CSC535: Probabilistic Graphical Models

Introduction and Course Overview

Prof. Jason Pacheco

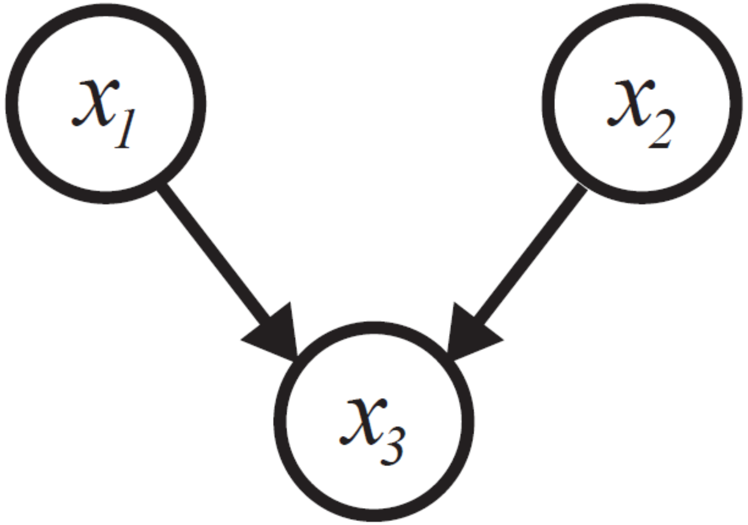
What is a Probabilistic Graphical Model?

*A probabilistic graphical model allows us to pictorially represent a probability distribution**

Probability Model:

$$p(x_1, x_2, x_3) = p(x_1)p(x_2)p(x_3 \mid x_1, x_2)$$

Graphical Model:

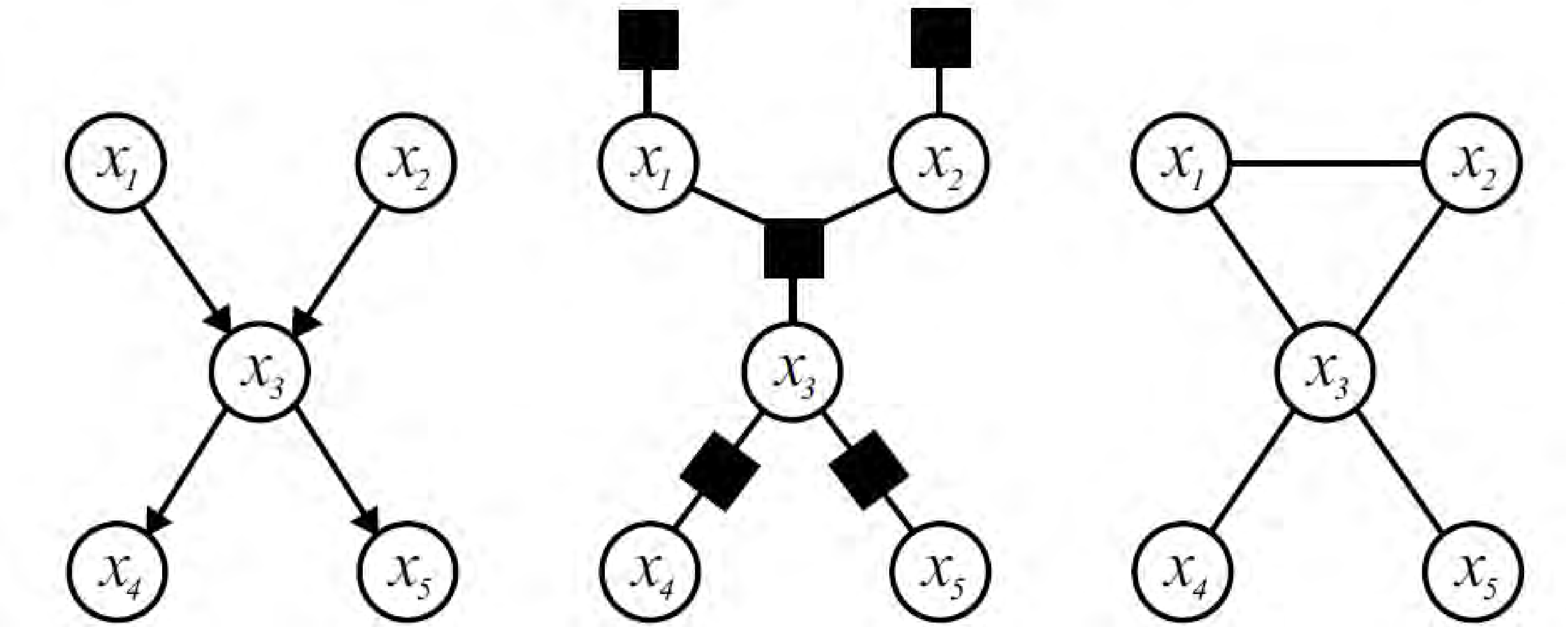


The graphical model structure *obeys* the factorization of the probability function in a sense we will formalize later

* We will use the term "distribution" loosely to refer to a CDF / PDF / PMF

Graphical Models

A variety of graphical models can represent the same probability distribution



Bayes Network **Factor Graph** **Markov Random Field**

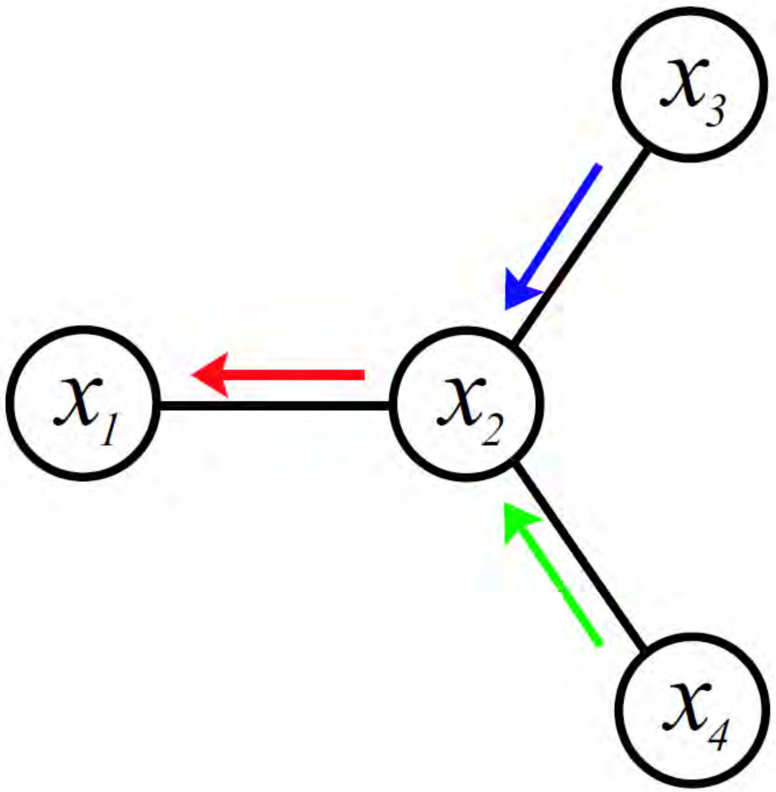
Directed Models

Undirected Models

Erik Sudderth, PhD Thesis

Why Graphical Models?

Structure simplifies both **representation** and **computation**



Representation

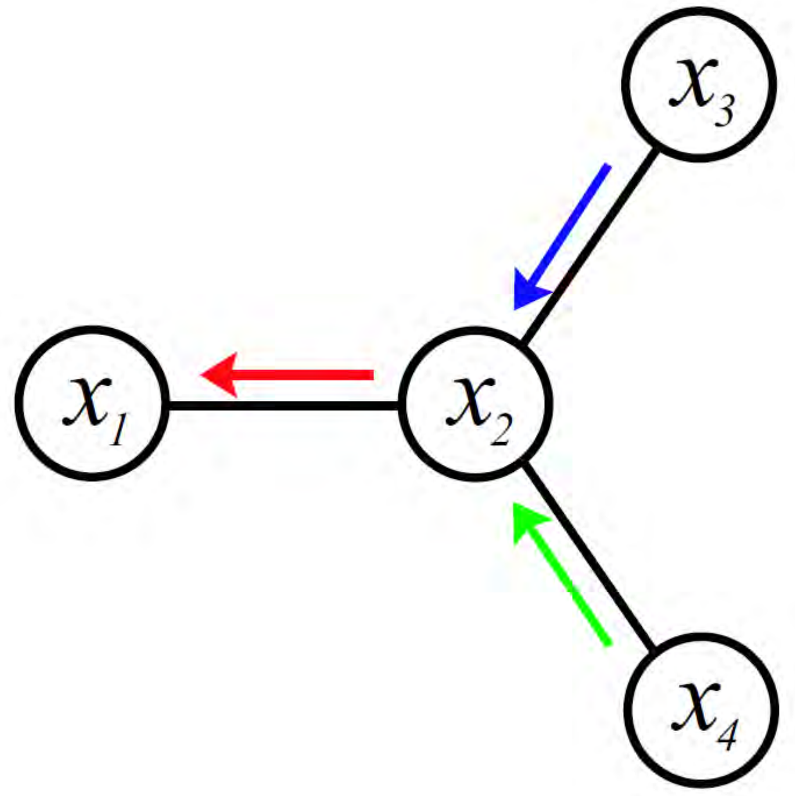
Complex global phenomena arise by simpler-to-specify local interactions

Computation

Inference / estimation depends only on subgraphs (e.g. dynamic programming, belief propagation, Gibbs sampling)

Why Graphical Models?

Structure simplifies both **representation** and **computation**



Representation
Complex global phenomena arise by simpler-to-specify local interactions

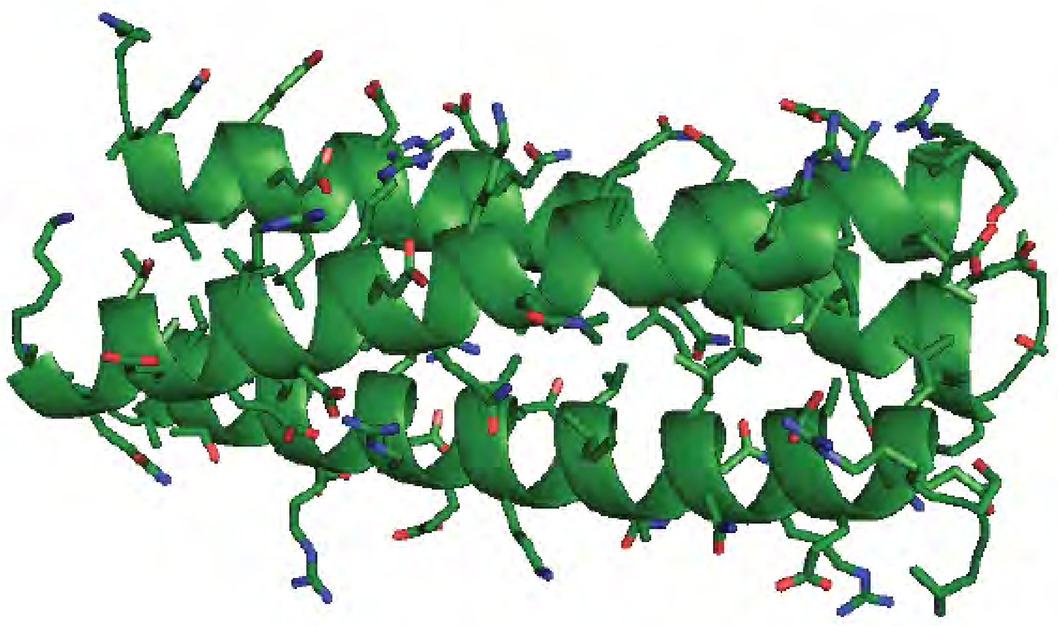
Computation
Inference / estimation depends only on subgraphs (e.g. dynamic programming, belief propagation, Gibbs sampling)

We will discuss inference later, but let's focus on representation…

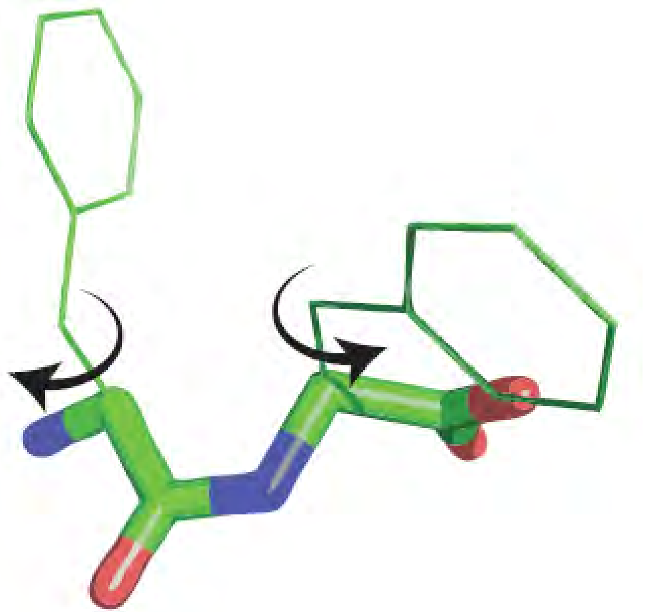
Protein Side Chain Prediction

Problem: Given 3D protein backbone structure, estimate orientation of every side chain molecule.

Backbone + Side Chains



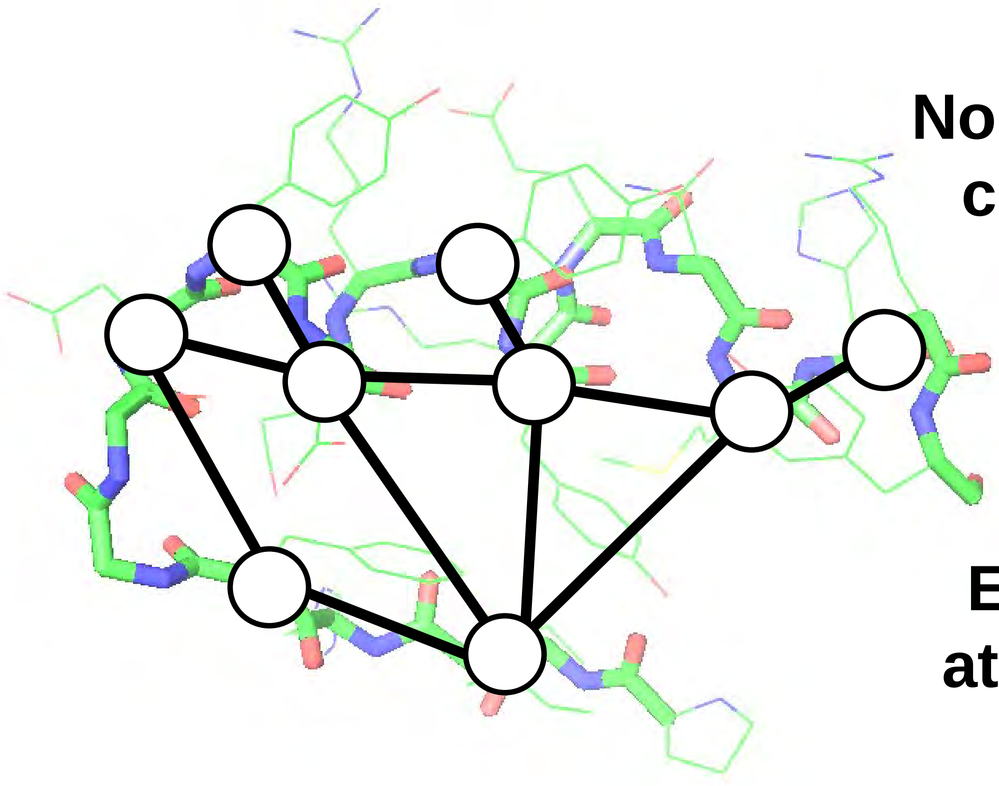
Side Chain Rotation



Solution: Just physics of atomic interaction. Easy, right!?

Protein Side Chain Prediction

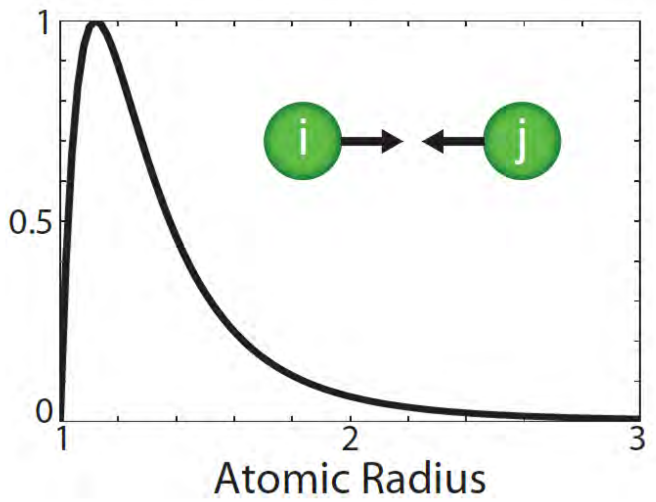
Graphical Model

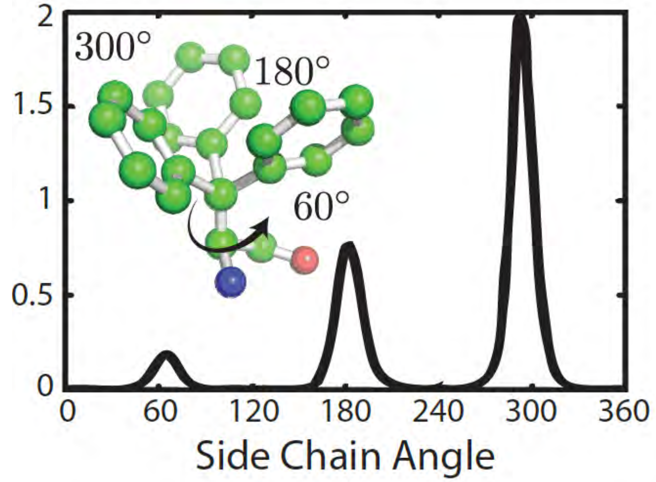


Nodes represent side chain orientations

Edges represent atomic interaction

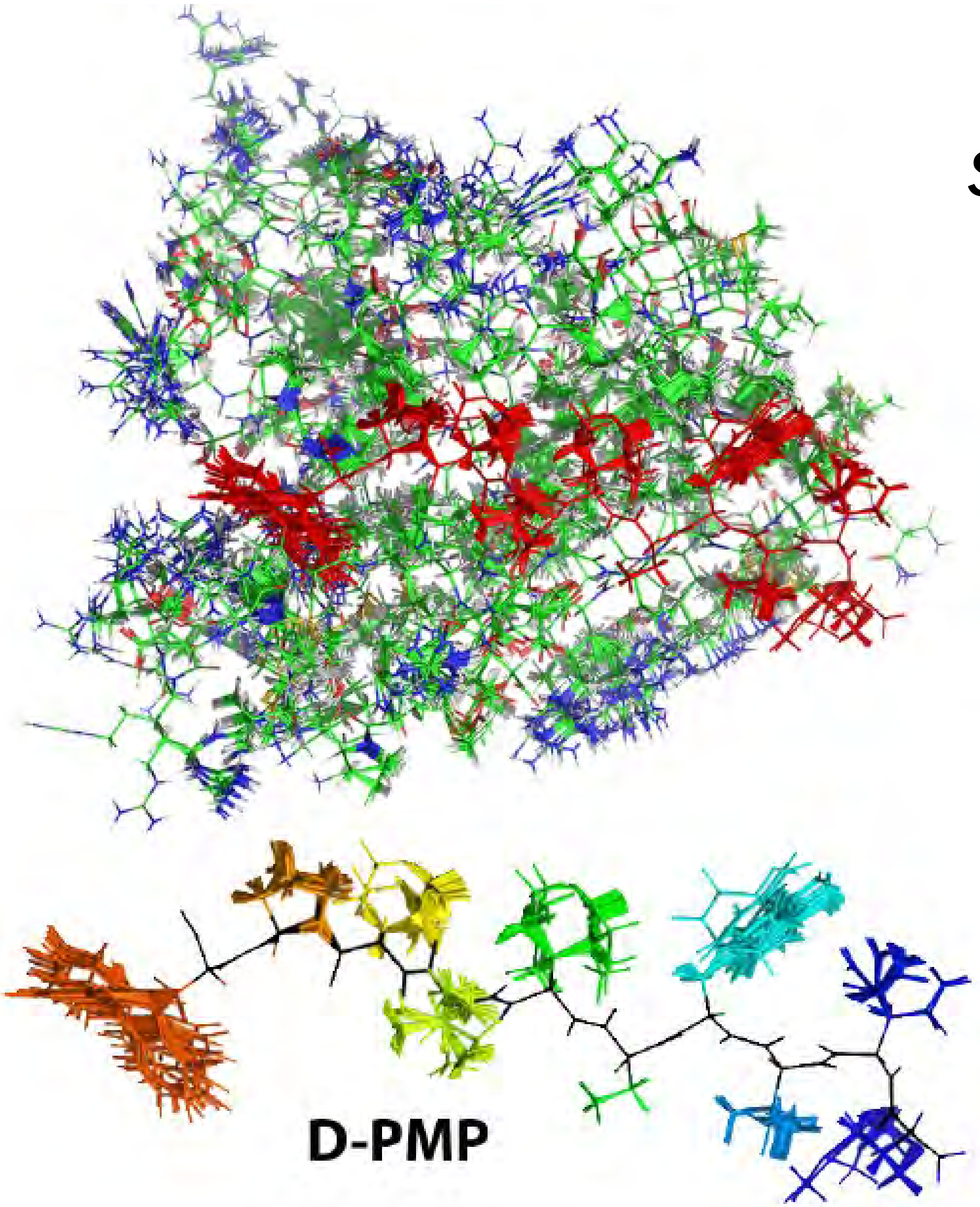
Configuration Likelihoods



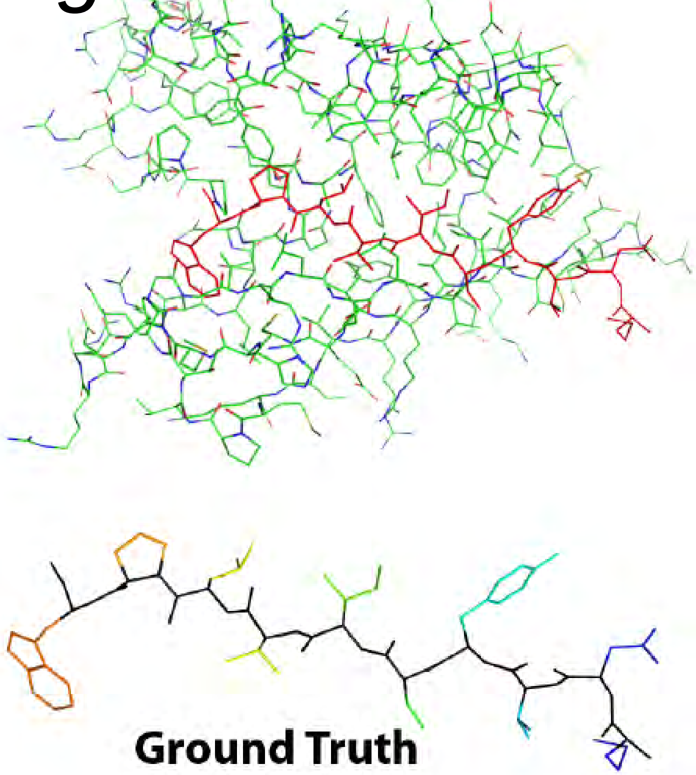


Complex phenomena specified by simpler atomic interactions

Protein Side Chain Prediction



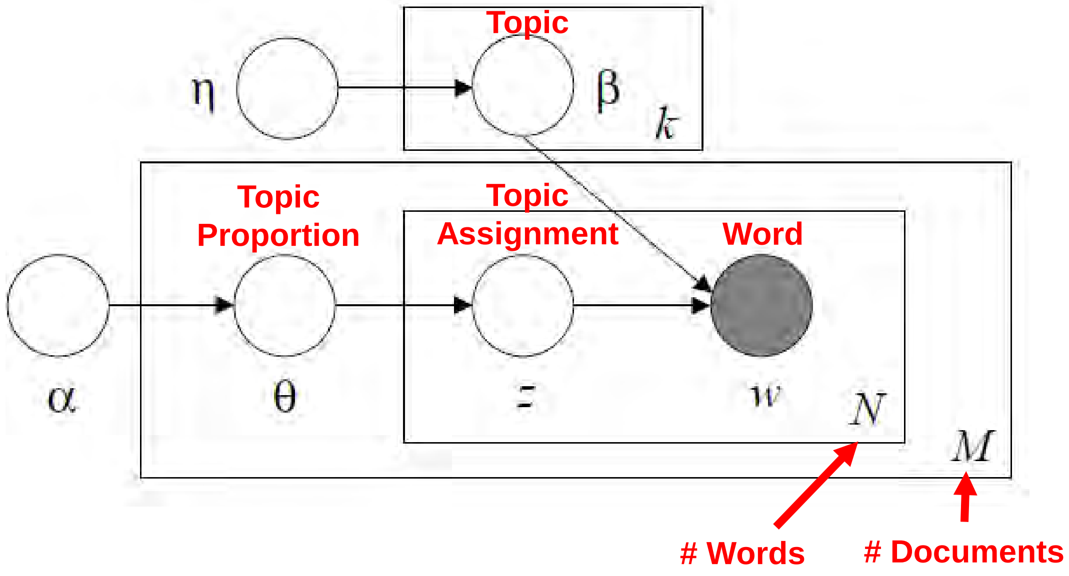
By exploiting graphical model structure we can scale computation to large macromolecules



Pacheco and Sudderth, ICML 2015

Topic Models

Latent Dirichlet Allocation (LDA)



Allows <u>unsupervised learning</u> of document corpus via mixture modeling

"Arts"	"Budgets"	"Children"	"Education"
NEW	MILLION	CHILDREN	SCHOOL
FILM	TAX	WOMEN	STUDENTS
SHOW	PROGRAM	PEOPLE	SCHOOLS
MUSIC	BUDGET	CHILD	EDUCATION
MOVIE	BILLION	YEARS	TEACHERS
PLAY	FEDERAL	FAMILIES	HIGH
MUSICAL	YEAR	WORK	PUBLIC
BEST	SPENDING	PARENTS	TEACHER
ACTOR	NEW	SAYS	BENNETT
FIRST	STATE	FAMILY	MANIGAT
YORK	PLAN	WELFARE	NAMPHY
OPERA	MONEY	MEN	STATE
THEATER	PROGRAMS	PERCENT	PRESIDENT
ACTRESS	GOVERNMENT	CARE	ELEMENTARY
LOVE	CONGRESS	LIFE	HAITI

The William Randolph Hearst Foundation will give $1.25 million to Lincoln Center, Metropolitan Opera Co., New York Philharmonic and Juilliard School. "Our board felt that we had a real opportunity to make a mark on the future of the performing arts with these grants an act every bit as important as our traditional areas of support in health, medical research, education and the social services," Hearst Foundation President Randolph A. Hearst said Monday in announcing the grants. Lincoln Center's share will be $200,000 for its new building, which will house young artists and provide new public facilities. The Metropolitan Opera Co. and New York Philharmonic will receive $400,000 each. The Juilliard School, where music and the performing arts are taught, will get $250,000. The Hearst Foundation, a leading supporter of the Lincoln Center Consolidated Corporate Fund, will make its usual annual $100,000 donation, too.

Blei, D. et al., JMLR 2003

Pose Estimation

Graphical Model

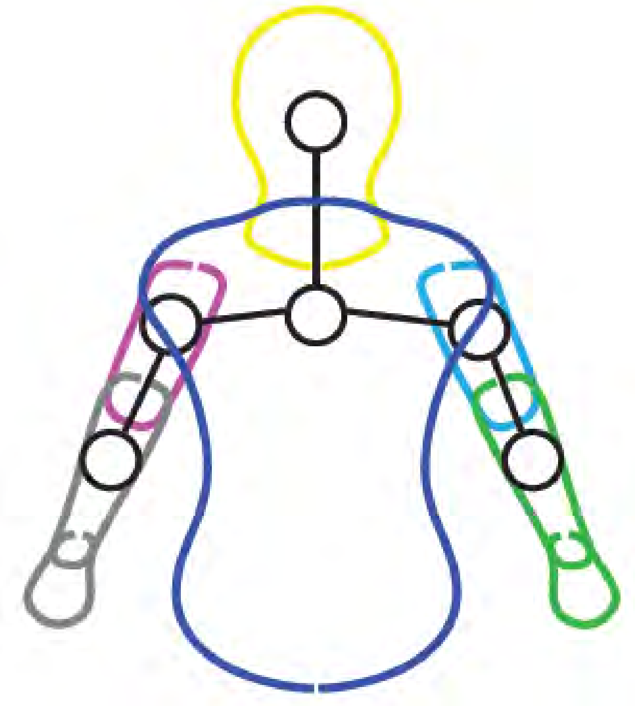
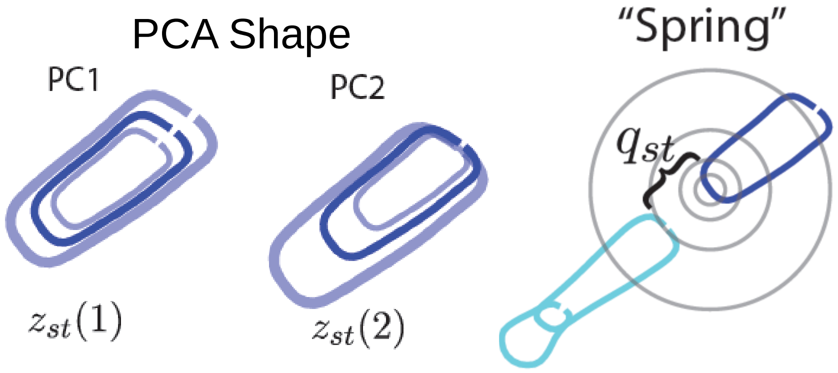
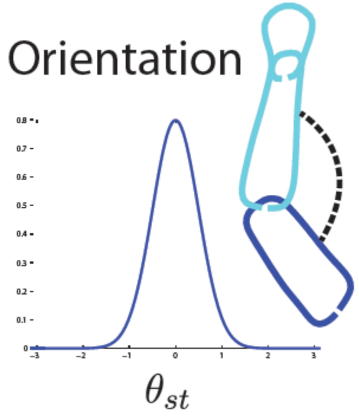


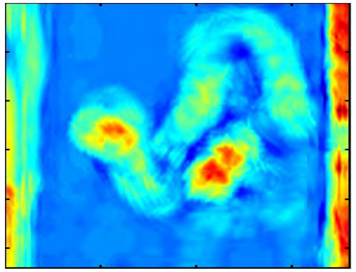
Image (Data / Observation)

Model encodes likelihood of shape / pose / image consistency (e.g. skin color)





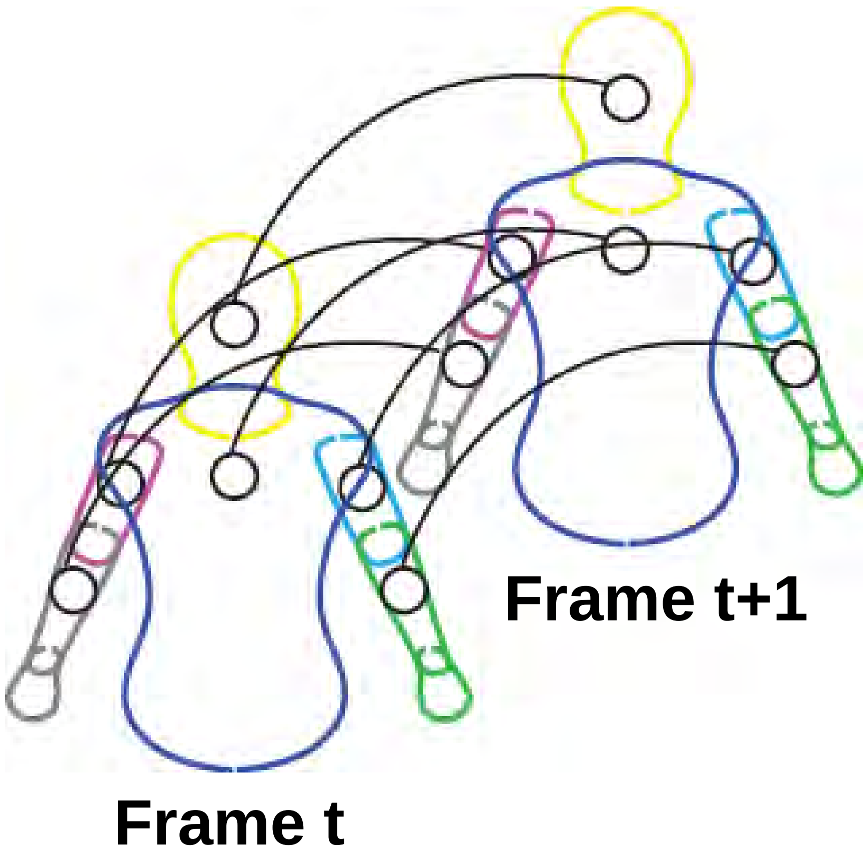
Skin Tone

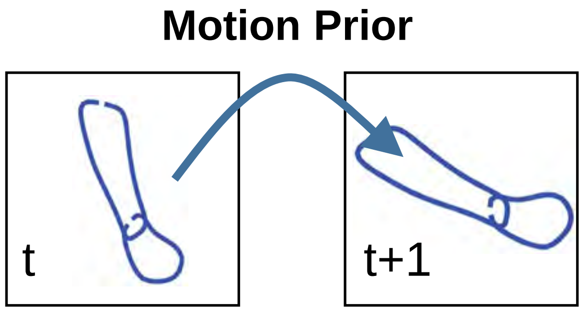


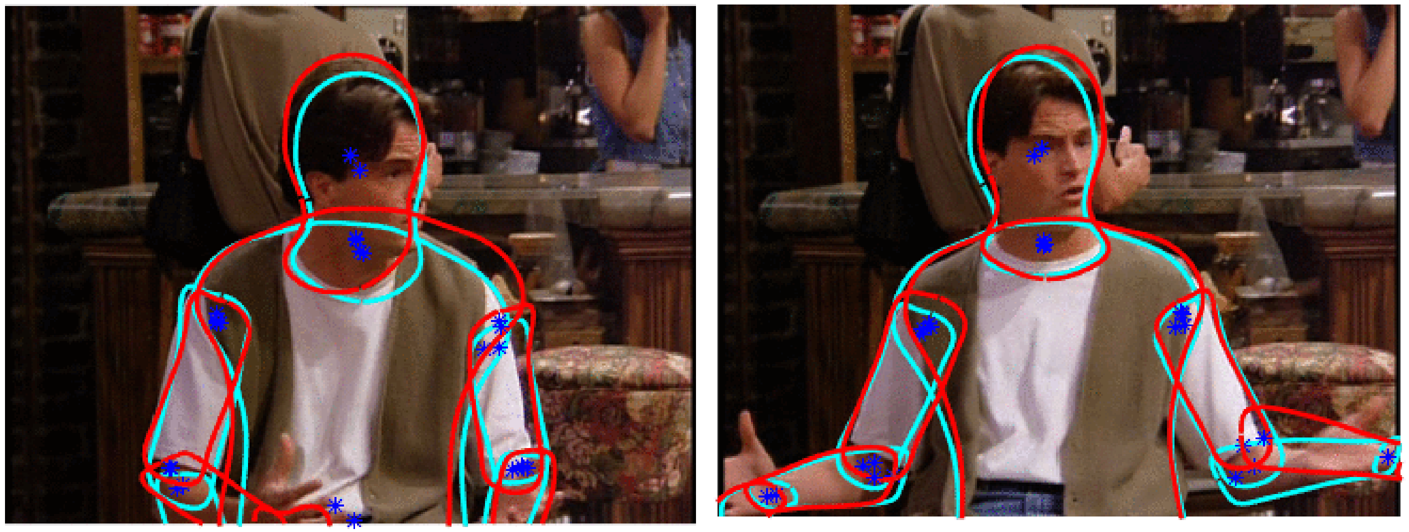
Problem: Estimate orientation / shape / pose of human figure from an image

Pacheco, et al., NeurIPs 2014

Pose Tracking

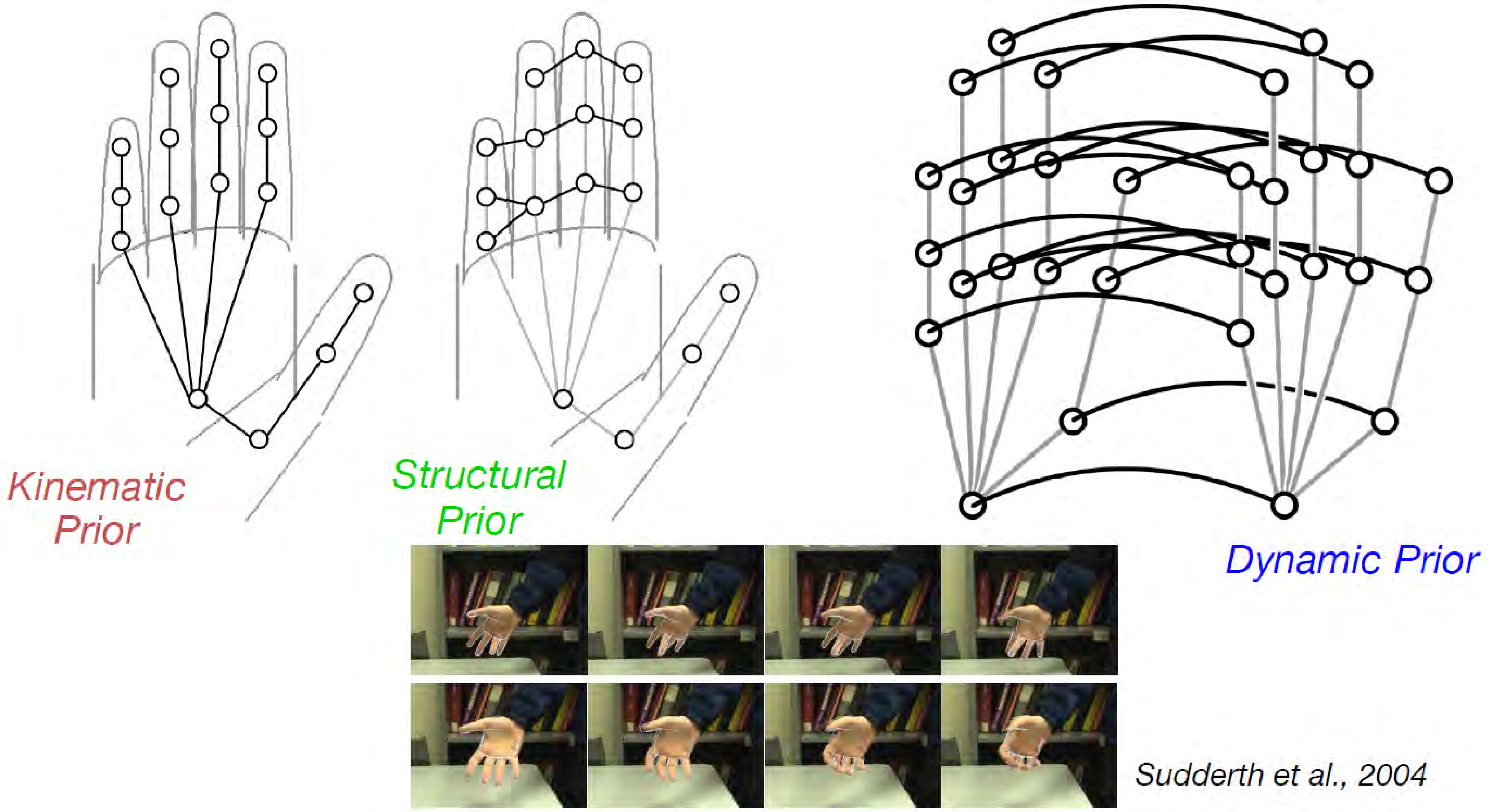






By composing single-frame model with temporal dynamics and motion prior we can do video tracking…

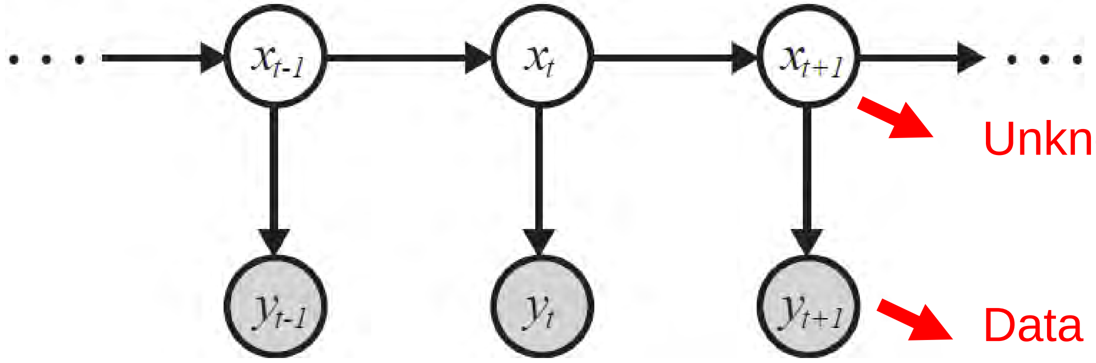
Kinematic Hand Tracking



Sudderth et al., 2004

Hidden Markov Models

Sequential models *of <u>discrete</u> quantities of interest*

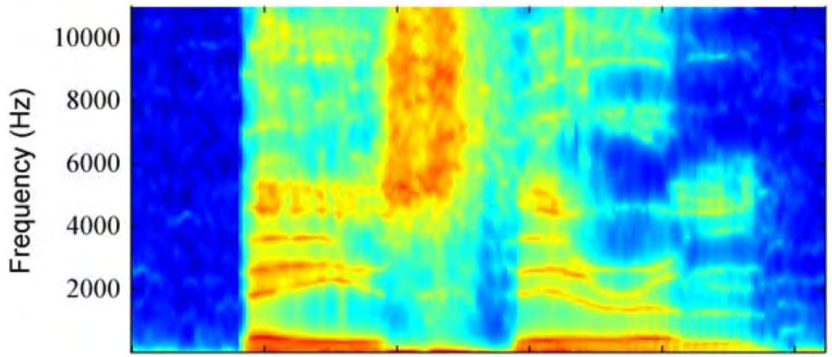


Unknowns

Data

Example: Speech Recognition

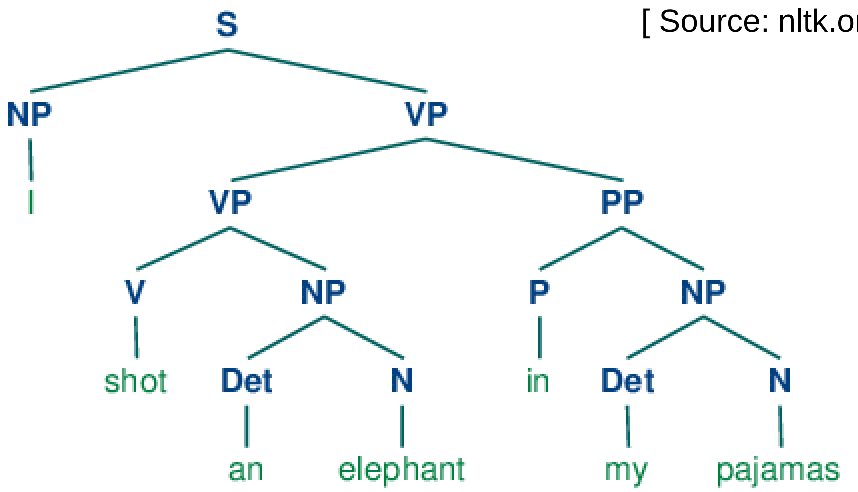
$y =$



x = b-ey-z-th-ih-er-em → Bayes' Theorem

[Source: Bishop, PRML]

[Source: nltk.org]



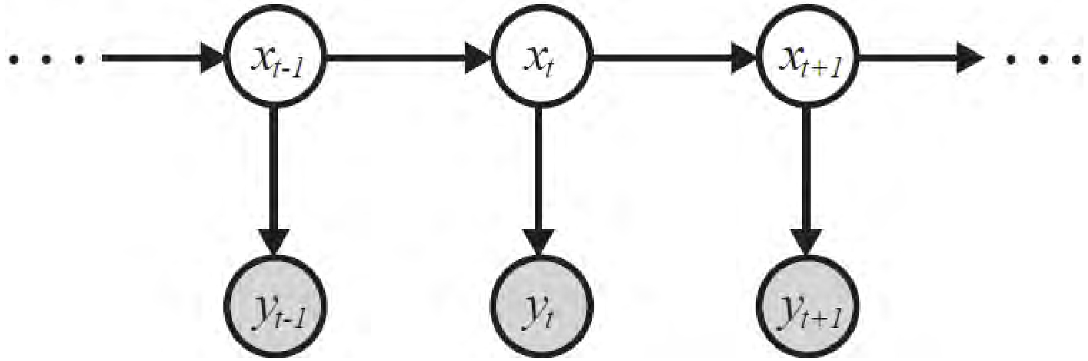
Example: Part-of-speech Tagging:

y = "I shot an elephant in my pajamas."

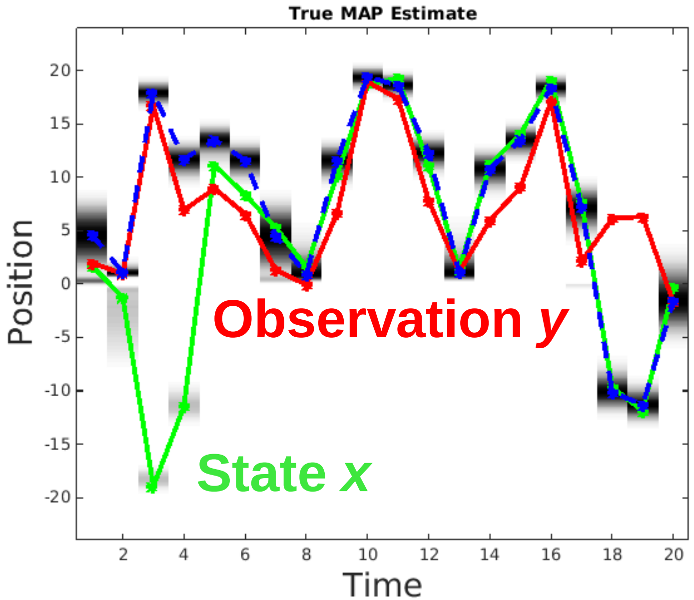
x = NP-V-Det-N-P-Det-N

Dynamical Models

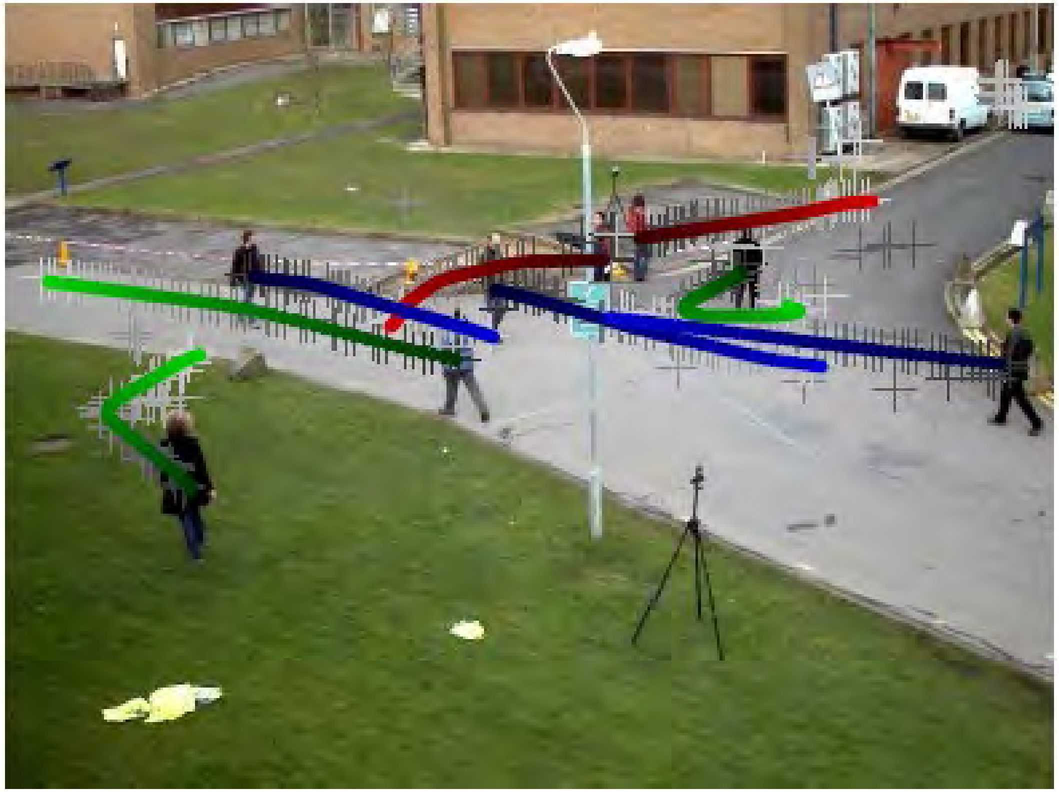
Sequential models of *<u>continuous</u> quantities of interest*



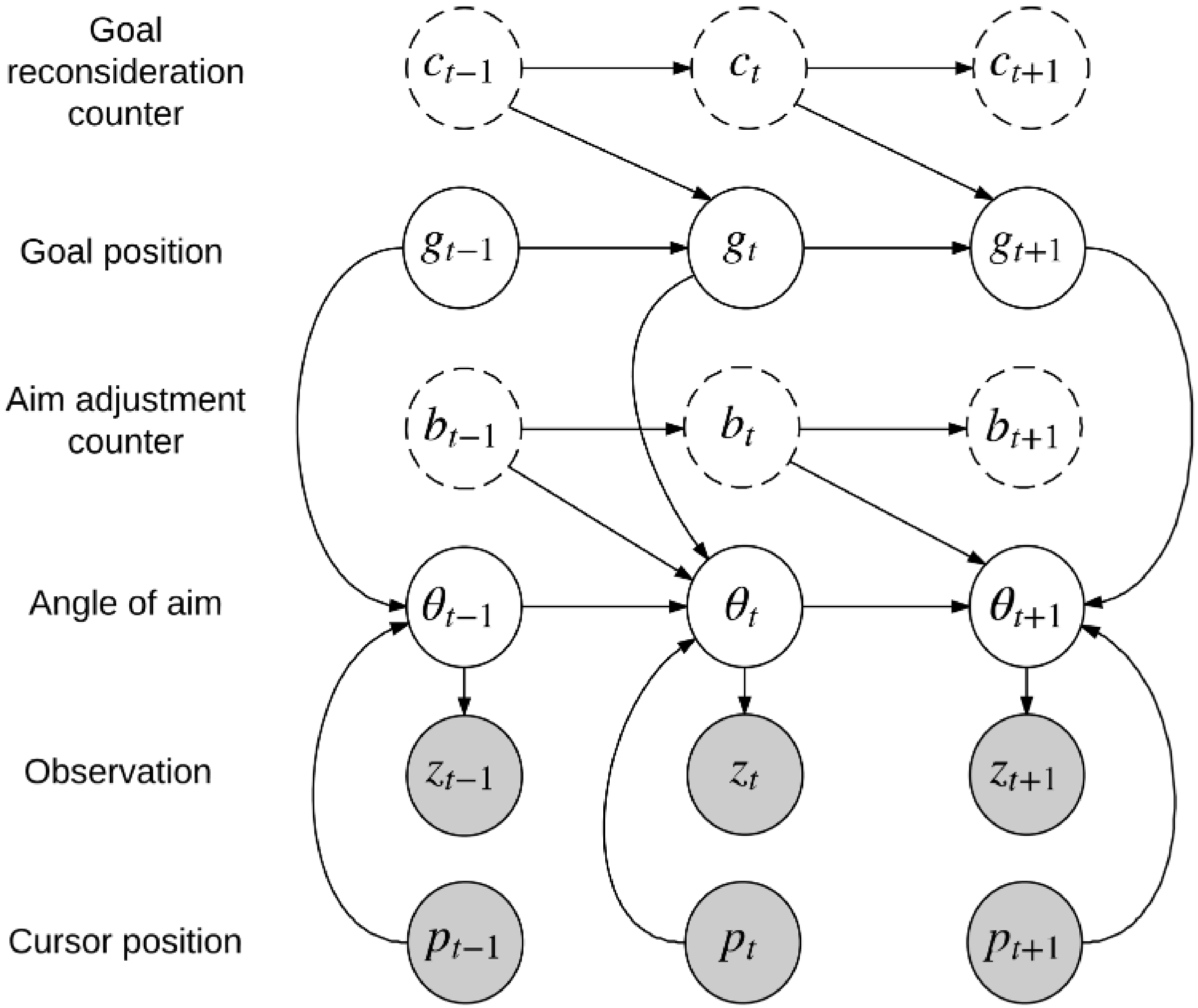
Example: Nonlinear Time Series



Example: Multitarget Tracking



State-Space Models



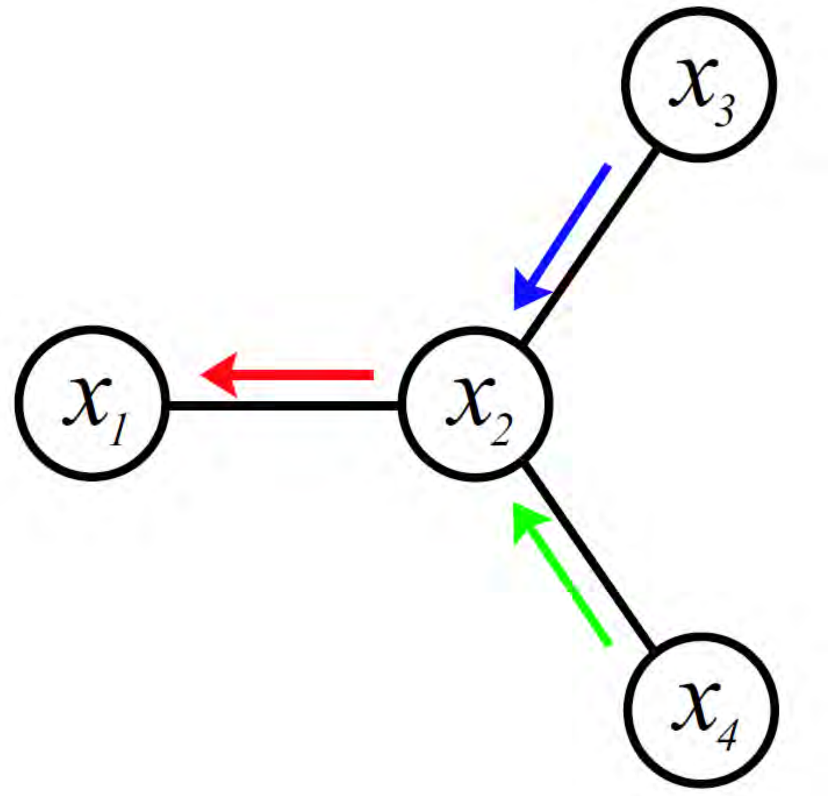
Intracortical Brain-Computer Interface

Block 12: "Multiscale Semi-Markov Model"

Milstein, Pacheco, et al., NeurIPs 2017

Why Graphical Models?

Structure simplifies both **representation** and **computation**



Representation

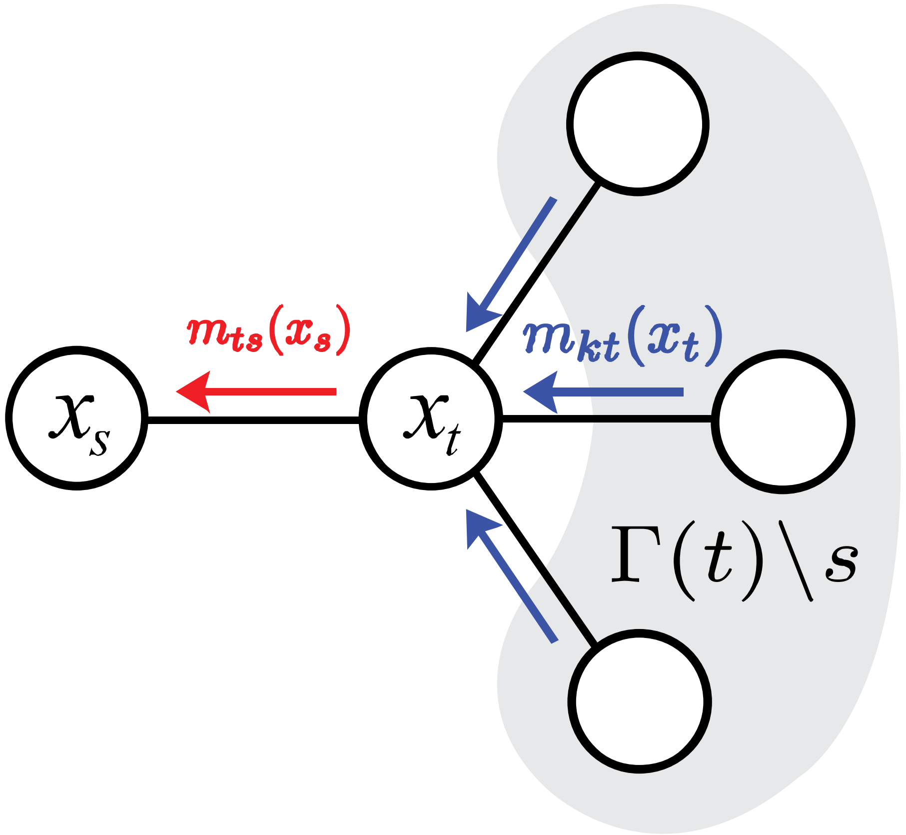
Complex global phenomena arise by simpler-to-specify local interactions

Computation

Inference / estimation depends only on subgraphs (e.g. dynamic programming, belief propagation, Gibbs sampling)

Computation in Graphical Models

This style of computation generalizes to all graphical models…

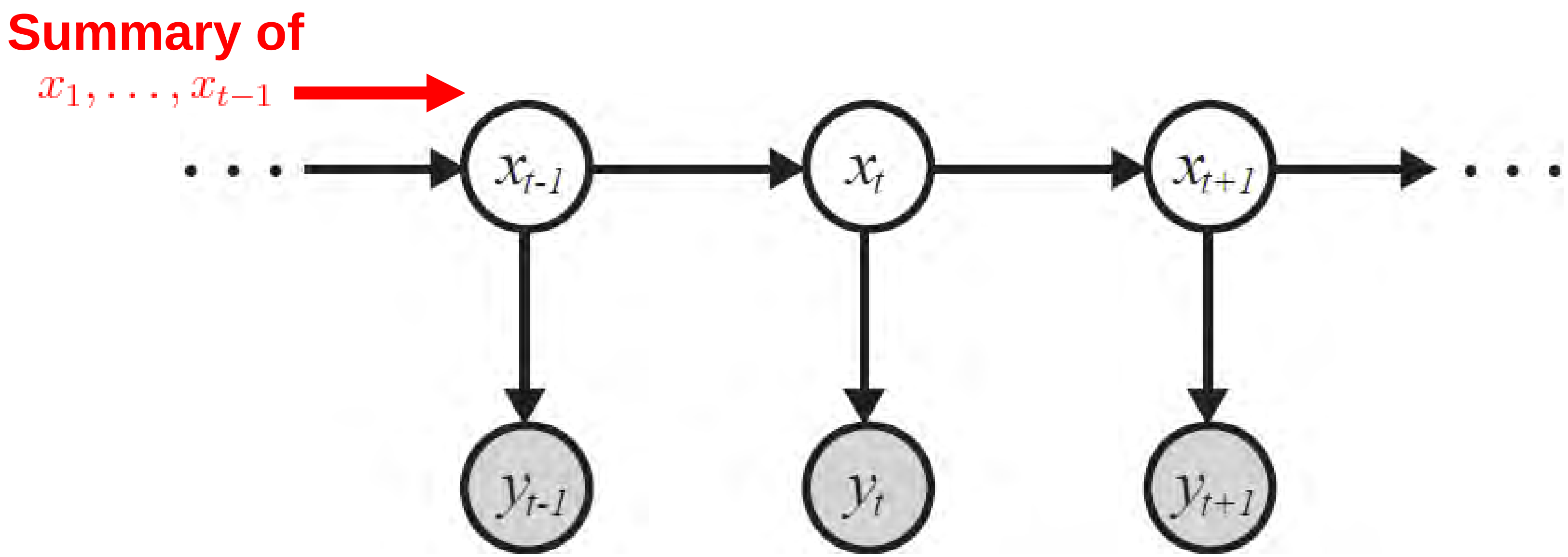


Example algorithms

- Belief propagation
- Gibbs sampling
- Particle filtering
- Viterbi decoder for HMMs
- Kalman filter (marginal inference)

Key Idea: Local computations only depend on the statistics of <u>the current node</u> and <u>neighboring interactions</u>

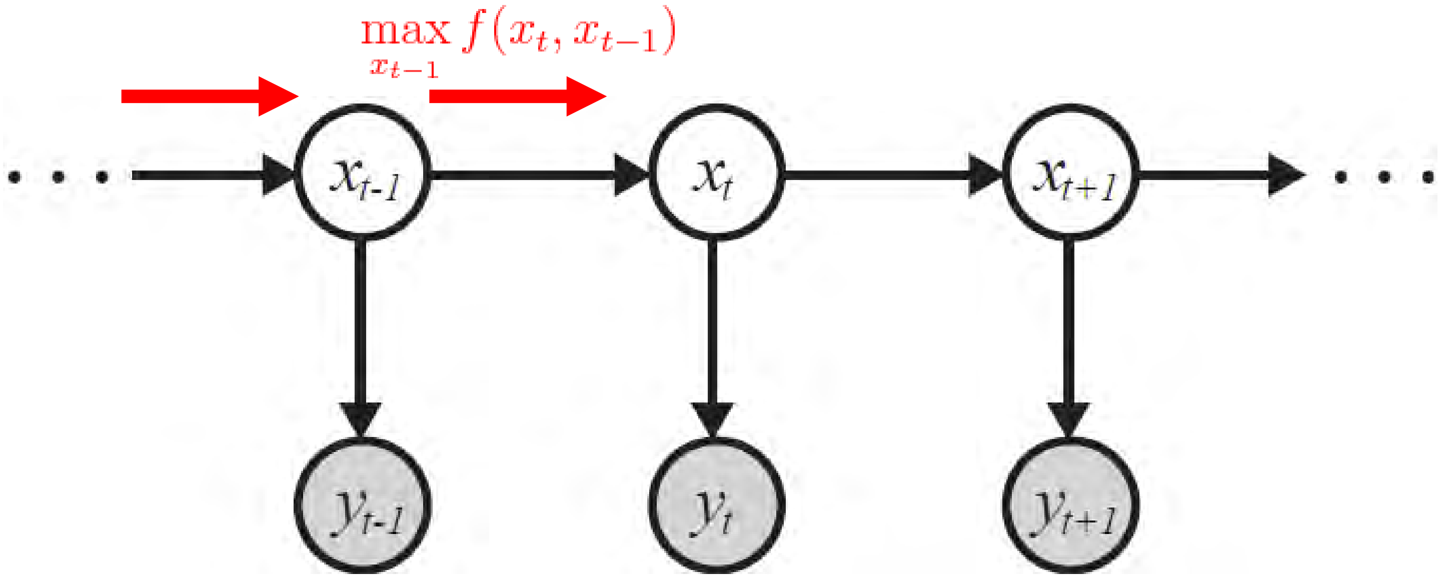
Viterbi Decoder



$$x^* = \underset{x}{\operatorname{argmax}}\, p(x \mid y)$$

Efficiently computes MAP estimate for state-space model by *passing messages* forward and backward along chain.

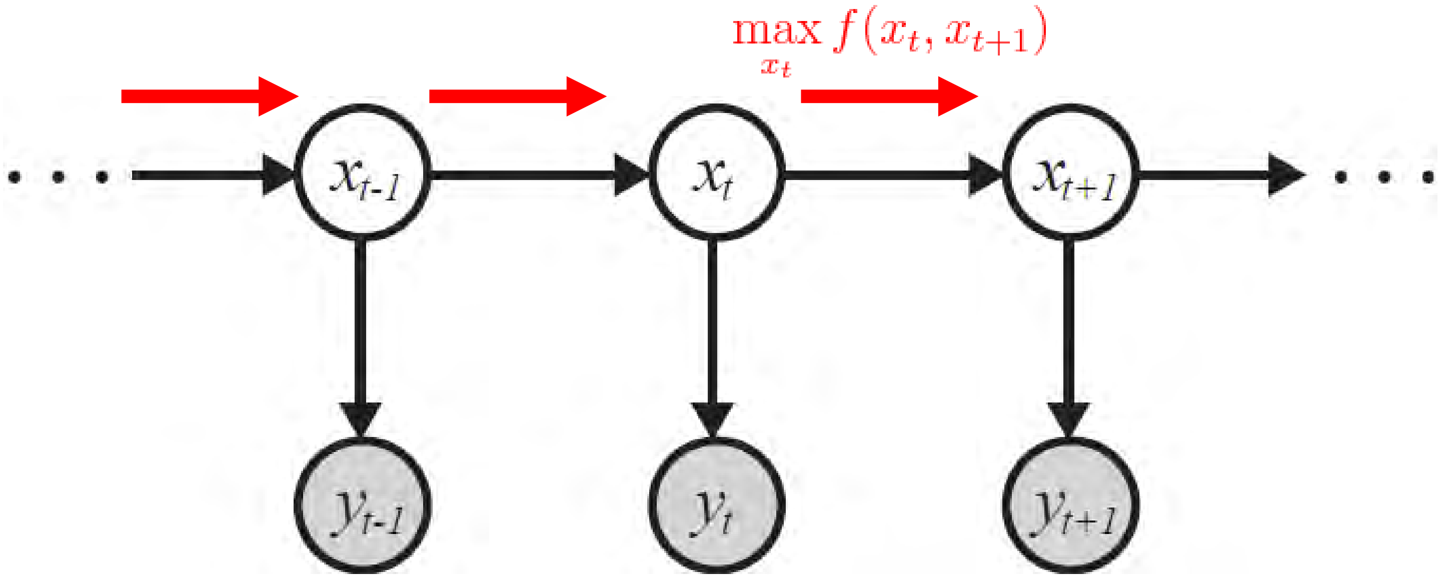
Viterbi Decoder



$$x^* = \operatorname*{argmax}_x p(x \mid y)$$

Efficiently computes MAP estimate for state-space model by *passing messages* forward and backward along chain.

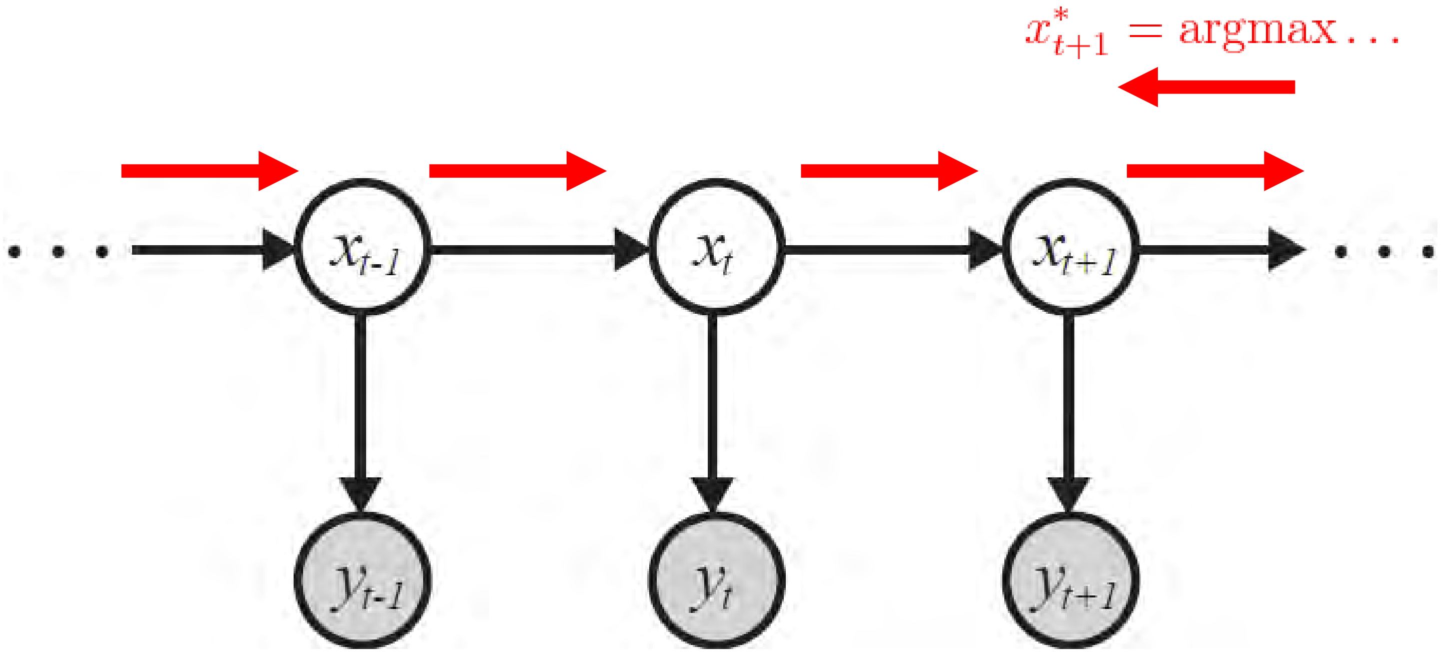
Viterbi Decoder



$$x^* = \underset{x}{\operatorname{argmax}}\, p(x \mid y)$$

Efficiently computes MAP estimate for state-space model by *passing messages* forward and backward along chain.

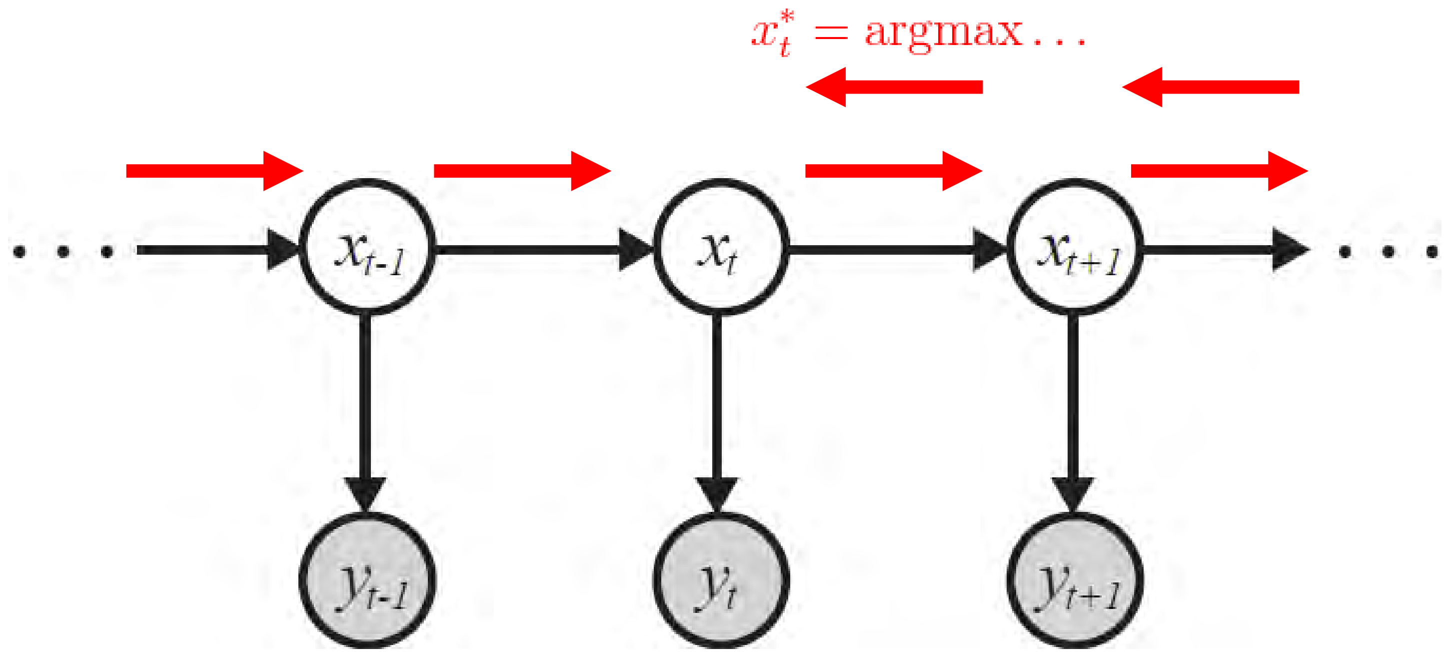
Viterbi Decoder



$$x^* = \underset{x}{\text{argmax}}\; p(x \mid y)$$

Efficiently computes MAP estimate for state-space model by *passing messages* forward and backward along chain.

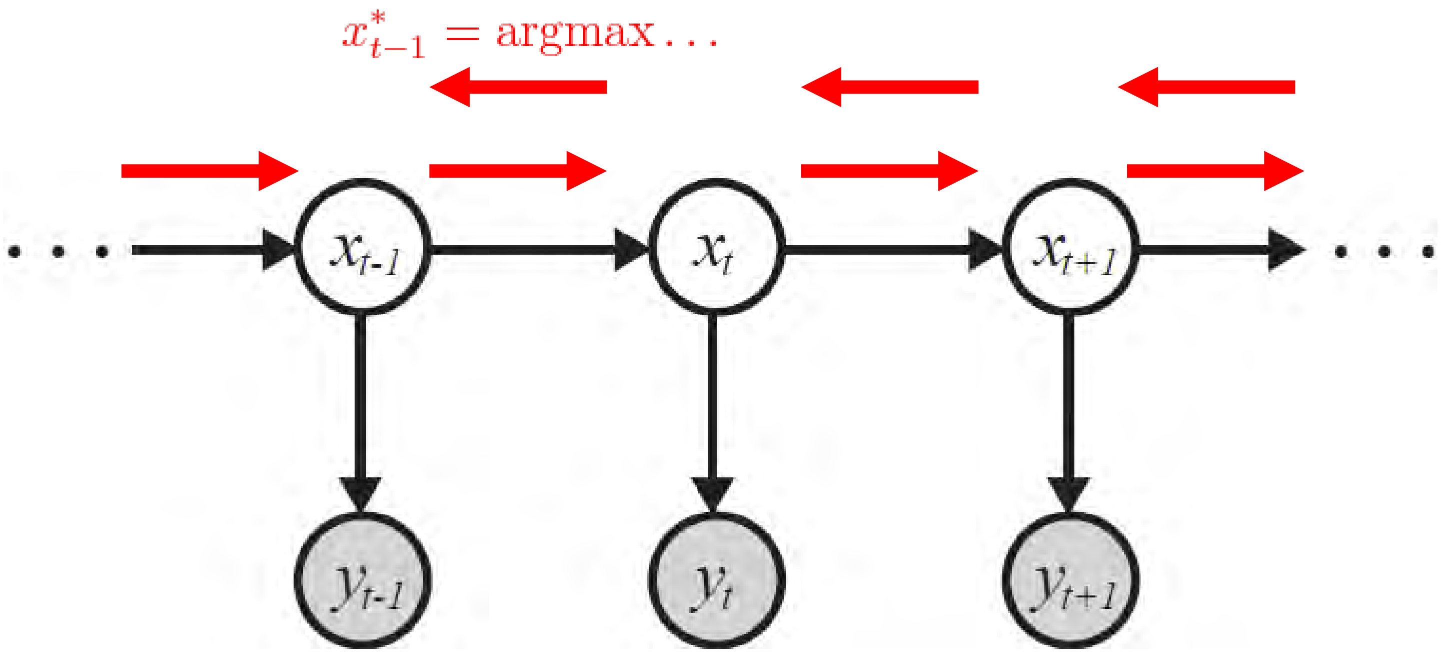
Viterbi Decoder



$$x^* = \underset{x}{\text{argmax}}\, p(x \mid y)$$

Efficiently computes MAP estimate for state-space model by *passing messages* forward and backward along chain.

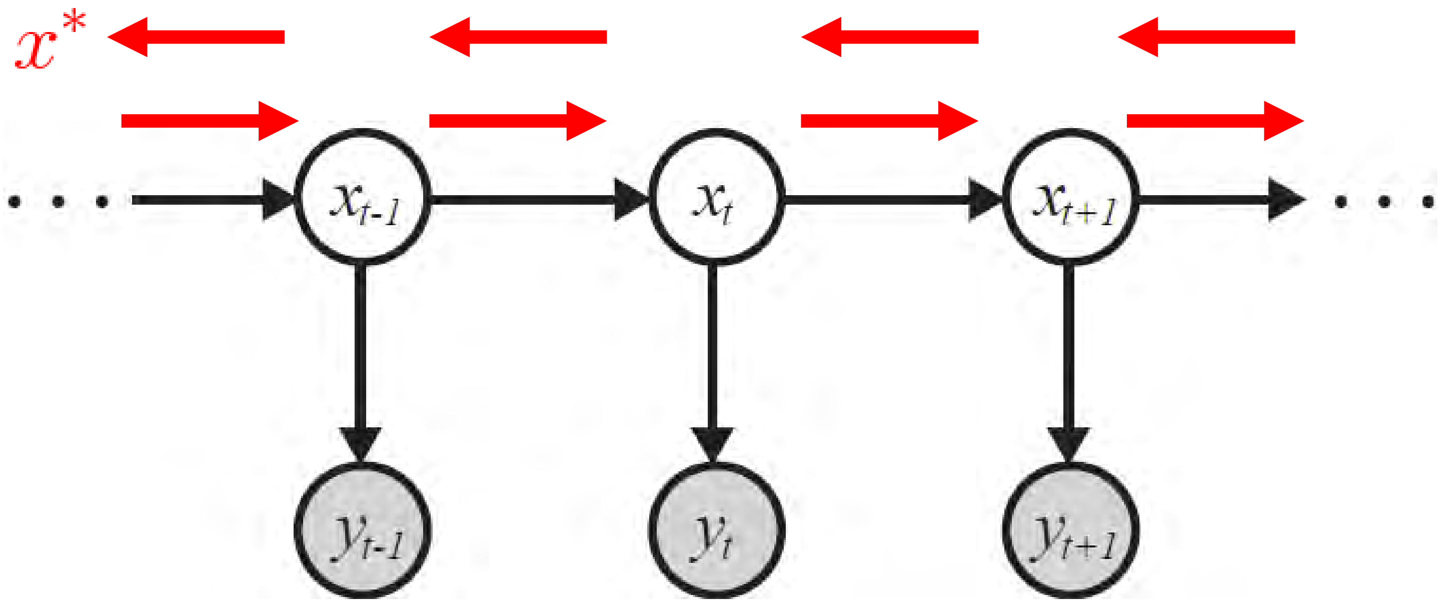
Viterbi Decoder



$$x^* = \underset{x}{\text{argmax}}\, p(x \mid y)$$

Efficiently computes MAP estimate for state-space model by *passing messages* forward and backward along chain.

Viterbi Decoder



$$x^* = \operatorname*{argmax}_x p(x \mid y)$$

Efficiently computes MAP estimate for state-space model by *passing messages* forward and backward along chain.

Course Objectives

Along with basic familiarity of PGMs, we will develop the following basic skills…

- Create directed / undirected graphical models of stochastic processes
- Identify conditional independencies in graphical models
- Apply exact inference to compute marginal probabilities and maximally probable configurations given a model (elimination, sum-product, and max-sum algorithms)
- Apply approximate inference to learn model parameters using expectation maximization (EM), variational inference, and various Monte Carlo methods

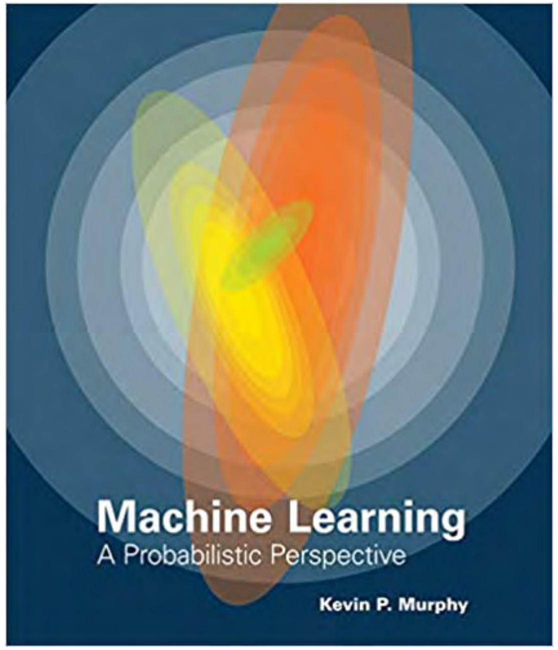
Course Overview

We will cover **six** primary topics…

Intro Sequence	Message Passing Algorithms	Parameter Learning	Monte Carlo Methods	Dynamical Systems	Variational Inference
Probability primer, Bayesian statistics, PGMs, Exponential families	Elimination, Junction tree, Sum-product / max-product, Belief propagation, Viterbi decoding	Maximum likelihood, Maximum a posteriori, Expectation Maximization (EM)	Rejection sampling, Importance sampling, Metropolis-Hastings, Gibbs	Linear and switching state-space models, Kalman filter, Particle filter	Mean field, Stochastic variational, Bethe energy methods

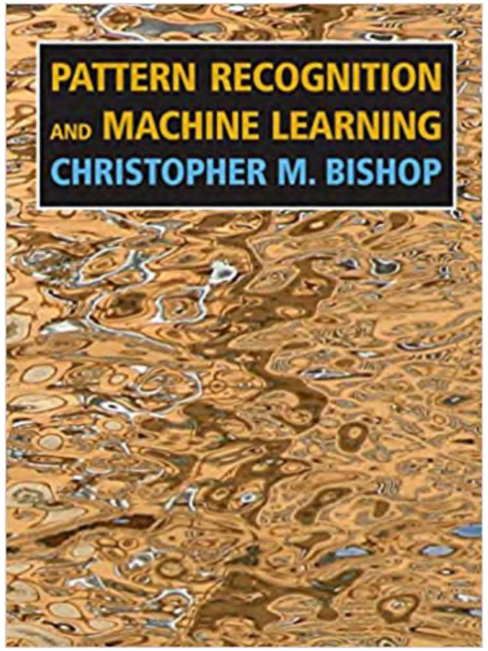
Textbooks

Both texts are available electronically online…



Murphy, K. "Machine Learning: A Probabilistic Perspective." MIT press, 2012

[(UA Library)]



Bishop, C. "Pattern Recognition and Machine Learning." Springer, 2006

[(Microsoft)]

Additional readings on the course webpage

Online Resources

All material (lectures / HWs / readings) are available on the **course webpage**:

http://pachecoj.com/courses/csc535_fall20/

We will use **D2L** for Zoom links, submitting assignments, grades:

https://d2l.arizona.edu/d2l/home/937505

We will use **Piazza** for discussion:

https://piazza.com/arizona/fall2020/csc535

Assignments

Current plan

- 10 HW assignments (every 2-2.5wks)
- *“Take-home”* midterm and final

Our plan may get modified as things progress…

Programming Language

- Some HWs will provide **Matlab** starter code
- You can use Python/R/etc. but will need to re-implement handout code
- If anyone *does* re-implement handout code in Python I *will* consider extra credit

Grading Breakdown

Grading Rubric

- 65% - Homeworks
- 10% - Midterm
- 25% - Final

Collaboration Policy:

- HWs should be done individually, but you may discuss with others
- Collaboration on HWs is fine, but <u>cheating is not</u>!
- I will ask you to disclose HW collaborators
- NO COLLABORATION on Midterm / Final exams

Lectures and Attendance

Synchronous (Live) Zoom Lectures

- ➢ Attendance during Zoom lectures is **highly encouraged**
- ➢ As per College of Science policy I <u>will not grade attendance</u>

Asynchronous Lectures

- ➢ All lectures will be recorded and posted online after-the-fact
- ➢ I will ensure personally identifying info is not recorded
- ➢ PRO: You can watch me at *“chipmunk speed”*
- ➢ CON: Can’t participate in discussion or ask questions

IMPORTANT: If anticipate that you will not be able to attend synchronous lectures (e.g. due to timezone constraints) message me on Piazza and *briefly* outline your situation so that I am aware

“Office” Hours

Bottom Line: Online learning is not ideal, and I plan to try and make up for this with flexible office hours.

Some details to be worked how, but here’s what I know…

- I will hold 1.5hrs/wk **minimum**
- Office hours will be held on Zoom, but schedule is TBD
- I plan to **be flexible** in trying to adapt to people’s scheduling constraints (message me on Piazza if you have difficult constraints)

Still deciding on scheduling and format (e.g. individual, group, etc.)

Mental Well-Being

Some level of stress / depression / anxiety is normal, but sometimes you may need extra help

- Non-emergency UA resources at Counseling & Psych Services Mon-Fri
  - Phone: 520-621-3334
  - Web: https://health.arizona.edu/counseling-psych-services
- Emergency resources in Tucson in this Google Doc

I am happy and point you in the right direction, but keep in mind that I am not a mental health professional

Inclusivity

I want to foster a comfortable and inclusive classroom experience

Please let me know if you feel excluded in any way, e.g.

- “Alice-and-Bob” style examples of material
- Improper use of pronouns
- Microagressions
- Miscellaneous statements / interactions

You can message me anonymously on Piazza