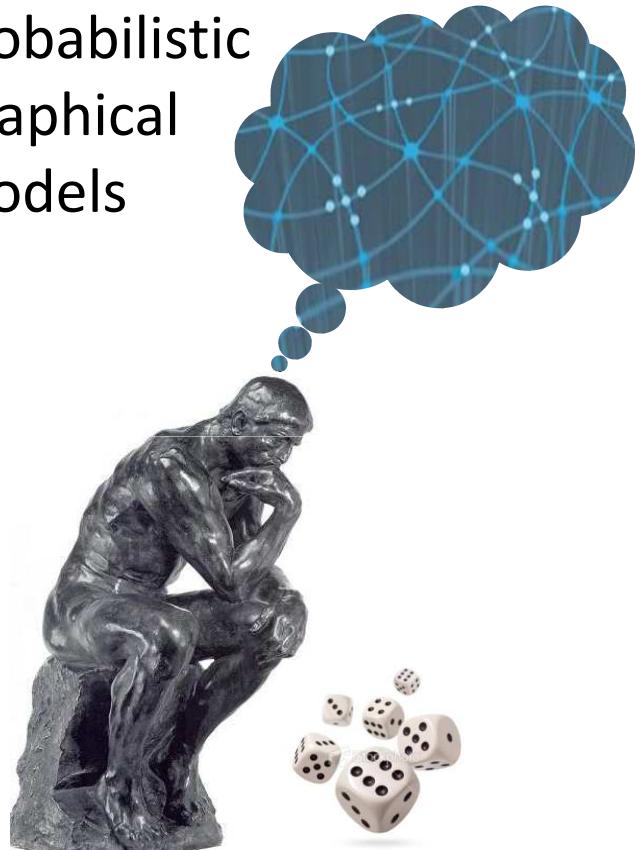
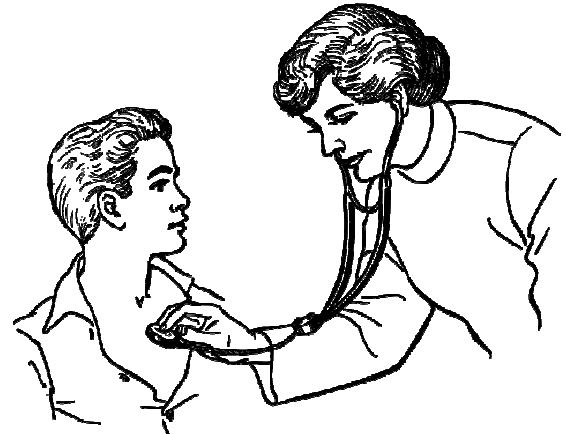


Probabilistic
Graphical
Models

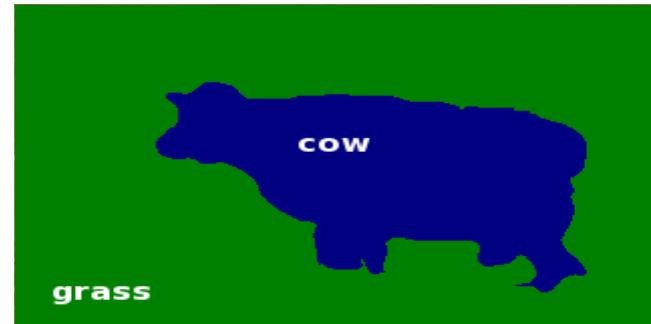


Introduction

Motivation and Overview



predisposing factors
symptoms
test results
diseases
treatment outcomes

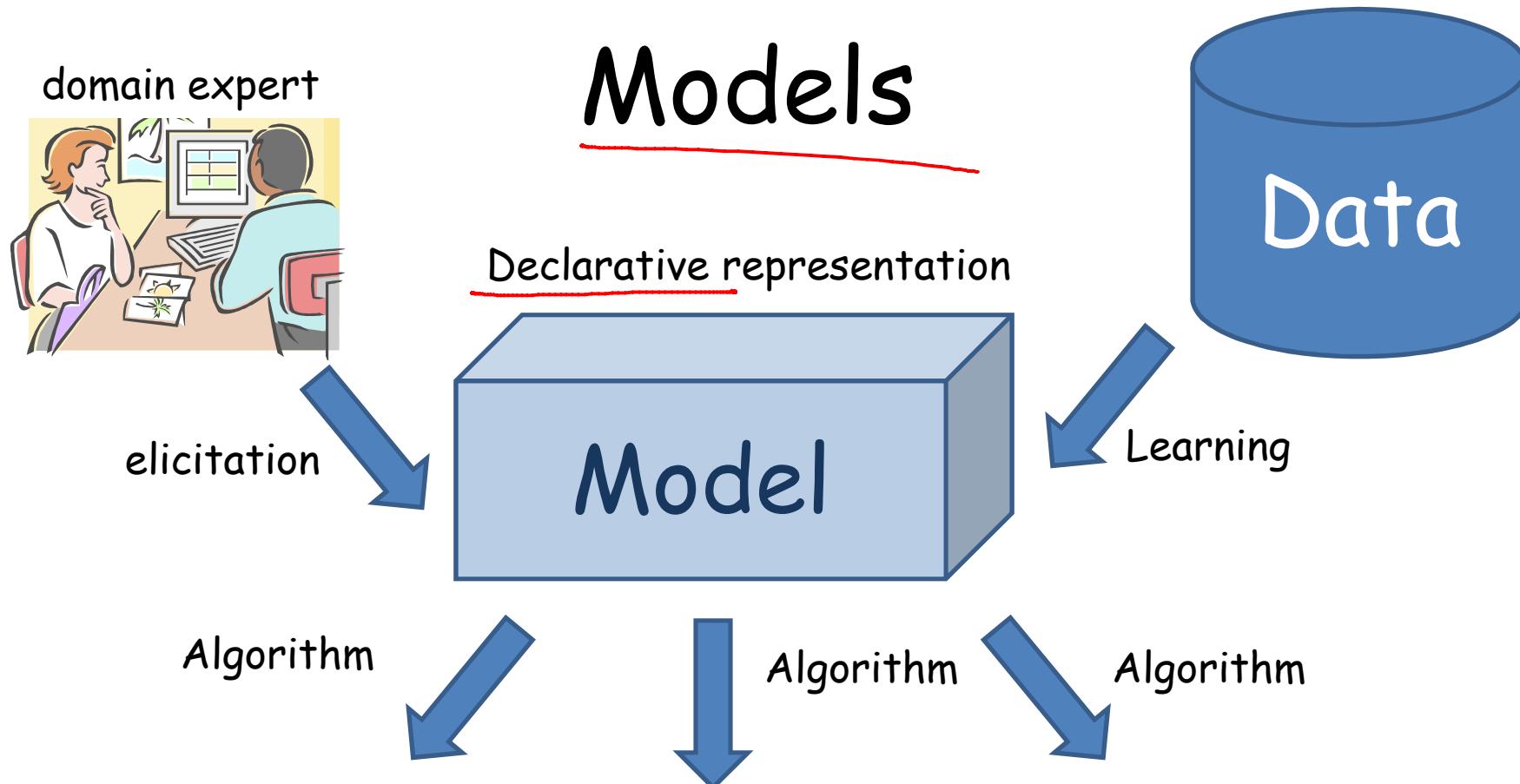


millions of pixels or
thousands of superpixels

each needs to be labeled
{grass, sky, water, cow, horse, ...}

Probabilistic Graphical Models

Daphne Koller



Uncertainty

- Partial knowledge of state of the world
- Noisy observations
- Phenomena not covered by our model
- Inherent stochasticity

Probability Theory

- Declarative representation with clear semantics
- Powerful reasoning patterns *conditioning
decision making*
- Established learning methods

Complex Systems

predisposing factors

symptoms

test results

diseases

treatment outcomes

class labels for
thousands of superpixels

Random variables

$$X_1, \dots, X_n$$

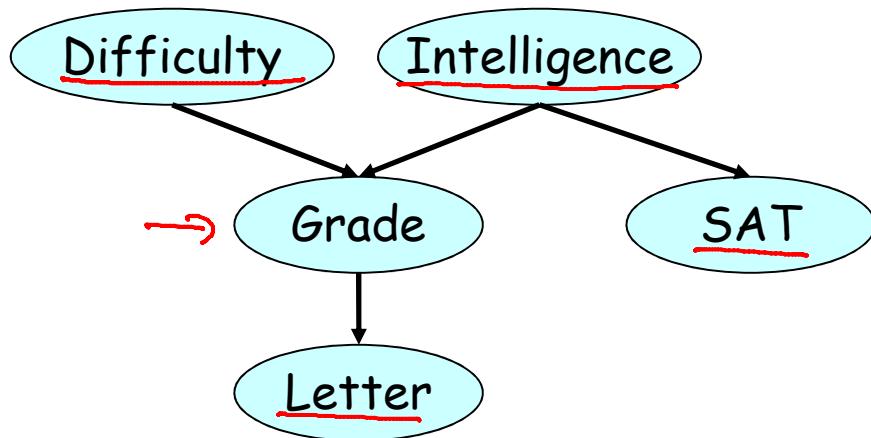
~ binary valued
distribution
over 2^n
possible states

Joint distribution $P(X_1, \dots, X_n)$

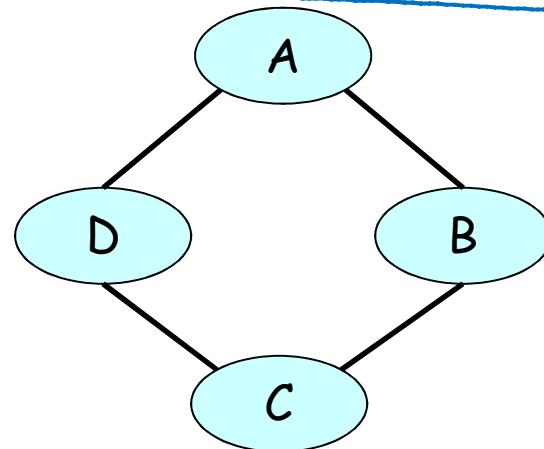
$x_1 \dots x_n$ —
nodes

Graphical Models

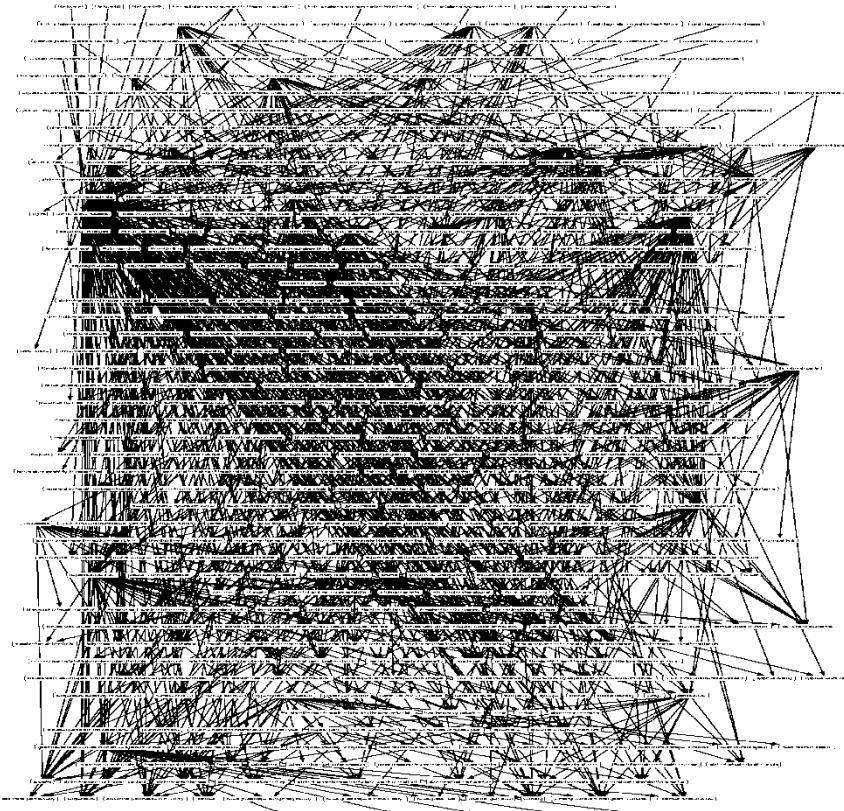
directed graph
Bayesian networks



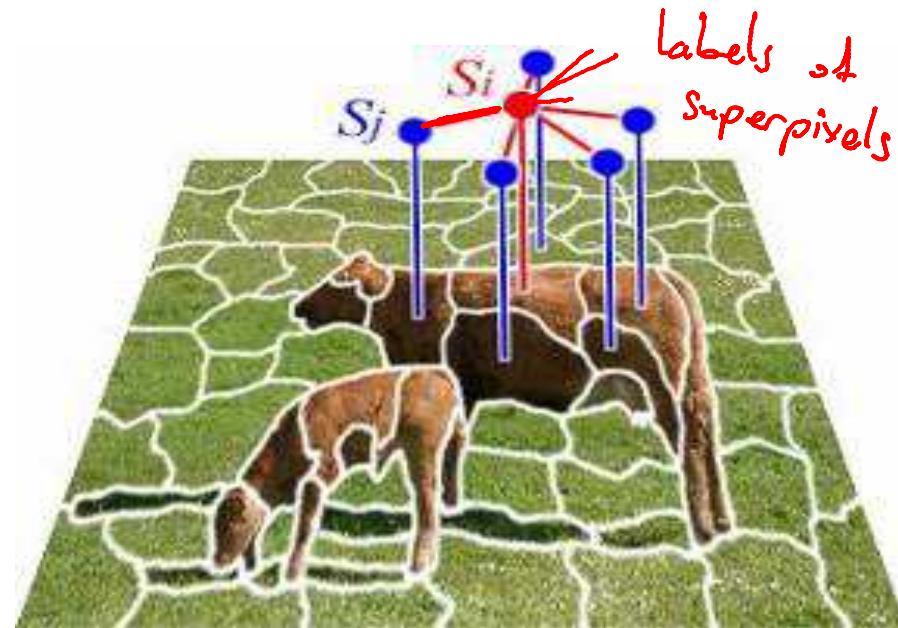
undirected graph
Markov networks



diagnosis
CPGS



Graphical Models



Daphne Koller

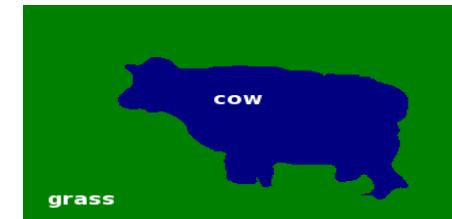
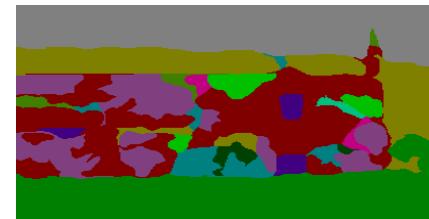
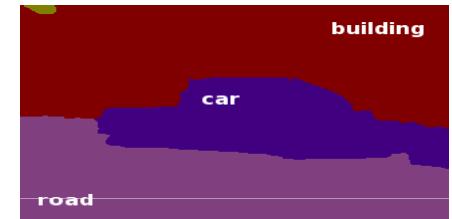
Graphical Representation

- Intuitive & compact data structure
- Efficient reasoning using general-purpose algorithms
- Sparse parameterization
 - feasible elicitation \leftarrow by hand
 - learning from data \leftarrow automatically

Many Applications

- Medical diagnosis
- Fault diagnosis
- Natural language processing
- Traffic analysis
- Social network models
- Message decoding
- Computer vision
 - Image segmentation
 - 3D reconstruction
 - Holistic scene analysis
- Speech recognition
- Robot localization & mapping

Image Segmentation



superpixels

machine learning
to separate superpixels

Daphne Koller

Medical Diagnosis

Applet started

MS on □ □ X C ON STAGE ESSENTIALS COMMUNICATE FIND ✓ ?

OnParenting May 14 - May 20, 1997 Fidelity Investments® Fidelity Distributors Corporation Our home on the web [is where] click here

cover contents news experts fun handbook talk find help feedback

There are two ways to search for specific information in **OnParenting**. In **Find by Word**, type the word(s) you want to find and get a list of titles relevant to that word. **Find by Symptom** will help you get information about children's symptoms. [Help](#) has tips to target your search.

Describe the child
in the drop-down boxes at the right. Relevant information will appear below.

Age: Toddler Sex: Female
Complaint: Abdominal pain

Localized pain: Can the child localize, or point to, the site of the pain?
 No, unable to localize
 Below the navel to the child's left
 Above the child's navel
 Either of the child's sides
 Below the navel to the child's right
 Above the navel to the child's right
 Above the navel to the child's left
 Don't Know

Results so far

Disorder	Relevance
Viral gastroenteritis	High
Psychosomatic pain	Medium
Urinary tract infection	Low
Other	Low

Start Over Review
Next>> Finish

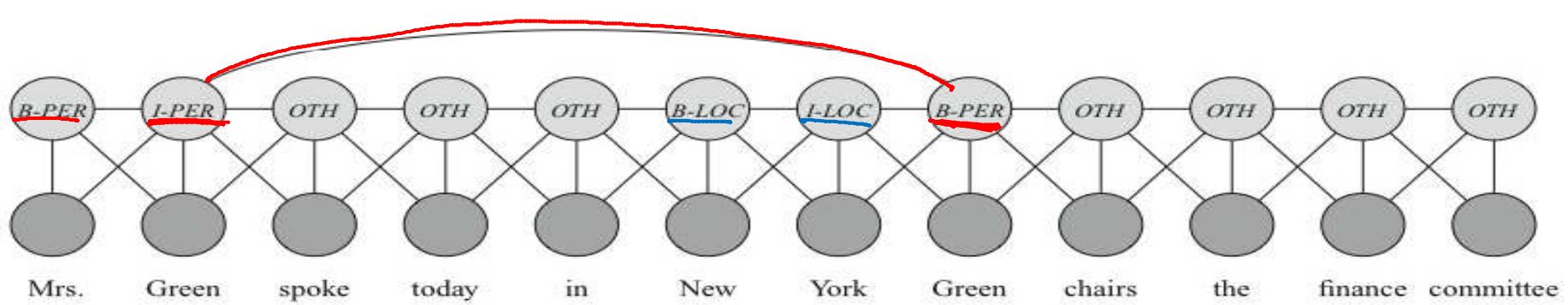


Daphne Koller

Textual Information Extraction

Mrs. Green spoke today in New York. Green chairs the finance committee.

Person Location Person Organization

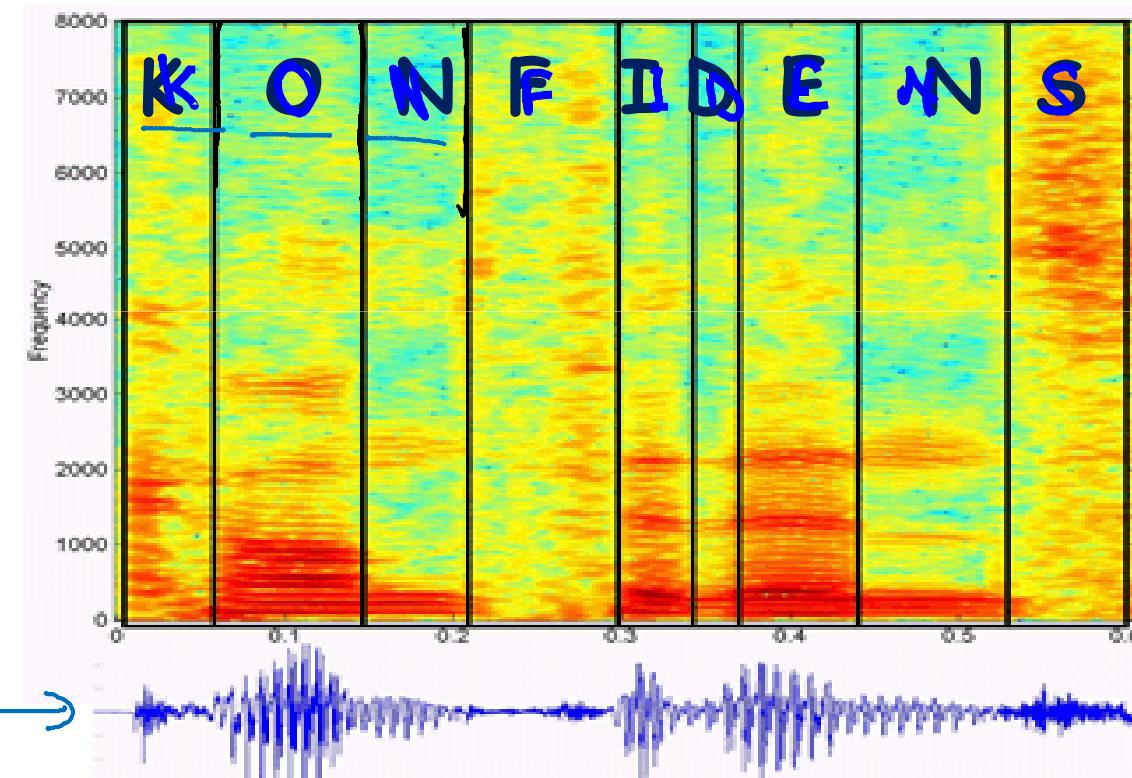


Multi-Sensor Integration: Traffic

Live Search Maps
http://maps.live.com/#/nEseXaud2fzaGiuZ3rvbtkTyU3ZXNzdC4wJTdIcGcuMSzIyJ0D0545NTvMtk5MD2NIIN2UtNkuNzg1MTU2MjUIN2UyMC44
copy to clipboard firebox
Web to PDF
Live Search Maps
Live Search | MSN | Windows Live
Live Search | Maps
Businesses People Collections Locations Web
Washington dc
2D 3D Road Aerial Bird's eye Calibration Tools Traffic
Washington, D.C.
I95 corridor experiment: accurate to ±5 MPH in 85% of cases
Fielded in 72 cities
Thanks to: Eric Horvitz, Microsoft Research

- I95 corridor experiment: accurate to ± 5 MPH in 85% of cases
- Fielded in 72 cities

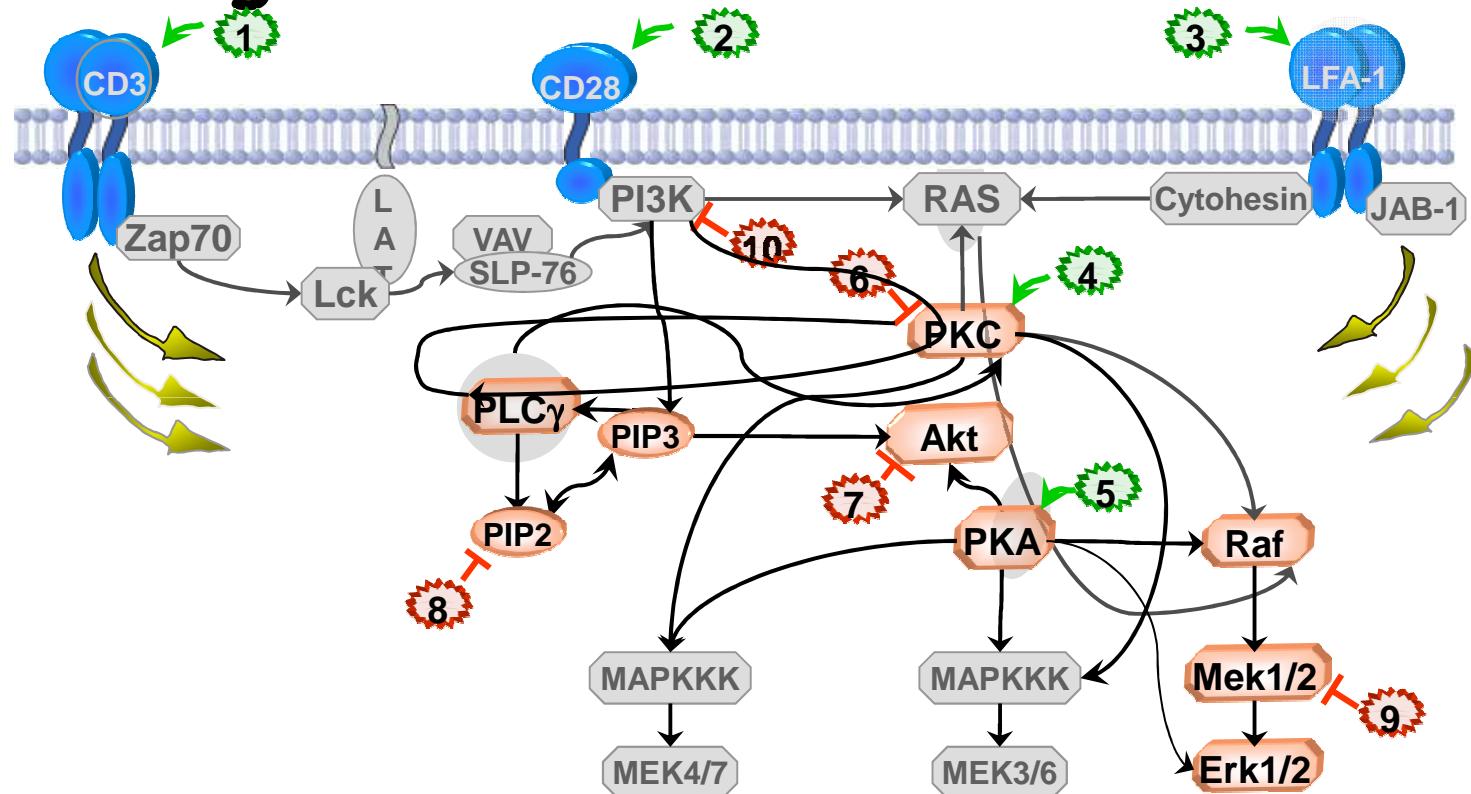
Speech Recognition



[Source: <http://www-g.eng.cam.ac.uk/125/now/speech2.html>]

Daphne Koller

Biological Network Reconstruction



Causal protein-signaling networks derived
from multiparameter single-cell data

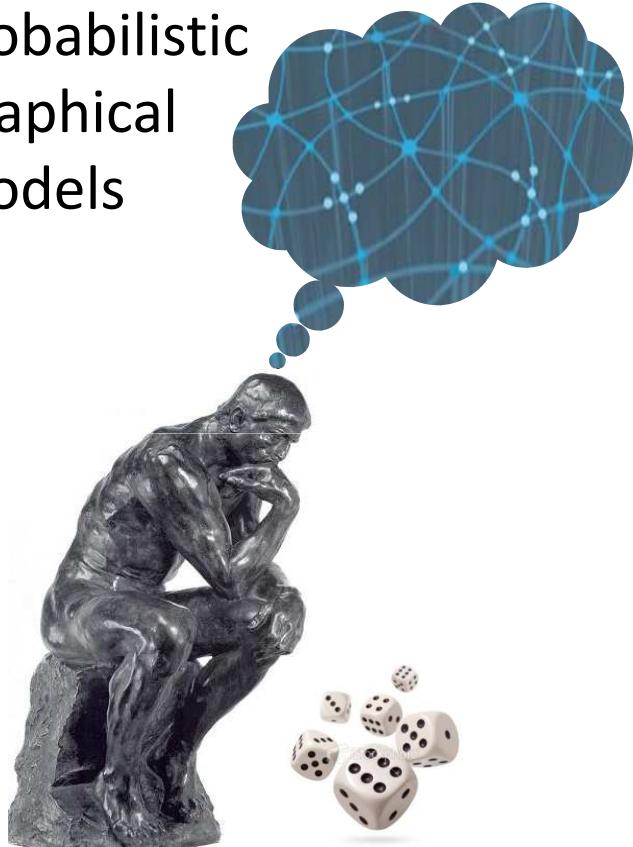
Sachs et al., *Science* 2005

Daphne Koller

Overview

- Representation
 - Directed and undirected
 - Temporal and plate models
- Inference *reasoning*
 - Exact and approximate
 - Decision making
- Learning
 - Parameters and structure
 - With and without complete data

Probabilistic
Graphical
Models



Introduction

Preliminaries: Distributions

Joint Distribution $P(I, D, G)$

- Intelligence (I) $\leftarrow 2$
 - i^0 (low), i^1 (high),
- Difficulty (D) $\leftarrow 2$
 - d^0 (easy), d^1 (hard)
- Grade (G) $\leftarrow 3$
 - $g^1 (A), g^2 (B), g^3 (C)$

$2 \times 2 \times 3 = 12$ parameters
independent params
!!

I	D	G	Prob.
i^0	d^0	g^1	0.126
i^0	d^0	g^2	0.168
i^0	d^0	g^3	0.126
i^0	d^1	g^1	0.009
i^0	d^1	g^2	0.045
i^0	d^1	g^3	0.126
i^1	d^0	g^1	0.252
i^1	d^0	g^2	0.0224
i^1	d^0	g^3	0.0056
i^1	d^1	g^1	0.06
i^1	d^1	g^2	0.036
i^1	d^1	g^3	0.024

Daphne Koller

Conditioning

condition on g^1

I	D	G	Prob.
i^0	d^0	g^1	0.126
i^0	d^0	g^2	0.168
i^0	d^0	g^3	0.126
i^0	d^1	g^1	0.009
i^0	d^1	g^2	0.045
i^0	d^1	g^3	0.126
i^1	d^0	g^1	0.252
i^1	d^0	g^2	0.0224
i^1	d^0	g^3	0.0056
i^1	d^1	g^1	0.06
i^1	d^1	g^2	0.036
i^1	d^1	g^3	0.024

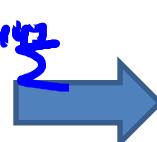
Conditioning: Reduction

I	D	G	Prob.
i ⁰	d ⁰	g ¹	0.126
i ⁰	d ¹	g ¹	0.009
i ¹	d ⁰	g ¹	0.252
i ¹	d ¹	g ¹	0.06

Daphne Koller

Conditioning: Renormalization

I	D	G	Prob.
i ⁰	d ⁰	g ¹	0.126
i ⁰	d ¹	g ¹	0.009
i ¹	d ⁰	g ¹	0.252
i ¹	d ¹	g ¹	0.06



I	D	Prob.
i ⁰	d ⁰	0.282
i ⁰	d ¹	0.02
i ¹	d ⁰	0.564
i ¹	d ¹	0.134

$$\frac{P(I, D, g^1)}{0.447}$$

unnormalized measure

$$P(\underline{I}, \underline{D} | g^1)$$

Marginalization

$P(I, D)$ Marginalize I

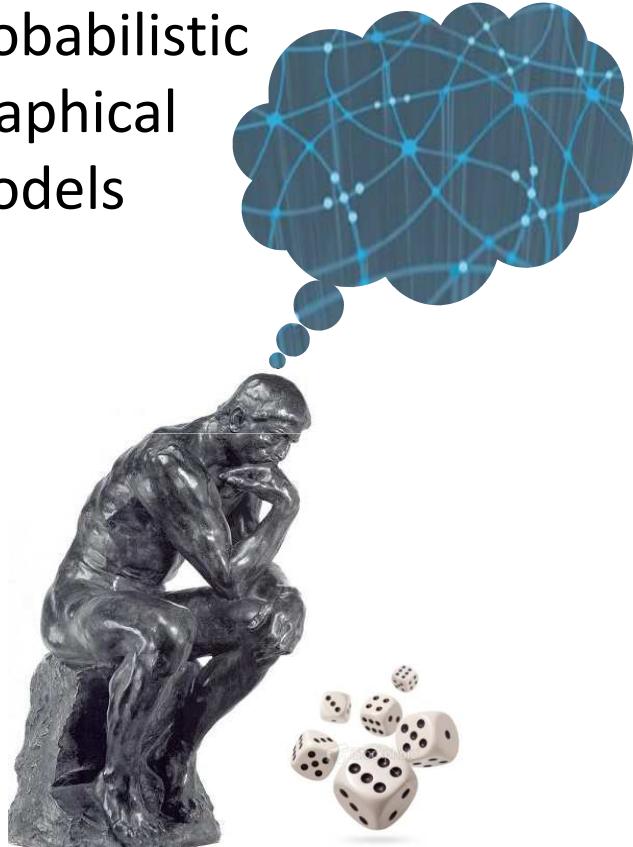
I	D	Prob.
i^0	d^0	0.282
i^0	d^1	0.02
i^1	d^0	0.564
i^1	d^1	0.134

Diagram illustrating marginalization:

- Arrows point from the **I** column to the **D** column in the first two rows.
- Arrows point from the **I** column to the **D** column in the last two rows.
- The resulting table shows the marginalized probabilities for **D**:

D	Prob.
d^0	0.846
d^1	0.154

Probabilistic
Graphical
Models



Introduction

Preliminaries: Factors

Factors

- A factor $\phi(\underline{X}_1, \dots, \underline{X}_k)$

$$\phi : \text{Val}(\underline{X}_1, \dots, \underline{X}_k) \rightarrow R$$

- Scope = $\{\underline{X}_1, \dots, \underline{X}_k\}$

Joint Distribution

$P(I, D, G)$

<u>I</u>	<u>D</u>	<u>G</u>	Prob.
i^0	d^0	g^1	0.126
i^0	d^0	g^2	0.168
i^0	d^0	g^3	0.126
i^0	d^1	g^1	0.009
i^0	d^1	g^2	0.045
i^0	d^1	g^3	0.126
i^1	d^0	g^1	0.252
i^1	d^0	g^2	0.0224
i^1	d^0	g^3	0.0056
i^1	d^1	g^1	0.06
i^1	d^1	g^2	0.036
i^1	d^1	g^3	0.024

Daphne Koller

Unnormalized measure $P(I, D, g^1)$

scope = {I, D}

$P(I, D, g^1)$

I	D	G	Prob.
i ⁰	d ⁰	g^1	0.126
i ⁰	d ¹	g^1	0.009
i ¹	d ⁰	g^1	0.252
i ¹	d ¹	g^1	0.06

Conditional Probability Distribution (CPD)

$P(G | I, D)$

constant

	g^1	g^2	g^3
i^0, d^0	0.3	0.4	0.3
i^0, d^1	0.05	0.25	0.7
i^1, d^0	0.9	0.08	0.02
i^1, d^1	0.5	0.3	0.2

$\frac{a}{a+b+c}$

General factors

Scope = {A, B}

A	B	ϕ
a ⁰	b ⁰	30
a ⁰	b ¹	5
a ¹	b ⁰	1
a ¹	b ¹	10

Factor Product

a^1	b^1	0.5
a^1	b^2	0.8
a^2	b^1	0.1
a^2	b^2	0
a^3	b^1	0.3
a^3	b^2	0.9

$\varphi_1(A, B)$

b^1	c^1	0.5
b^1	c^2	0.7
b^2	c^1	0.1
b^2	c^2	0.2

$\varphi_2(B, C)$

a^1	b^1	c^1	$0.5 \cdot 0.5 = 0.25$
a^1	b^1	c^2	$0.5 \cdot 0.7 = 0.35$
a^1	b^2	c^1	$0.8 \cdot 0.1 = 0.08$
a^1	b^2	c^2	$0.8 \cdot 0.2 = 0.16$
a^2	b^1	c^1	$0.1 \cdot 0.5 = 0.05$
a^2	b^1	c^2	$0.1 \cdot 0.7 = 0.07$
a^2	b^2	c^1	$0 \cdot 0.1 = 0$
a^2	b^2	c^2	$0 \cdot 0.2 = 0$
a^3	b^1	c^1	$0.3 \cdot 0.5 = 0.15$
a^3	b^1	c^2	$0.3 \cdot 0.7 = 0.21$
a^3	b^2	c^1	$0.9 \cdot 0.1 = 0.09$
a^3	b^2	c^2	$0.9 \cdot 0.2 = 0.18$

$A \otimes B \otimes C$

Scope A, B, C

Factor Marginalization

a ¹	b ¹	c ¹	0.25
a ¹	b ¹	c ²	0.35
a ¹	b ²	c ¹	0.08
a ¹	b ²	c ²	0.16
a ²	b ¹	c ¹	0.05
a ²	b ¹	c ²	0.07
a ²	b ²	c ¹	0
a ²	b ²	c ²	0
a ³	b ¹	c ¹	0.15
a ³	b ¹	c ²	0.21
a ³	b ²	c ¹	0.09
a ³	b ²	c ²	0.18

A, C

a ¹	c ¹	<u>0.33</u>
a ¹	c ²	0.51
a ²	c ¹	0.05
a ²	c ²	0.07
a ³	c ¹	0.24
a ³	c ²	0.39

Factor Reduction

a ¹	b ¹	c ¹	0.25
a ¹	b ¹	c ²	0.35
a ¹	b ²	c ¹	0.08
a ¹	b ²	c ²	0.16
a ²	b ¹	c ¹	0.05
a ²	b ¹	c ²	0.07
a ²	b ²	c ¹	0
a ²	b ²	c ²	0
a ³	b ¹	c ¹	0.15
a ³	b ¹	c ²	0.21
a ³	b ²	c ¹	0.09
a ³	b ²	c ²	0.18



a ¹	b ¹	c ¹	0.25
a ¹	b ²	c ¹	0.08
a ²	b ¹	c ¹	0.05
a ²	b ²	c ¹	0
a ³	b ¹	c ¹	0.15
a ³	b ²	c ¹	0.09

A, B

Why factors?

- Fundamental building block for defining distributions in high-dimensional spaces
- Set of basic operations for manipulating these probability distributions