Probabilistic Graphical Models

10-708

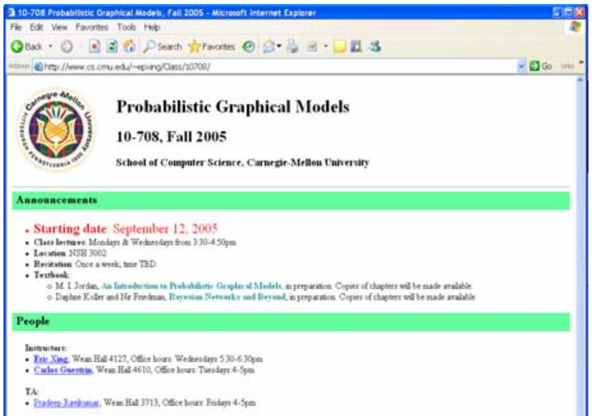
Eric Xing and Carlos Guestrin

Lecture 1

Sep 12, 2005

Logistics

- Class webpage:
  - http://www.cs.cmu.edu/~epxing/Class/10708/
  - http://www.cs.cmu.edu/~guestrin/Class/10708/



Logistics

- Bi-weekly homework: 50% of grade
  - Theory exercises
  - Implementation exercises
- Final project: 30% of grade
  - Applying PGM to your research area
    - NLP, IR, Computational Biology, vision, graphics …
  - Theoretical and/or algorithmic work
    - a more efficient approximate inference algorithm
    - a new sampling scheme for a non-trivial model …
- Take home final: 20% of grade
  - Theory exercises and/or analysis
- Policies …

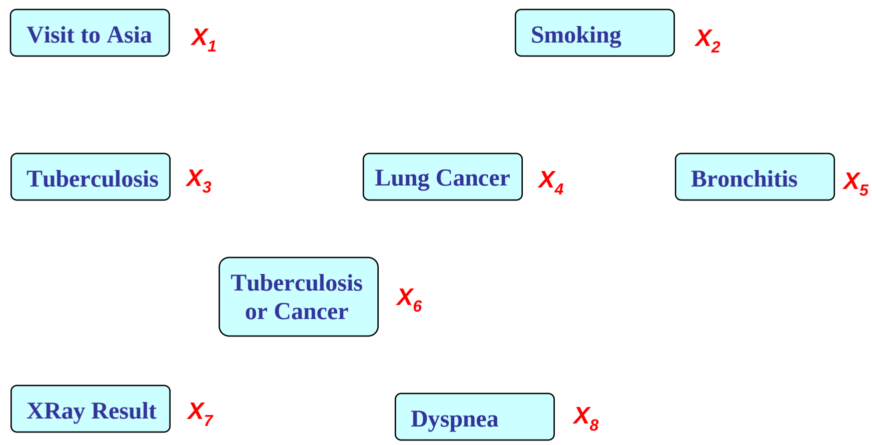
Logistics

- No formal text book, but draft chapters will be handed out in class:
  - M. I. Jordan, **An Introduction to Probabilistic Graphical Models**
  - Daphne Koller and Nir Friedman, **Bayesian Networks and Beyond**
- Mailing Lists:
  - To contact the instructors: 10708-instr@cs.cmu.edu
  - Class announcements list: 10708-announce@cs.cmu.edu.
- TA:
  - Pradeep Ravikumar, Wean Hall 3713, Office hours: Fridays 4-5pm
- Class Assistant:
  - Monica Hopes, Wean Hall 4616, x8-5527

What is a graphical model?

--- example from medical diagnostics

- A possible world for a patient with lung problem:



Recap of Basic Prob. Concepts

- Representation: what is the joint probability dist. on multiple variables?

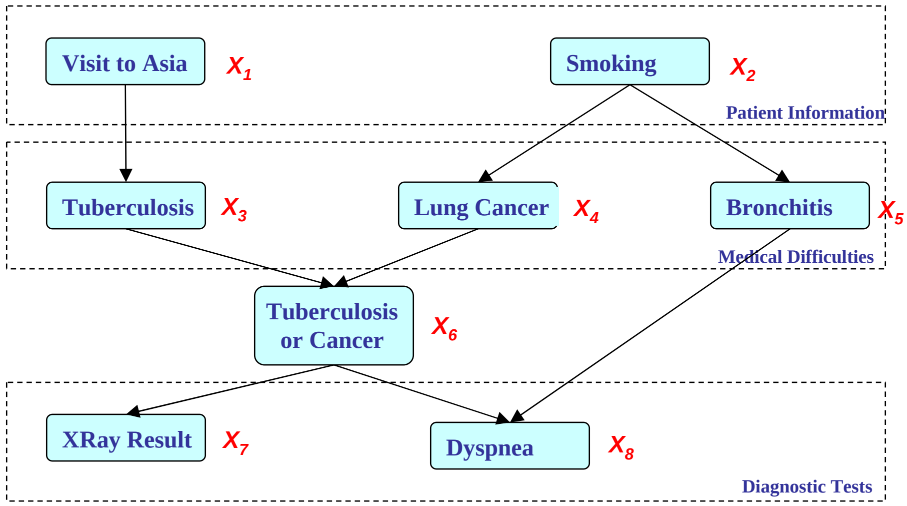
$$P(X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8,)$$

  - How many state configurations in total? --- 2^8
  - Are they all needed to be represented?
  - Do we get any scientific/medical insight?

- Learning: where do we get all this probabilities?
  - Maximal-likelihood estimation? but how many data do we need?
  - Where do we put domain knowledge in terms of plausible relationships between variables, and plausible values of the probabilities?

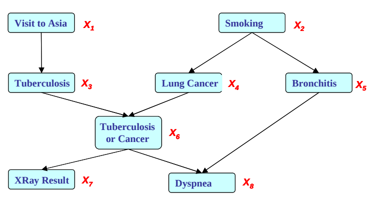
- Inference: If not all variables are observable, how to compute the conditional distribution of latent variables given evidence?

Dependencies among variables



Probabilistic Graphical Models

- If X_i's are **conditionally independent** (as described by a **PGM**), the joint can be factored to a product of simpler terms, e.g.,



$$P(X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8)$$
$$= P(X_1)\, P(X_2)\, P(X_3| X_1)\, P(X_4| X_2)\, P(X_5| X_2) P(X_6| X_3, X_4)\, P(X_7| X_6)\, P(X_8| X_5, X_6)$$

- Why we may favor a PGM?
  - Representation cost: how many probability statements are needed?
    2+2+4+4+4+8+4+8=36, an 8-fold reduction from 2^8!
  - Algorithms for systematic and efficient inference/learning computation
    - Exploring the graph structure and probabilistic (e.g., Bayesian, Markovian) semantics

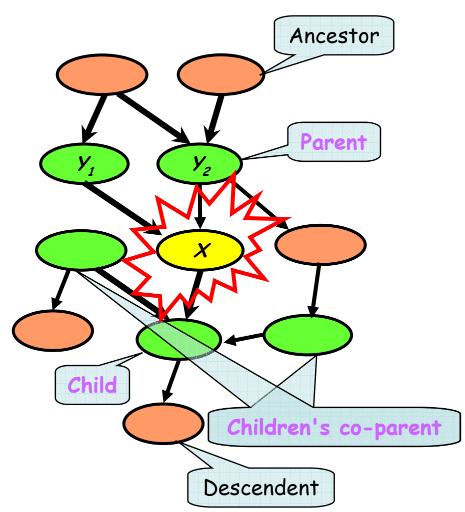
Two types of GMs

- Directed edges give causality relationships (**Bayesian Network** or **Directed Graphical Model**):
- Undirected edges simply give correlations between variables (**Markov Random Field** or **Undirected Graphical model**):

Bayesian Network

Structure: ***DAG***

- Meaning: a node is **conditionally independent** of every other node in the network outside its **Markov blanket**
- Local conditional distributions (**CPD**) and the **DAG** completely determine the **joint** dist.
- Give **causality** relationships

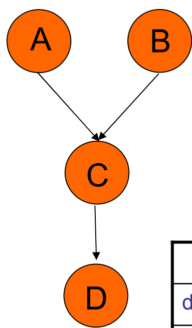


Conditional probability tables (CPTs)

a^0	0.75
a^1	0.25

b^0	0.33
b^1	0.67

P(a,b,c.d) =
P(a)P(b)P(c|a,b)P(d|c)



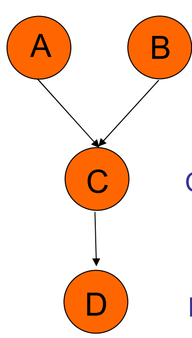
	a^0b^0	a^0b^1	a^1b^0	a^1b^1
c^0	0.45	1	0.9	0.7
c^1	0.55	0	0.1	0.3

	c^0	c^1
d^0	0.3	0.5
d^1	07	0.5

Conditional probability density func. (CPDs)

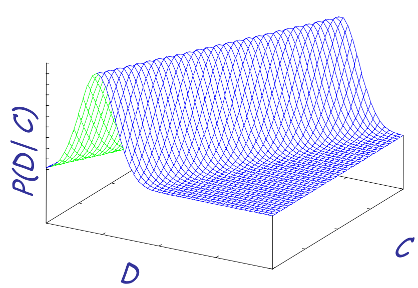
$A \sim N(\mu_a, \Sigma_a)$ $B \sim N(\mu_b, \Sigma_b)$

P(a,b,c.d) =
P(a)P(b)P(c|a,b)P(d|c)



$C \sim N(A+B, \Sigma_c)$

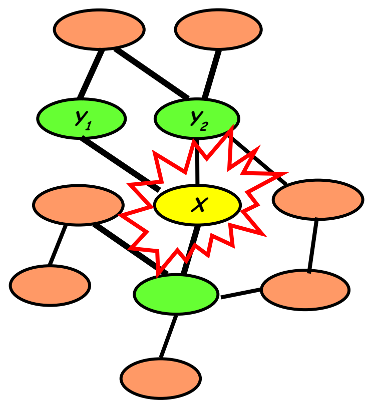
$D \sim N(\mu_a+C, \Sigma_a)$



Markov Random Fields

Structure: an
undirected graph

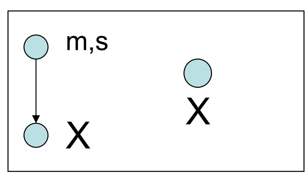
- Meaning: a node is **conditionally independent** of every other node in the network given its **direct neighbors**
- Local contingency functions (**potentials**) and the **cliques** in the graph completely determine the **joint** dist.
- Give **correlations** between variables



Simplest GMs: the building blocks

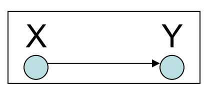
Density estimation

Parametric and nonparametric methods



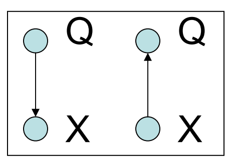
Regression

Linear, conditional mixture, nonparametric

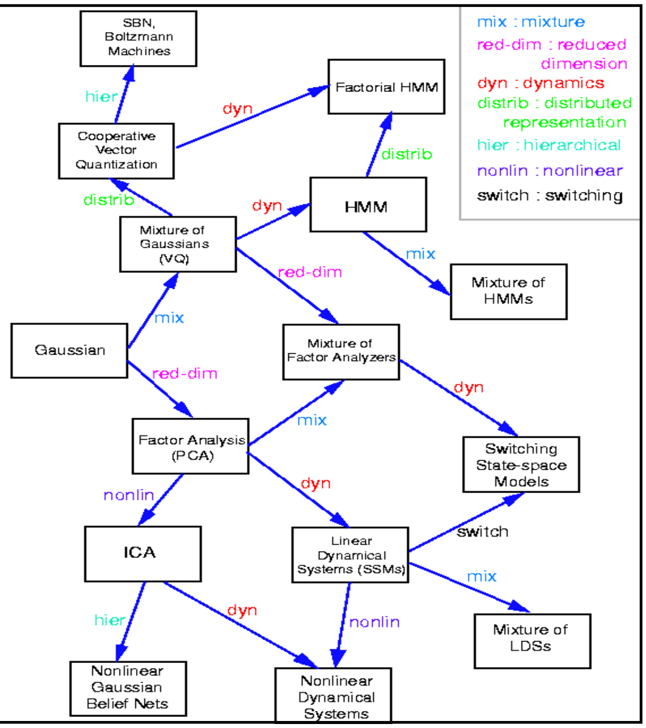


Classification

Generative and discriminative approach



An (incomplete) genealogy of graphical models



(Picture by Zoubin Ghahramani and Sam Roweis)

Probabilistic Inference

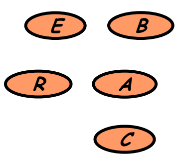
- Computing statistical queries regarding the network, e.g.:
  - Is node X independent on node Y given nodes Z,W ?
  - What is the probability of X=true if (Y=false and Z=true)?
  - What is the joint distribution of (X,Y) if R=false?
  - What is the likelihood of some full assignment?
  - What is the most likely assignment of values to all or a subset the nodes of the network?

- General purpose algorithms exist to fully automate such computation
  - Computational cost depends on the topology of the network
  - Exact inference:
    - The junction tree algorithm
  - Approximate inference;
    - Loopy belief propagation, variational inference, Monte Carlo sampling

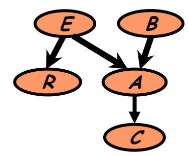
Learning Graphical Models

The goal:

Given set of independent samples (***assignments*** of random variables), find the ***best*** (the most likely?) Bayesian Network (both DAG and CPDs)



(B,E,A,C,R)=(T,F,F,T,F)
(B,E,A,C,R)=(T,F,T,T,F)
……..
(B,E,A,C,R)=(F,T,T,T,F)



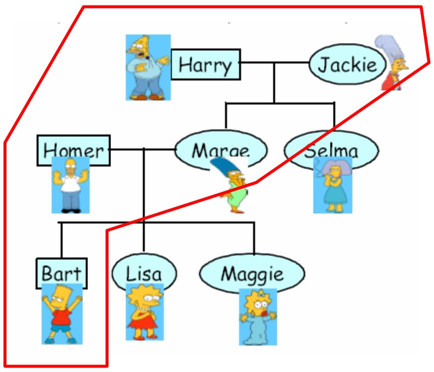
E	B	P(A | E,B)	
e	b	0.9	0.1
e	$\bar{b}$	0.2	0.8
$\bar{e}$	b	0.9	0.1
$\bar{e}$	$\bar{b}$	0.01	0.99

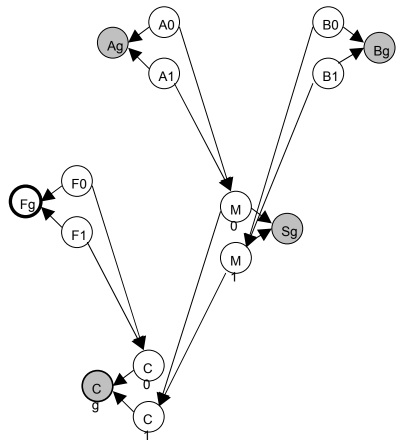
Application of GMs

- Machine Learning
- Computational statistics

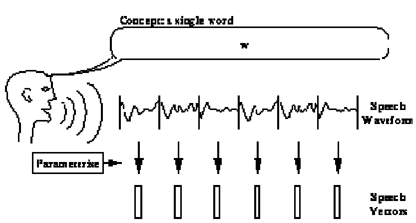
- Computer vision and graphics
- Natural language processing
- Informational retrieval
- Robotic control
- Decision making under uncertainty
- Error-control codes
- Computational biology
- Genetics and medical diagnosis/prognosis
- Finance and economics
- Etc.

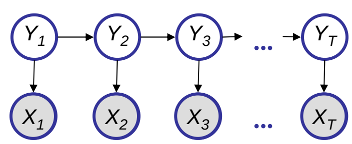
Genetic Pedigree





Speech recognition

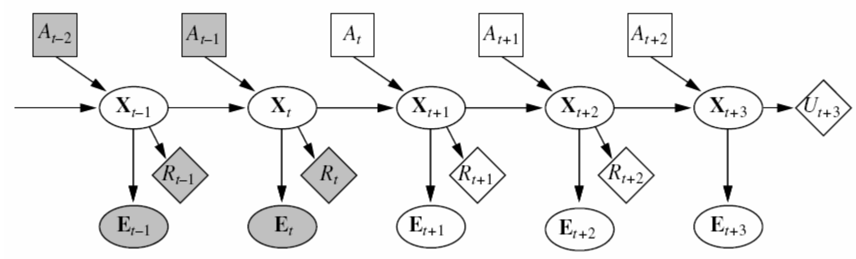




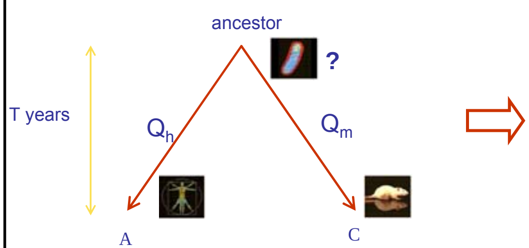
Hidden Markov Model

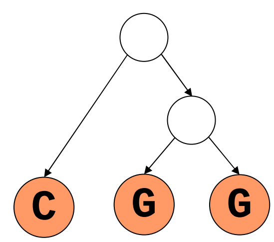
Reinforcement learning

- Partially observed Markov decision processes (POMDP)

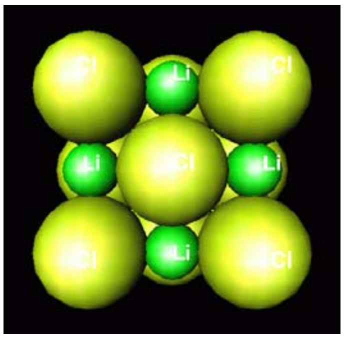


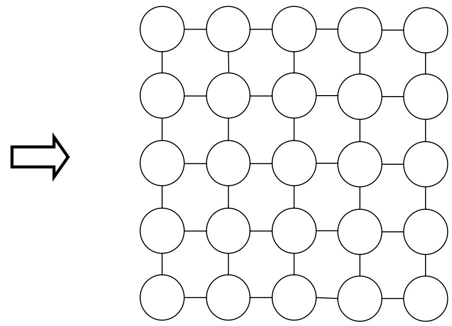
Evolution





Solid State Physics





Ising/Potts model

Why graphical models

- **Probability theory** provides the **glue** whereby the parts are combined, ensuring that the system as a whole is consistent, and providing ways to interface models to data.
- The **graph theoretic** side of graphical models provides both an intuitively appealing interface by which humans can model highly-interacting sets of variables as well as a data structure that lends itself naturally to the design of efficient general-purpose algorithms.
- **Many of the classical multivariate probabilistic systems** studied in fields such as statistics, systems engineering, information theory, pattern recognition and statistical mechanics **are special cases of the general graphical model formalism**
  - -- examples include mixture models, factor analysis, hidden Markov models, Kalman filters and Ising models.
- The graphical model framework provides a way to view all of these systems as instances of a **common underlying formalism**.

--- M. Jordan

Plan for the Class

- Bayesian Network and Markov Random Fields
  - Representation
  - Inference
  - learning
- Approximate inference
  - Monte Carlo algorithms
  - Vatiational methods
- Continuous and Hybrid models, exponential family, GLIM
- Temporal models
  - HMM and Kalman Filtering
  - Dynamic Bayesian networks
- Advanced topics
  - Probabilistic relational models
  - Applications
  - Causal learning
  - Decision making