Computer Science

CSC 665-1: Advanced Topics in Probabilistic Graphical Models

Introduction and Course Overview

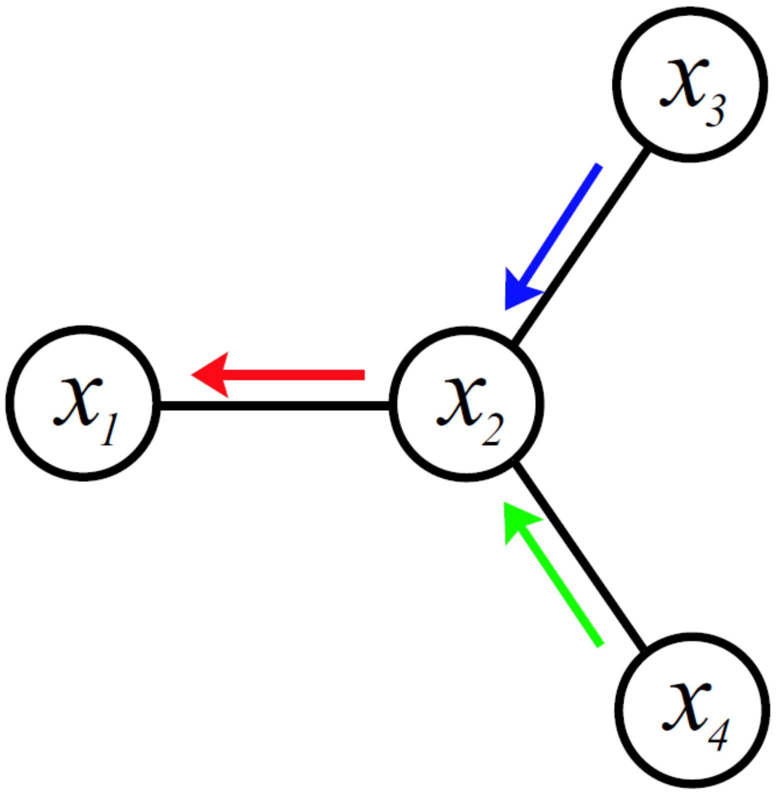
Instructor: Prof. Jason Pacheco

Outline

- Motivating examples of representation
- Efficient computation on graphical models
- Overview of course topics
- Course details (attendance, grading, etc.)

Why Graphical Models?

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



Representation

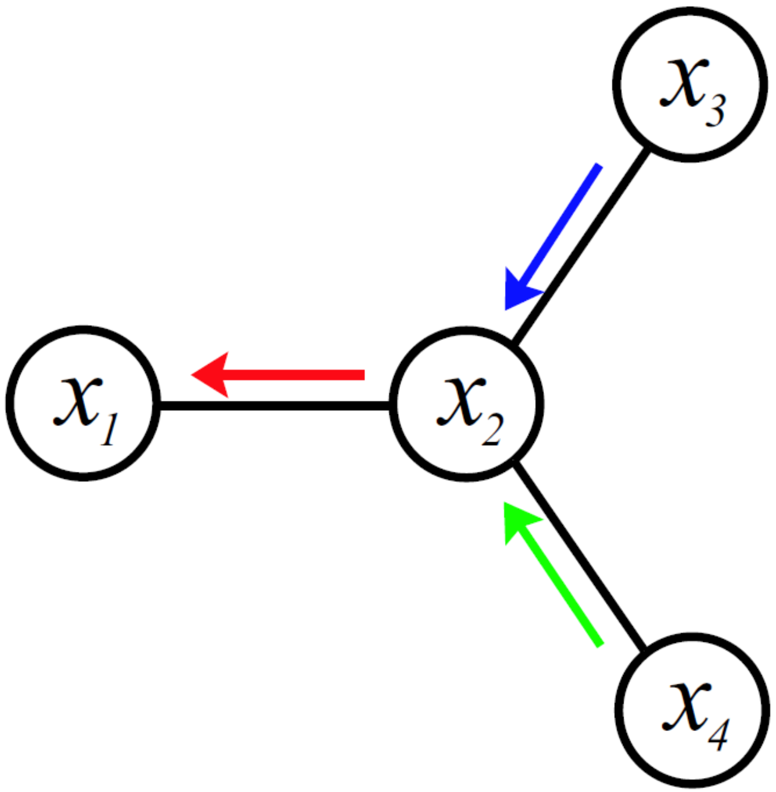
Complex global phenomena arise by simpler-to-specify local interactions

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…

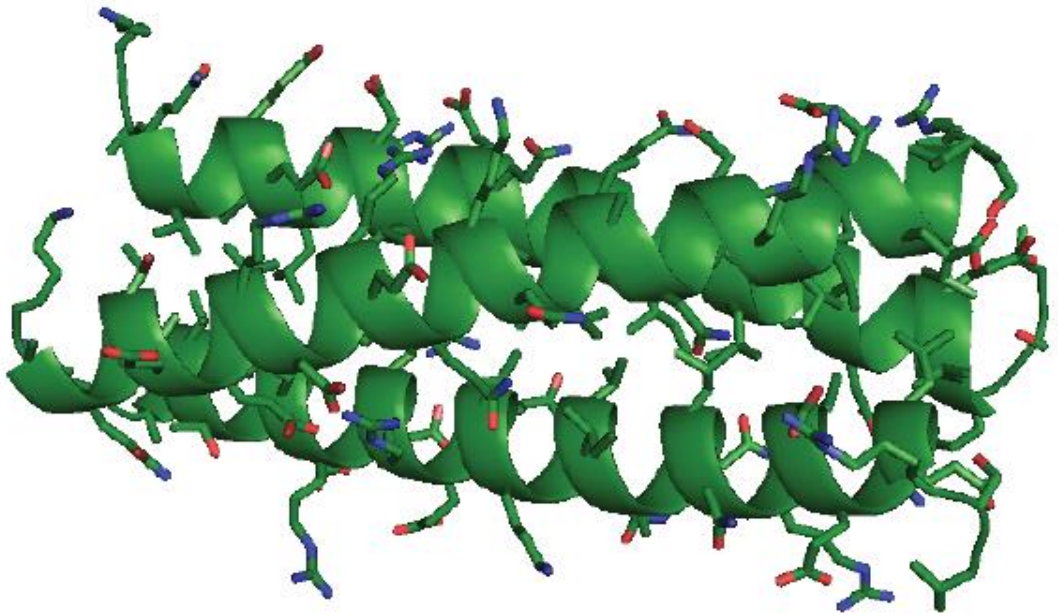
Outline

- Motivating examples of representation
- Efficient computation on graphical models
- Overview of course topics
- Course details (attendance, grading, etc.)

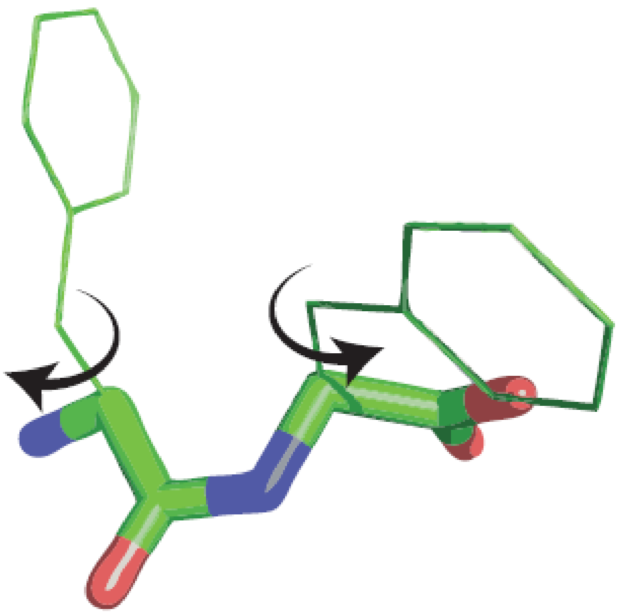
Protein Side Chain Prediction

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

Backbone + Side Chains



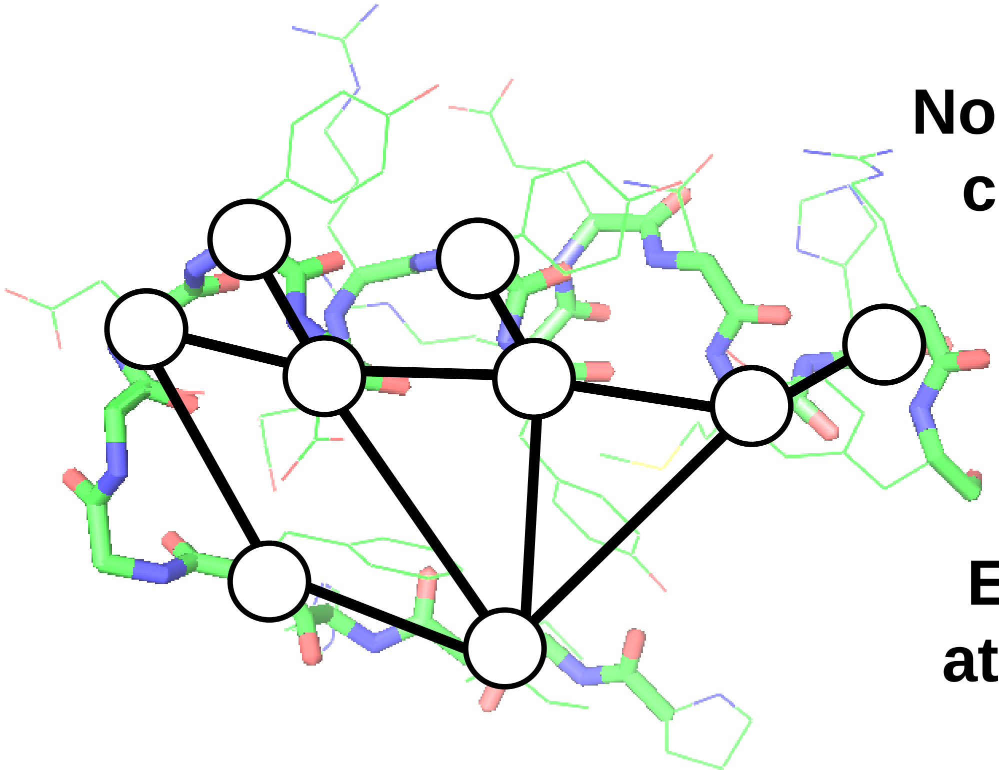
Side Chain Rotation



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

Protein Side Chain Prediction

Graphical Model

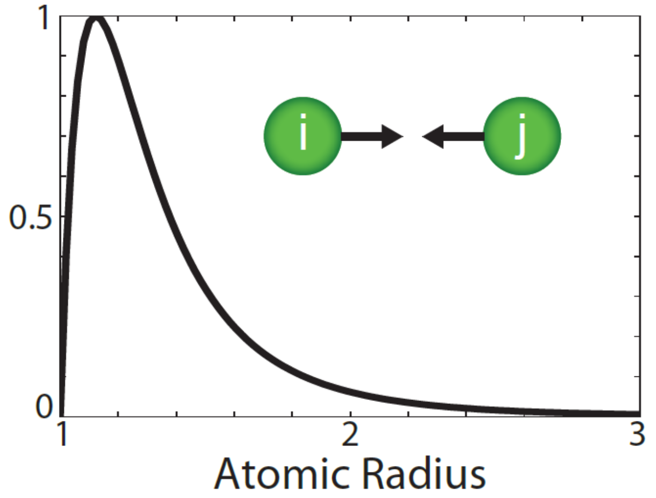


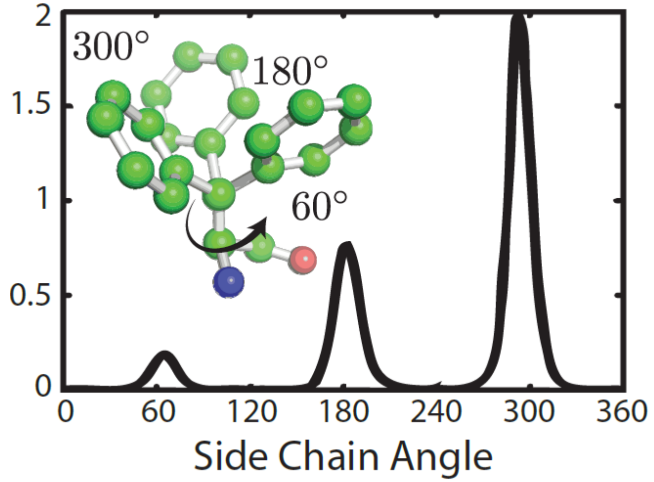
Nodes represent side chain orientations

Edges represent atomic interaction

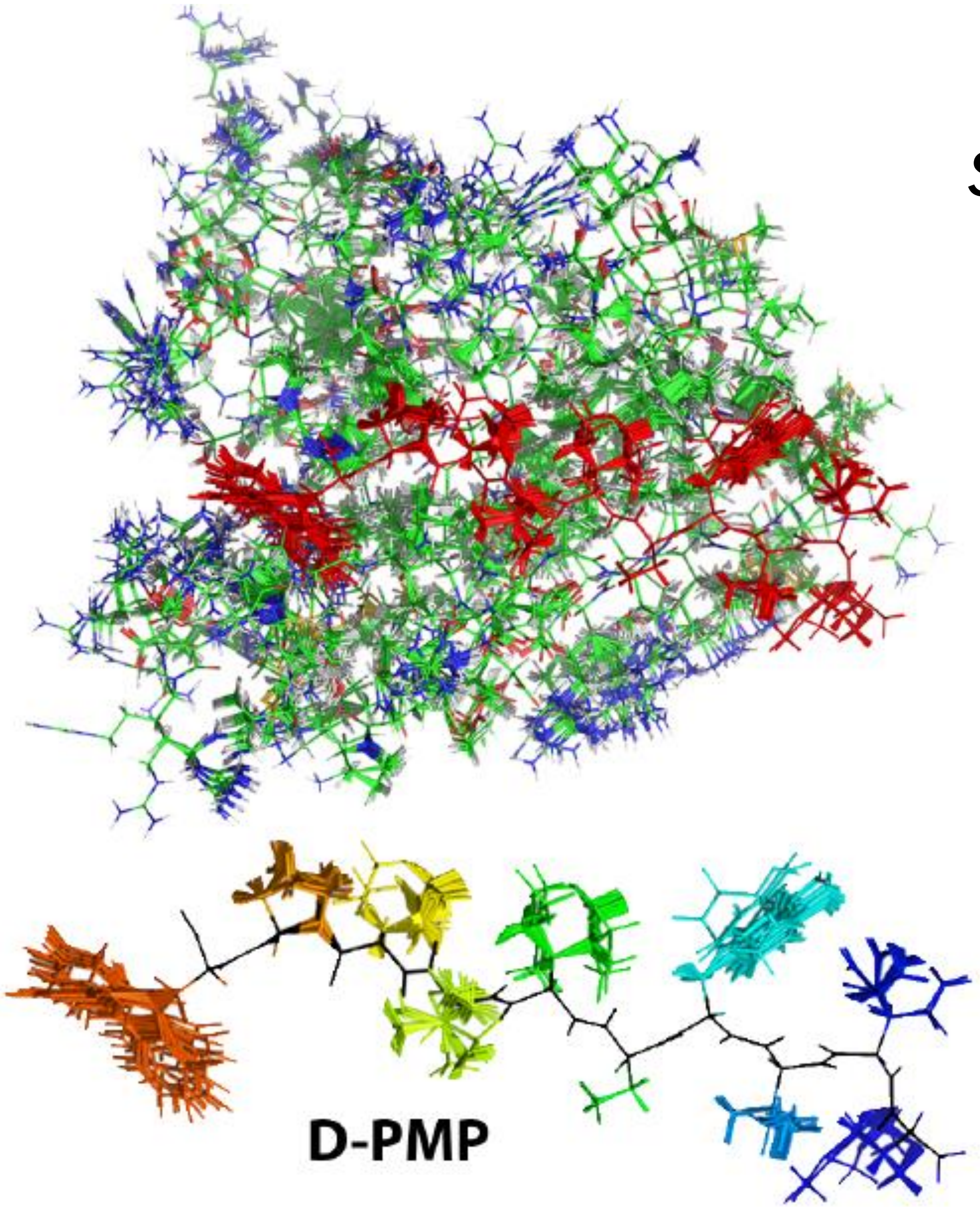
Complex phenomena specified by simpler atomic interactions

Configuration Likelihoods

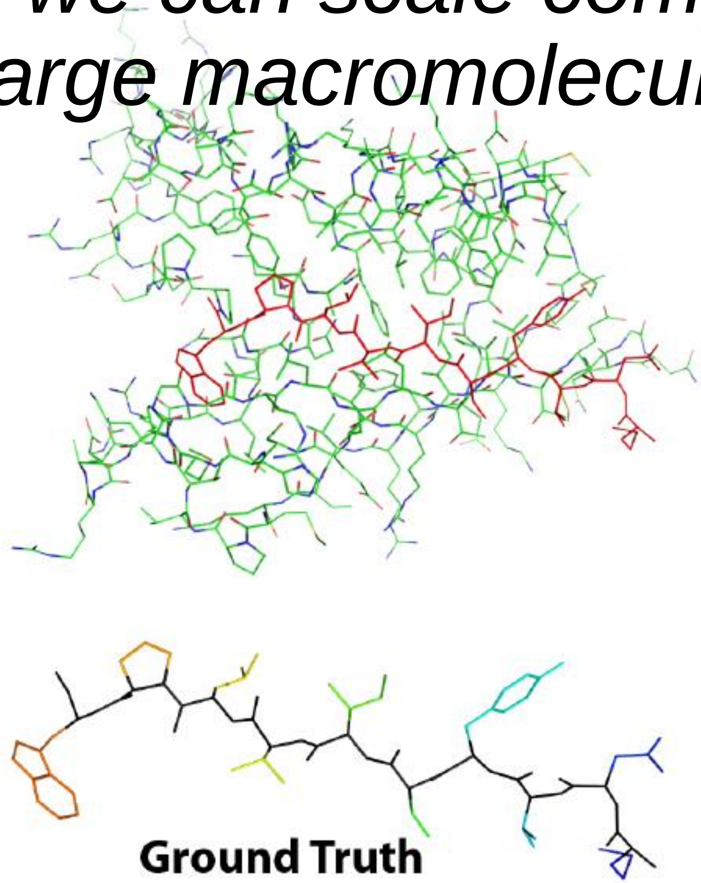




Protein Side Chain Prediction



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



[Pacheco and Sudderth, ICML 2015]

Pose Estimation

Graphical Model

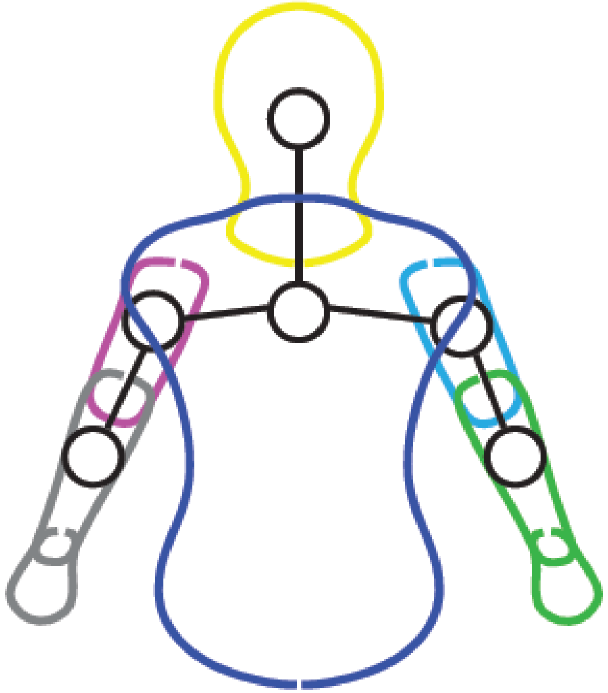
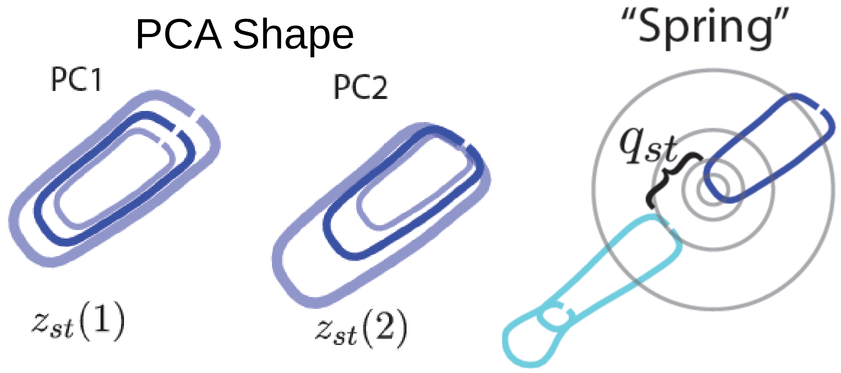
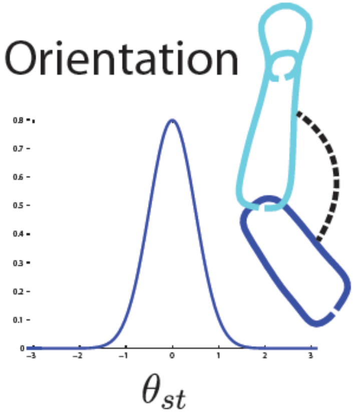


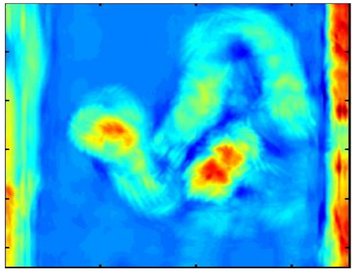
Image (Data / Observation)

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





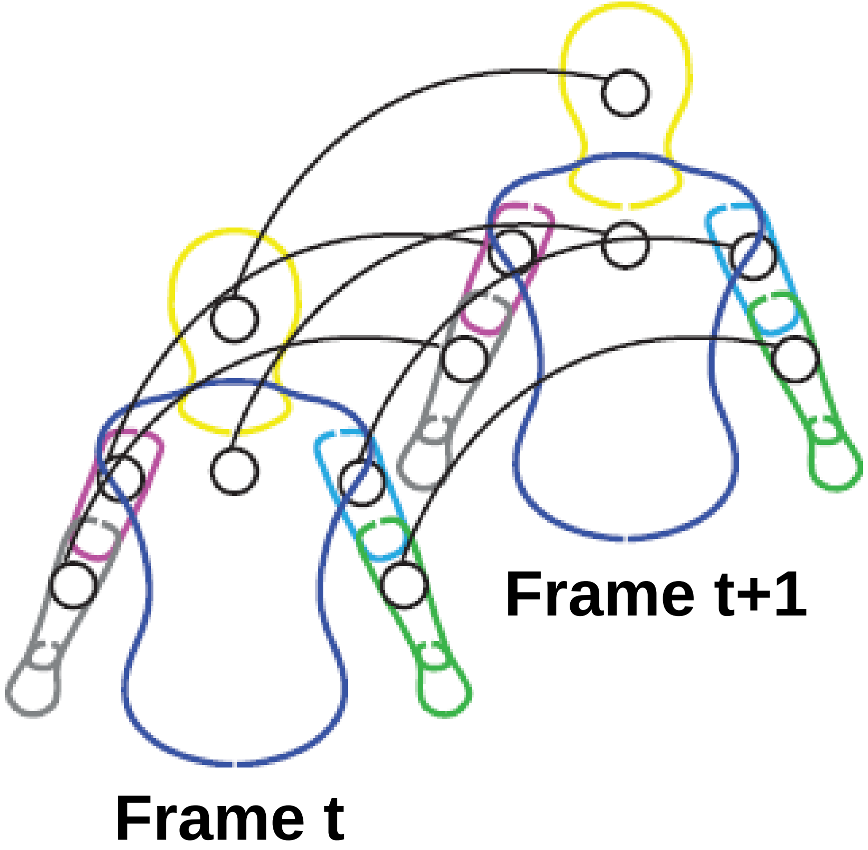
Skin Tone

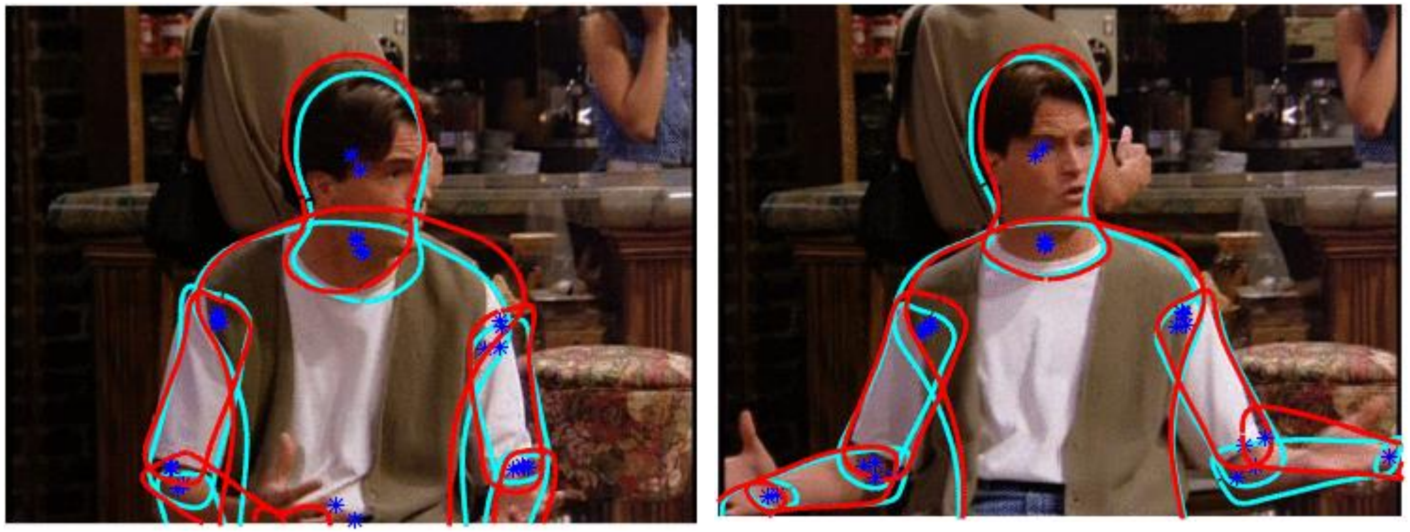


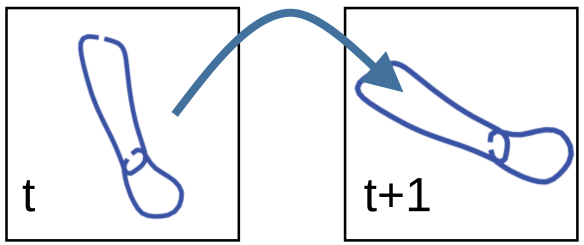
Problem: Estimate orientation / shape / pose of human figure from an image

[Pacheco, et al., NeurIPs 2014]

Pose Tracking

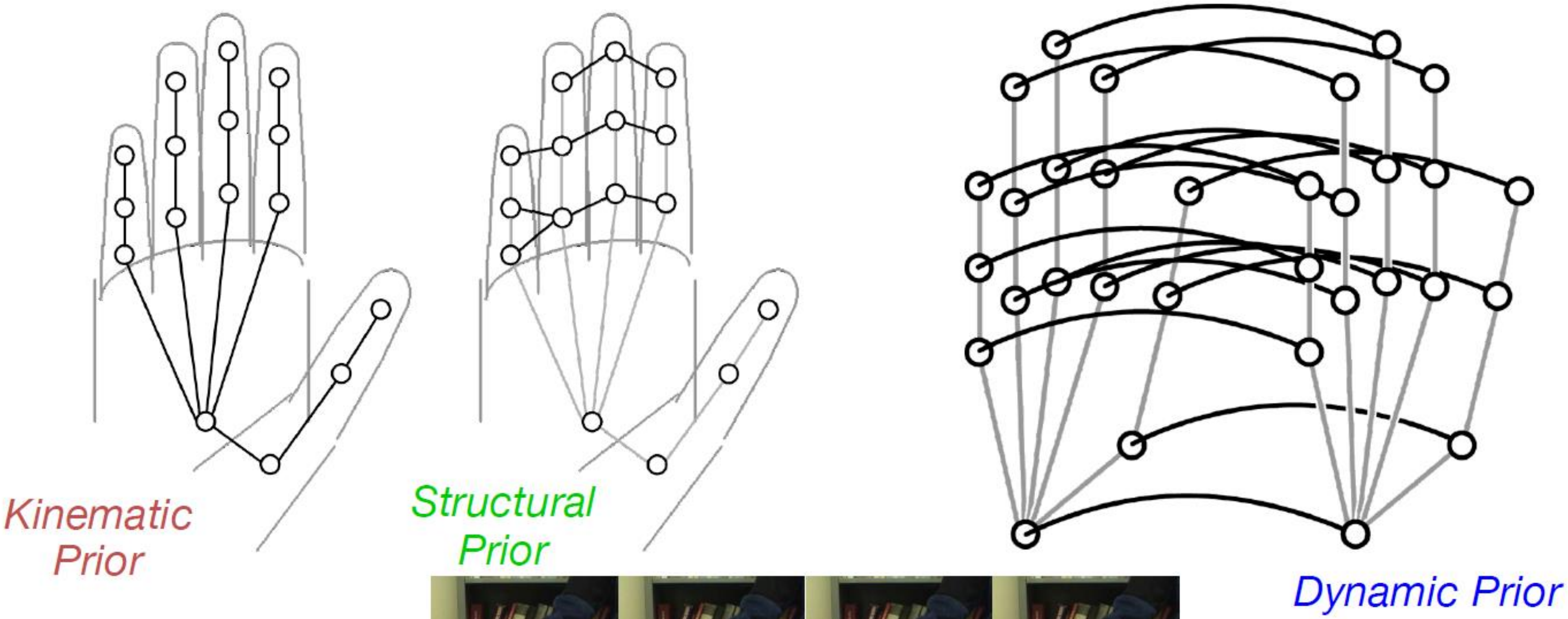


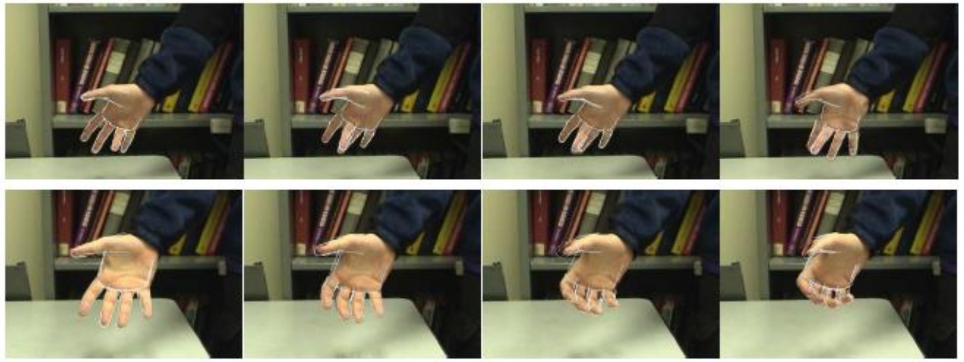




By composing single-frame model with temporal dynamics and motion prior we can do video tracking…

Kinematic Hand Tracking

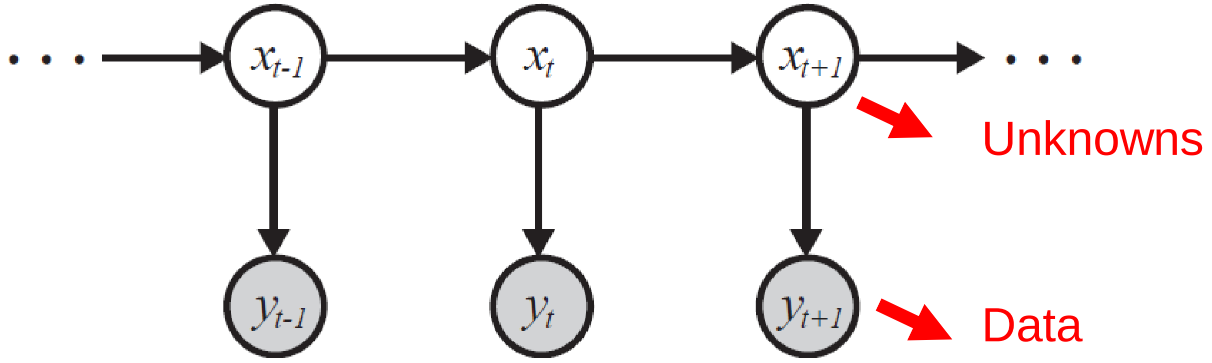




Sudderth et al., 2004

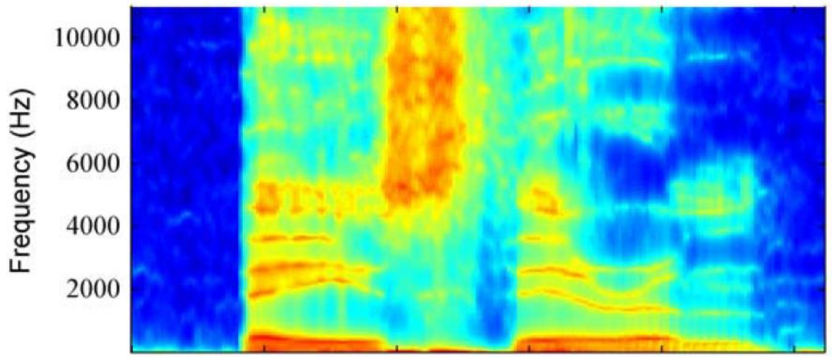
Hidden Markov Models

Sequential models *of discrete quantities of interest*



Example: Speech Recognition

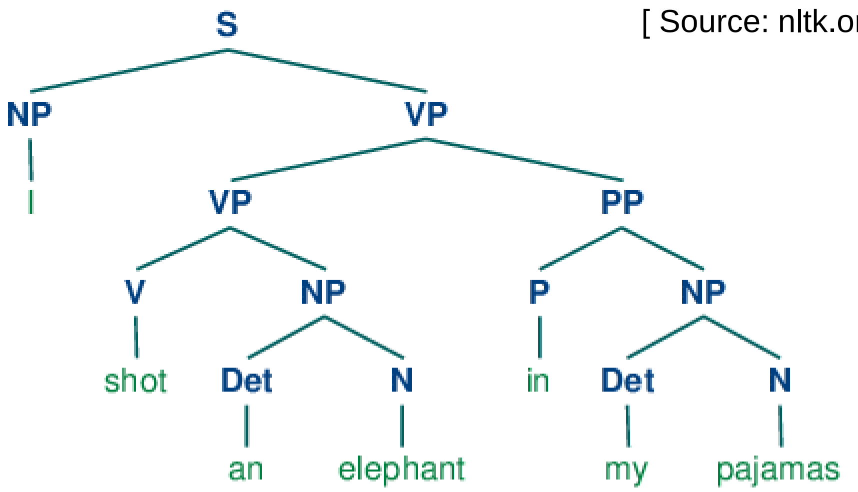
$y =$



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

[Source: Bishop, PRML]

[Source: nltk.org]



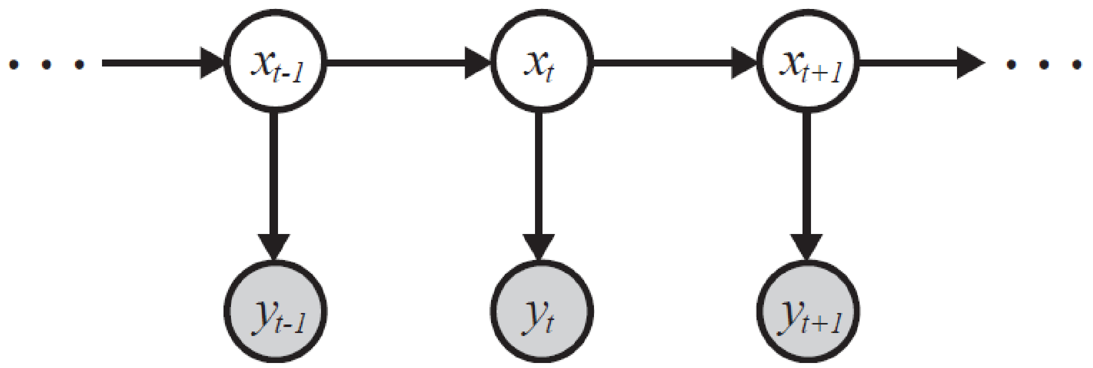
Example: Part-of-speech Tagging:

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

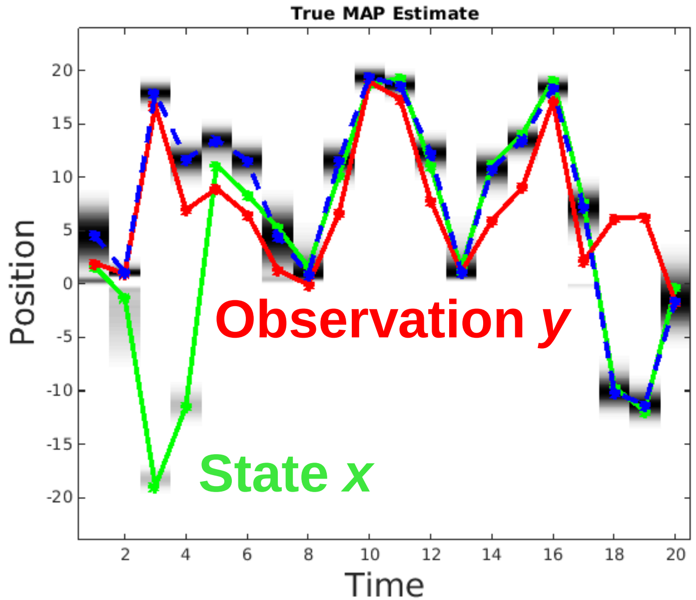
x = NP-V-Det-N-P-Det-N

Dynamical Models

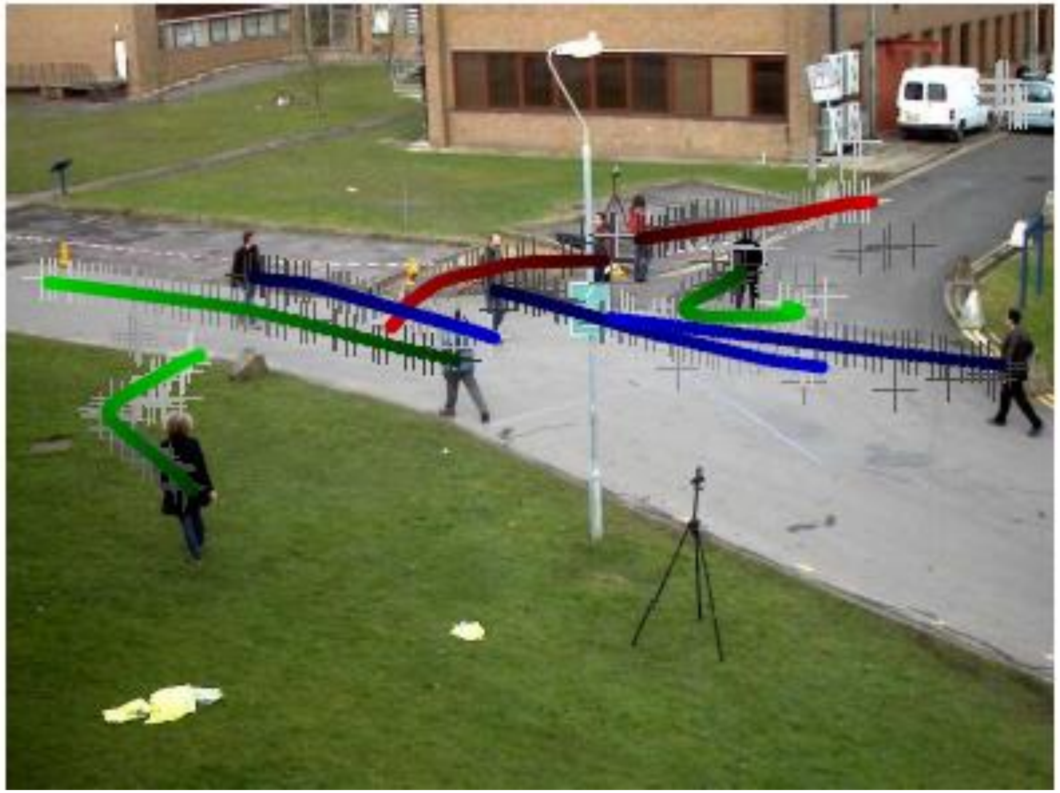
Sequential models of <u>*continuous*</u> *quantities of interest*



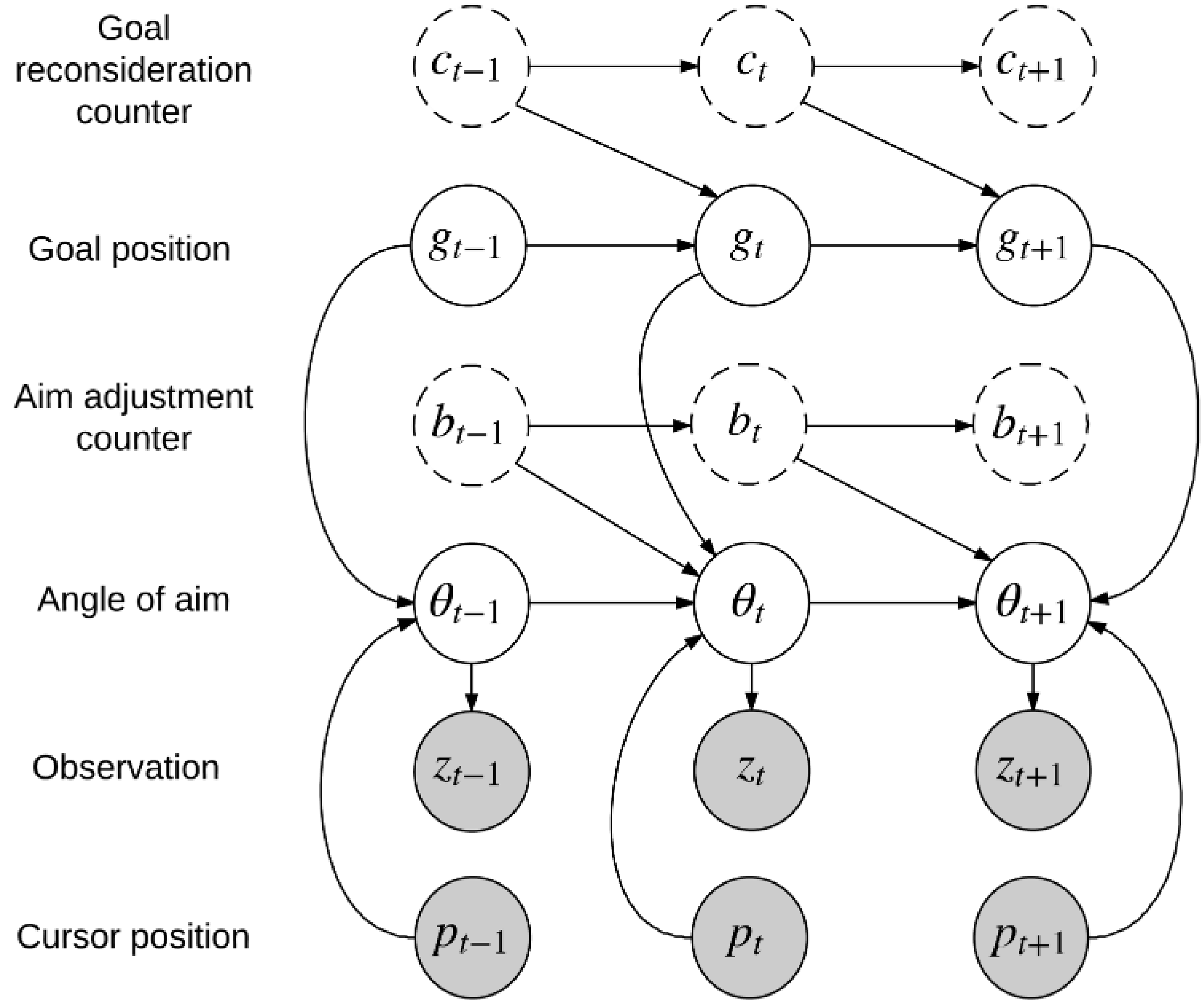
Example: Nonlinear Time Series



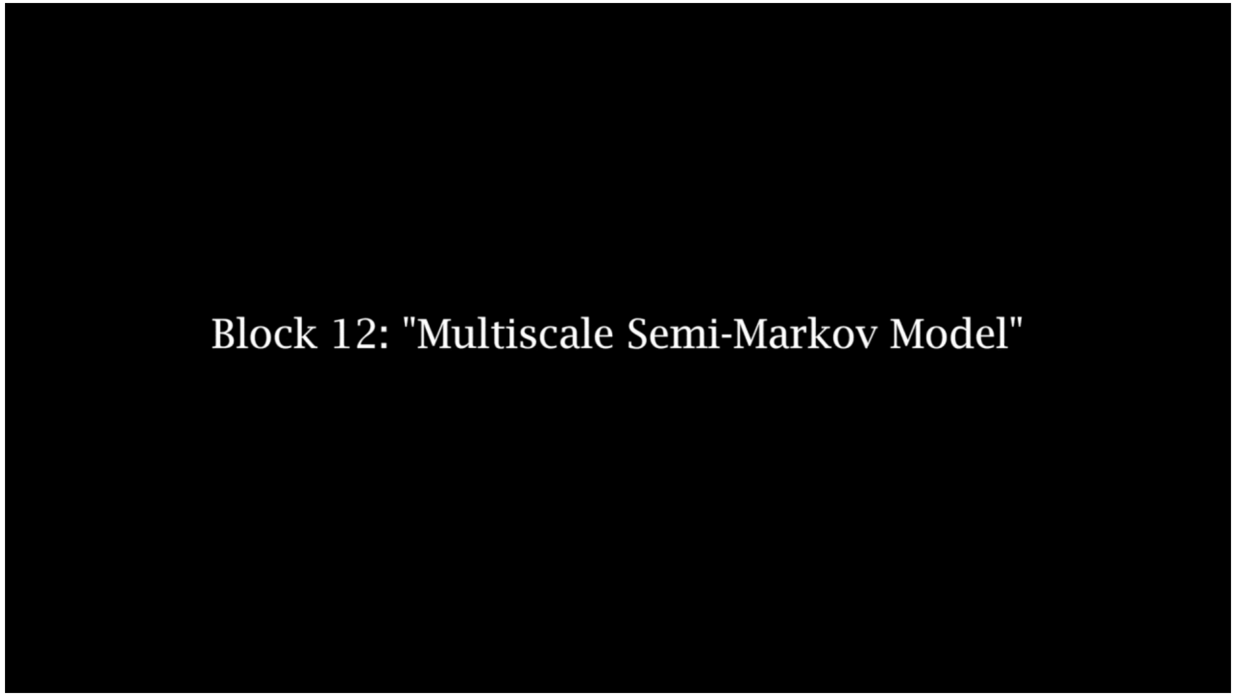
Example: Multitarget Tracking



State-Space Models



Intracortical Brain-Computer Interface



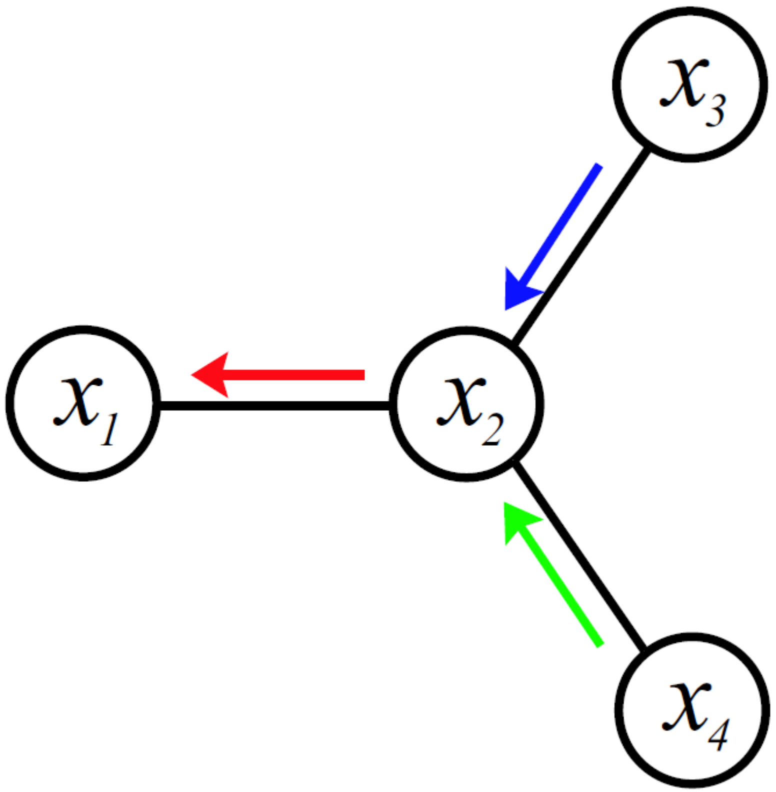
[Milstein, Pacheco, et al., NeurIPs 2017]

Outline

- Motivating examples of representation
- Efficient computation on graphical models
- Overview of course topics
- Course details (attendance, grading, etc.)

Why Graphical Models?

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



Representation

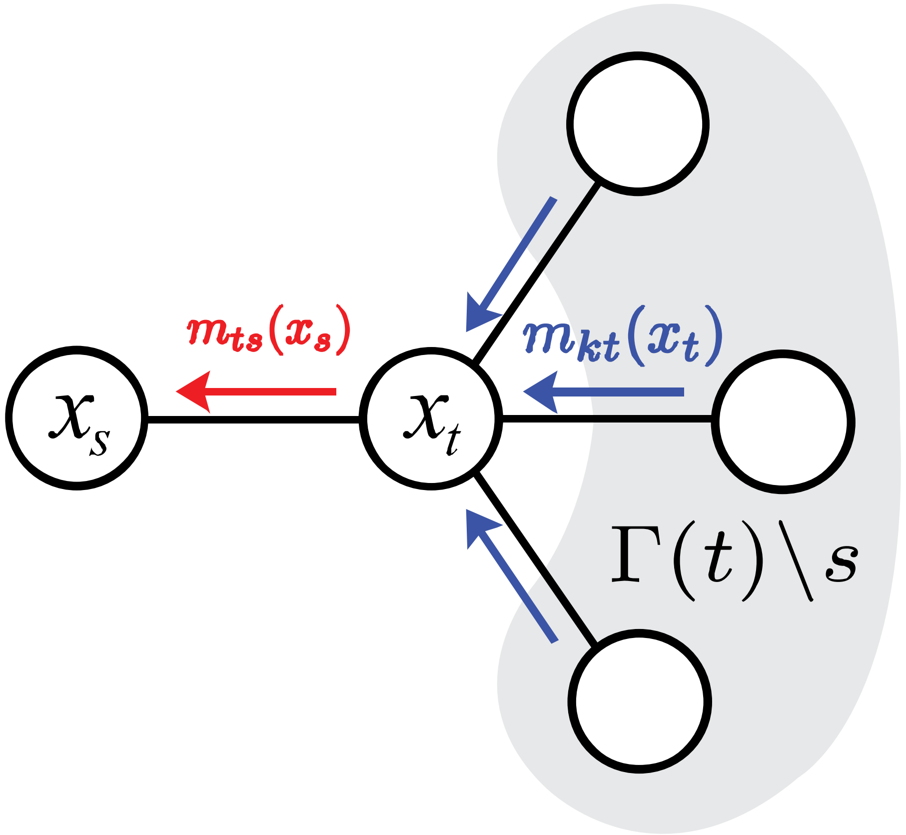
Complex global phenomena arise by simpler-to-specify local interactions

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…

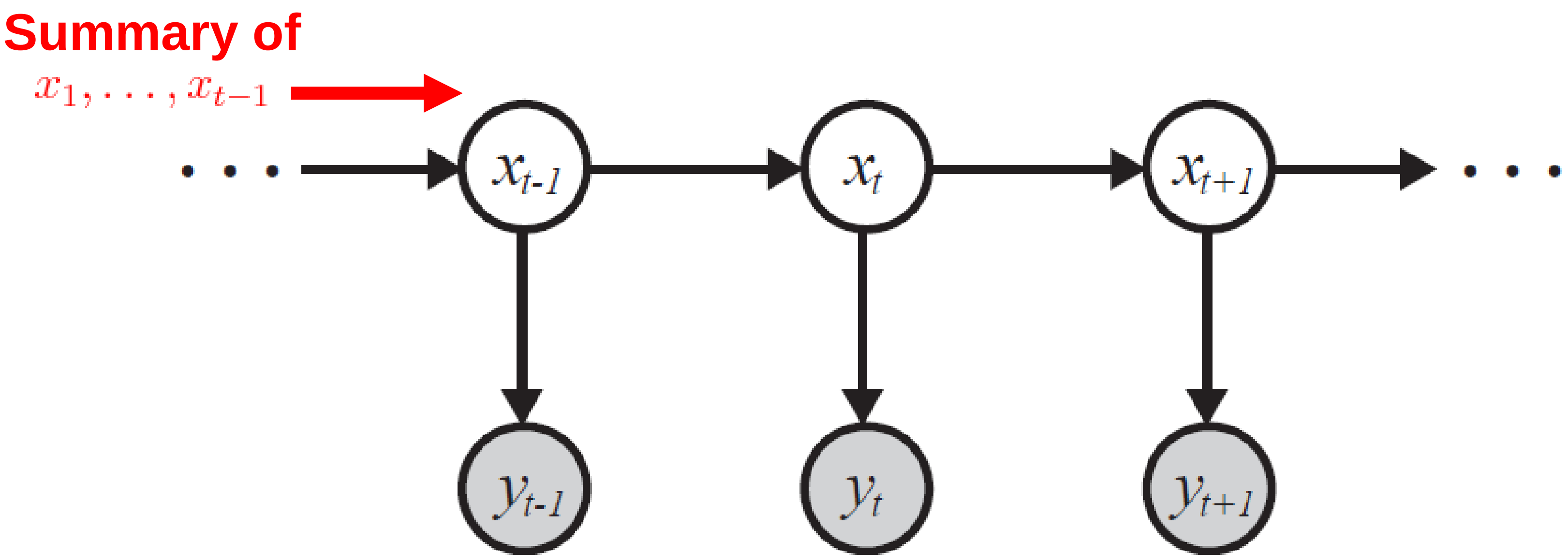


Example algorithms

- 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>

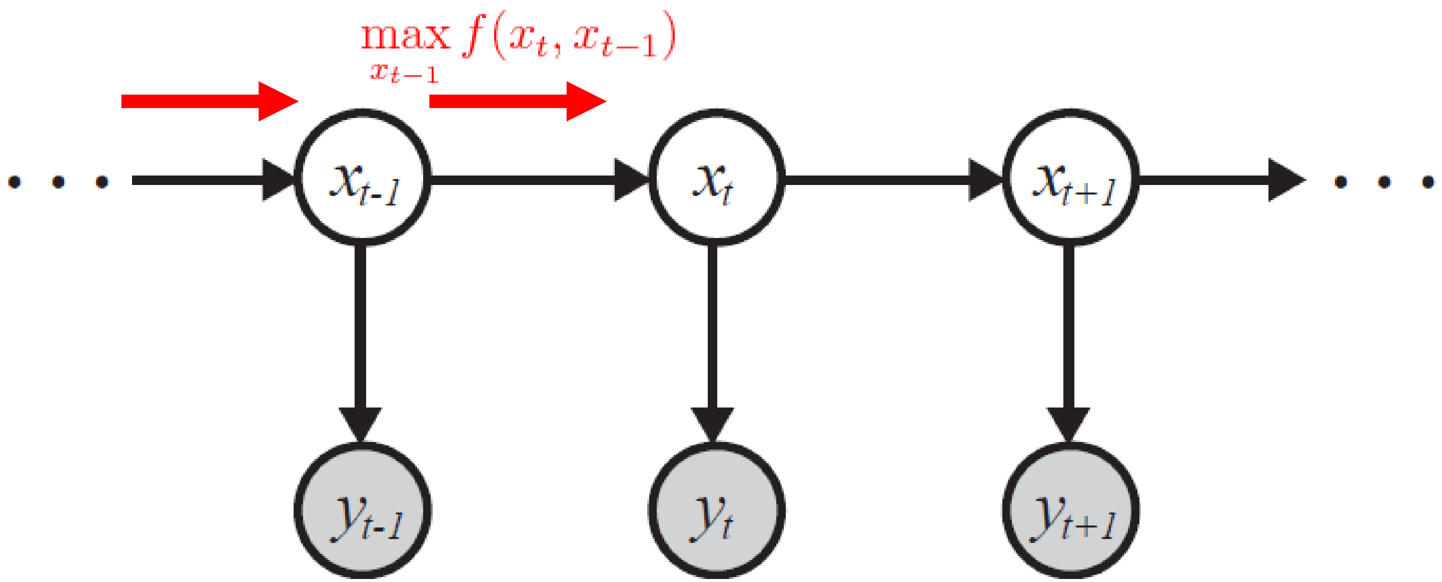
Viterbi Decoder



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

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

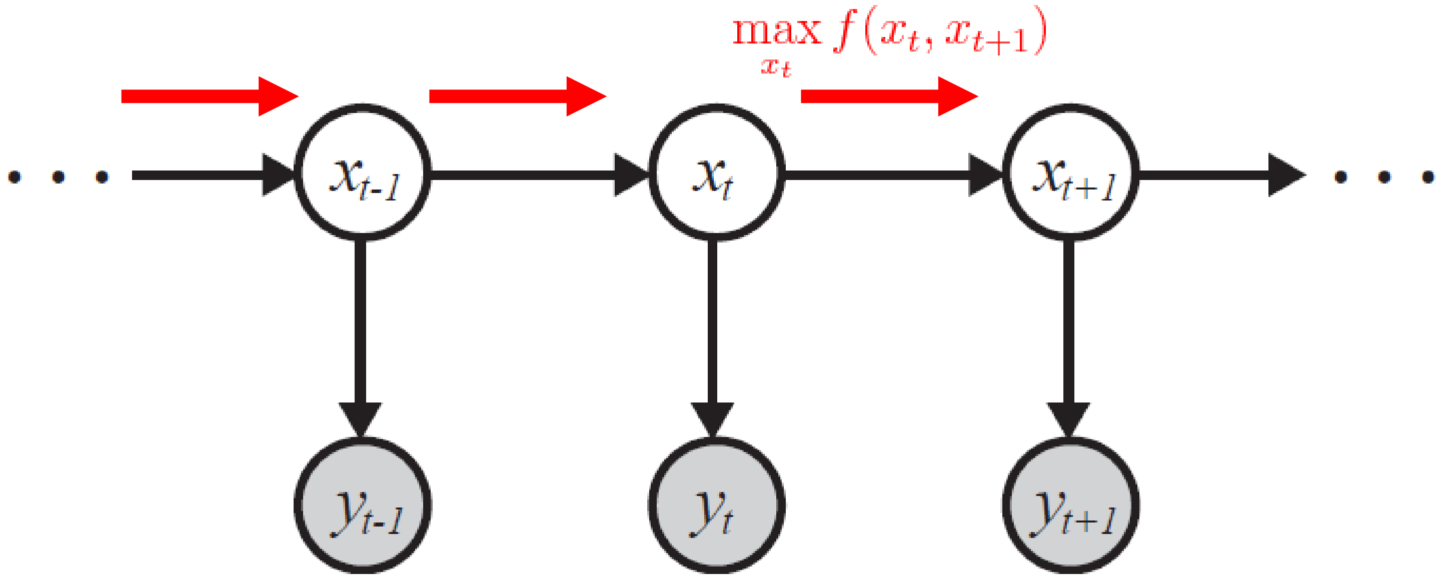
Viterbi Decoder



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

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

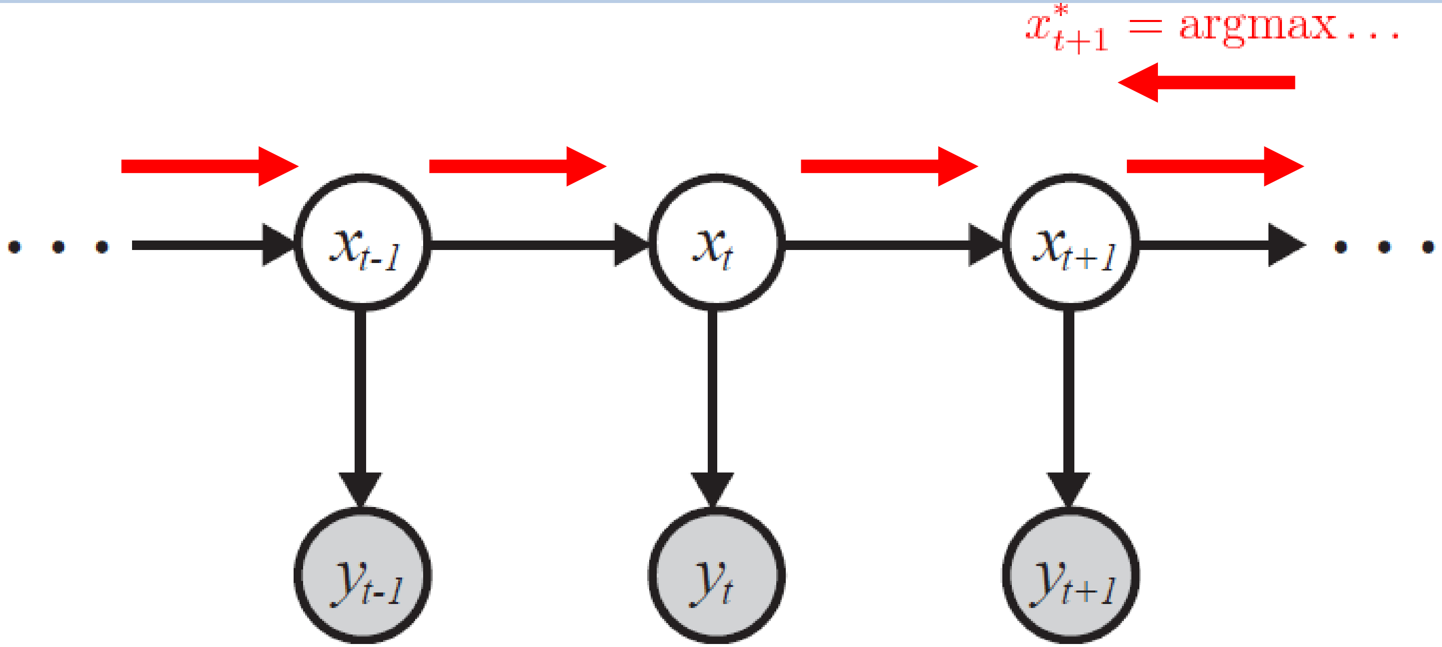
Viterbi Decoder



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

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

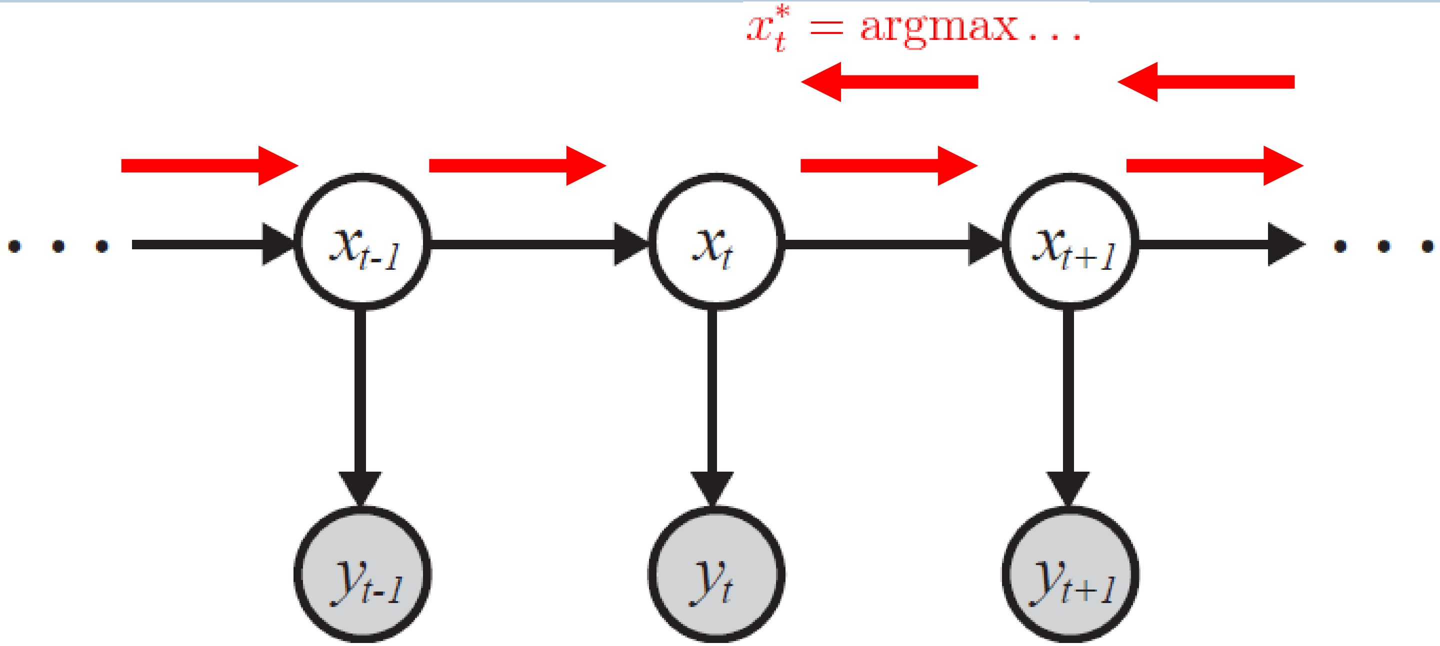
Viterbi Decoder



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

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

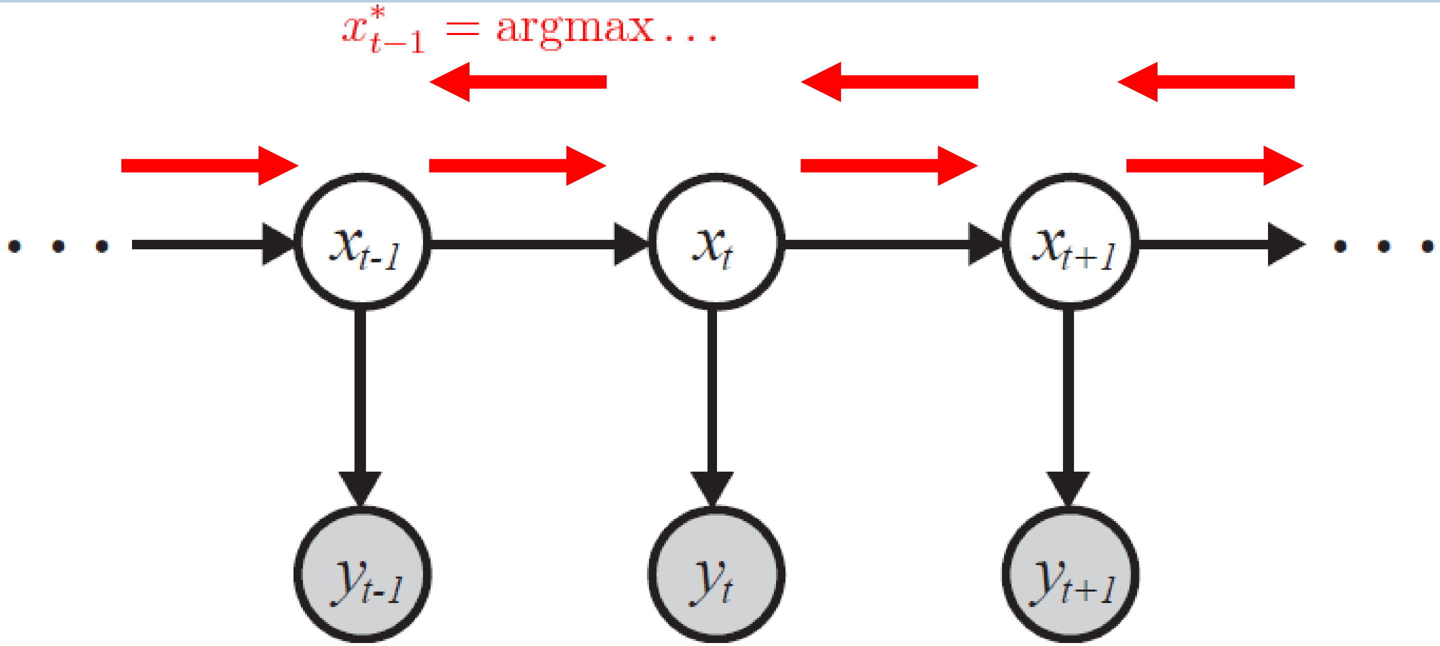
Viterbi Decoder



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

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

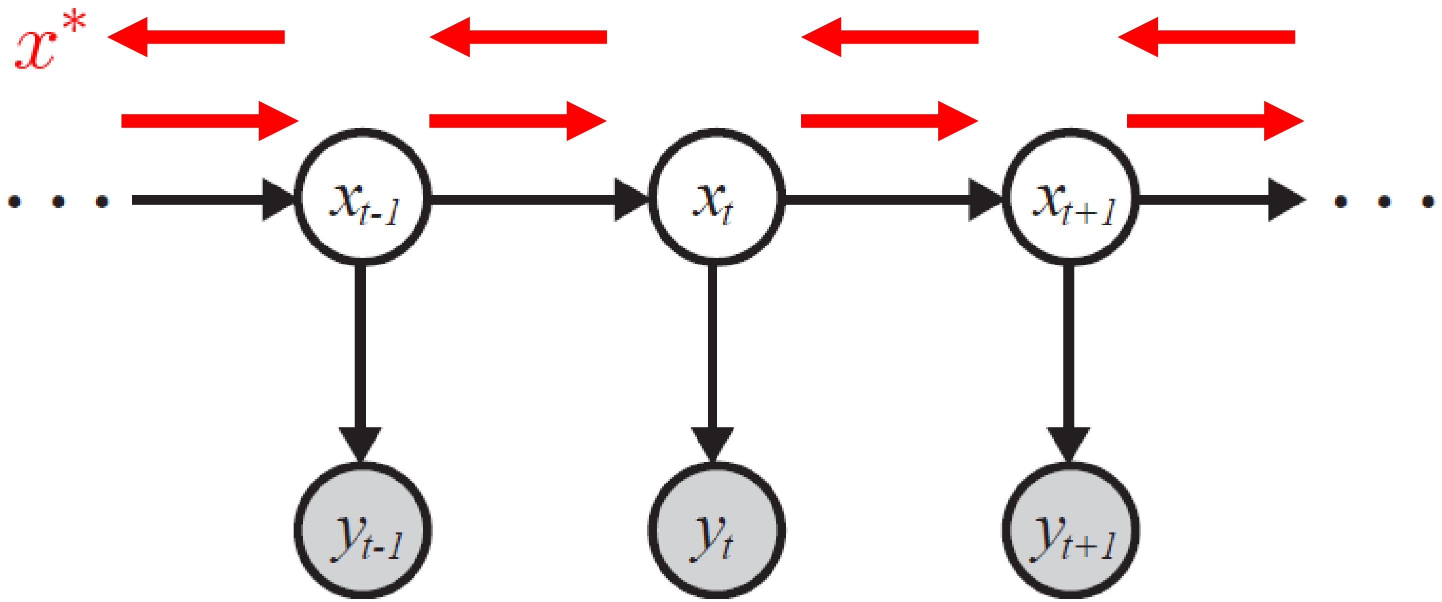
Viterbi Decoder



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

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.

Outline

- Motivating examples of representation
- Efficient computation on graphical models
- Overview of course topics
- Course details (attendance, grading, etc.)

Course Overview

We will cover **five** primary topics…

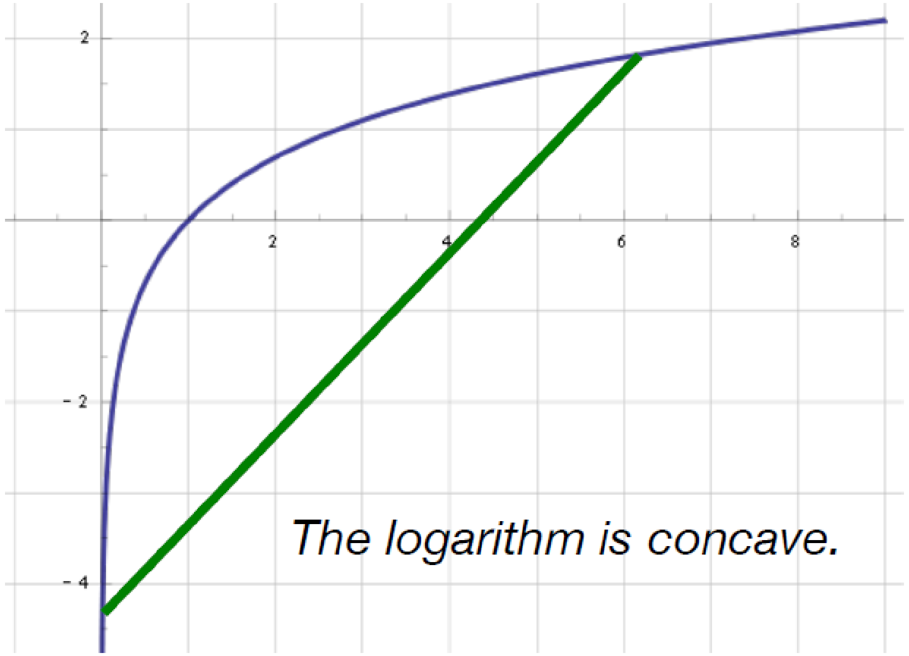
Variational Inference	Advanced Markov chain Monte Carlo	Bayesian Nonparametrics	Bayesian Optimization	Bayesian Deep Learning
Efficient methods for approximate posterior inference	Techniques for obtaining asymptotically exact inference while avoiding local optima	A class of probability models where model complexity is inferred from the data	Probabilistic methods for global optimization of smooth functions	Probabilistic uncertainty models for deep learning

Variational Inference

Uses Jensen's inequality to bound quantities of inference

Jensen's Inequality
(for concave functions)

$$f(\mathbb{E}[x]) \geq \mathbb{E}[f(x)]$$



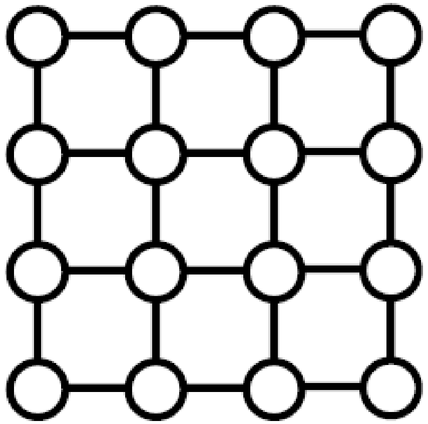
Variational Lower Bound

$$\log p(y) \geq \mathbb{E}_q \left[\log \frac{p(x,y)}{q(x)} \right]$$

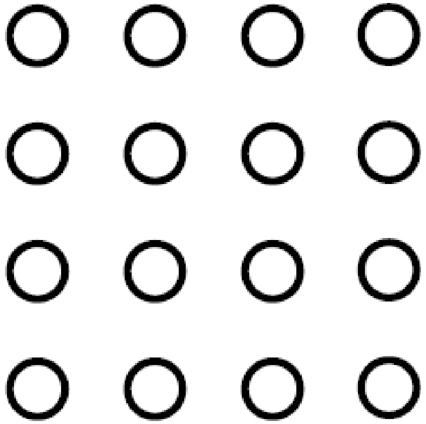
- Partition Function
- Marginal likelihood

Variational Approximation

$p(x)$



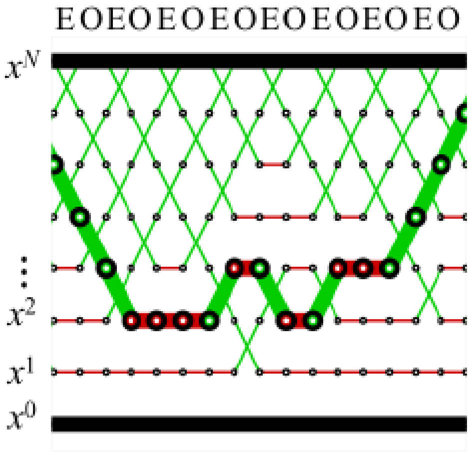
$q(x)$

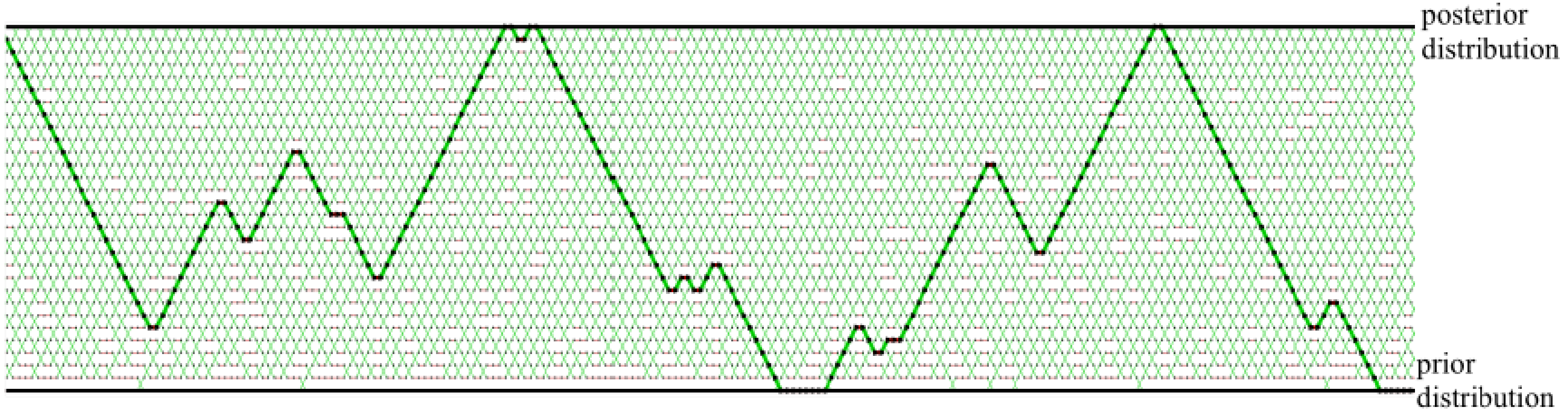


Advanced Markov Chain Monte Carlo

Advanced MCMC techniques reduce sample complexity and avoid getting stuck in local energy minima

[Source: Syed et al, 2019]

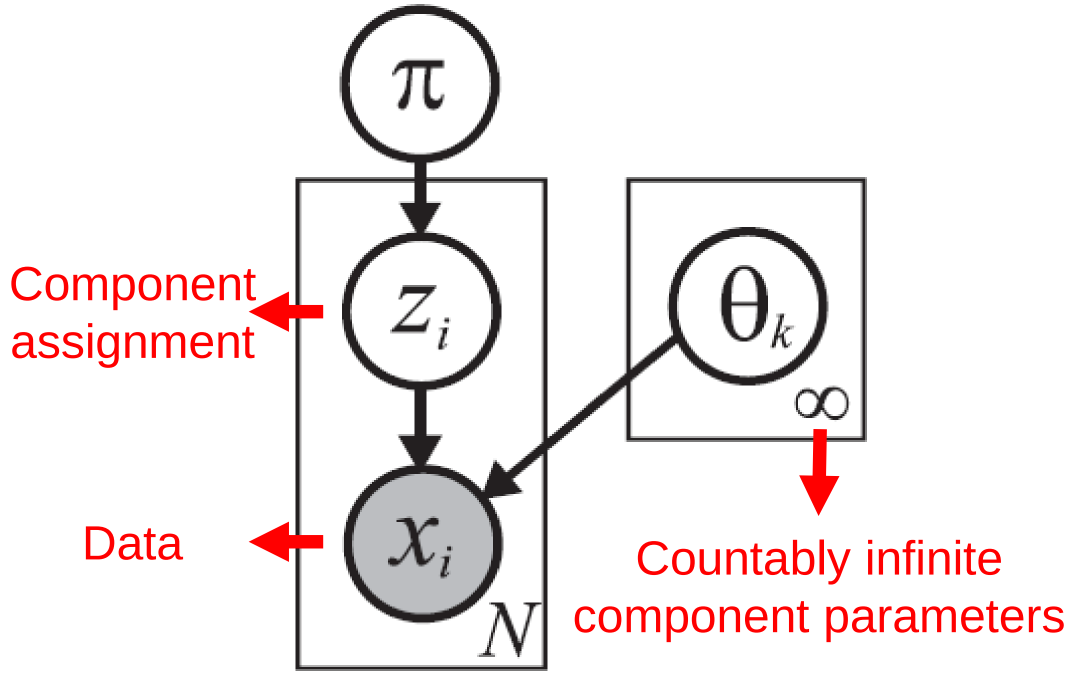


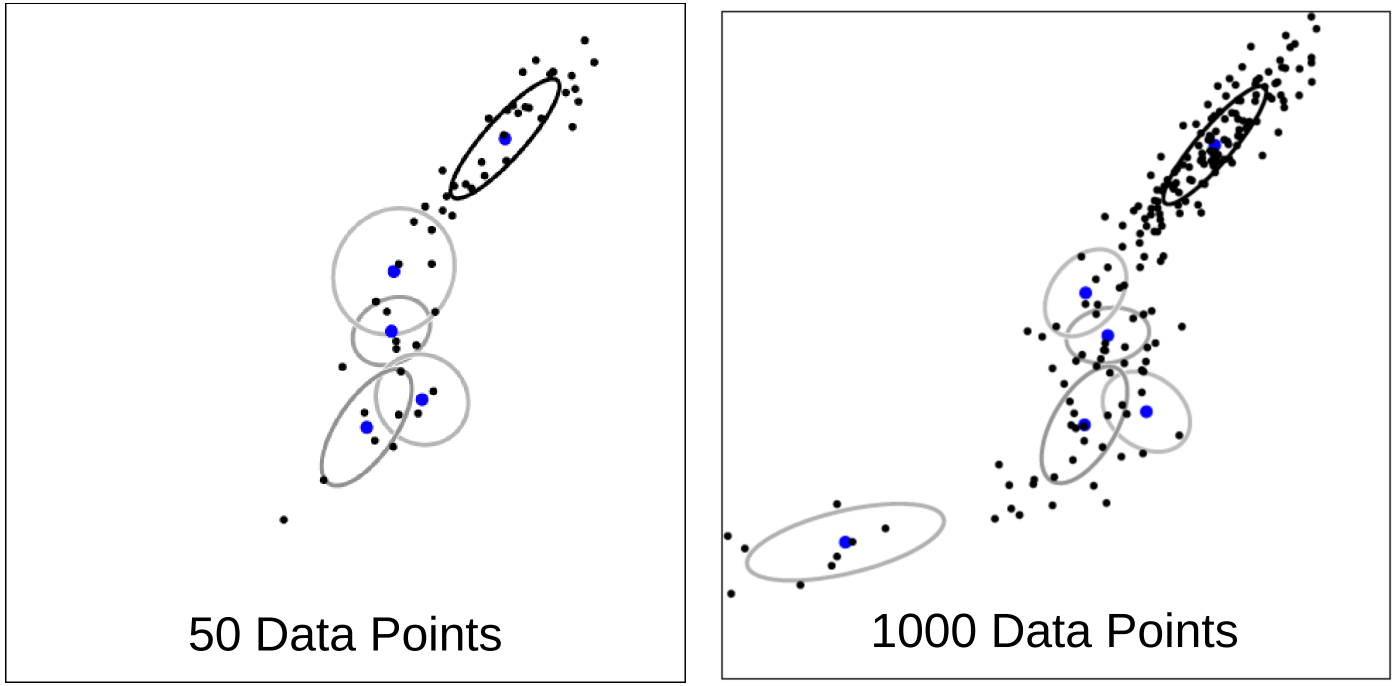


Example: Parallel tempering exchange replicates across multiple MCMC chains running in (embarrassingly) parallel

Bayesian Nonparametrics

Amount and nature of data drive model complexity

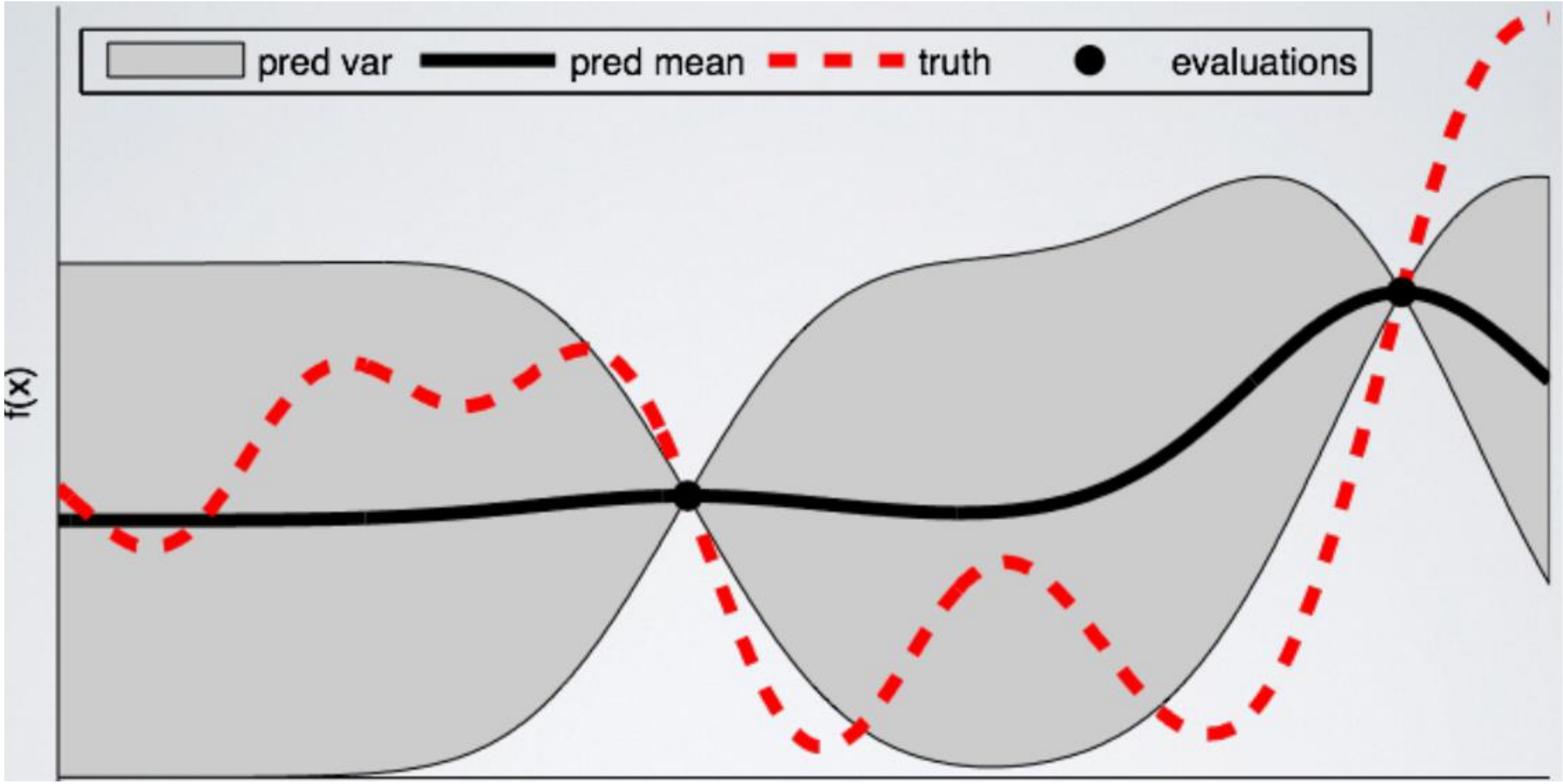




Example: Dirichlet process mixture models a distribution over an <u>infinite</u> number of mixture components

Bayesian Optimization

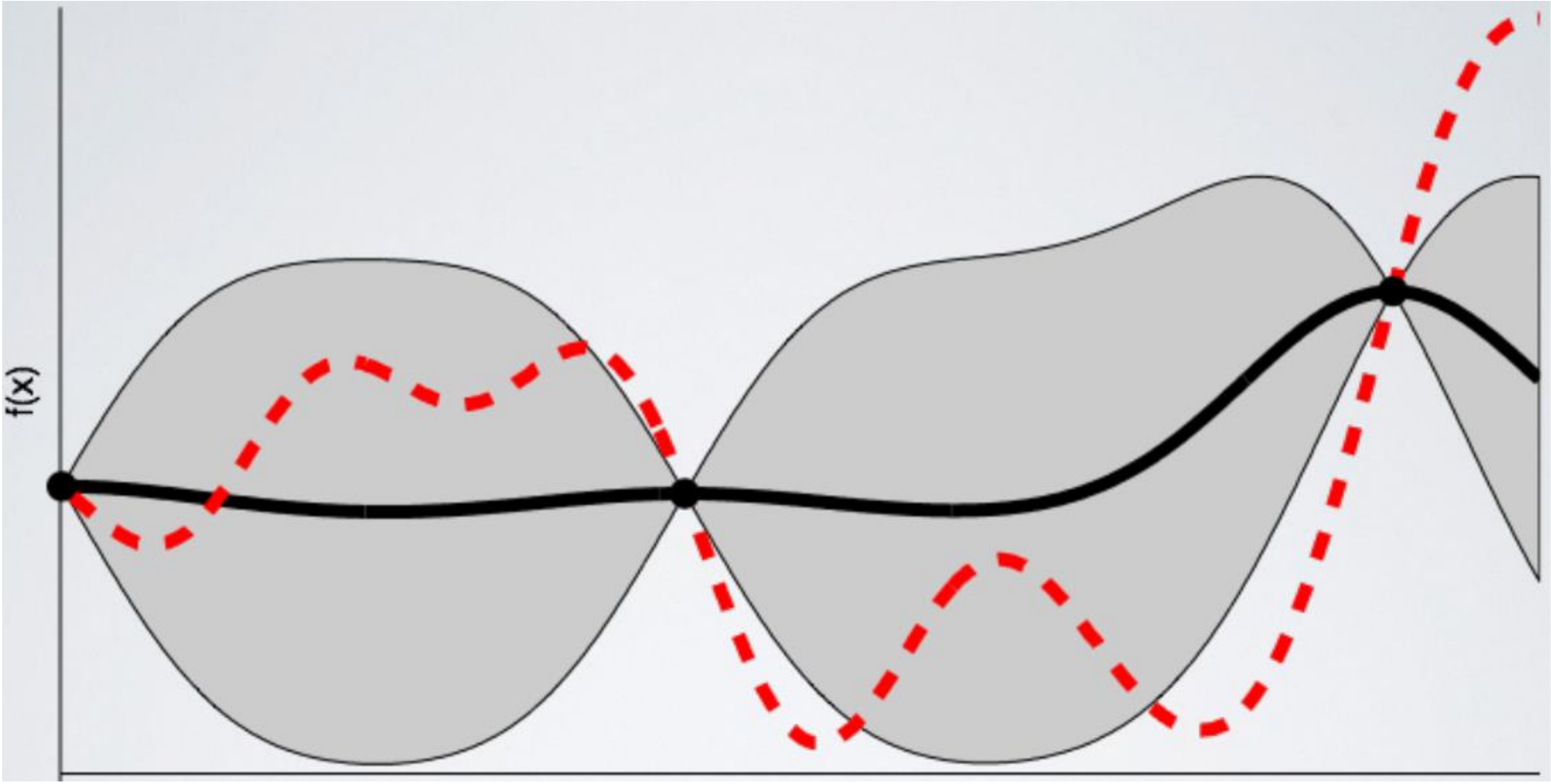
Global optimization of <u>random functions</u>: $\min_x f(x)$



[Source: Ryan Adams]

Bayesian Optimization

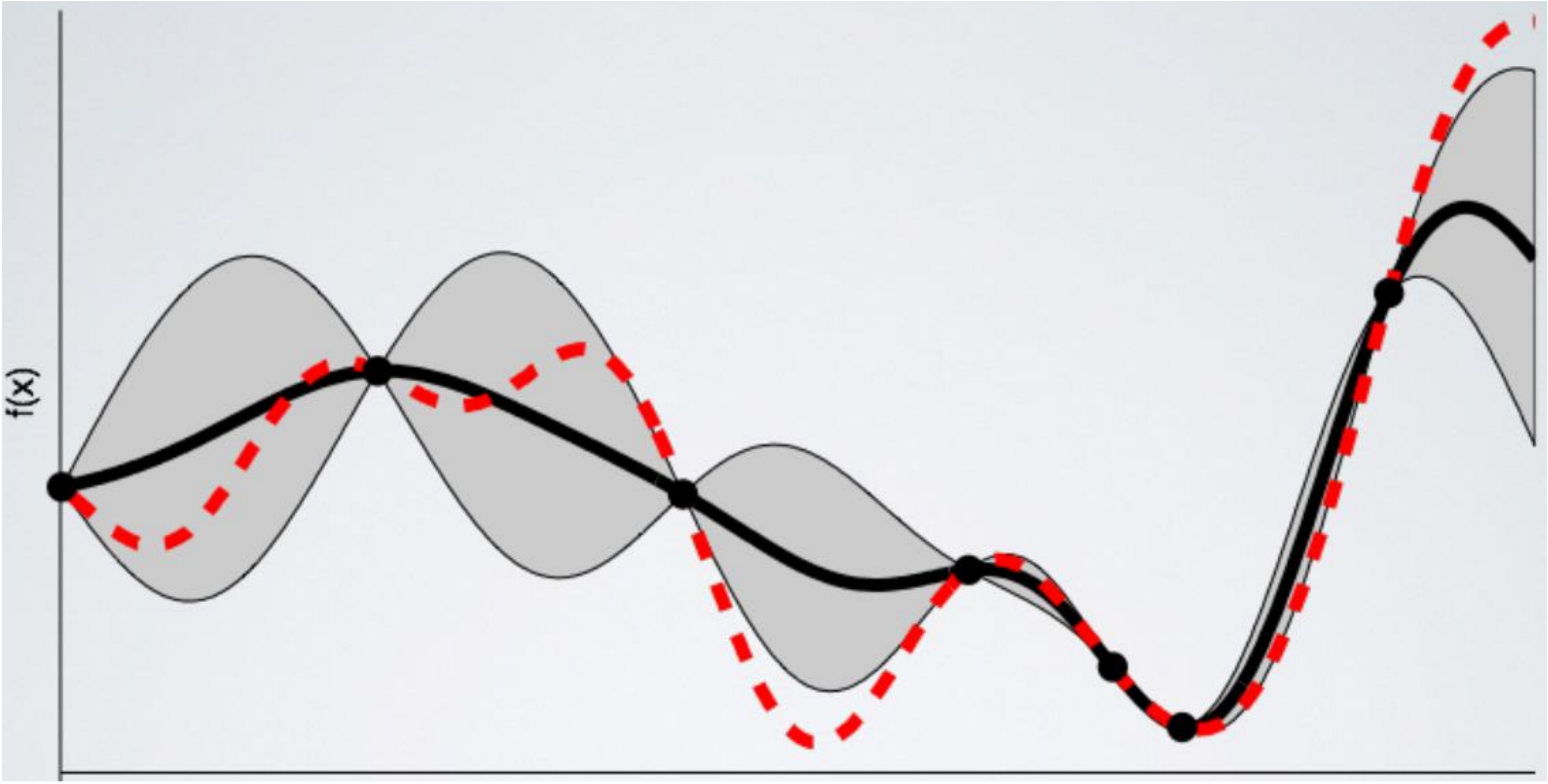
Iteratively updates distribution over function value (regression)



[Source: Ryan Adams]

Bayesian Optimization

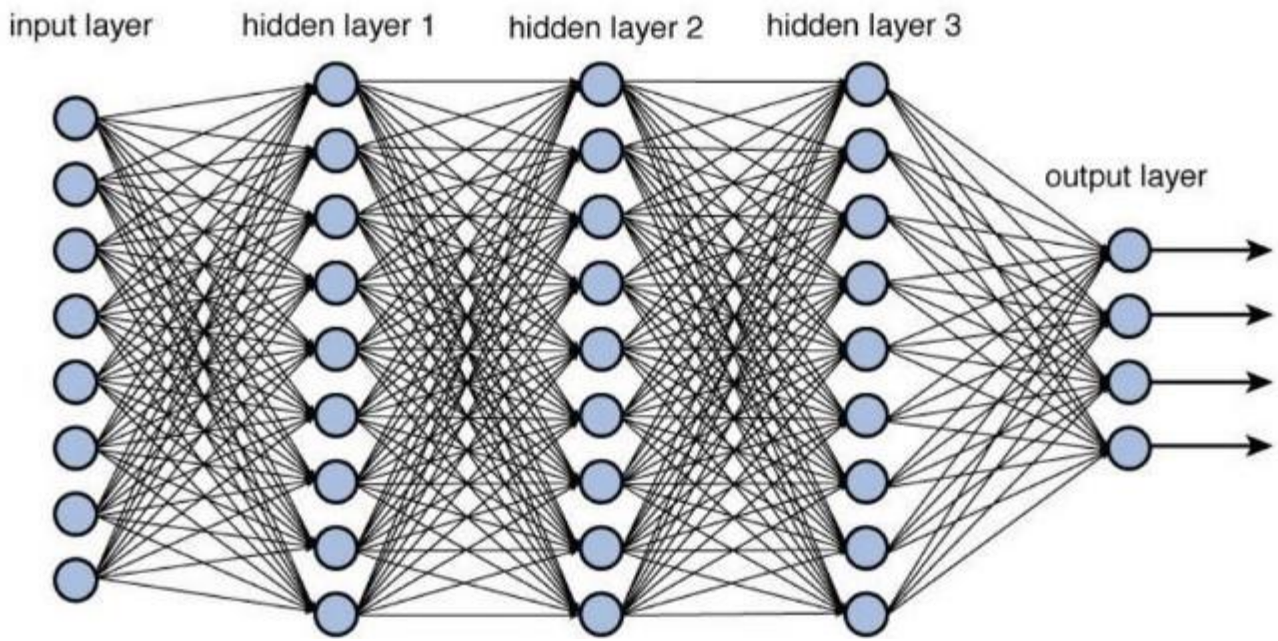
The function is well-approximated around the minimizer



[Source: Ryan Adams]

Bayesian Deep Learning

Neural networks are graphical models too…

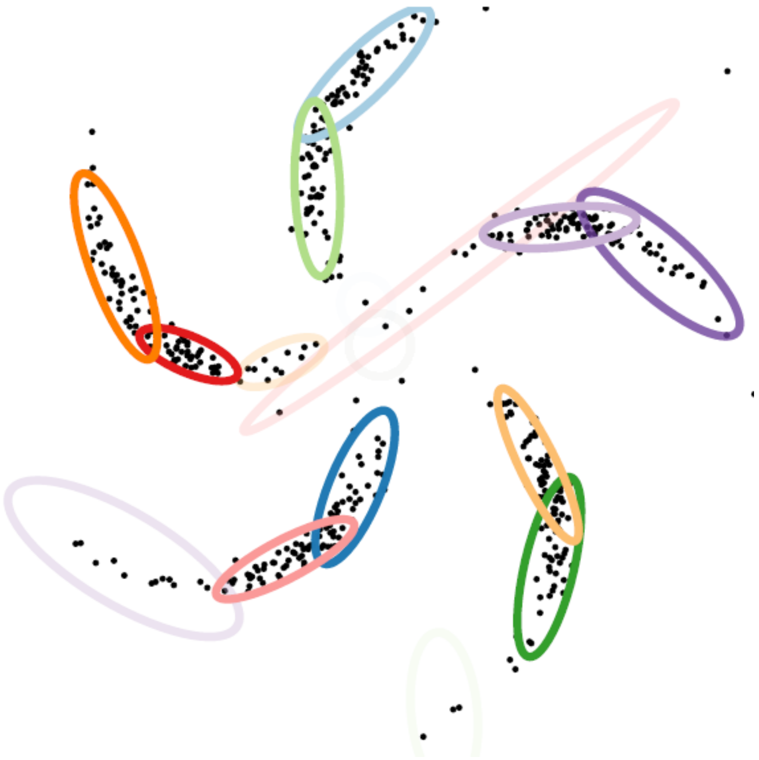


…but they are *typically* not probabilistic

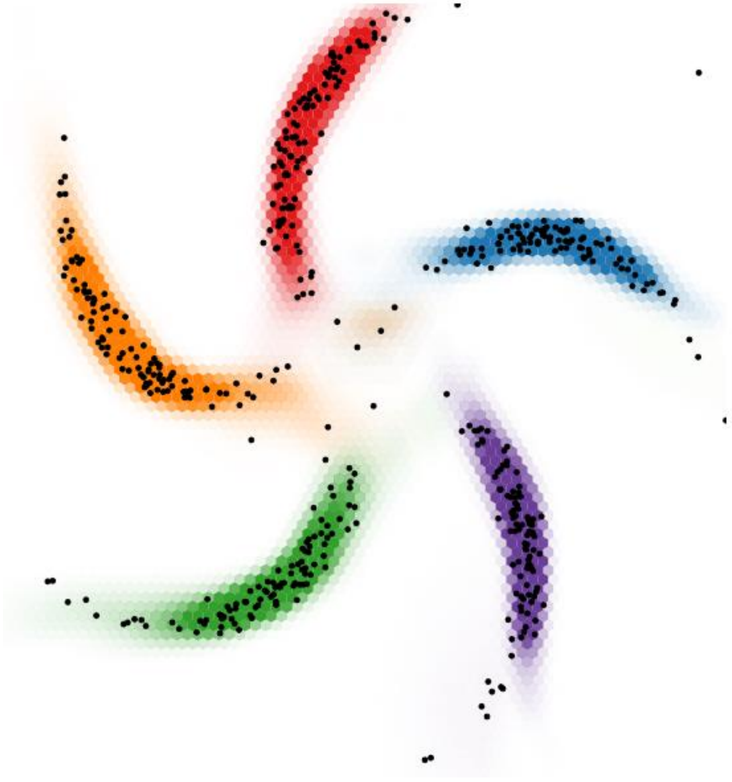
Bayesian Deep Learning

Combines deep learning with uncertainty models

Data



Gaussian Mixture Model (GMM)



GMM Structured Variational Autoencoder

[Source: Johnson et al., NIPS 2016]

Outline

- Motivating examples of representation
- Efficient computation on graphical models
- Overview of course topics
- Course details (attendance, grading, etc.)

Now for the bulleted lists of stuff…