Probabilistic
Graphical
Models

Introduction

Welcome to the PGM Class

Probabilistic Graphical Models

Daphne Koller

Course Structure

- 10 weeks + final
- Videos + quizzes
- 9 problem sets
  - 25% of score
  - Multiple submissions

Course Structure

- 9 programming assignments
  - Genetically inherited diseases
  - Optical character recognition
  - Recognizing activities from Kinect sensor
  - 9 x 7% = 63% of score
- Final exam
  - 12% of score

Background

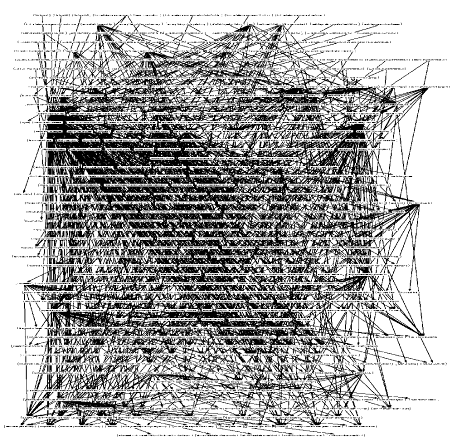
- Required
  - Basic probability theory
  - Some programming
  - Some algorithms and data structures
- Recommended
  - Machine learning
  - Simple optimization
  - Matlab or Octave

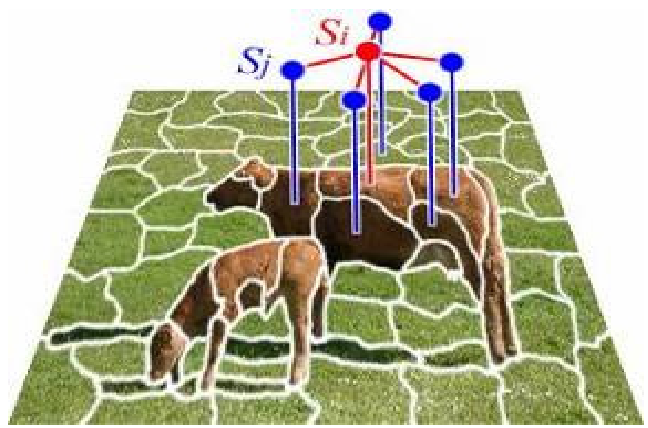
Other Issues

- Honor code
- Time management (10-15 hrs / week)
- Discussion forum & study groups

What you'll learn

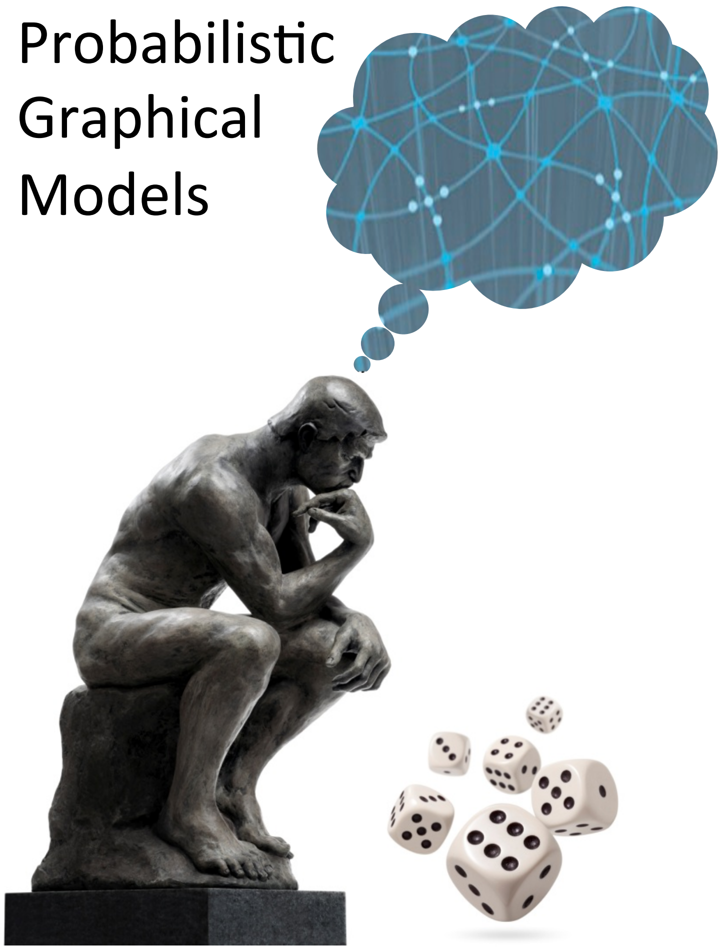
- Fundamental methods
- Real-world applications
- How to use these methods in your work





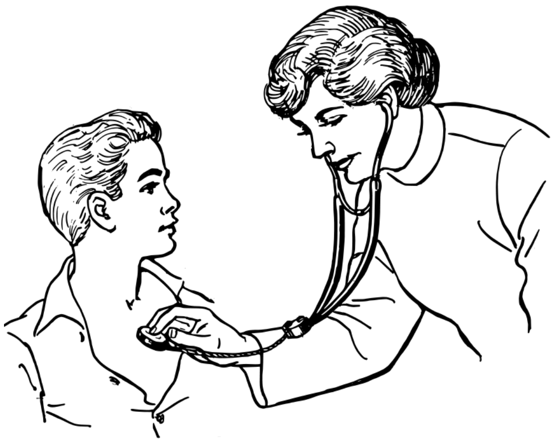
M. Pradhan , G. Provan , B. Middleton , M.Henrion, UAI 94

Probabilistic
Graphical
Models

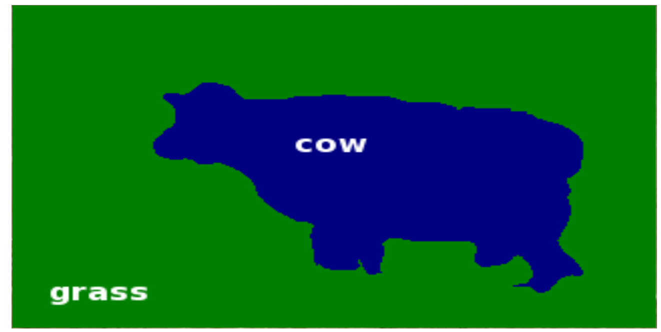


Introduction

Motivation and Overview



predisposing factors
symptoms
test results
diseases
treatment outcomes



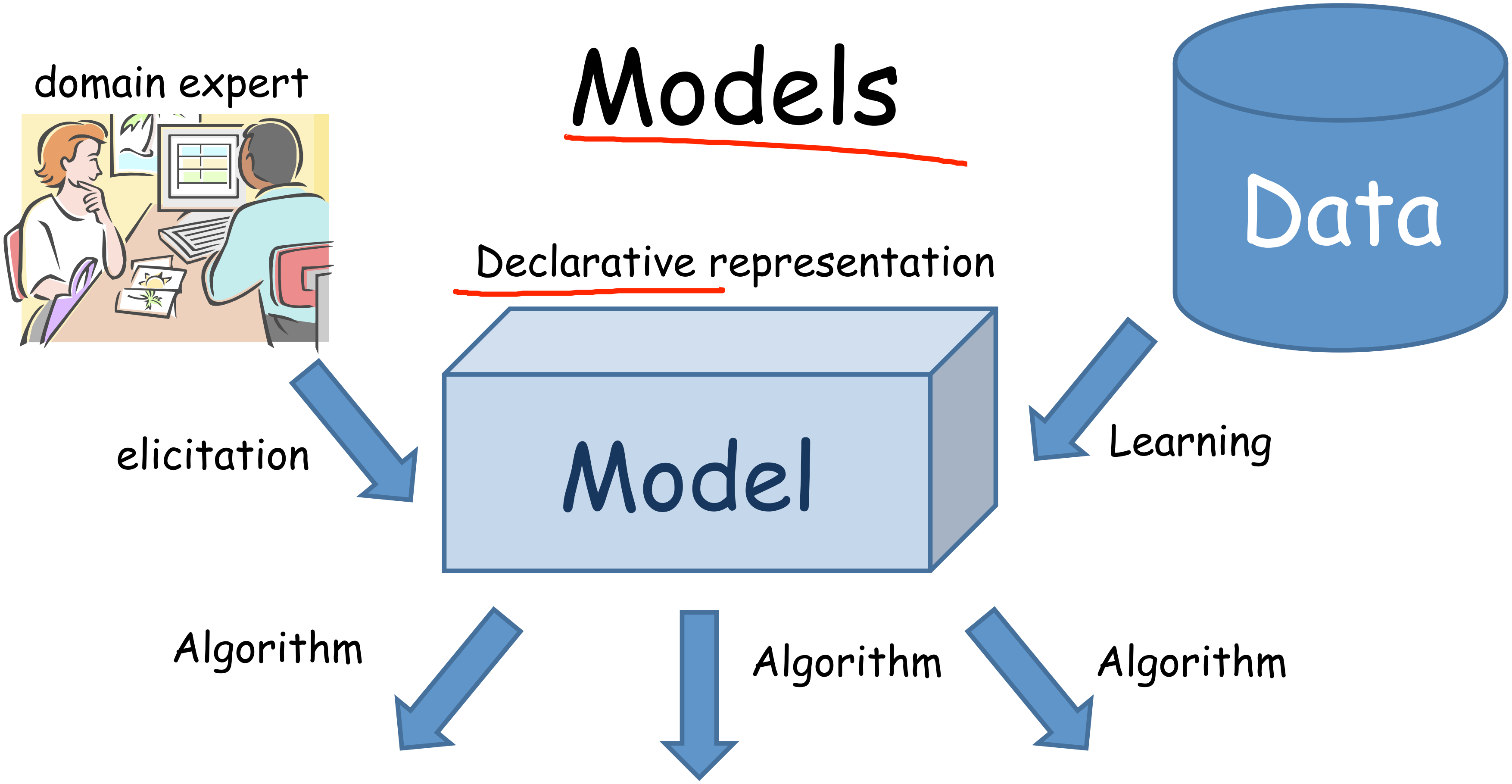
millions of pixels or
thousands of superpixels

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

Probabilistic Graphical Models

Daphne Koller

Models


domain expert
Declarative representation
Data
elicitation
Model
Learning
Algorithm
Algorithm
Algorithm

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 $\boxed{X_1, \ldots, X_n}$ n binary valued

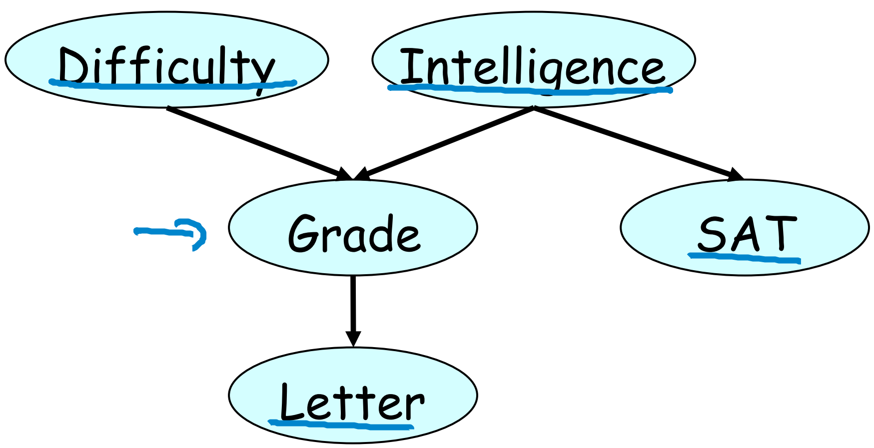
Joint distribution $\underline{\underline{P(X_1, \ldots, X_n)}}$ distribution over 2^n possible states

$X_1 \ldots X_n$ — nodes

Graphical Models

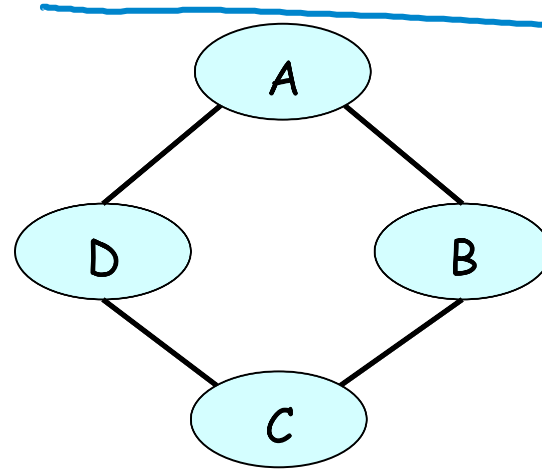
directed graph

Bayesian networks



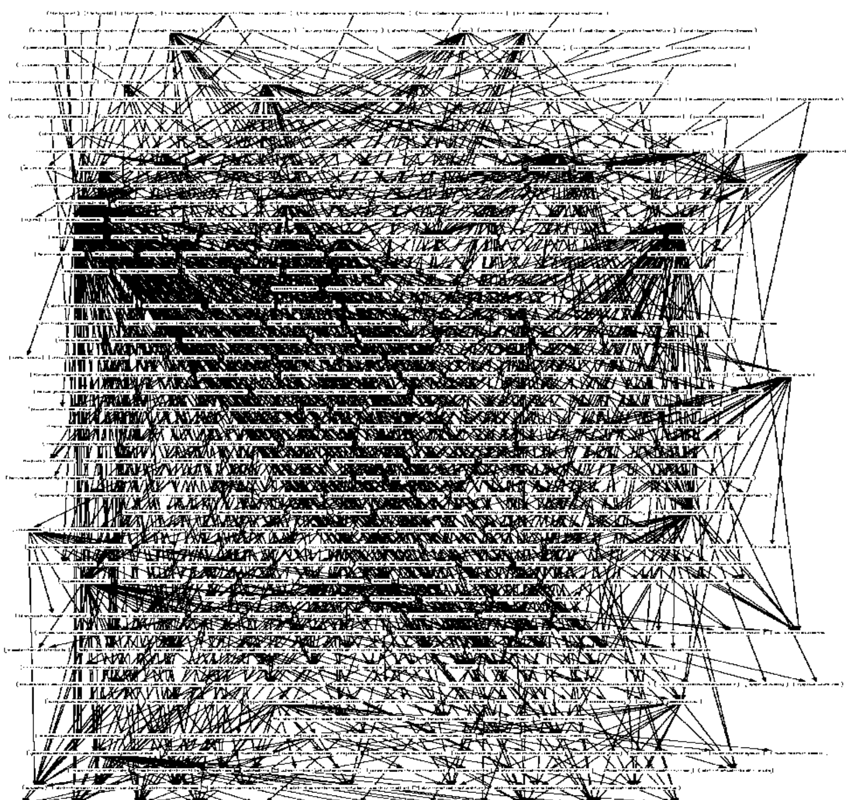
undirected graph

Markov networks



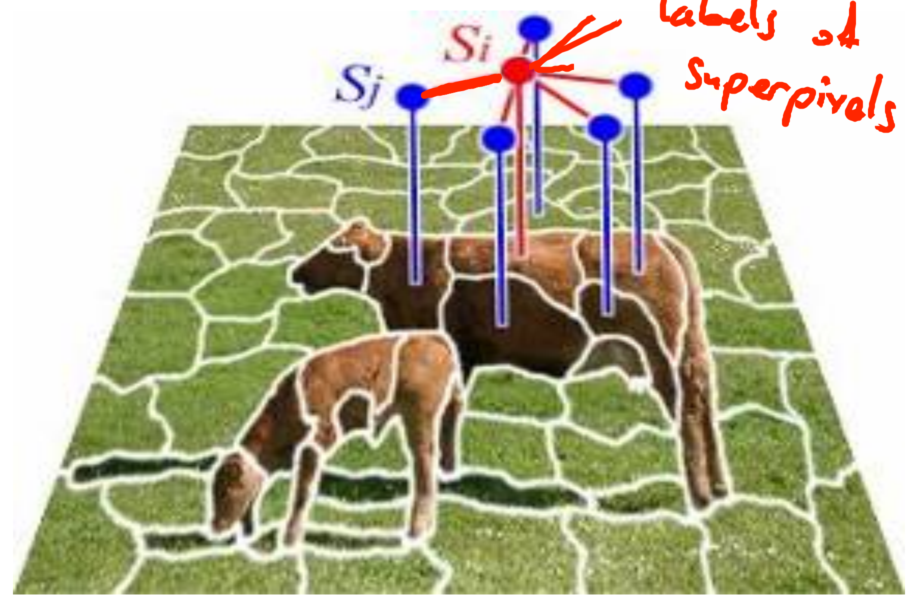
diagnosis
CPCS

Graphical Models



M. Pradhan, G. Provan, B. Middleton, M. Henrion, UAI 94

labels of superpixels



Daphne Koller

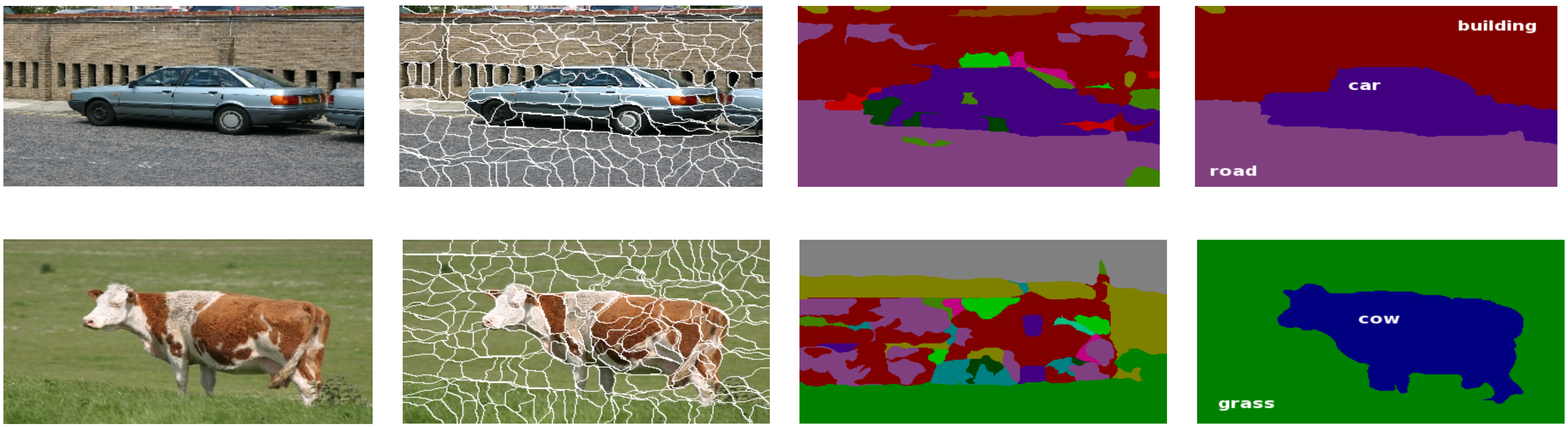
Graphical Representation

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



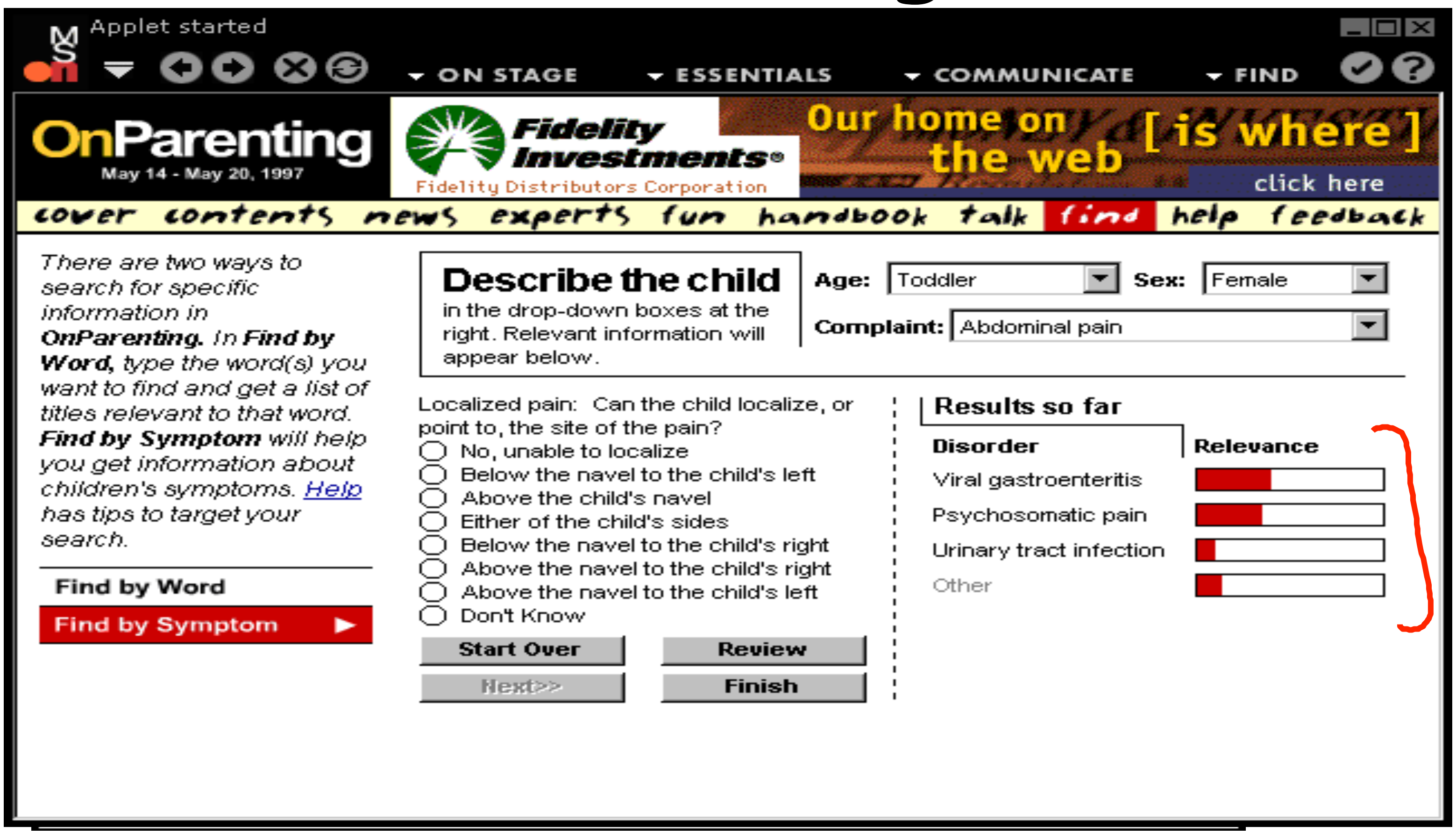
superpixels

machine learning to separate superpixels

Daphne Koller

Thanks to: Eric Horvitz, Microsoft Research

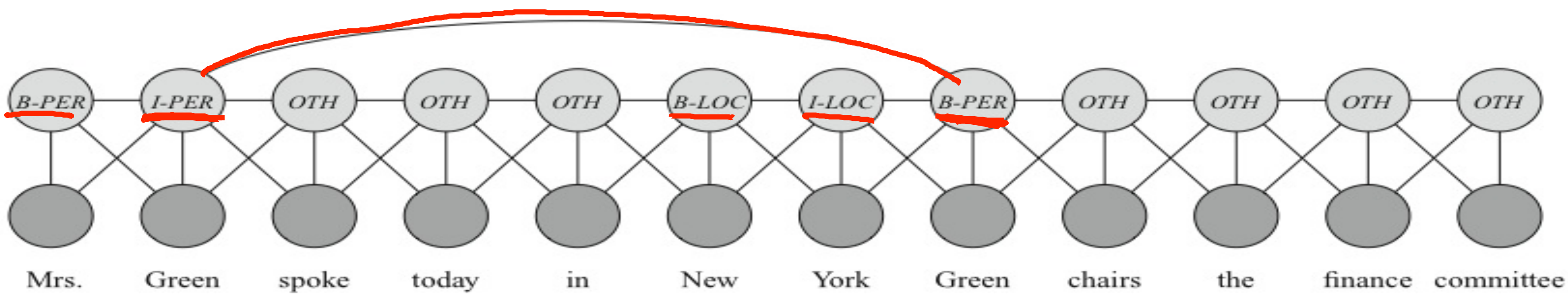
Medical Diagnosis



Textual Information Extraction

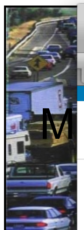
Mrs. Green spoke today in New York. Green chairs the finance committee.

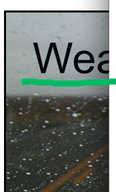
Person — Location — Person — organization

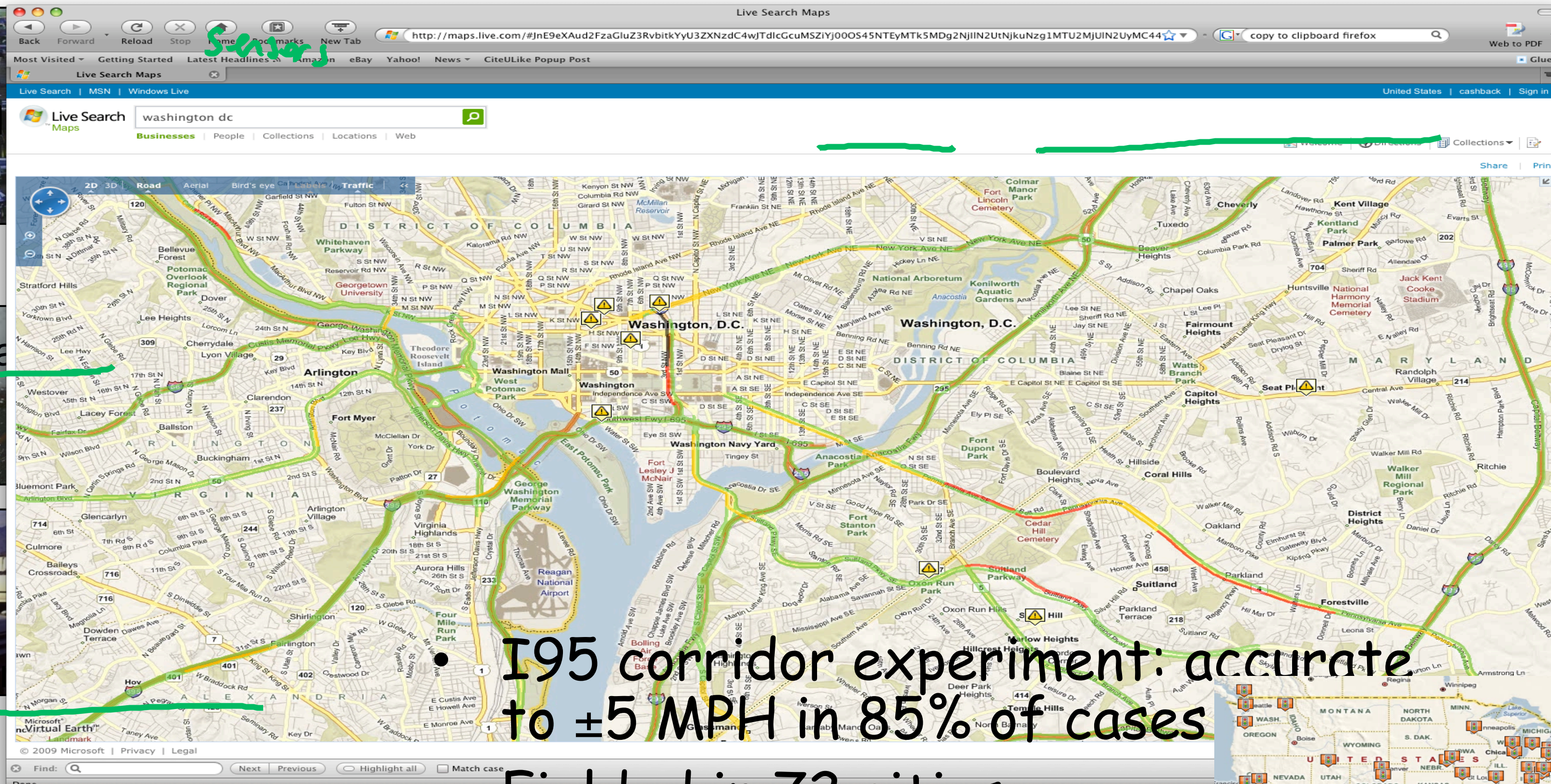


Daphne Koller

Multi-Sensor Integration: Traffic



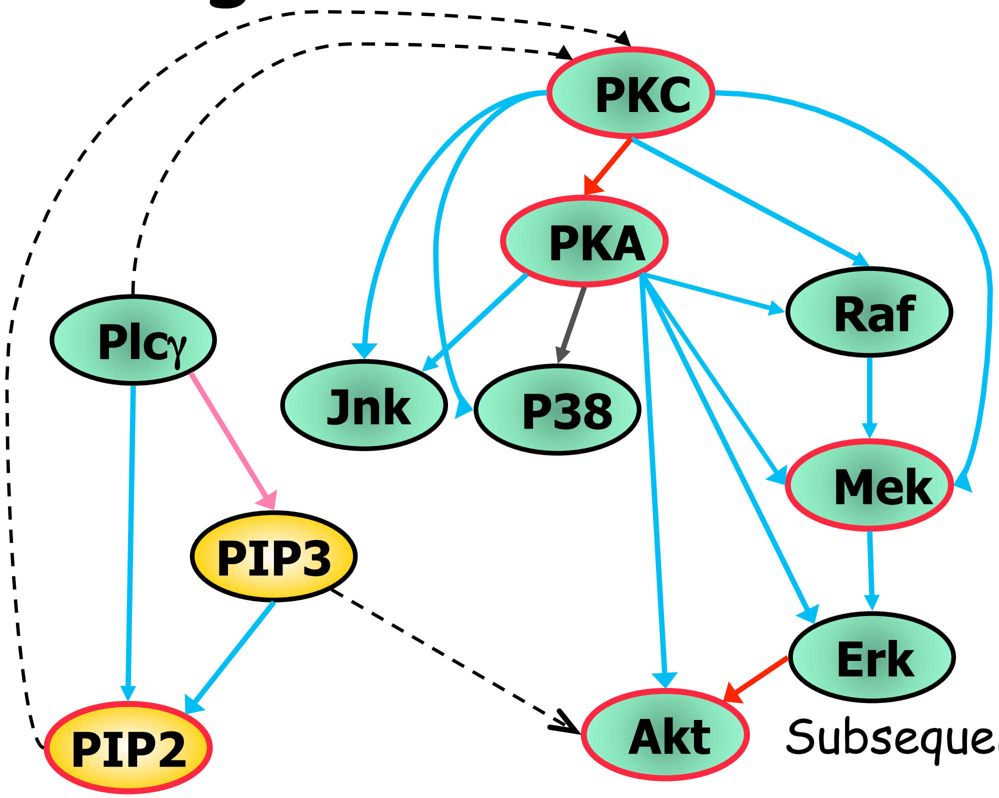




- I95 corridor experiment: accurate to ±5 MPH in 85% of cases
- Fielded in 72 cities

Thanks to: Eric Horvitz, Microsoft Research

Biological Network Reconstruction



Phospho-Proteins
Phospho-Lipids
Perturbed in data

Known	15/17
Supported	2/17
Reversed	1
Missed	3

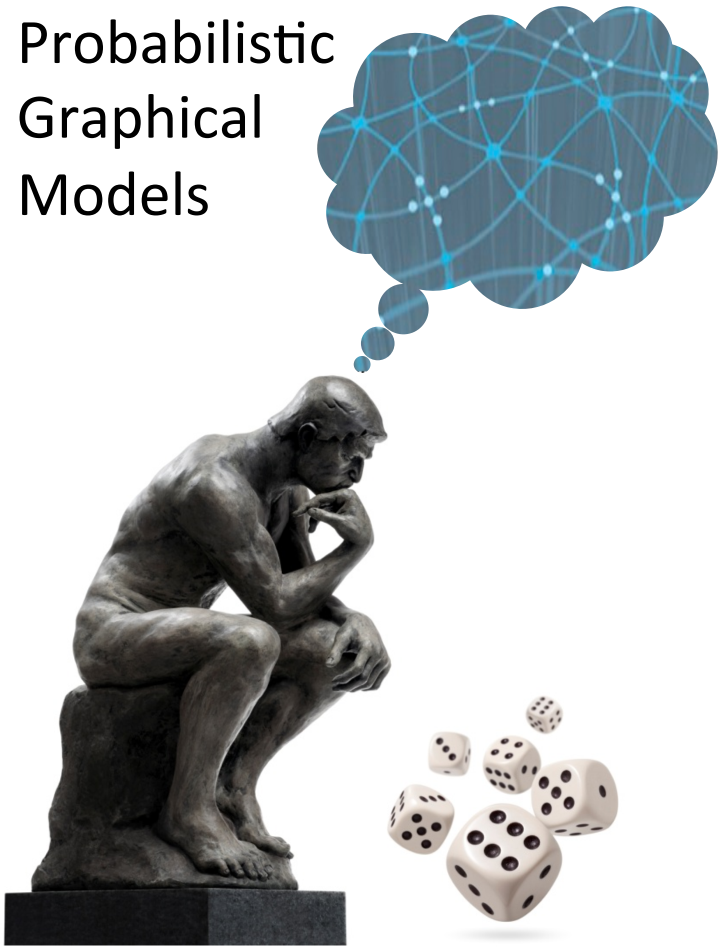
Causal protein-signaling networks derived from multiparameter single-cell data
Sachs et al., *Science* 2005

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) ← 2
  - i^0 (low), i^1 (high),
- Difficulty (D) ← 2
  - d^0 (easy), d^1 (hard)
- Grade (G) ← 3
  - g^1 (A), g^2 (B), g^3 (C)

parameters
$2 \times 2 \times 3 = 12$

independent params
11

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

1

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

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

/0.447

$P(I, D, g^1)$ 0.447

unnormalized measure

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

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

Marginalization

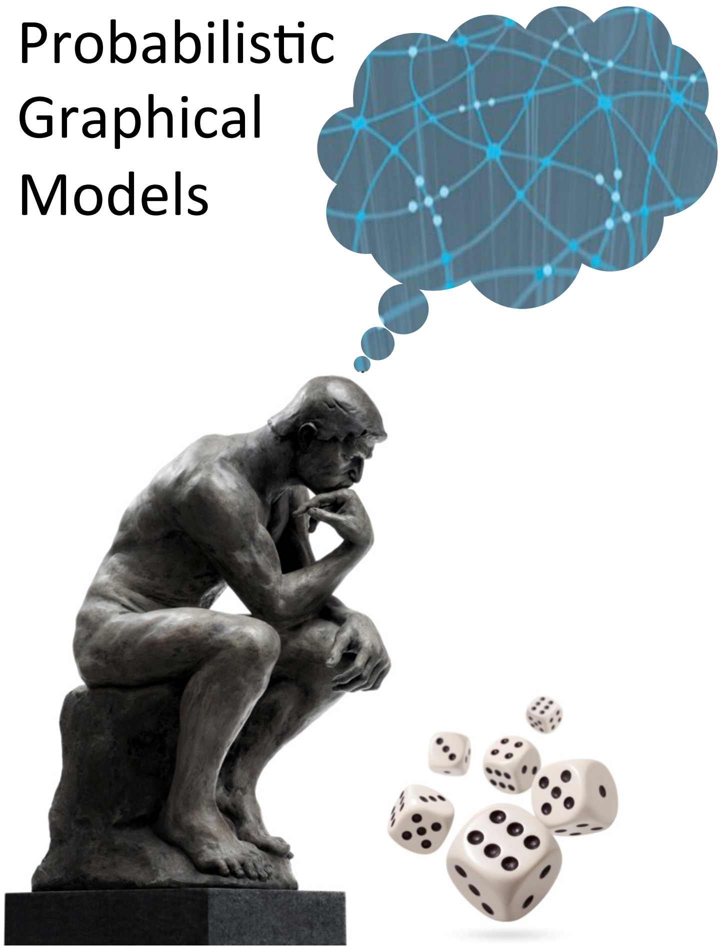
Marginalize I

P(I, D)

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

D	Prob.
d^0	0.846
d^1	0.154

Probabilistic
Graphical
Models



Introduction

Preliminaries: Factors

Factors

- A factor $\phi(X_1,\ldots,X_k)$

  $\phi : \text{Val}(X_1,\ldots,X_k) \rightarrow \mathbb{R}$

- Scope = $\{X_1,\ldots,X_k\}$

Joint Distribution

$P(I,D,G)$

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

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)

$P(G \mid I, D)$

	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
	A	B	C

context

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

$\phi_1(A,B)$

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

$\phi_2(B,C)$

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

A,B,C

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

Daphne Koller

Factor Marginalization

Scope A,B,C

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	0.33
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

$\Rightarrow$

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