INDEPENDENCE AND BAYESIAN NETWORKS

Today

- Reading
  - AIMA Ch. 14.1-14.4 (Bayesian Networks)
- Goals
  - Independence
  - Bayesian networks

Test Details

- Next Wednesday October 15th
- In-class, closed book, closed notes
- Aiming for 30-50 minute exam
- Types of questions
  - Short answer
  - True/False
  - Carry out an algorithm by hand
- List of topics posted on Piazza

Inference Practice

Step One: select the entries in the table consistent with the evidence (this becomes our world)

Step Two: sum over the H variables to get the joint distribution of the query and evidence variables

Step Three: Normalize the joint to get the conditional

Cavity	Catch	ToothAche	P
yes	yes	yes	0.16
yes	yes	no	0.08
yes	no	yes	0.01
yes	no	no	0.01
no	yes	yes	0.02
no	yes	no	0.16
no	no	yes	0.06
no	no	no	0.5

Queries:
p(Catch | Toothache = false)
p(Toothache | Cavity = true)
p(Cavity, Catch)
p(Cavity | Toothache = false, Catch = true)

Notions of Independence

Lecture proceeds on the whiteboard!

Checking independence

- Are X and Y independent?

p(X,Y)		X		
		1	2	3
Y	1	0	1/6	1/12
	2	1/5	1/9	0
	3	2/15	1/4	1/18

Moving away from numerical quantities

"The traditional definition of independence uses equality of numerical quantities, as in

$$p(x, y) = p(x)p(y)$$

suggesting that one must test whether the joint distribution of X and Y is equal to the product of their marginals in order to determine whether X and Y are independent. By contrast people can easily and confidently detect dependencies, even though they may not be able to provide precise numerical estimates of probabilities. A person who is reluctant to estimate the probability of being burglarized the next day or of having a nuclear war within five years can nevertheless state with ease whether the two events are dependent, namely, whether knowing the truth of one proposition will alter the belief of the other."

- Judea Pearl

Moving away from numerical quantities

"It is usually easy for a domain expert to decide what direct influences exist in the domain – much easier, in fact, than actually specifying the probabilities themselves"

- Humans can "easily and confidently" detect dependencies
- Move away from numerical representation of the joint distribution (or the conditional distributions) to a representation that encodes dependencies

Probabilistic Inference

Joint distribution → Answer any query but exponential in the number of variables

↓

(Conditional) independence → Reduces the space requirements but necessitates numerical checking

↓

??? → Reduces the space requirements and uses a human's intuitive ability to detect dependencies rather than numerical checking

Probabilistic Inference

Joint distribution → Answer any query but exponential in the number of variables

↓

(Conditional) independence → Reduces the space requirements but necessitates numerical checking

↓

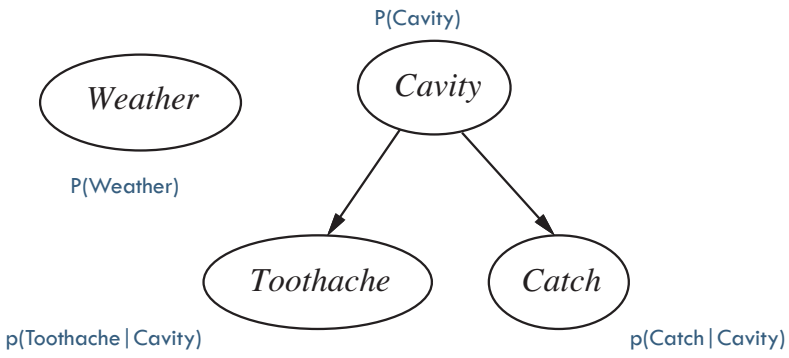
Graphical Models → Reduces the space requirements and uses a human's intuitive ability to detect dependencies rather than numerical checking

Bayesian Networks

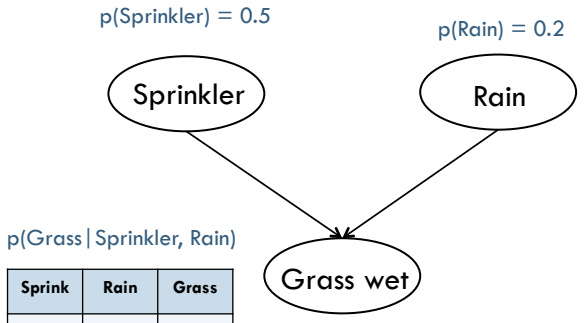
- A Bayesian Network is a directed acyclic graph that represents the conditional independencies of a set of random variables
  - Each r.v. corresponds to a node
  - A directed edge represents a direct influence
  - The conditional distribution of each node given its parents must be explicitly specified

Bayesian Network Examples

- Weather = {rainy, sunny, cloudy, snowy}
- Cavity = { yes, no}
- Toothache = {yes, no}
- Catch = {yes, no}



Bayesian Network Examples



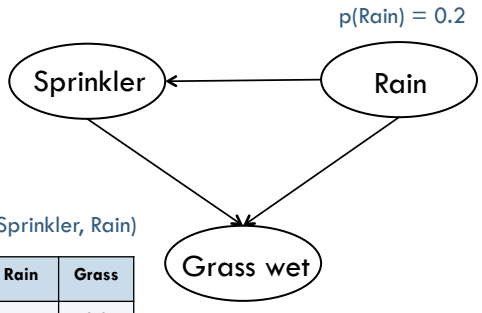
p(Grass | Sprinkler, Rain)

Sprink	Rain	Grass
F	F	0.0
F	T	0.8
T	F	0.9
T	T	0.99

Bayesian Network Examples

p(Sprinkler | Rain)

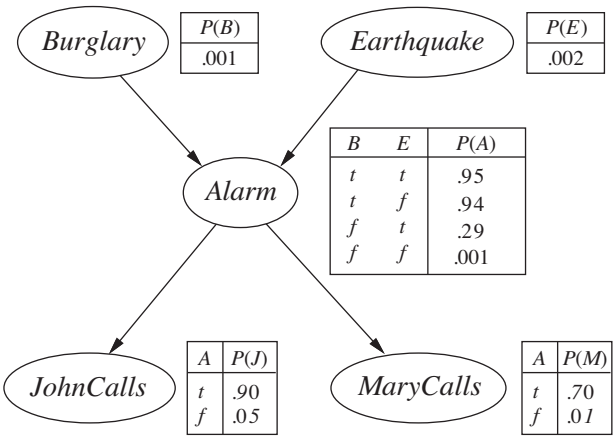
Rain	Sprink
F	0.5
T	0.01



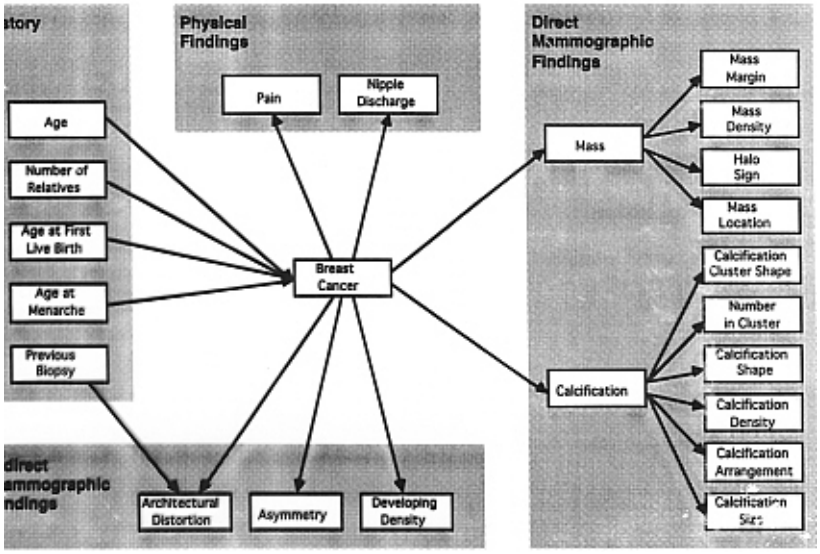
p(Grass | Sprinkler, Rain)

Sprink	Rain	Grass
F	F	0.0
F	T	0.8
T	F	0.9
T	T	0.99

Bayesian Network Examples

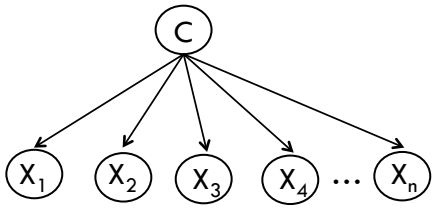


Example: MammoNet



Commonly used Bayesian Networks

- Naïve (Idiot) Bayes Classifier
  - Commonly used for text classification
  - C is the class (topic or label) of the document
  - The X variables represent the words in the document



Today an earthquake occurred in Southern California…

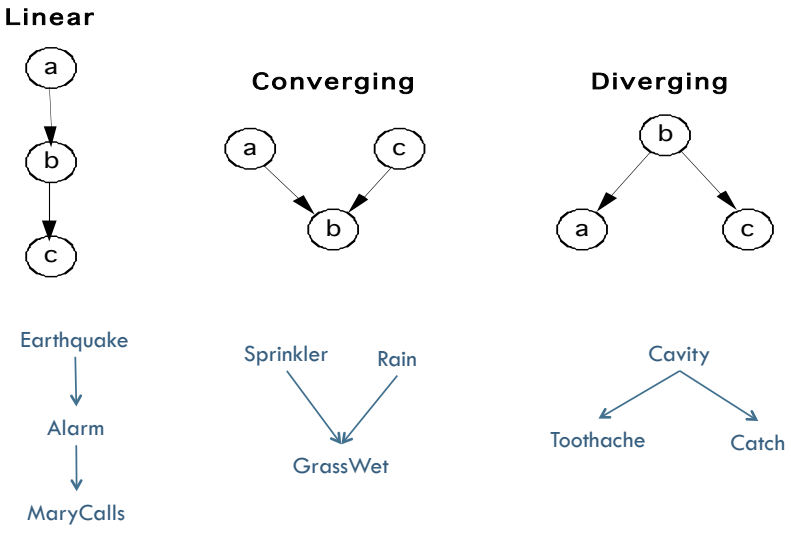
Representing the joint using a BN

- Given a BN over a set of random variables, the joint distribution can be factor as

$$p(x_1, \ldots, x_N) = \prod_{i=1}^{N} p(x_i | \text{parents}(x_i))$$

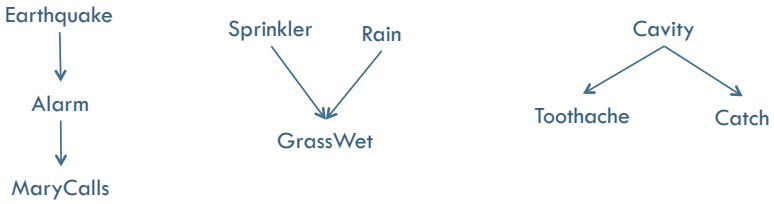
Independence in BNs

Three Types of Connections



Connection patterns and independence

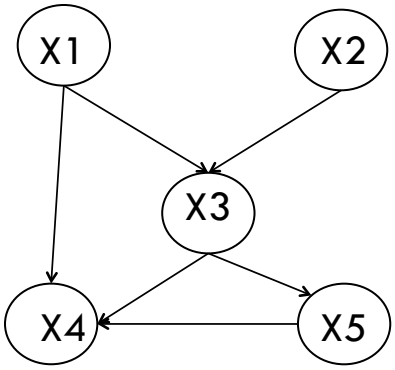
- **Linear connection:** The two end variables are dependent on each other. The middle variable renders them independent.
- **Converging connection:** The two end variables are independent of each other. The middle variable renders them dependent.
- **Divergent connection:** The two end variables are dependent on each other. The middle variable renders them independent.



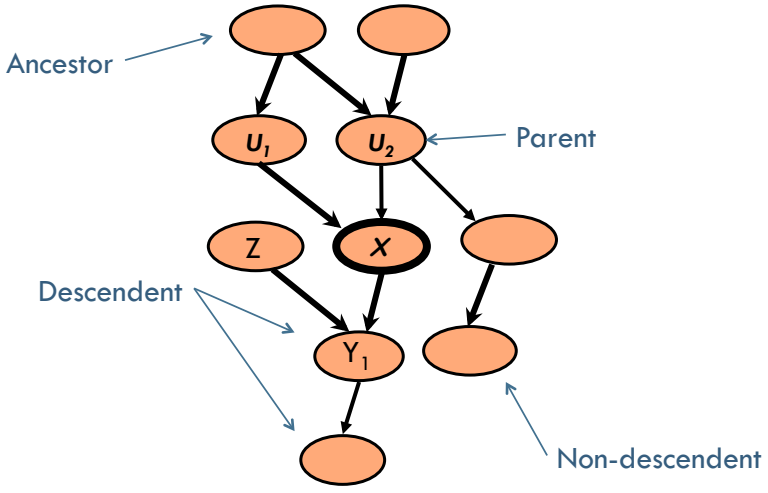
D-Separation

- Algorithm to determine independencies in BN
- Query: Are two variables X_i and X_j independent?
- Check all paths between X_i and X_j
  - If all paths are blocked, then independent
  - If any path is not blocked then not independent

List the independencies in the following Bayesian Network

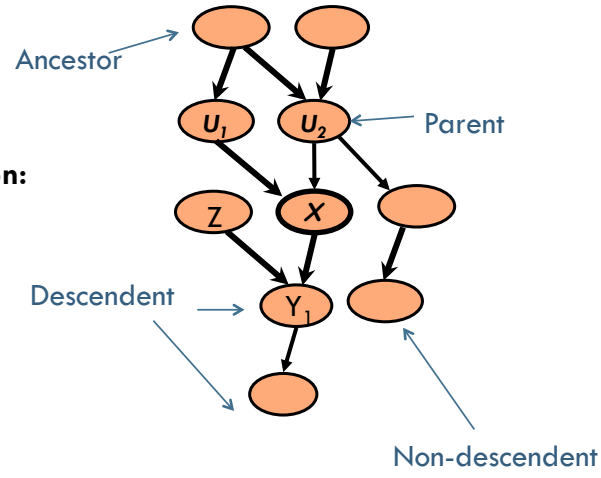


Bayesian Networks terminology



Independence assumptions encoded in the Bayesian Network

Local Markov Assumption:
A node X is independent of its non-descendants given its parents



Independence assumptions encoded in the Bayesian Network

Markov Blanket:
A node X is conditionally independent of all other nodes given its parents, children, and children's parents