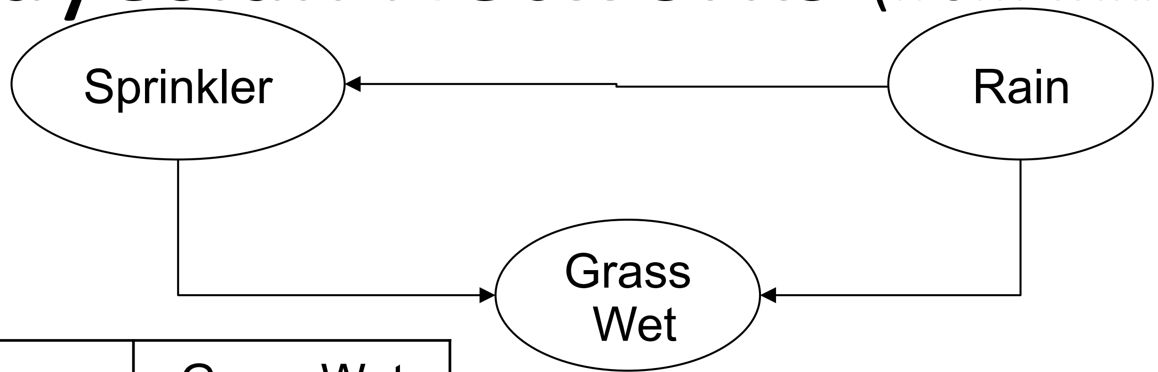
	Sprinkler	
Rain	true	false
false	0.4	0.6
true	0.01	0.99



Rain	true	false
	0.2	0.8

		Grass Wet	
Sprinkler	Rain	true	false
false	false	0.0	1.0
false	true	0.8	0.2
true	false	0.9	0.1
true	true	0.99	0.01

Bayesian Networks (from wikipedia)



		Grass Wet	
Sprinkler	Rain	true	false
false	false	0.0	1.0
false	true	0.8	0.2
true	false	0.9	0.1
true	true	0.99	0.01

	Sprinkler	
Rain	true	false
false	0.4	0.6
true	0.01	0.99

Rain	true	false
	0.2	0.8

What is the probability it is raining and the grass is wet and the sprinkler is off?

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

Bayesian Networks (from wikipedia)

	Sprinkler	
Rain	true	false
false	0.4	0.6
true	0.01	0.99

		Grass Wet	
Sprinkler	Rain	true	false
false	false	0.0	1.0
false	true	0.8	0.2
true	false	0.9	0.1
true	true	0.99	0.01

Rain	true	false
	0.2	0.8

What is the probability it is raining given that the grass is wet?

Bayesian Networks (from wikipedia)

	Sprinkler	
Rain	true	false
false	0.4	0.6
true	0.01	0.99

		Grass Wet	
Sprinkler	Rain	true	false
false	false	0.0	1.0
false	true	0.8	0.2
true	false	0.9	0.1
true	true	0.99	0.01

Rain	true	false
	0.2	0.8

To make inference we use the conditional probability equation

$$P(A|B) = \frac{P(A \wedge B)}{P(B)}$$

and we bear in mind that we multiply entries instead of adding them up.

Bayesian Networks (from wikipedia)

	Sprinkler	
Rain	true	false
false	0.4	0.6
true	0.01	0.99

		Grass Wet	
Sprinkler	Rain	true	false
false	false	0.0	1.0
false	true	0.8	0.2
true	false	0.9	0.1
true	true	0.99	0.01

Rain	true	false
	0.2	0.8

What is the probability it is raining given that the grass is wet?

$$P(R_t \mid G_t) = \frac{P(R_t \cap G_t)}{P(G_t)} = \frac{\sum_{S=\{t,f\}} P(R_t, S, G_t)}{\sum_{S,R=\{t,f\}} P(G_t, S, R)}$$

Bayesian Networks (from wikipedia)

	Sprinkler	
Rain	true	false
false	0.4	0.6
true	0.01	0.99

		Grass Wet	
Sprinkler	Rain	true	false
false	false	0.0	1.0
false	true	0.8	0.2
true	false	0.9	0.1
true	true	0.99	0.01

Rain	true	false
	0.2	0.8

What is the probability it is raining given that the grass is wet?

$$= \frac{P\left(G_t \mid S_t \cap R_t\right)P(S_t \mid R_t)P(R_t) + P\left(G_t \mid S_f \cap R_t\right)P(S_f \mid R_t)P(R_t)}{M + N}$$

$$M = P\left(G_t \mid S_t \cap R_t\right)P(S_t \mid R_t)P(R_t) + P\left(G_t \mid S_f \cap R_t\right)P(S_f \mid R_t)P(R_t)$$

$$N = P\left(G_t \mid S_t \cap R_f\right)P(S_t \mid R_f)P(R_f) + P\left(G_t \mid S_f \cap R_f\right)P(S_f \mid R_f)P(R_f)$$

$$\frac{(0.99\text{x}0.01\text{x}0.2)+(0.8\text{x}0.99\text{x}0.2)}{\underset{\text{TTT}}{(0.99\text{x}0.01\text{x}0.2)} + \underset{\text{TFT}}{(0.8\text{x}0.99\text{x}0.2)} + \underset{\text{TTF}}{(0.9\text{x}0.4\text{x}0.8)} + \underset{\text{TFF}}{0}}$$

Exercise

Consider the following example: Metastatic cancer is a possible cause of a brain tumor and is also an explanation for an increased total serum calcium. In turn, either of these could cause a patient to fall into occasional coma. Severe headache could also be explained by a brain tumor.

Exercise

a) Represent these causal links in a belief network. Let (a) stand for ‘metastatic cancer’, (b) for ‘increased total serum calcium’, (c) for ‘brain tumor’, (d) for ‘occasional coma’, and (e) for ‘severe headaches’.

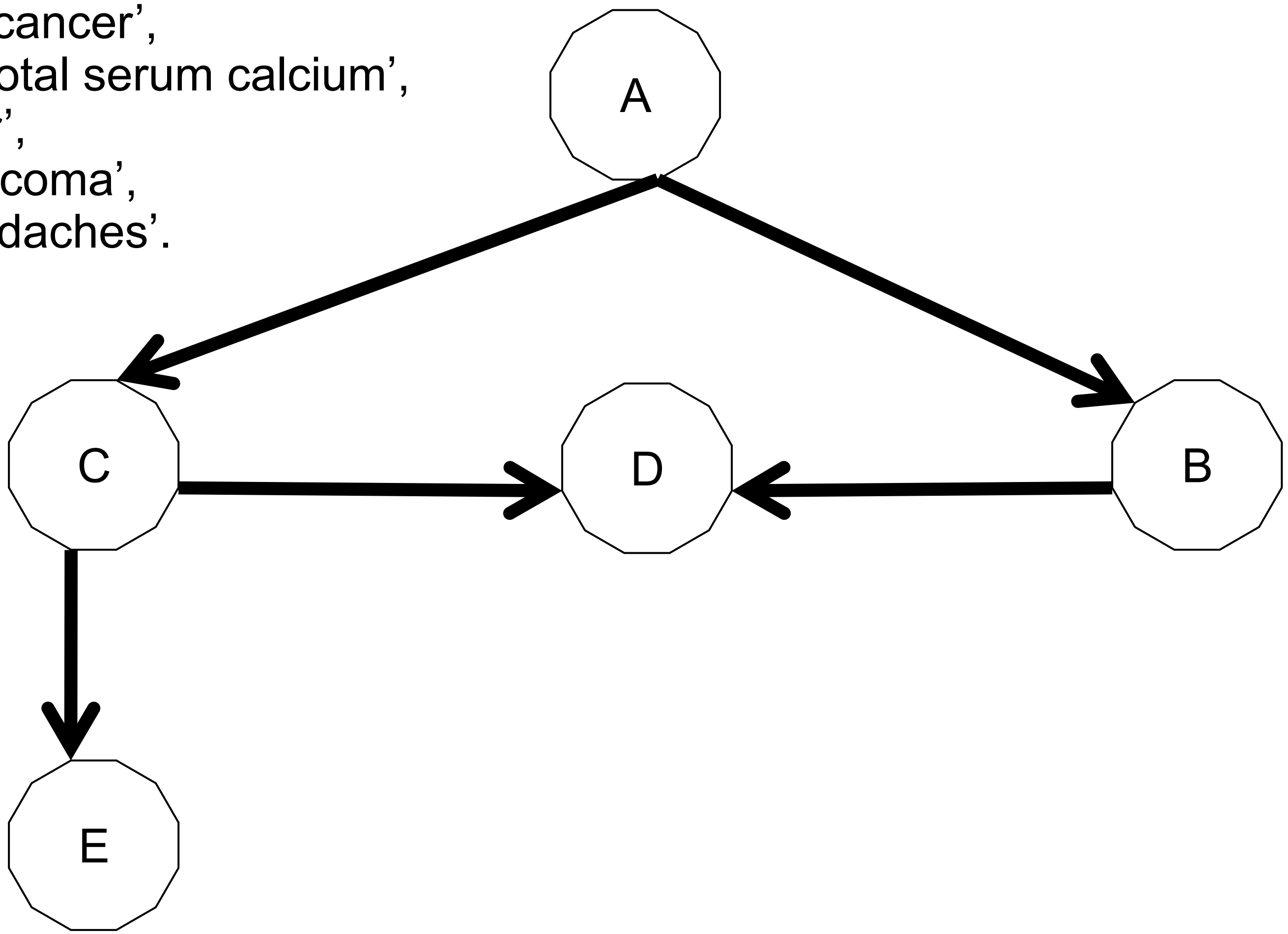
b) Give an example of an independence assumption that is implicit in this network.

c) Suppose the following probabilities are given: $Pr(a) = 0.2$, $Pr(b|a) = 0.8$, $Pr(b|\neg a) = 0.2$, $Pr(c|a) = 0.2$, $Pr(c|\neg a) = 0.05$, $Pr(e|c) = 0.8$, $Pr(e|\neg c) = 0.6$, $Pr(d|b \wedge c) = 0.8$, $Pr(d|b \wedge \neg c) = 0.8$, $Pr(d|\neg b \wedge c) = 0.8$, $Pr(d|\neg b \wedge \neg c) = 0.05$ and assume that it is also given that some patient is suffering from severe headaches but has not fallen into a coma. Calculate joint probabilities for the eight remaining possibilities (that is, according to whether a, b, and c are true or false).

d) According to the numbers given, the a priori probability that the patient has metastatic cancer is 0.2. Given that the patient is suffering from severe headaches but has not fallen into a coma, are we now more or less inclined to believe that the patient has cancer? Explain.

Answer (a)

(a) ‘metastatic cancer’,
(b) ‘increased total serum calcium’,
(c) ‘brain tumor’,
(d) ‘occasional coma’,
(e) ‘severe headaches’.



Answer (b)

Examples are:

$Pr(c \mid a \land b) = Pr(c \mid a)$, $Pr(c \mid \neg a \land b) = Pr(c \mid \neg a)$ etc.

$Pr(d \mid a \land b \land c) = Pr(d \mid b \land c)$

$Pr(e \mid a \land b \land c \land d) = Pr(e \mid c)$

Answer (c)

STEP 1
$\Pr(a \land b \land c \land \neg d \land e) =$
$\Pr(a) \cdot \Pr(b \mid a) \cdot \Pr(c \mid a \land b) \cdot \Pr(\neg d \mid a \land b \land c) \cdot \Pr(e \mid a \land b \land c \land \neg d) =$
substituting conditional probabilities using independence assumptions of the network
$\Pr(a) \cdot \Pr(b \mid a) \cdot \Pr(c \mid a) \cdot \Pr(\neg d \mid b \land c) \cdot \Pr(e \mid c) =$
(using the negation rule $\Pr(\neg d \mid b \land c) = 1 - \Pr(d \mid b \land c)$)
$\Pr(a) \cdot \Pr(b \mid a) \cdot \Pr(c \mid a) \cdot (1 - \Pr(d \mid b \land c)) \cdot \Pr(e \mid c) =$
$0.2 \cdot 0.8 \cdot 0.2 \cdot 0.2 \cdot 0.8 = 0.00512$

STEP 2
$\Pr(a \land b \land \neg c \land \neg d \land e) =$
$\Pr(a) \cdot \Pr(b \mid a) \cdot (1 - \Pr(c \mid a)) \cdot (1 - \Pr(d \mid b \land \neg c)) \cdot \Pr(e \mid \neg c) =$
$0.2 \cdot 0.8 \cdot 0.8 \cdot 0.2 \cdot 0.6 = 0.01536$

Answer (c)

STEP 3
$\Pr(a \wedge \neg b \wedge c \wedge \neg d \wedge e) =$
$\Pr(a) \cdot (1 - \Pr(b \mid a)) \cdot \Pr(c \mid a) \cdot (1 - \Pr(d \mid \neg b \wedge c)) \cdot \Pr(e \mid c) =$
$0.2 \cdot 0.2 \cdot 0.2 \cdot 0.2 \cdot 0.8 = 0.00128$

STEP 4
$\Pr(a \wedge \neg b \wedge \neg c \wedge \neg d \wedge e) =$
$\Pr(a) \cdot (1 - \Pr(b \mid a)) \cdot (1 - \Pr(c \mid a)) \cdot (1 - \Pr(d \mid \neg b \wedge \neg c)) \cdot \Pr(e \mid \neg c) =$
$0.2 \cdot 0.2 \cdot 0.8 \cdot 0.95 \cdot 0.6 = 0.01824$

STEP 5
$\Pr(\neg a \wedge b \wedge c \wedge \neg d \wedge e) =$
$(1 - \Pr(a)) \cdot \Pr(b \mid \neg a) \cdot \Pr(c \mid \neg a) \cdot (1 - \Pr(d \mid b \wedge c)) \cdot \Pr(e \mid c) =$
$0.8 \cdot 0.2 \cdot 0.05 \cdot 0.2 \cdot 0.8 = 0.00128$

Answer (c)

STEP 6
$\Pr(\neg a \wedge b \wedge \neg c \wedge \neg d \wedge e) =$
$(1 - \Pr(a)) \cdot \Pr(b \mid \neg a) \cdot (1 - \Pr(c \mid \neg a)) \cdot (1 - \Pr(d \mid b \wedge \neg c)) \cdot \Pr(e \mid \neg c) =$
$0.8 \cdot 0.2 \cdot 0.95 \cdot 0.2 \cdot 0.6 = 0.01824$

STEP 7
$\Pr(\neg a \wedge \neg b \wedge c \wedge \neg d \wedge e) =$
$(1 - \Pr(a)) \cdot (1 - \Pr(b \mid \neg a)) \cdot \Pr(c \mid \neg a) \cdot (1 - \Pr(d \mid \neg b \wedge c)) \cdot \Pr(e \mid c) =$
$0.8 \cdot 0.8 \cdot 0.05 \cdot 0.2 \cdot 0.8 = 0.00512$

STEP 8
$\Pr(\neg a \wedge \neg b \wedge \neg c \wedge \neg d \wedge e) =$
$(1 - \Pr(a)) \cdot (1 - \Pr(b \mid \neg a)) \cdot (1 - \Pr(c \mid \neg a)) \cdot (1 - \Pr(d \mid b \wedge \neg c)) \cdot \Pr(e \mid \neg c) =$
$0.8 \cdot 0.8 \cdot 0.95 \cdot 0.95 \cdot 0.6 = 0.34656$

Answer (d)

We are asked whether $Pr(a \mid \neg d \wedge e)$ is greater or smaller than $Pr(a)$.

$Pr(a \mid \neg d \wedge e) = Pr(a \wedge \neg d \wedge e)/Pr(\neg d \wedge e)$ (conditional probability definition). We need to compute $Pr(a \wedge \neg d \wedge e)$ and $Pr(\neg d \wedge e)$, and to do that we use the probabilities we computed above. They describe all 8 possible states of the world given that $\neg d$ and e are true, and they are all disjoint. We are using $Pr(X) = Pr(X \wedge Y) + Pr(X \wedge \neg Y)$, or that the probability of the union of disjoint events equals to the sum of probabilities of those events.

So $Pr(a \wedge \neg d \wedge e) = Pr(a \wedge b \wedge c \wedge \neg d \wedge e) + Pr(a \wedge b \wedge \neg c \wedge \neg d \wedge e) + Pr(a \wedge \neg b \wedge c \wedge \neg d \wedge e) + Pr(a \wedge \neg b \wedge \neg c \wedge \neg d \wedge e)$ and $Pr(\neg d \wedge e)$ is the sum of all 8 numbers above.

$Pr(a \wedge \neg d \wedge e) = 0.04$

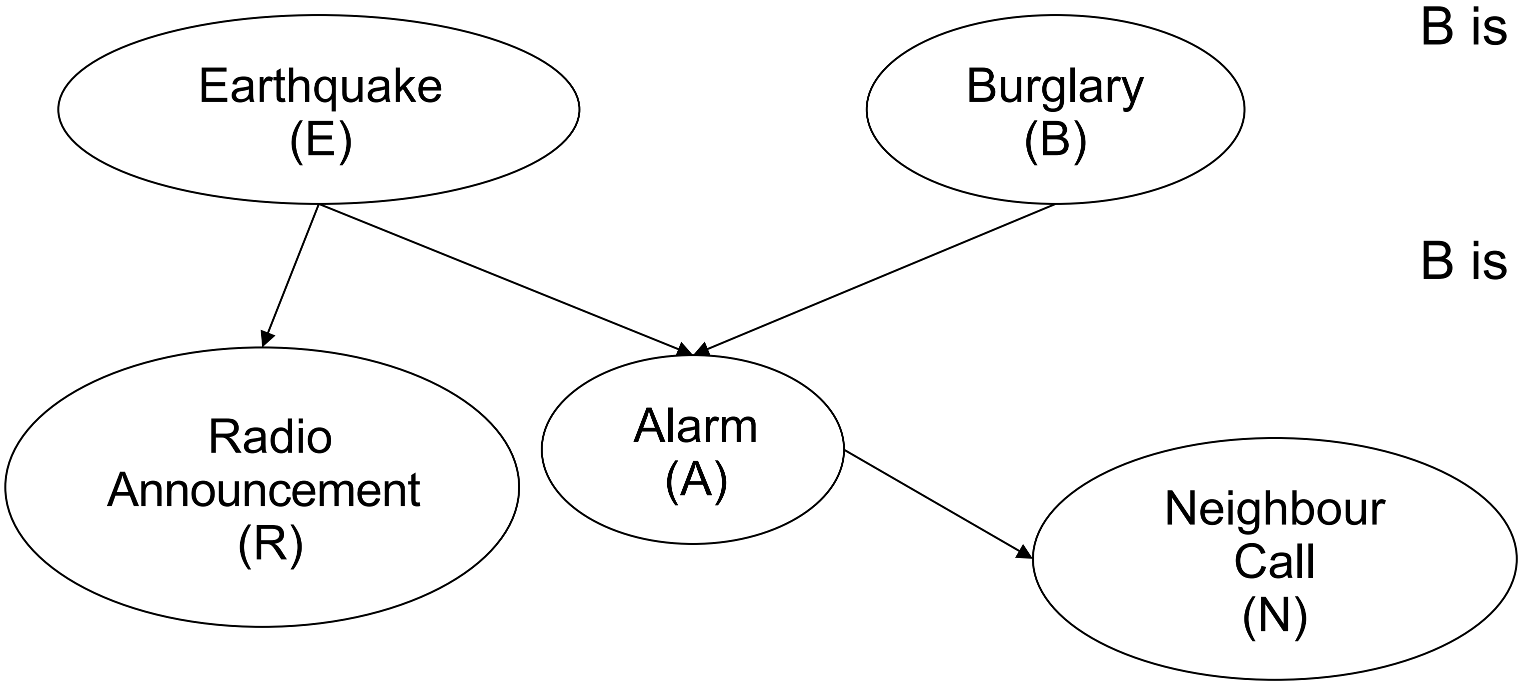
$Pr(\neg d \wedge e) = 0.04 + 0.00128 + 0.01824 + 0.00512 + 0.34656 = 0.4112$

$Pr(a \mid \neg d \wedge e) = 0.04/0.4112$ which is approximately 0.1. So the probability got smaller.

Bayesian Networks second example

The joint probability

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



A is independent from R

B is independent from R

Given E

B is Independent from E