Intelligent Systems (AI-2)

Computer Science cpsc422, Lecture 17

Oct, 19, 2015

Slide Sources
D. Koller, Stanford CS - Probabilistic Graphical Models
D. Page, Whitehead Institute, MIT

Several Figures from
"Probabilistic Graphical Models: Principles and Techniques" *D. Koller, N. Friedman* 2009

Simple but Powerful Approach: Particle Filtering

Idea from Exact Filtering**:** should be able to compute $P(X_{t+1} | e_{1:t+1})$ from $P(X_t | e_{1:t})$

".. One slice from the previous slice…"

Idea from Likelihood Weighting

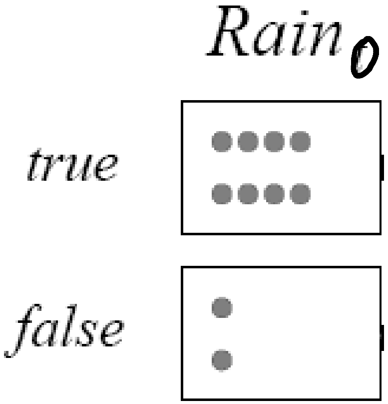
- Samples should be weighted by the probability of evidence given parents

New Idea: run multiple samples simultaneously through the network

Particle Filtering

- Run all **N samples together** through the network, one slice at a time

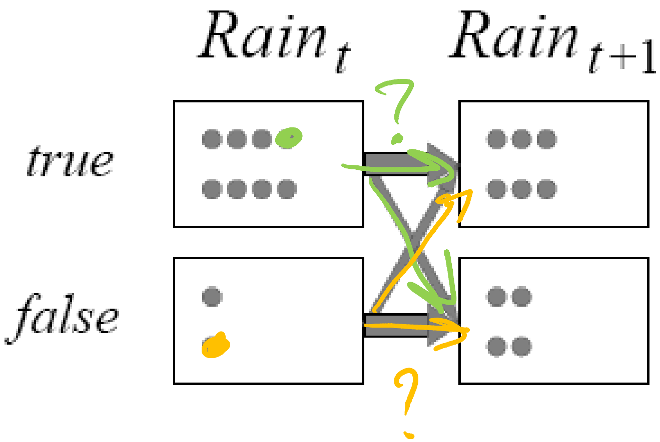
STEP 0: Generate a population on N initial-state samples by sampling from initial state distribution $P(X_0)$



N = 10

Particle Filtering

STEP 1: Propagate each sample for x_t forward by sampling the next state value x_{t+1} based on $P(X_{t+1}|X_t)$



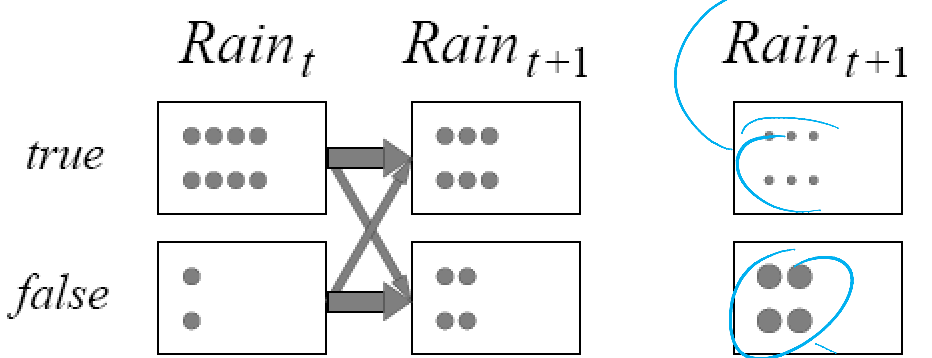
(a) Propagate

R_t	$P(R_{t+1}=t)$
t	0.7
f	0.3

Particle Filtering

STEP 2: **Weight each sample by the likelihood** it assigns to the evidence

- E.g. assume we observe ***not umbrella*** at t+1

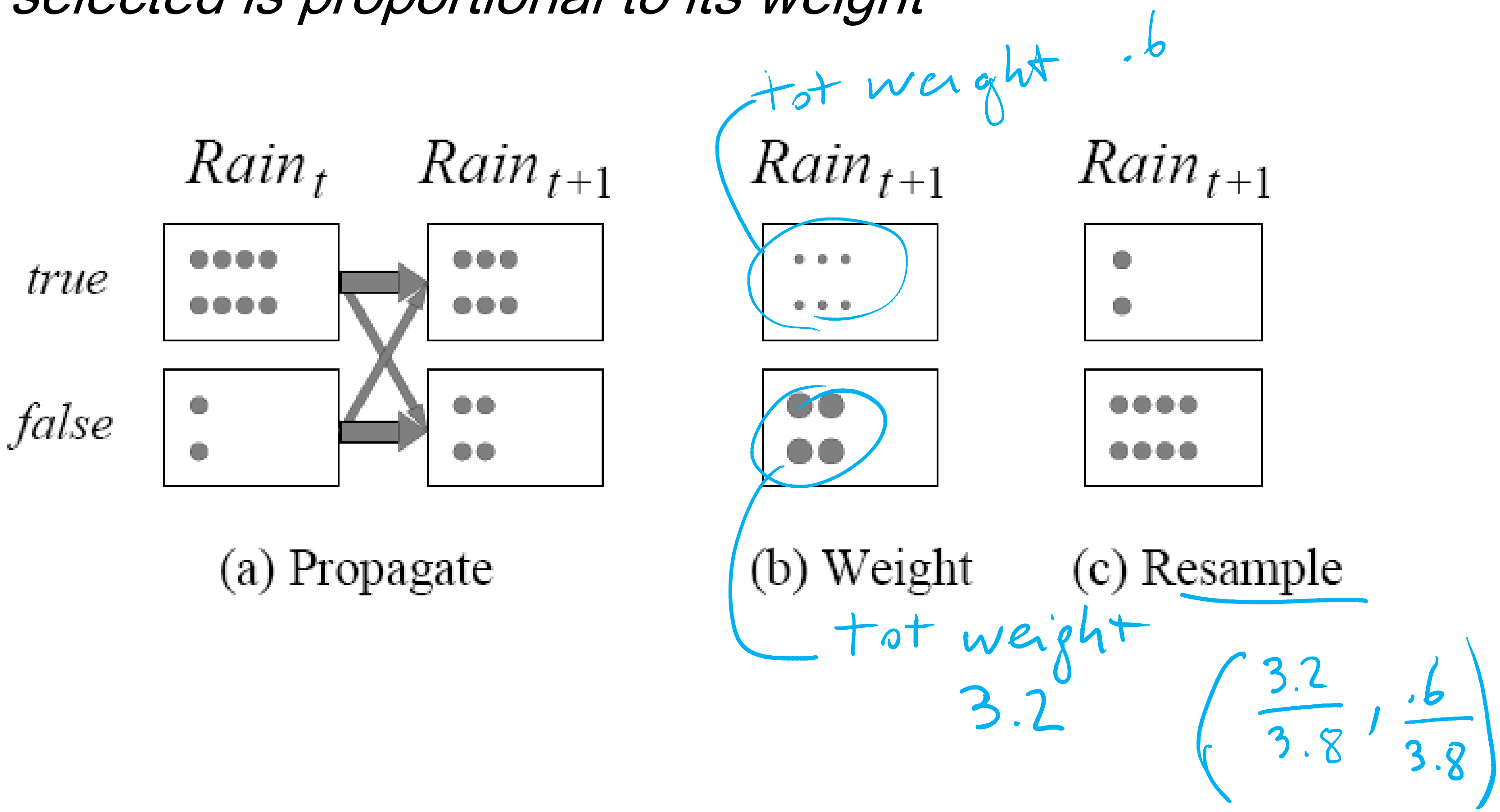


(a) Propagate (b) Weight

R_t	$P(u_t)$	$P(\neg u_t)$
t	0.9	0.1
f	0.2	0.8

Particle Filtering

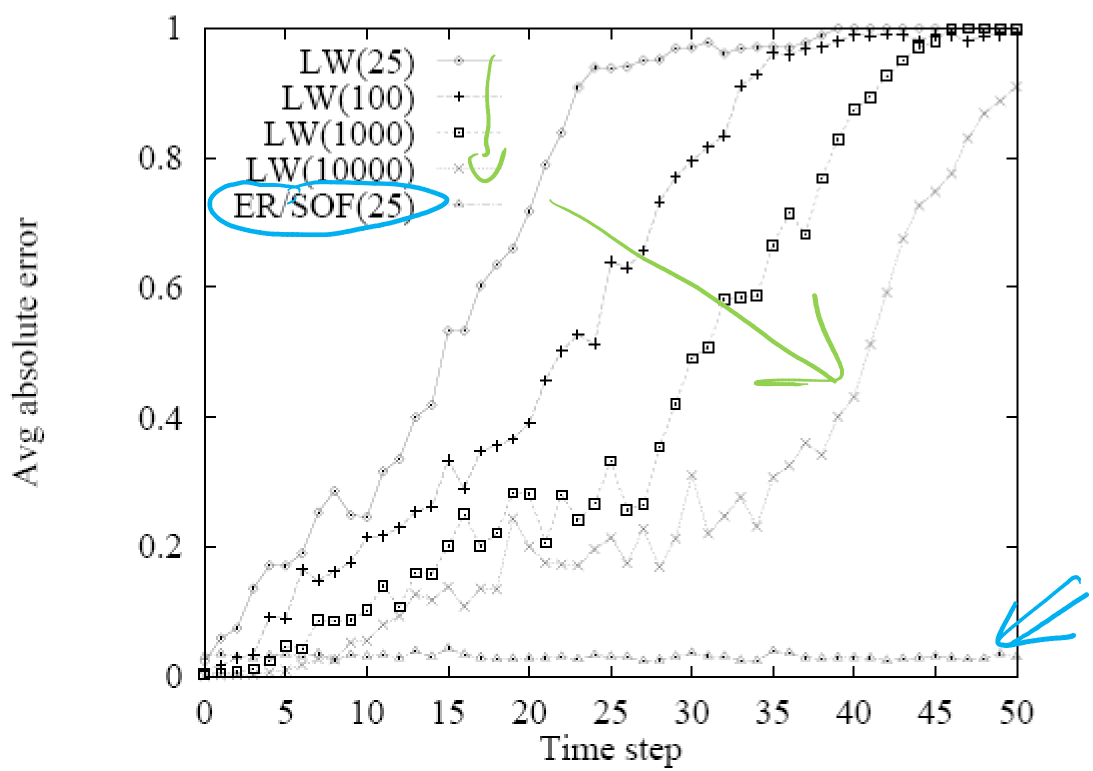
STEP 3: Create a new population from the population at X_{t+1}, *i.e. resample the population so that the probability that each sample is selected is proportional to its weight*



- ➢ Start the Particle Filtering cycle again from the new sample

Is PF Efficient?

In practice, approximation error of particle filtering remains bounded overtime



It is also possible to prove that the approximation maintains bounded error with high probability

(with specific assumption: probs in transition and sensor models >0 and <1)

422 big picture: Where are we?

Hybrid: Det +Sto

Prob CFG
Prob Relational Models
Markov Logics

Deterministic | **Stochastic**

Query

Deterministic:

Logics
First Order Logics
Ontologies
Temporal rep.

- Full Resolution
- SAT

Stochastic:

Belief Nets

Approx. : Gibbs

Markov Chains and HMMs

Forward, Viterbi….
Approx. : Particle Filtering

Undirected Graphical Models
Markov Networks
Conditional Random Fields

Planning

Stochastic:

Markov Decision Processes and Partially Observable MDP

- Value Iteration
- Approx. Inference

Reinforcement Learning

Applications of AI

Representation

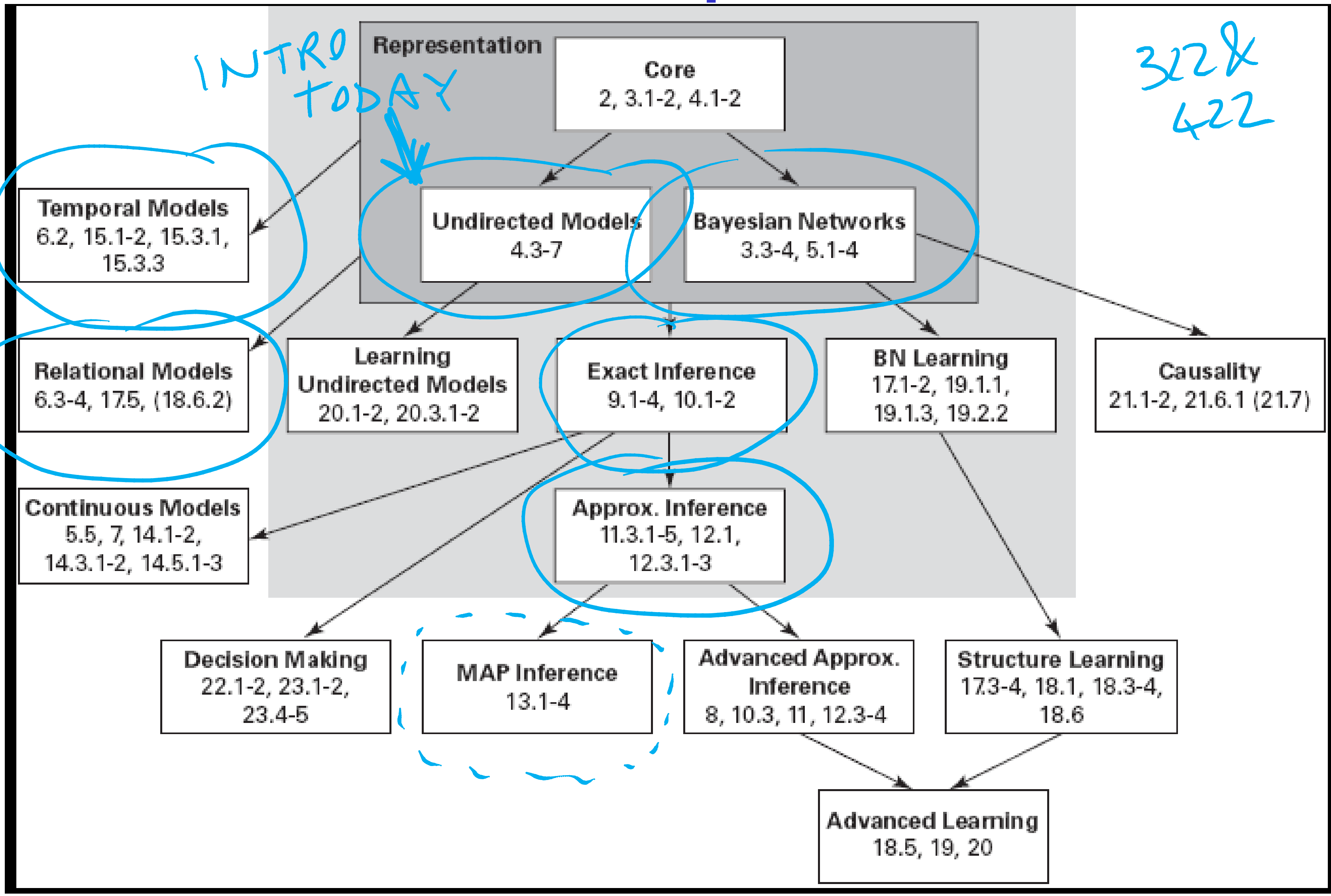
Reasoning Technique

Lecture Overview

Probabilistic Graphical models

- Intro
- Example
- Markov Networks Representation (vs. Belief Networks)
- Inference in Markov Networks (Exact and Approx.)
- Applications of Markov Networks

Probabilistic Graphical Models

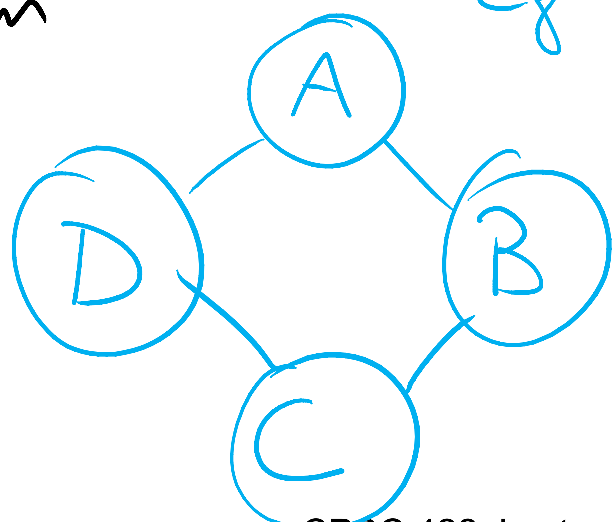


From "Probabilistic Graphical Models: Principles and Techniques" *D. Koller, N. Friedman* 2009

Misconception Example

- Four students (Alice, Bill, Debbie, Charles) get together in pairs, to work on a homework
- But only in the following pairs: AB AD DC BC
- Professor misspoke and might have generated misconception
- A student might have figured it out later and told study partner

Four random vars



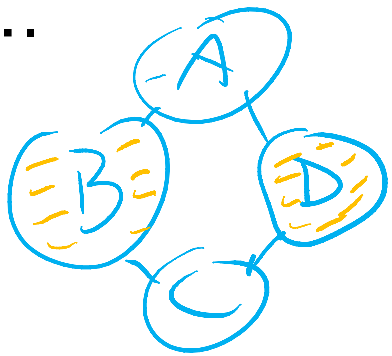
eg A random var
two values
a^1 Alice has the misc.
a^0 Alice doesn't have the misc.

Example: In/Depencencies

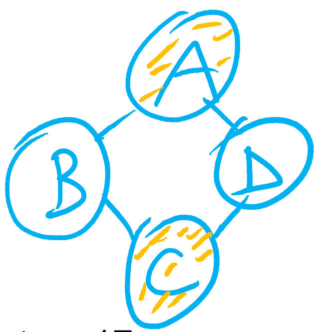
Are A and C independent because they never spoke?

No, because A might have figure it out and told B who then told C

But if we know the values of B and D….

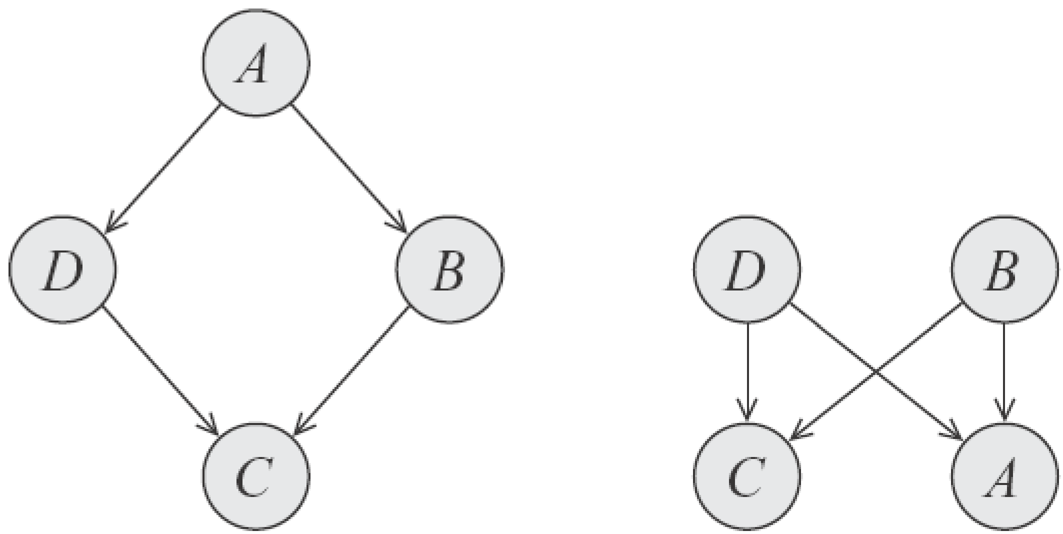


And if we know the values of A and C



Which of these two Bnets captures the two independencies of our example?

iclicker



a.

b.

c. Both

d. None

$(A \perp C \mid B D)$

$(B \perp D \mid A C)$

in a. $B \not\perp D \mid C$

in b. same

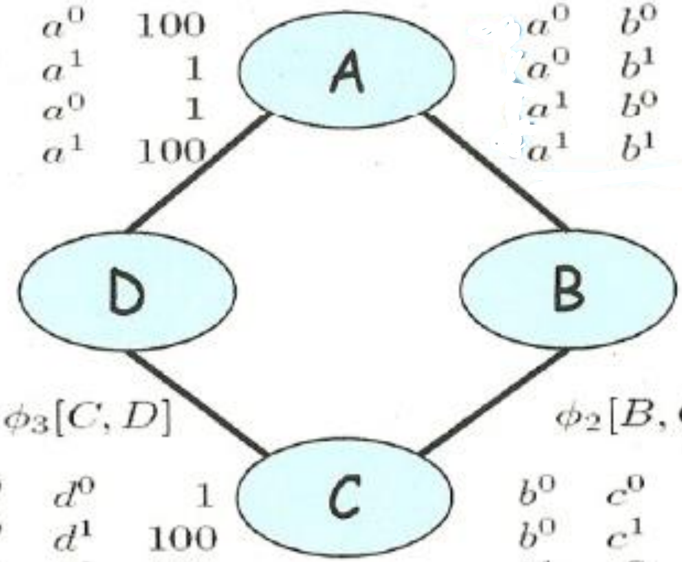
Parameterization of Markov Networks

$\phi_4[D, A]$

d^0	a^0	100
d^0	a^1	1
d^1	a^0	1
d^1	a^1	100

$\phi_1[A, B]$

a^0	b^0	30
a^0	b^1	5
a^1	b^0	1
a^1	b^1	10



$\phi_3[C, D]$

c^0	d^0	1
c^0	d^1	100
c^1	d^0	100
c^1	d^1	1

$\phi_2[B, C]$

b^0	c^0	100
b^0	c^1	1
b^1	c^0	1
b^1	c^1	100

X set of random vars: A factor is $\Phi(Val(X)) \rightarrow \mathbb{R}$

Factors define the local interactions (like CPTs in Bnets)

What about the global model? What do you do with Bnets?

How do we combine local models?

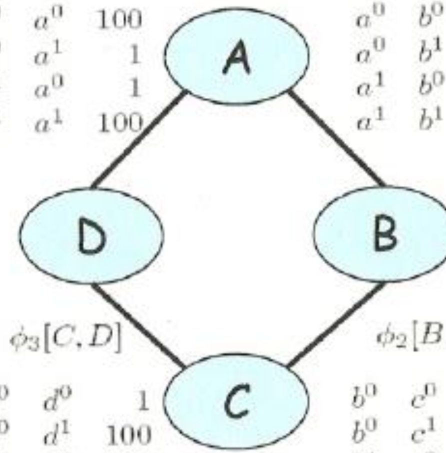
As in BNets by multiplying them!

$$\tilde{P}(A,B,C,D) = \phi_1(A,B) \times \phi_2(B,C) \times \phi_3(C,D) \times \phi_4(A,D)$$

$$P(A,B,C,D) = \frac{1}{Z}\tilde{P}(A,B,C,D)$$

P(A, B) ?

Assignment				Unnormalized	Normalized
a^0	b^0	c^0	d^0	300000	.04
a^0	b^0	c^0	d^1	300000	.04
a^0	b^0	c^1	d^0	300000	.04
a^0	b^0	c^1	d^1	30	4.1×10^{-6}
a^0	b^1	c^0	d^0	500	
a^0	b^1	c^0	d^1	500	
a^0	b^1	c^1	d^0	5000000	.69
a^0	b^1	c^1	d^1	500	
a^1	b^0	c^0	d^0	100	
a^1	b^0	c^0	d^1	1000000	
a^1	b^0	c^1	d^0	100	
a^1	b^0	c^1	d^1	100	
a^1	b^1	c^0	d^0	10	
a^1	b^1	c^0	d^1	100000	
a^1	b^1	c^1	d^0	100000	
a^1	b^1	c^1	d^1	100000	



$\phi_4[D, A]$

d^0	a^0	100
d^0	a^1	1
d^1	a^0	1
d^1	a^1	100

$\phi_1[A, B]$

a^0	b^0	30
a^0	b^1	5
a^1	b^0	1
a^1	b^1	10

$\phi_3[C, D]$

c^0	d^0	1
c^0	d^1	100
c^1	d^0	100
c^1	d^1	1

$\phi_2[B, C]$

b^0	c^0	100
b^0	c^1	1
b^1	c^0	1
b^1	c^1	100

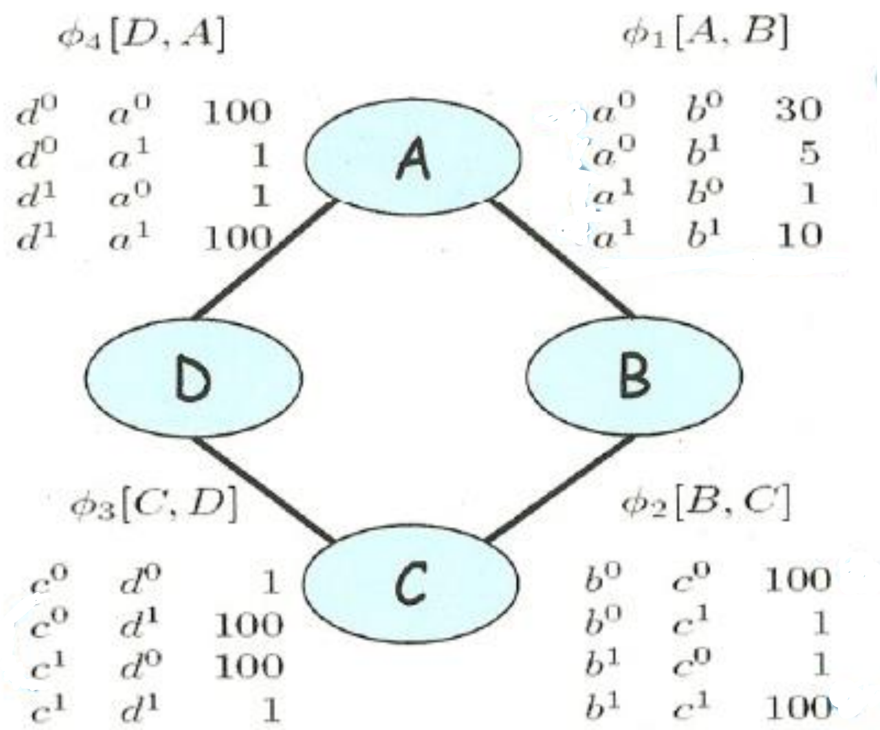
Multiplying Factors (same seen in 322 for VarElim)

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

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^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	$0.9 \cdot 0.2 = 0.18$

Factors do not represent marginal probs. !



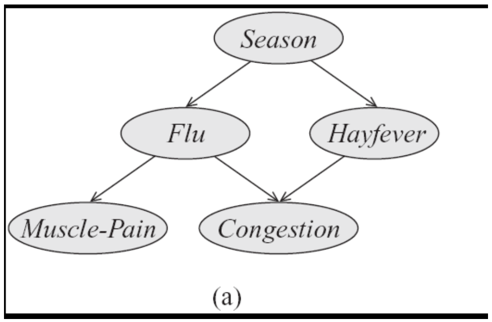
a^0 b^0	0.13
a^0 b^1	0.69
a^1 b^0	0.14
a^1 b^1	0.04

Marginal P(A,B)

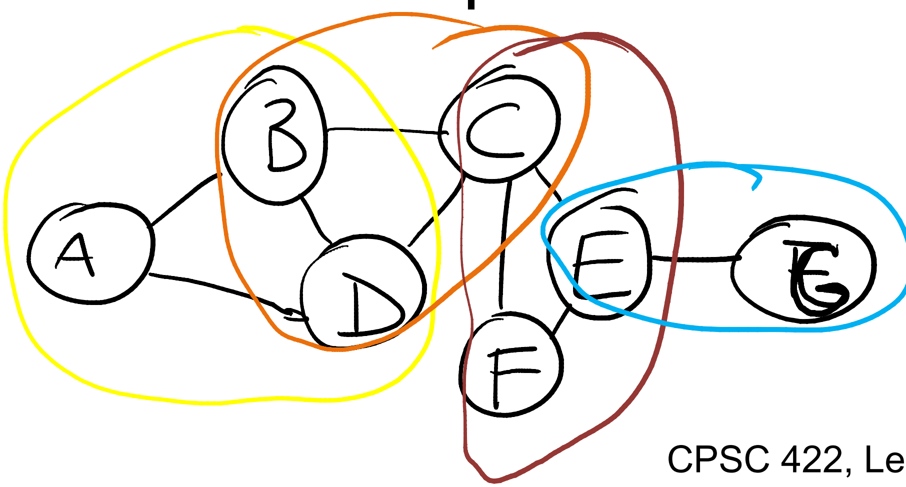
Computed from the joint

Step Back…. From structure to factors/potentials

In a Bnet the joint is factorized….



In a Markov Network you have one factor for each maximal clique

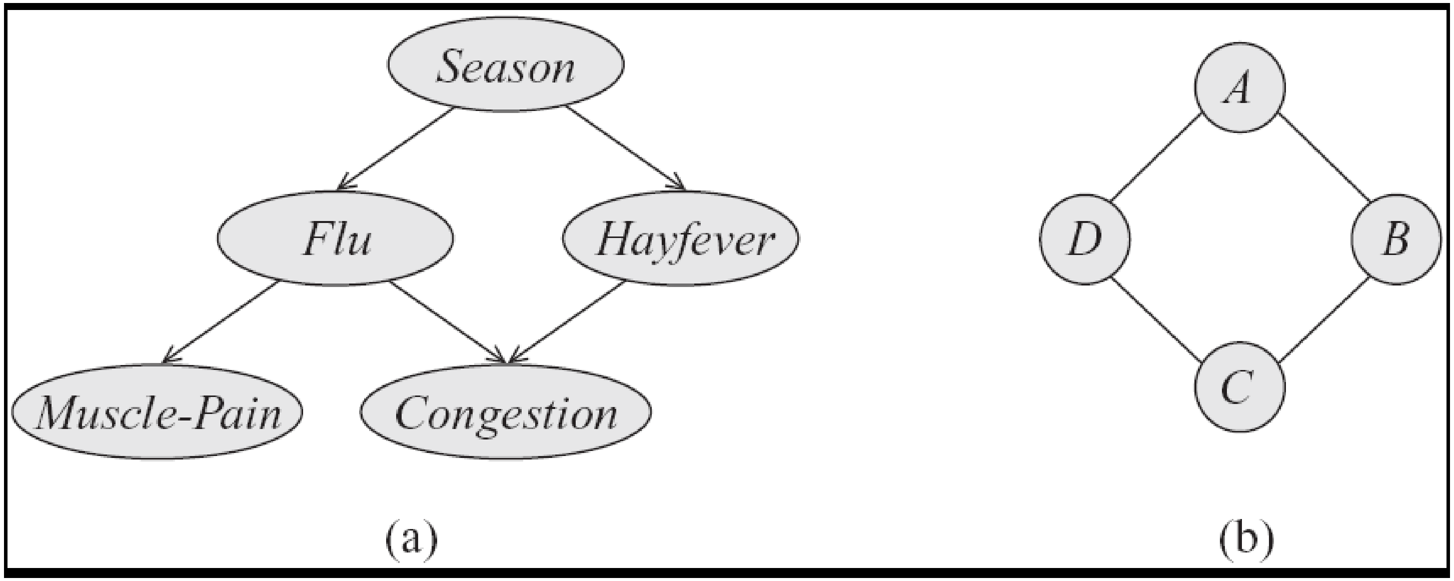


$\Phi_1(A\ B\ D)$

$\Phi_2(B\ D\ C)$ $\Phi_4(E\ G)$

$\Phi_3(C\ E\ F)$

Directed vs. Undirected



Independencies

$(F \perp H \mid S)$

$(C \perp S \mid F, H)$

$(M \perp C, H, S \mid F)$

$(A \perp C \mid B D)$

$(B \perp D \mid A C)$

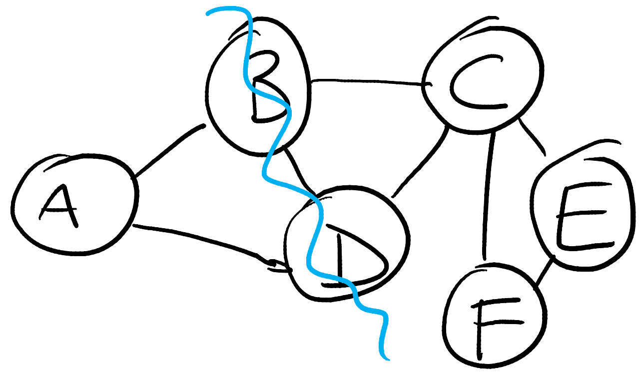
Factorization

$P(S, F, H, M, C) = P(S) * P(F \mid S) * P(H \mid S) * P(M \mid F) * P(C \mid F H)$

$P(ABCD) = \frac{1}{Z} \Phi_1(AB) * \Phi_2(BC) * \Phi_3(CD) * \Phi_4(AD)$

General definitions

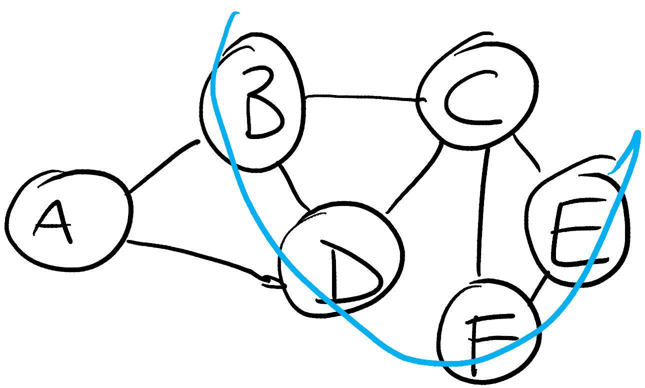
Two nodes in a Markov network are **independent** if and only if every path between them is cut off by evidence



eg for A C

So the **markov blanket** of a node is…?

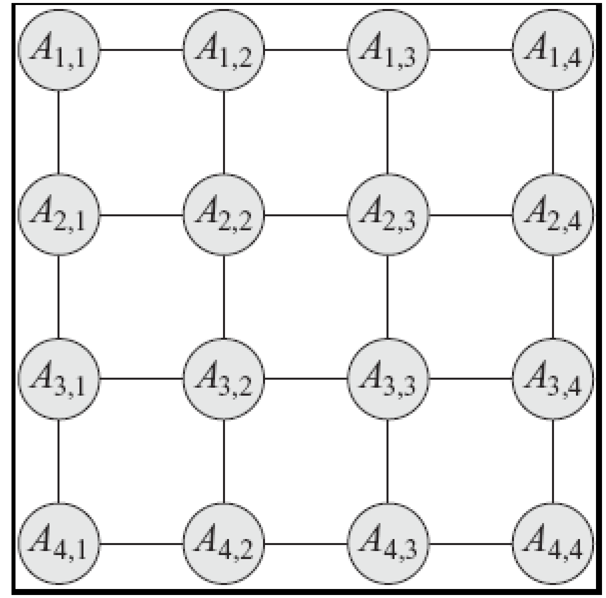
eg for C



Markov Networks Applications (1): Computer Vision

Called **Markov Random Fields**

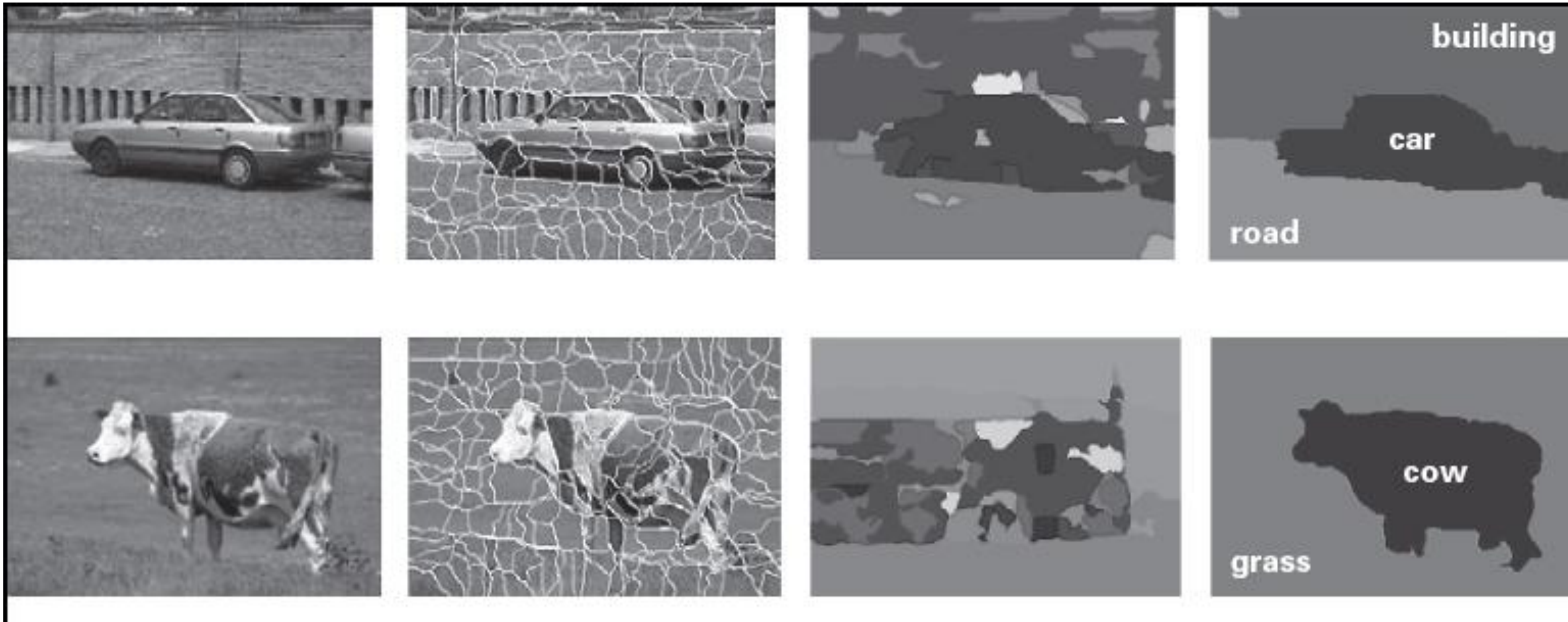
- Stereo Reconstruction
- Image Segmentation
- Object recognition



Typically **pairwise MRF**

- Each *vars* correspond to a *pixel* (or *superpixel*)
- Edges (factors) correspond to interactions between adjacent pixels in the image
  - E.g., in segmentation: from generically penalize discontinuities, to road under car

Image segmentation



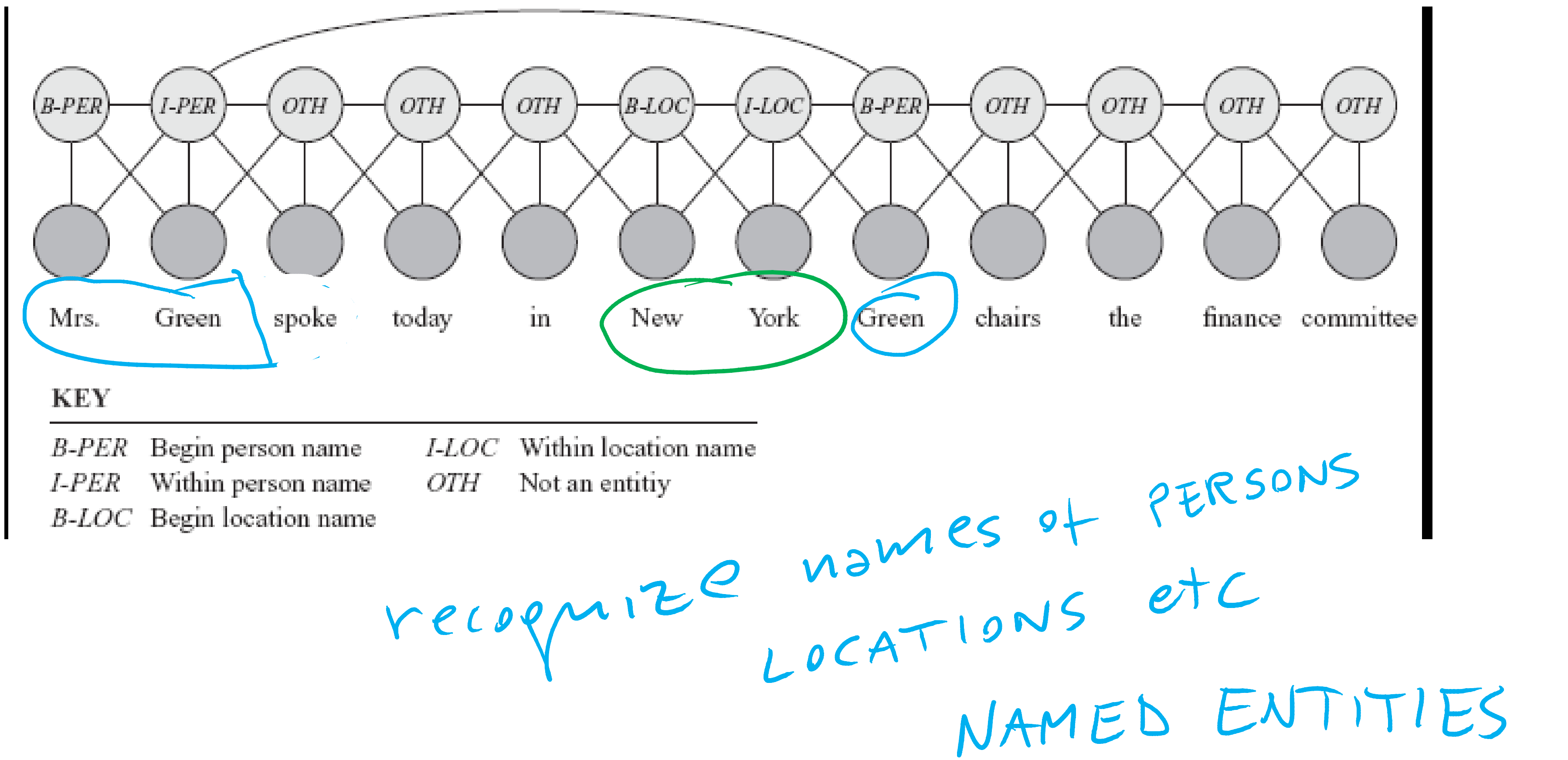
(a) (b) (c) (d)

classifying each superpixel independently

with a Markov Random Field!

Markov Networks Applications (2): Sequence Labeling in NLP and BioInformatics

Conditional random fields (next class Wed)



Learning Goals for today's class

➢ You can:

- Justify the need for undirected graphical model (Markov Networks)
- Interpret local models (factors/potentials) and combine them to express the joint
- Define independencies and Markov blanket for Markov Networks
- Perform Exact and Approx. Inference in Markov Networks
- Describe a few applications of Markov Networks

Midterm, Mon, Oct 26,
we will start at 9am sharp

How to prepare….

- Keep Working on **assignment-2** !
- Go to **Office Hours**
- **Learning Goals** (look at the end of the slides for each lecture – will post complete list)
- Revise all the **clicker questions** and **practice exercises**
- Will post **more practice material** today

How to acquire factors?