Computer Science

CSC696H: Advanced Topics in Probabilistic Graphical Models

Introduction and Course Overview

Prof. Jason Pacheco

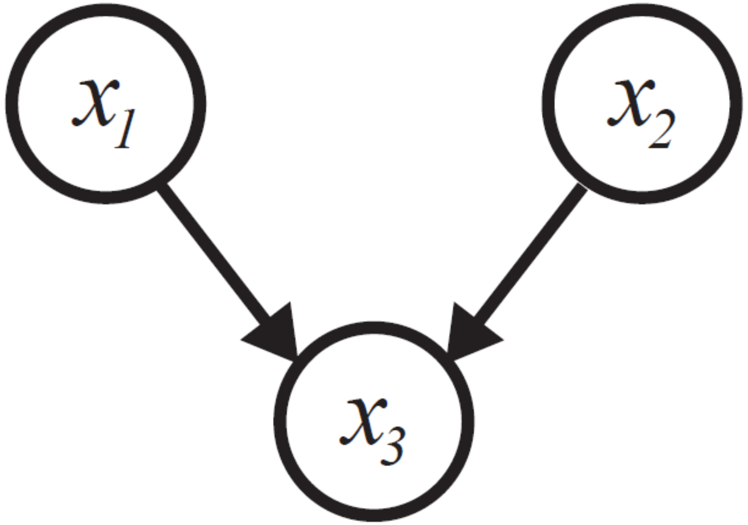
What is a Probabilistic Graphical Model?

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

Graphical Model:

Probability Model:

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

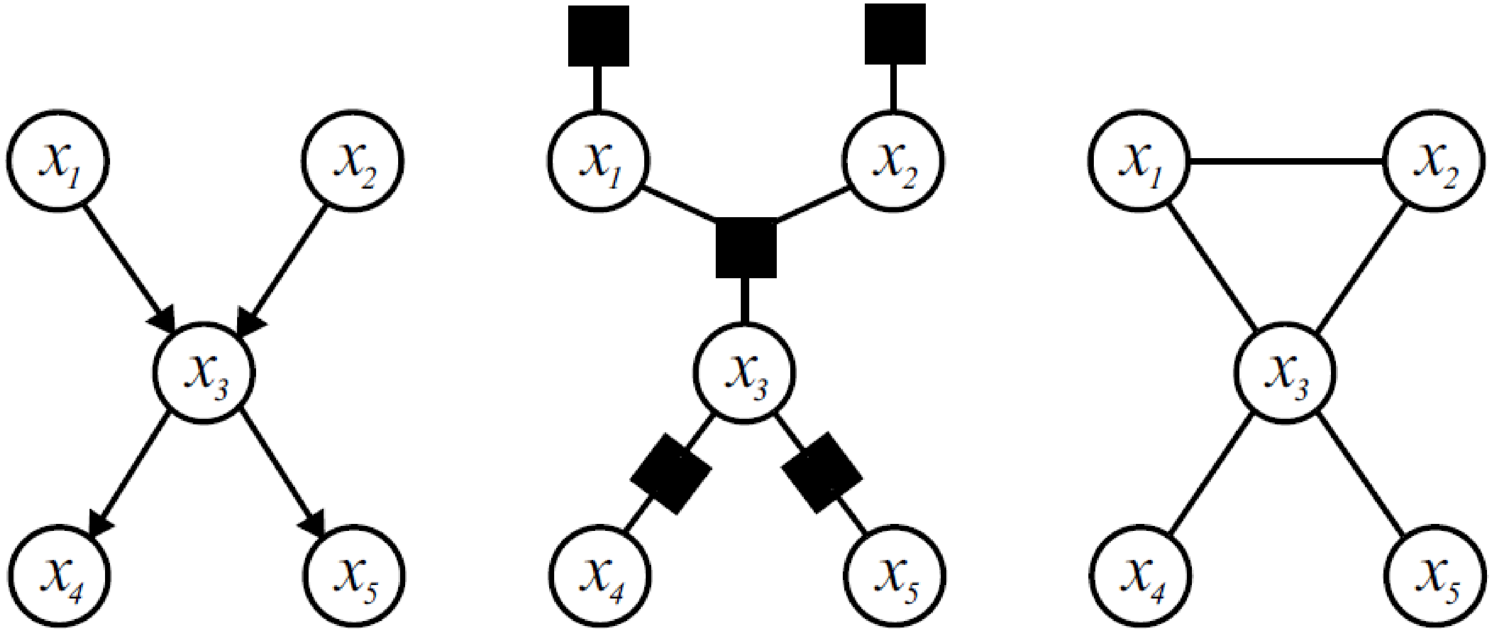


The graphical model structure *obeys* the factorization of the probability function in a sense we will formalize later

* 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



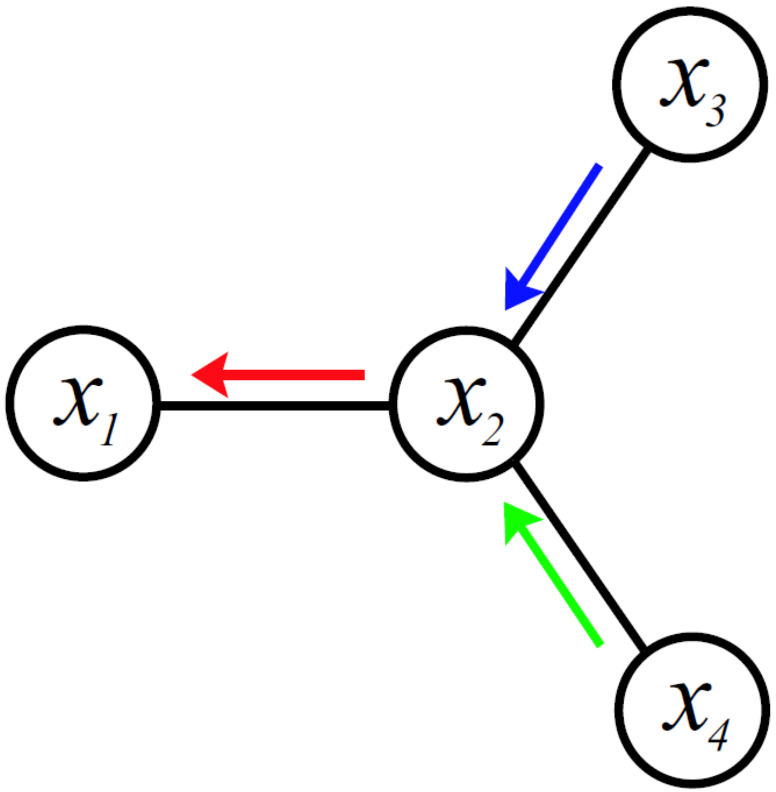
Bayes Network **Factor Graph** **Markov Random Field**

Directed Models **Undirected Models**

Erik Sudderth, PhD Thesis

Why Graphical Models?

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



Representation

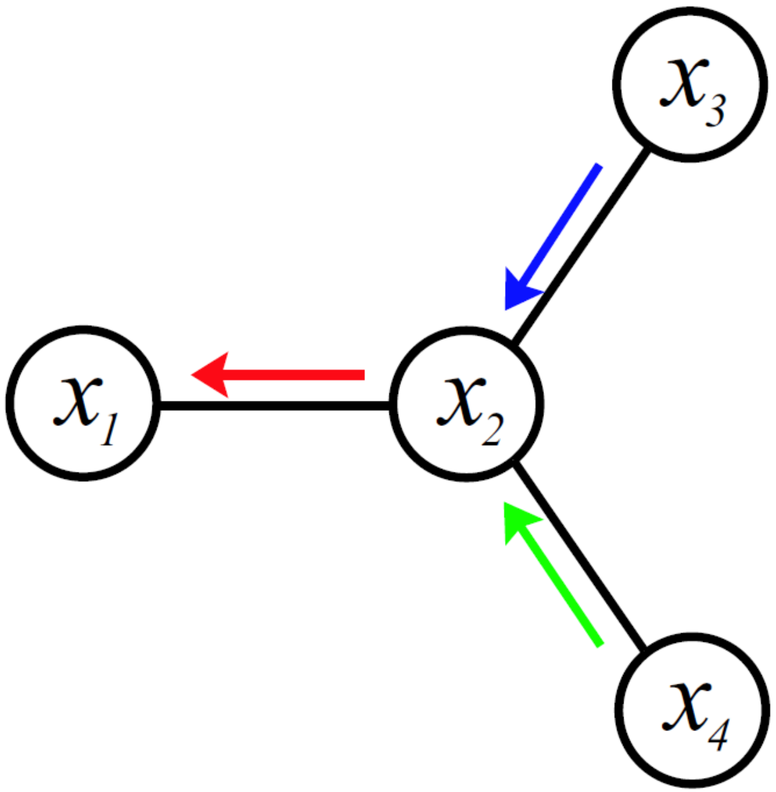
Complex global phenomena arise by simpler-to-specify local interactions

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

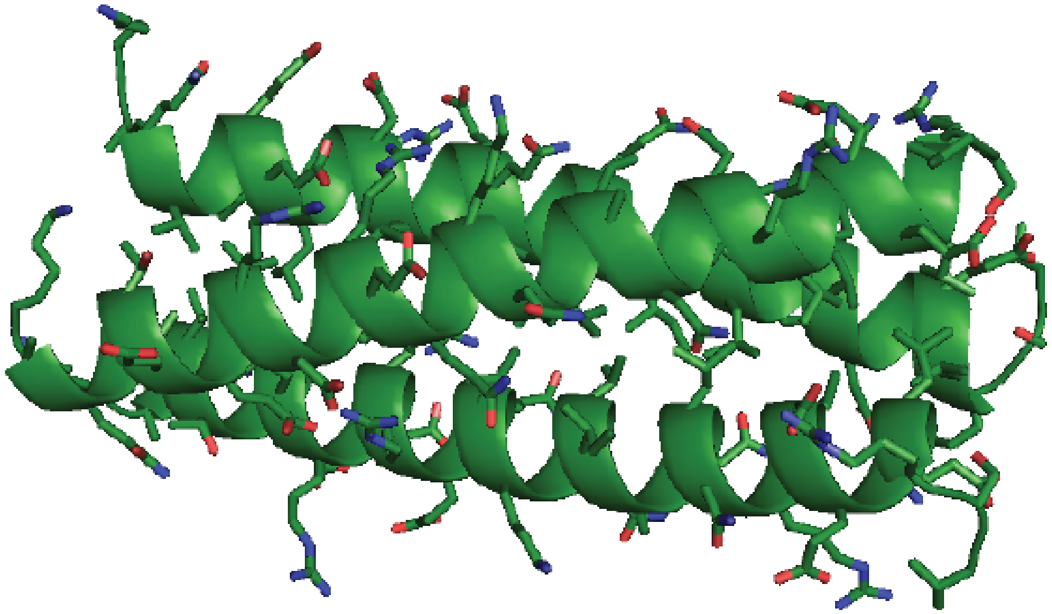
Inference / estimation depends only on subgraphs (e.g. dynamic programming, belief propagation, Gibbs sampling)

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

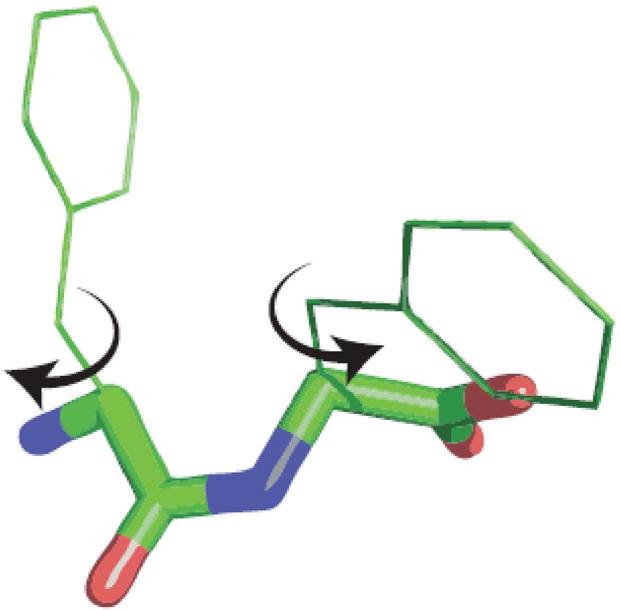
Protein Side Chain Prediction

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

Backbone + Side Chains



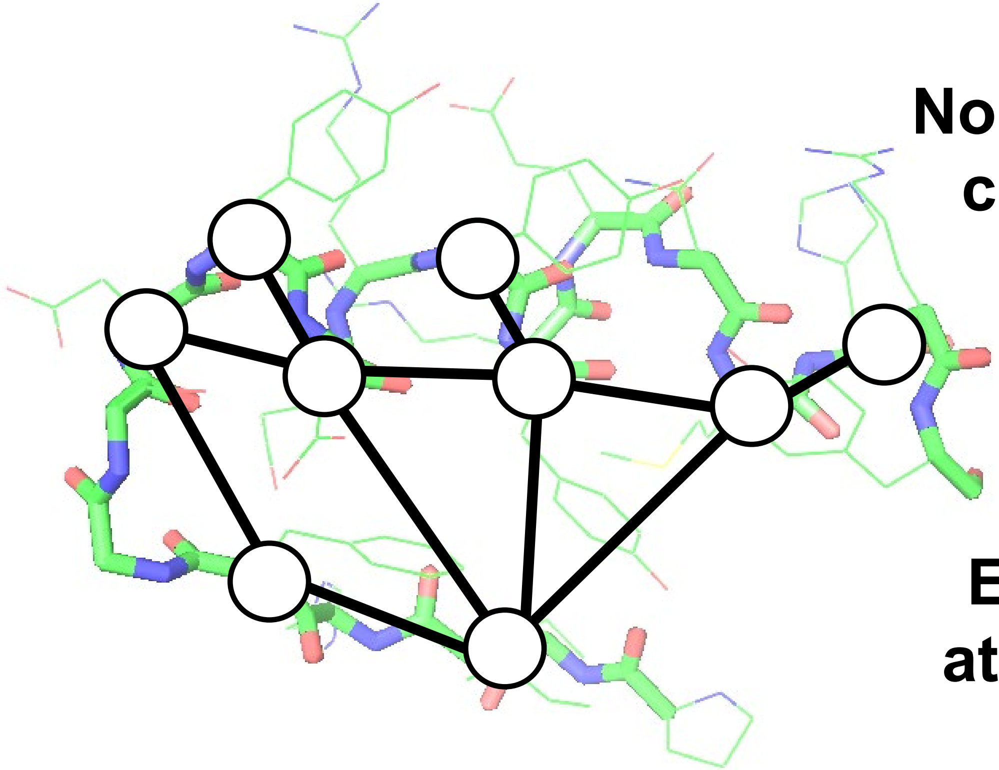
Side Chain Rotation



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

Protein Side Chain Prediction

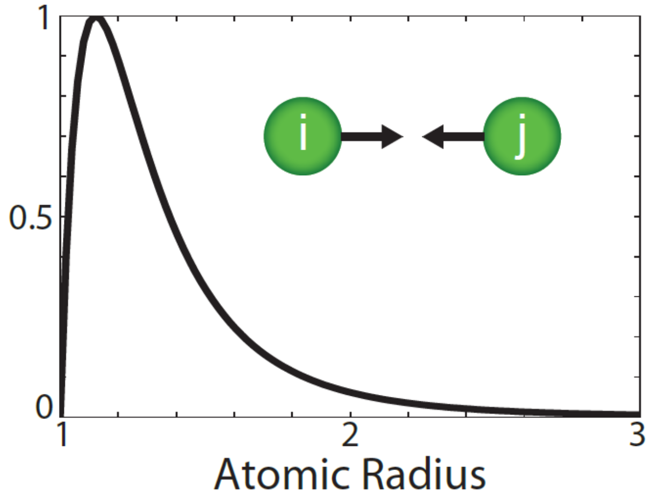
Graphical Model

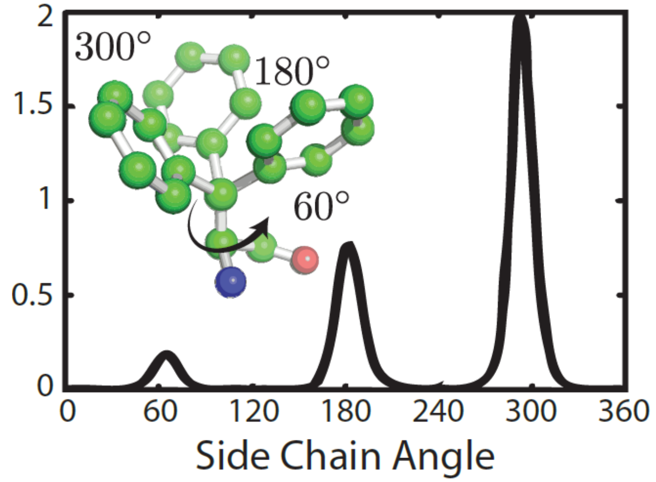


Nodes represent side chain orientations

Edges represent atomic interaction

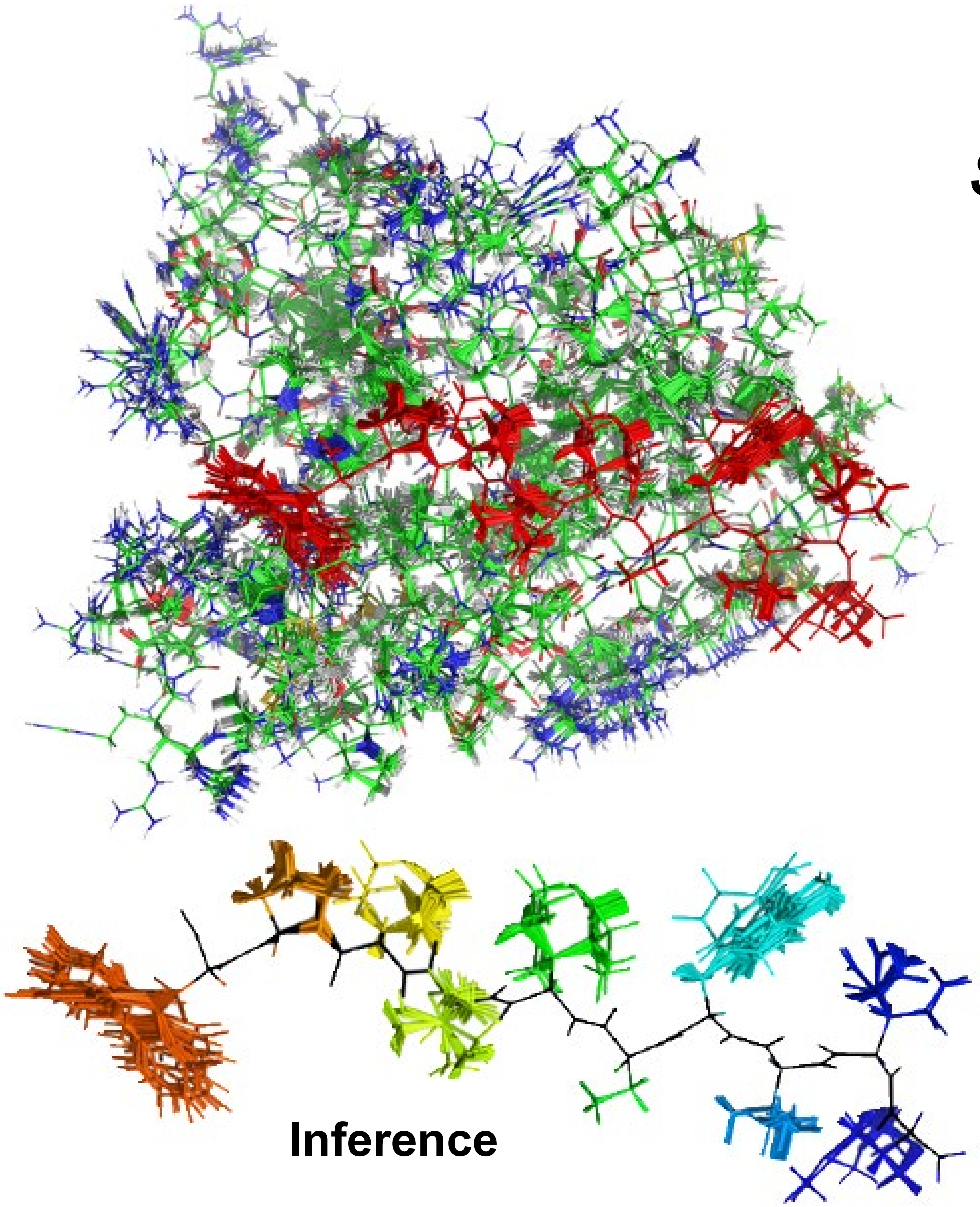
Configuration Likelihoods



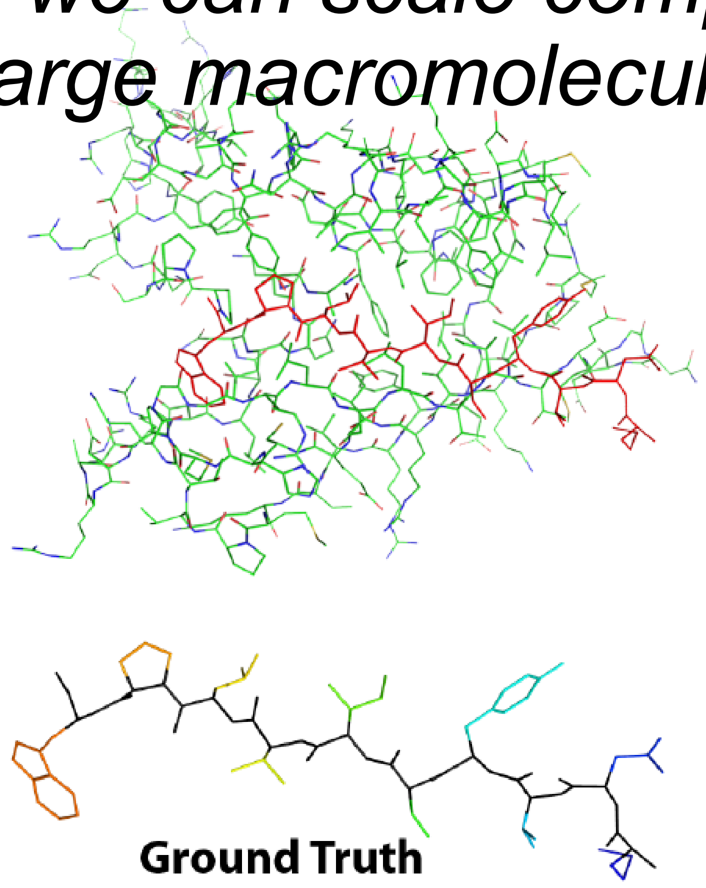


Complex phenomena specified by simpler atomic interactions

Protein Side Chain Prediction



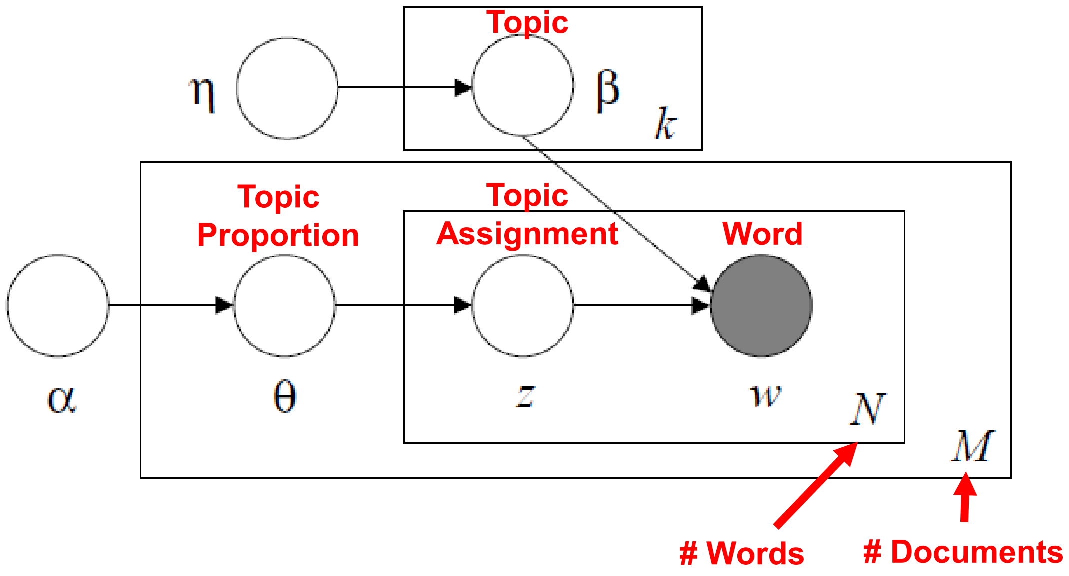
By exploiting graphical model structure we can scale computation to large macromolecules



Pacheco and Sudderth, ICML 2015

Topic Models

Latent Dirichlet Allocation (LDA)



Allows *unsupervised learning* 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

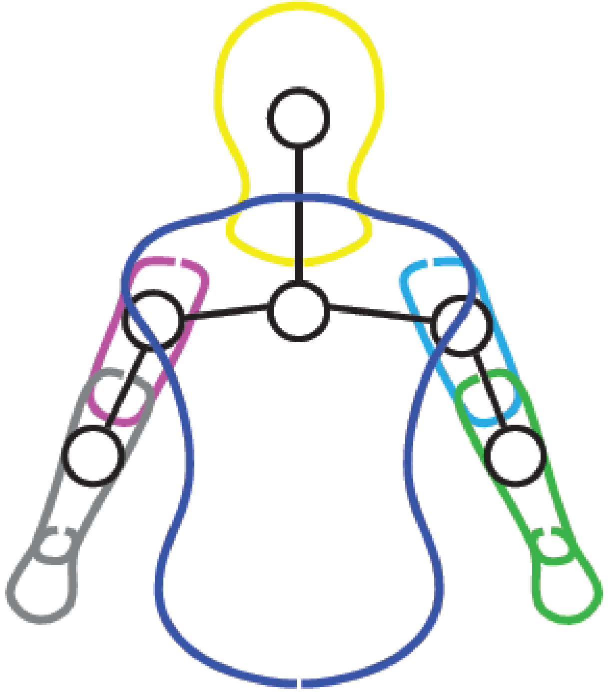
The William Randolph Hearst Foundation will give $1.25 million to Lincoln Center, Metropolitan Opera Co., New York Philharmonic and Juilliard School. "Our board felt that we had a real opportunity to make a mark on the future of the performing arts with these grants an act every bit as important as our 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

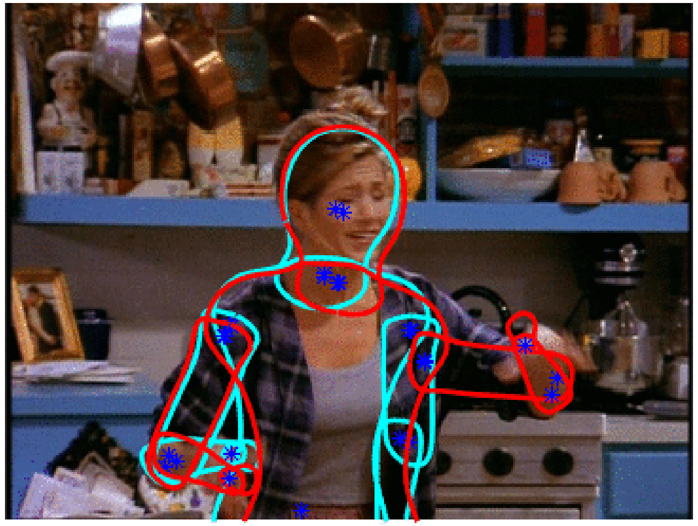
Estimate orientation / shape / pose of human figure from an image

Graphical Model

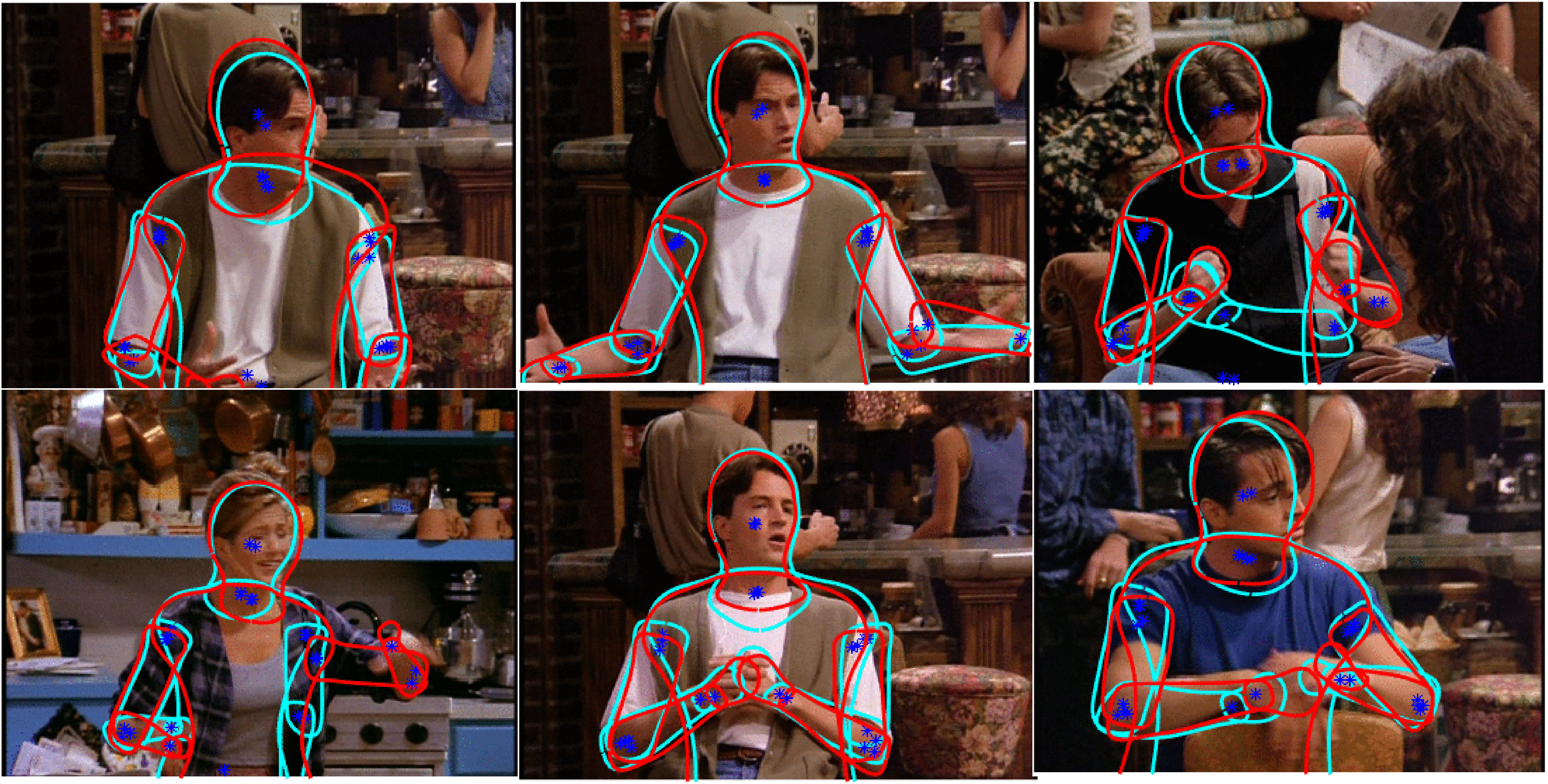


Data

Estimates

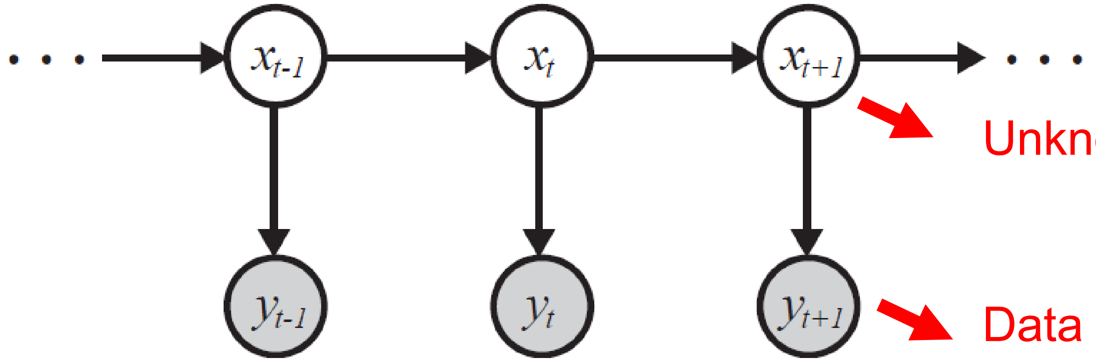


Pacheco, et al., ICML 2014



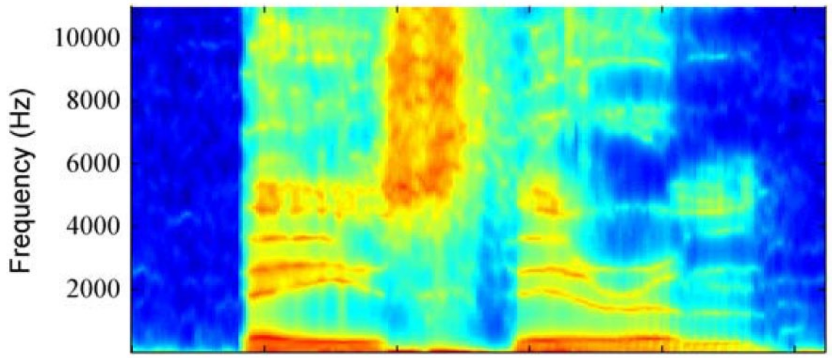
Hidden Markov Models

Sequential models of *discrete quantities of interest*



Example: Speech Recognition

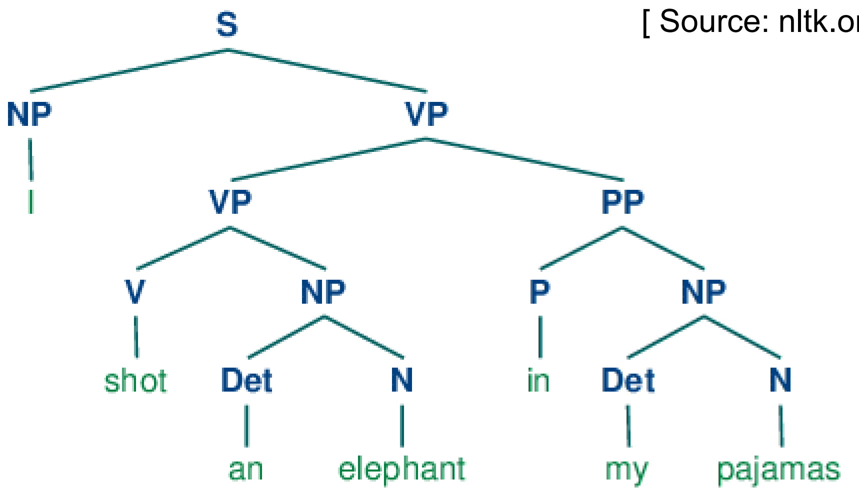
$y =$



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

[Source: Bishop, PRML]

[Source: nltk.org]



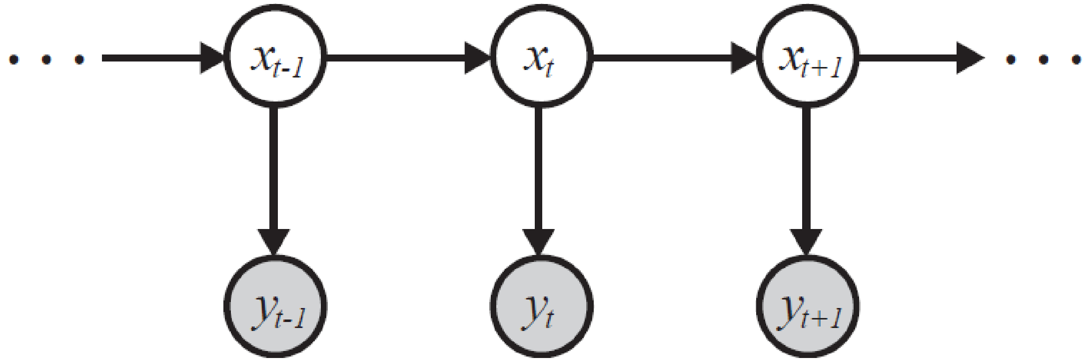
Example: Part-of-speech Tagging:

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

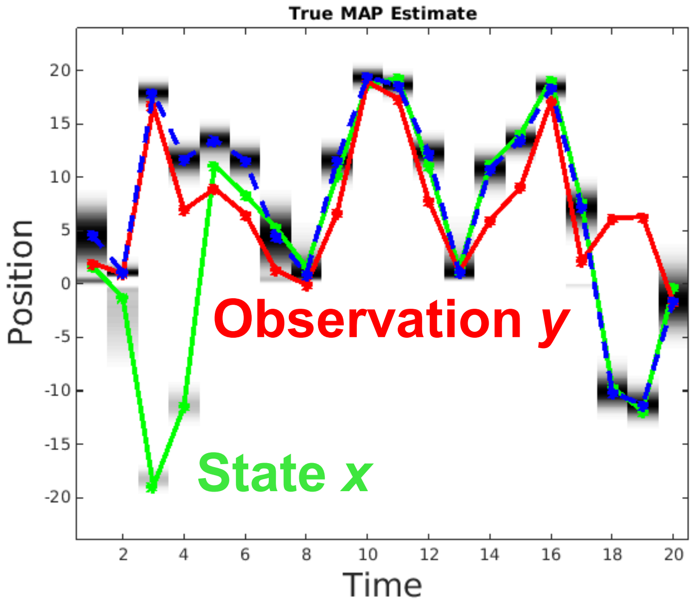
x = NP-V-Det-N-P-Det-N

Dynamical Models

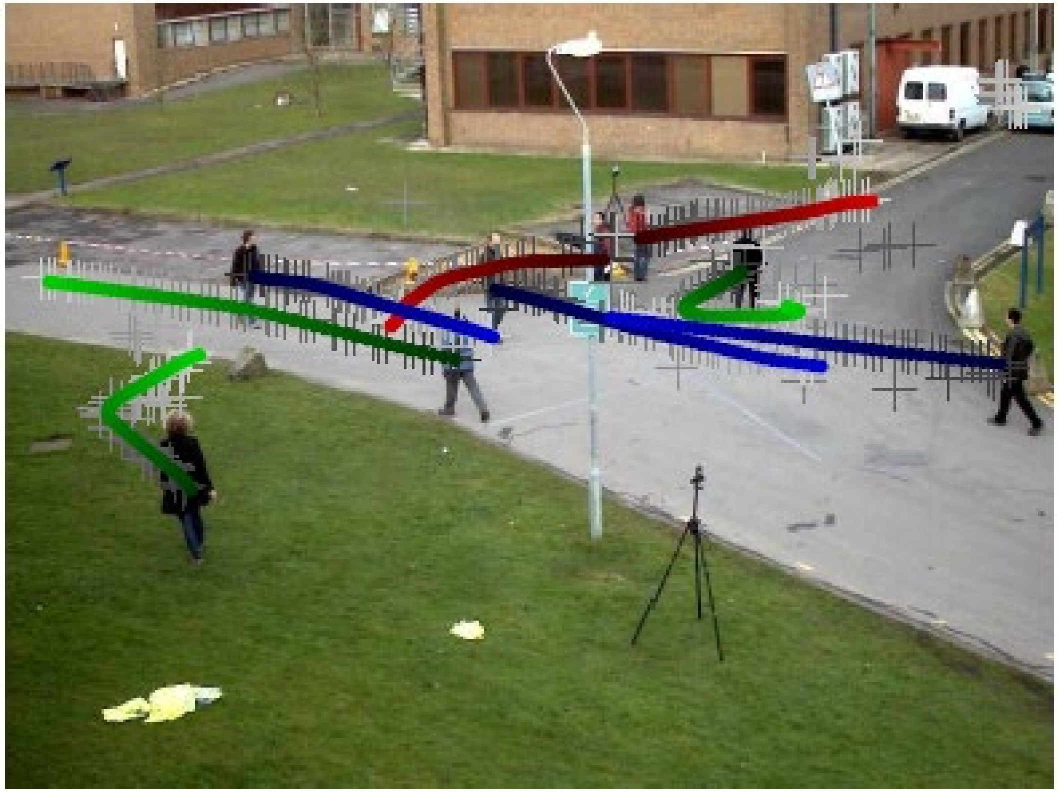
Sequential models of continuous *quantities of interest*



Example: Nonlinear Time Series

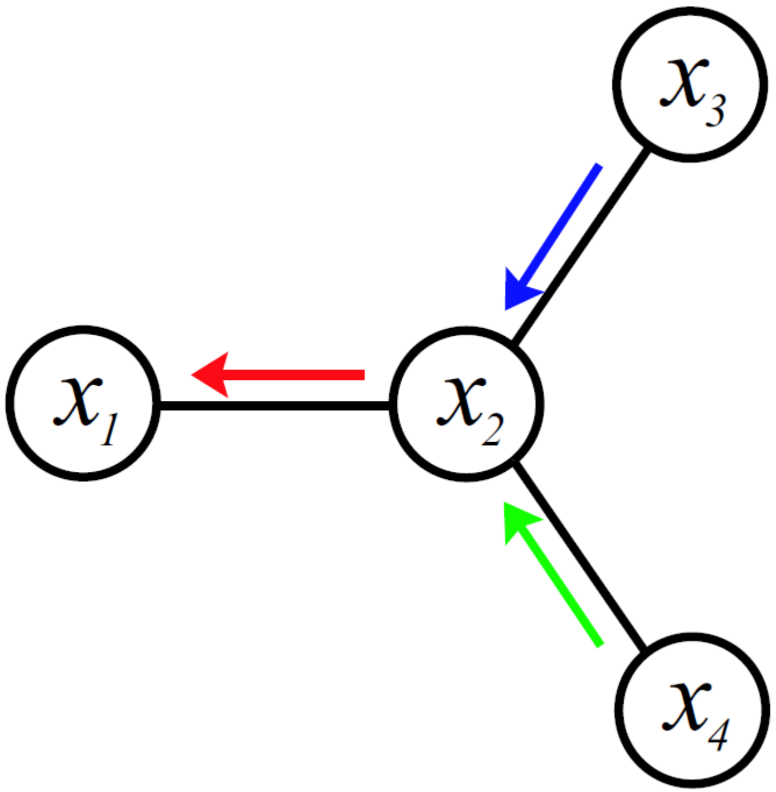


Example: Multitarget Tracking



Why Graphical Models?

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



<u>Representation</u>
Complex global phenomena arise by simpler-to-specify local interactions

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

Bayesian Inference

X **Quantity to be inferred**

D **Observed Data**

prior belief

likelihood

$$p(x \mid D) = \frac{\boxed{p(x)p(D \mid x)}}{p(D)}$$

model

Typically hard to compute

posterior belief

marginal likelihood

Posterior encodes our *belief* about unknowns *given* data

Marginal Likelihood

Inference typically involves solving high-dimensional integrals that lack a closed-form in non-trivial models… e.g. marginal likelihood:

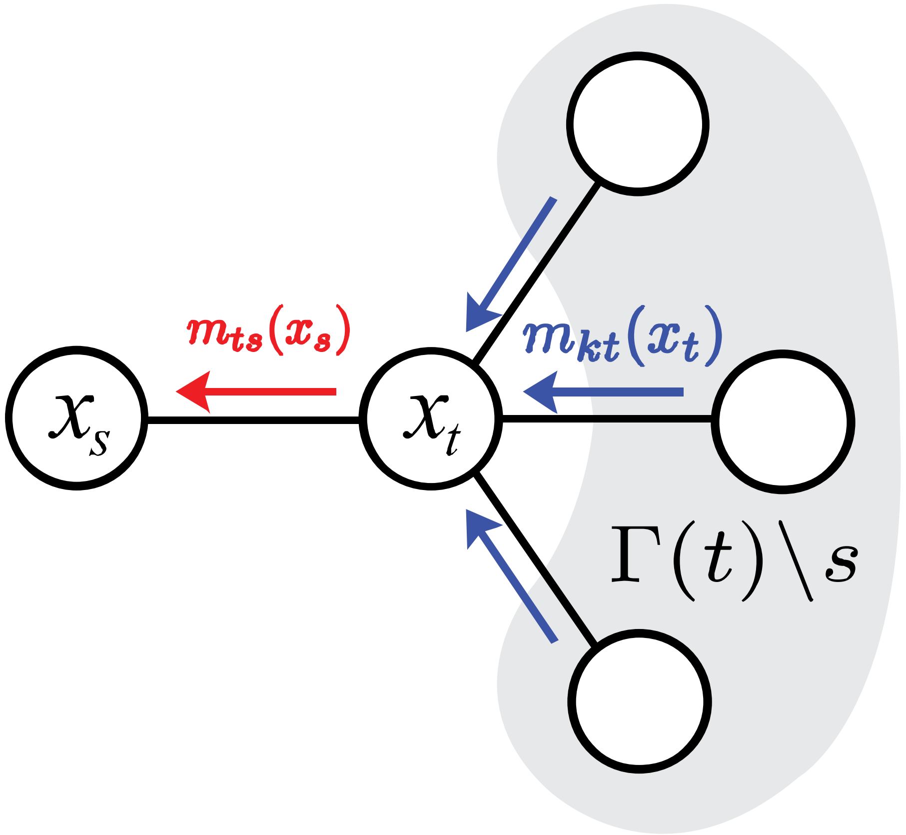
$$p(D) = \int p(x)p(D \mid x)\, dx$$

Example Pose estimation inference requires marginalizing over **every possible pose** that could ever occur. This is NP-hard…

As computer scientists, we will exploit graph structure to develop efficient algorithms…

Dynamic Programming (DP)

Breaks difficult global computations into simpler local updates

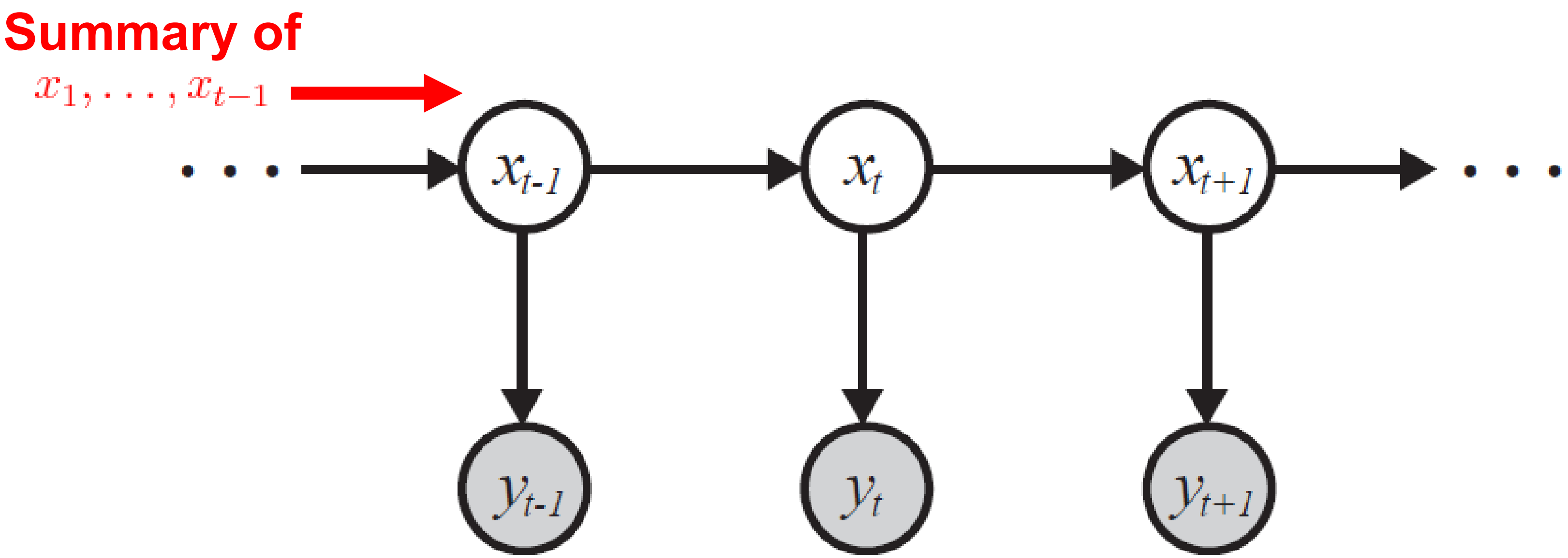


Many algorithms use some form of DP

- 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>

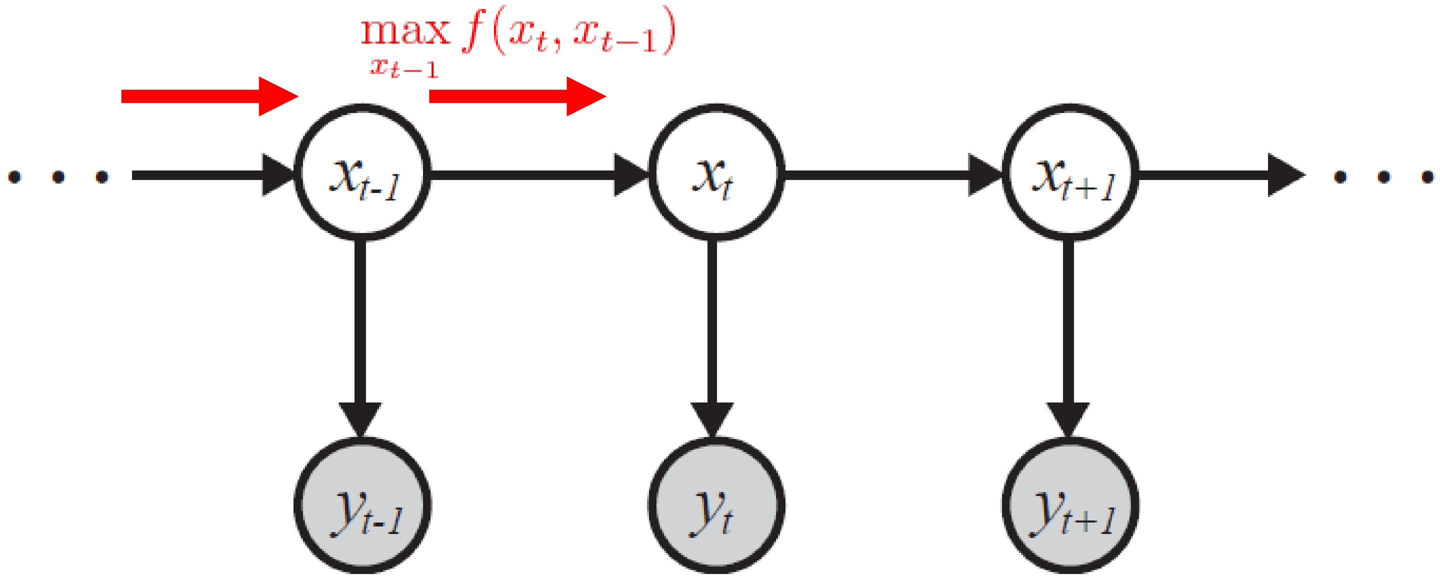
Viterbi Decoder



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

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

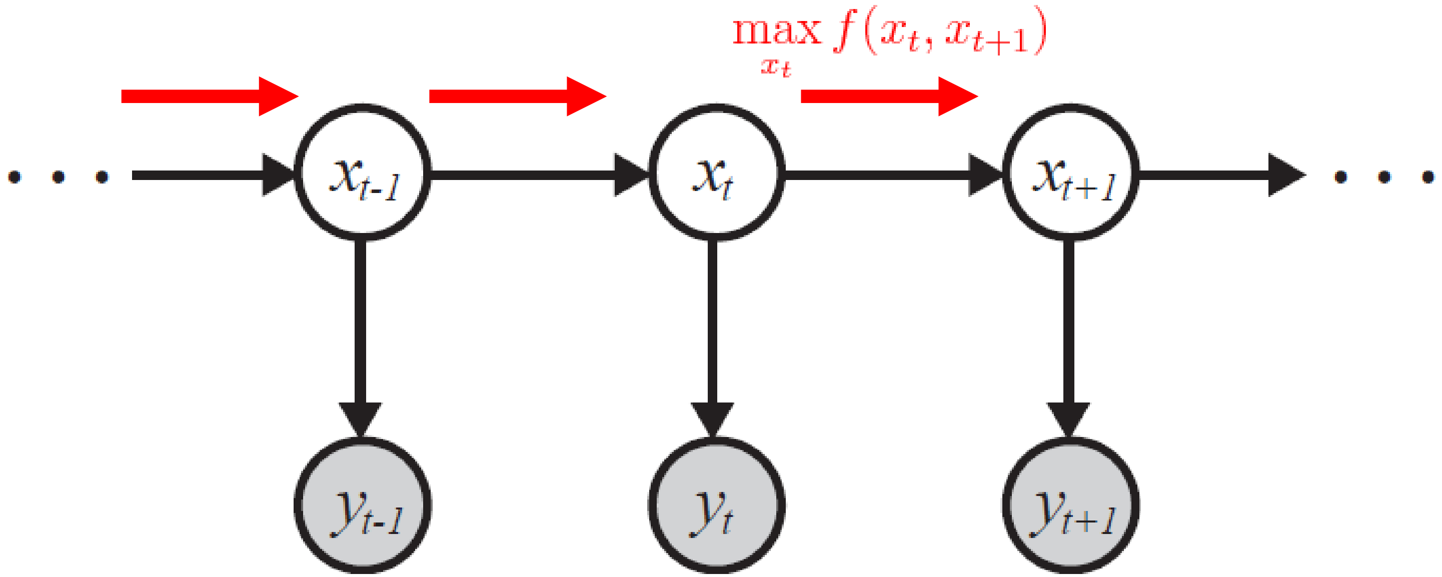
Viterbi Decoder



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

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

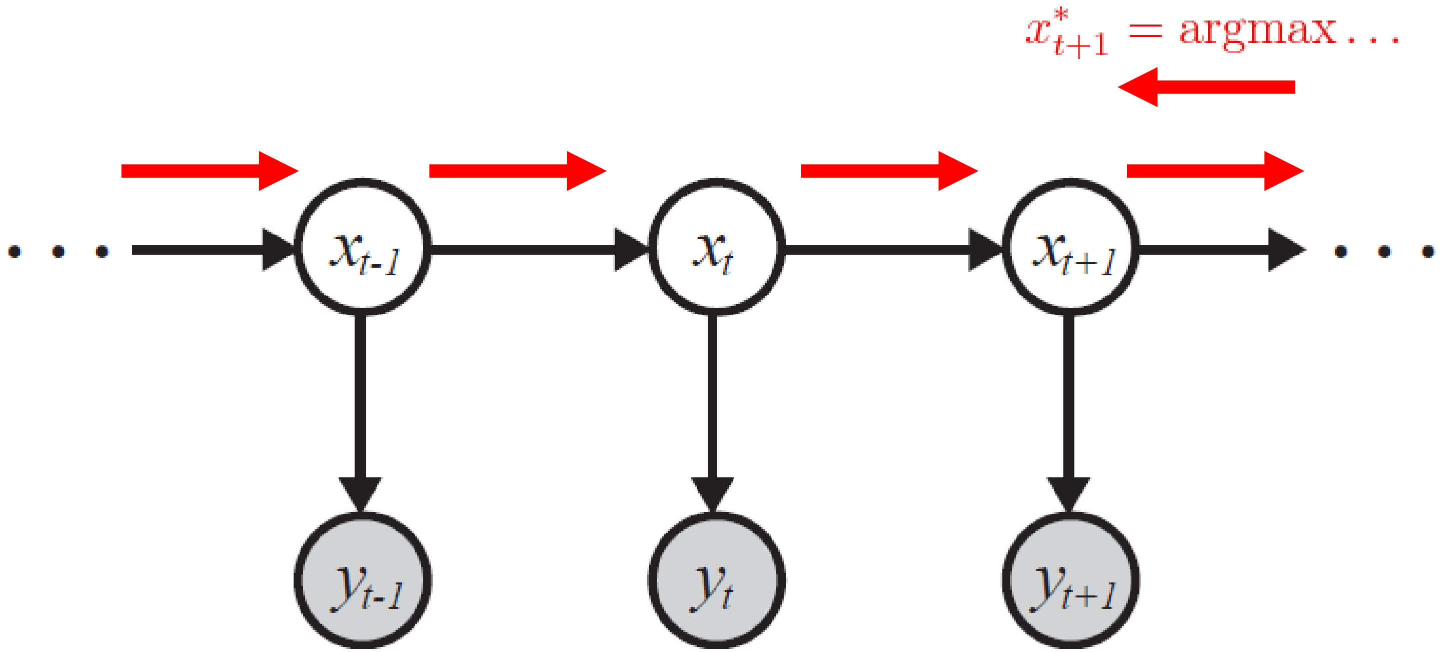
Viterbi Decoder



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

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

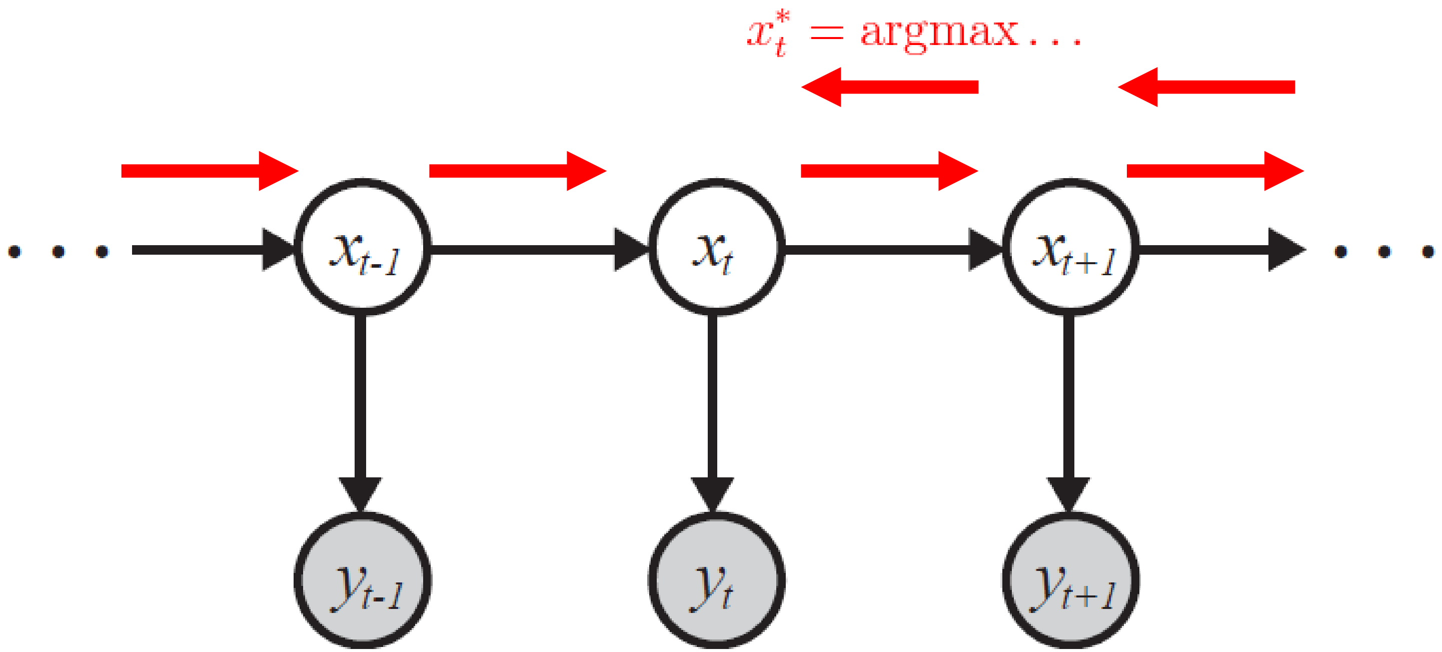
Viterbi Decoder



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

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

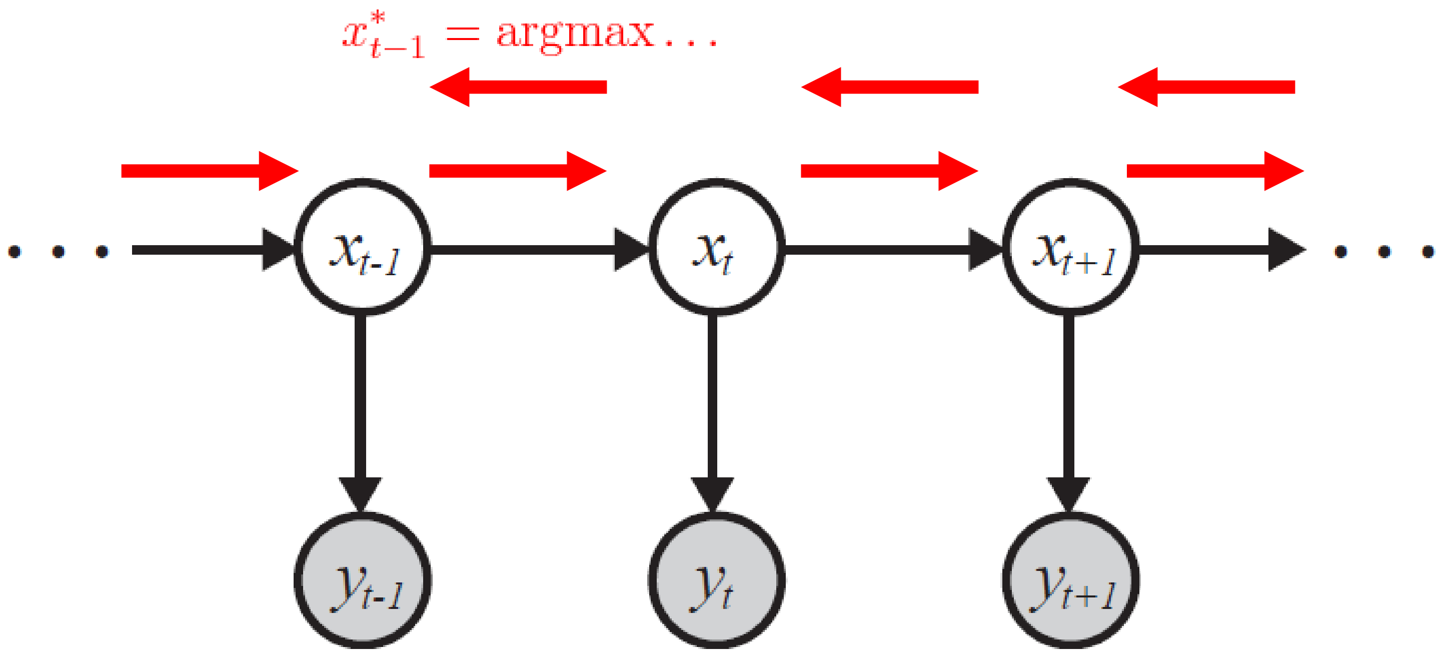
Viterbi Decoder



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

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

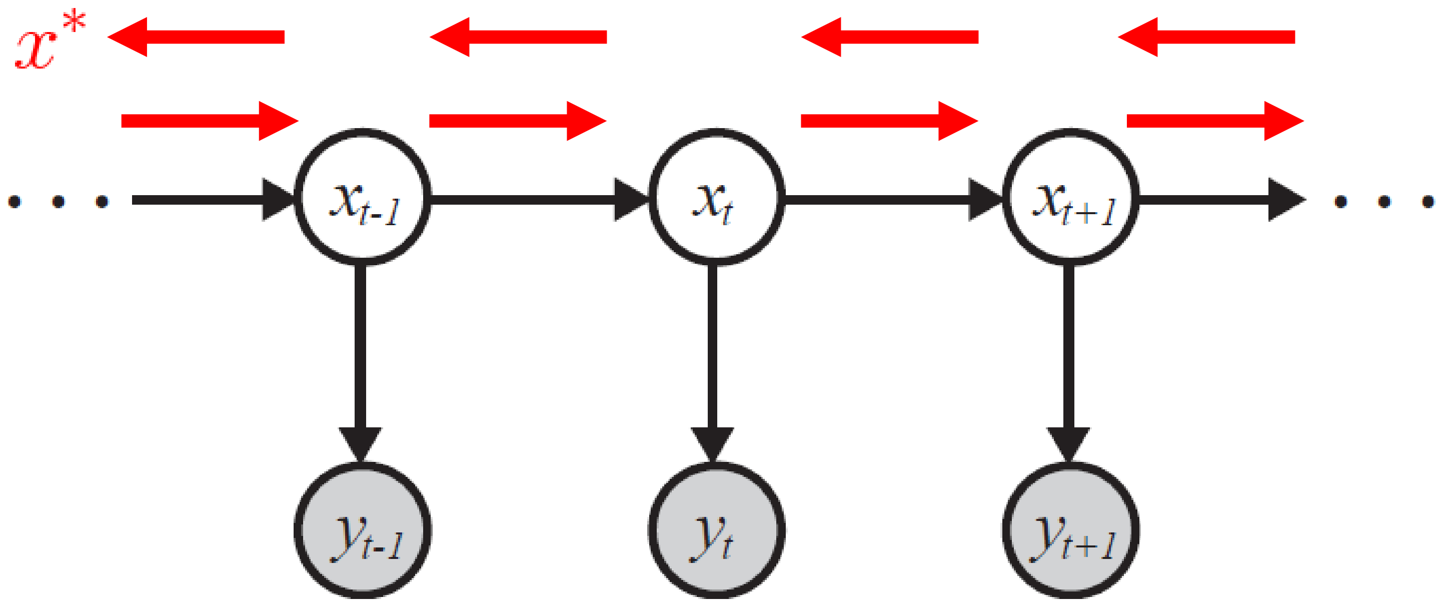
Viterbi Decoder



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

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

Viterbi Decoder



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

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

Course Objectives

In this course you will learn the following…

- Build on graphical models concepts from CSC 535
- Provide a deeper understanding with advanced algorithms for statistical inference in PGMs
- Understand approximate statistical inference methods
- Provide familiarity with recent PGMs research more advanced than CSC 535
- Ability to read, critique, and present research in PGMs
- Apply these methods to your research or a project of your choosing
- Ability to advance state-of-the-art in PGMs research

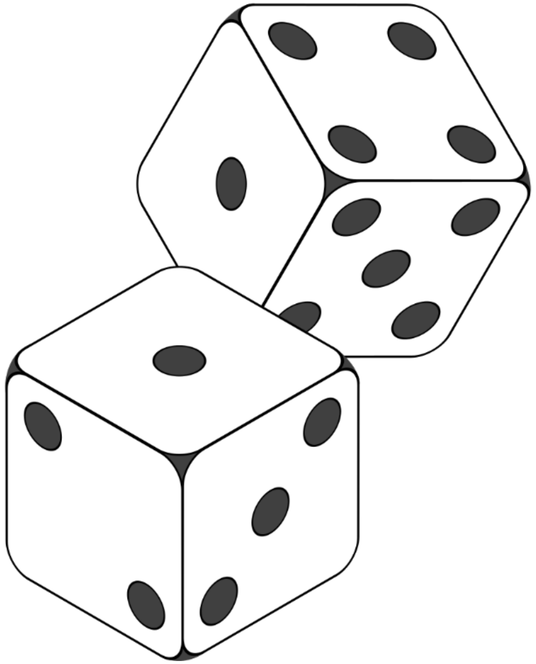
Course Overview

Course is broken down into **six** primary topics…

Probability and Statistics	Variational Inference	Advanced Monte Carlo Methods	Implicit Models	Bayesian Deep Learning	Gaussian Processes and Bayesian Optimization
Probability primer, Bayesian statistics, PGMs, Exponential families	Mean Field, Stochastic Variational, Expectation Propagation	Advanced Particle Filtering, Hamiltonian Monte Carlo, No U-Turn Sampler	Approximate Bayesian Computation, Neural Likelihood Free Inference	Bayesian Neural Network, Variational Autoencoder, Dropout Monte Carlo	Gaussian Process Regression, Bayesian Optimization

Probability and Statistics

Suppose we roll <u>two fair dice</u>…



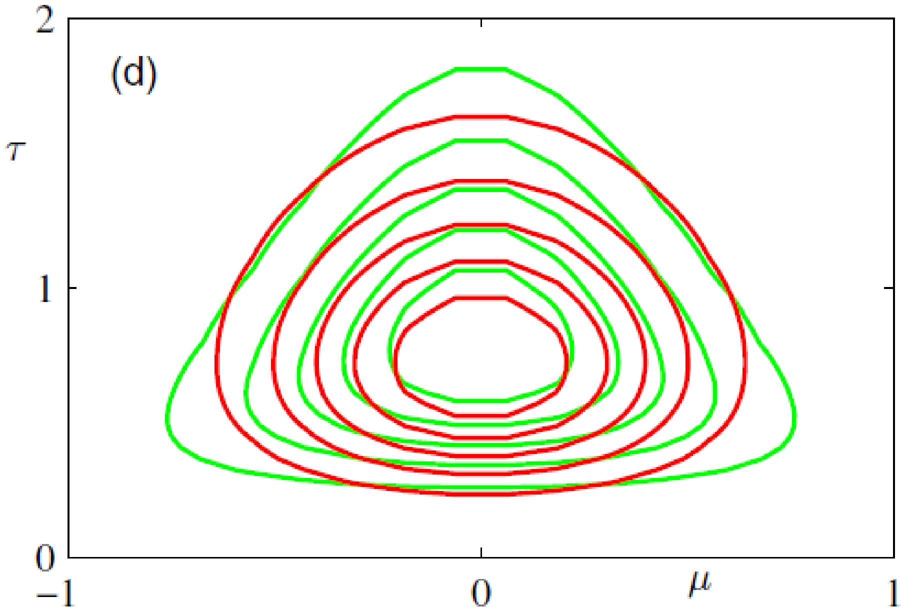
- What are the possible outcomes?
- What is the *probability* of rolling **even** numbers?

*… this is an **experiment** or **random process**.*

We will learn how to…

- Mathematically formulate outcomes and their probabilities?
- Describe characteristics of random processes
- Estimate unknown quantities (e.g. are the dice actually fair?)
- Characterize the uncertainty in random outcomes
- Identify and measure dependence among random quantities

Variational Inference



$$min_q \, \mathrm{KL}(q(x) \| p(x))$$

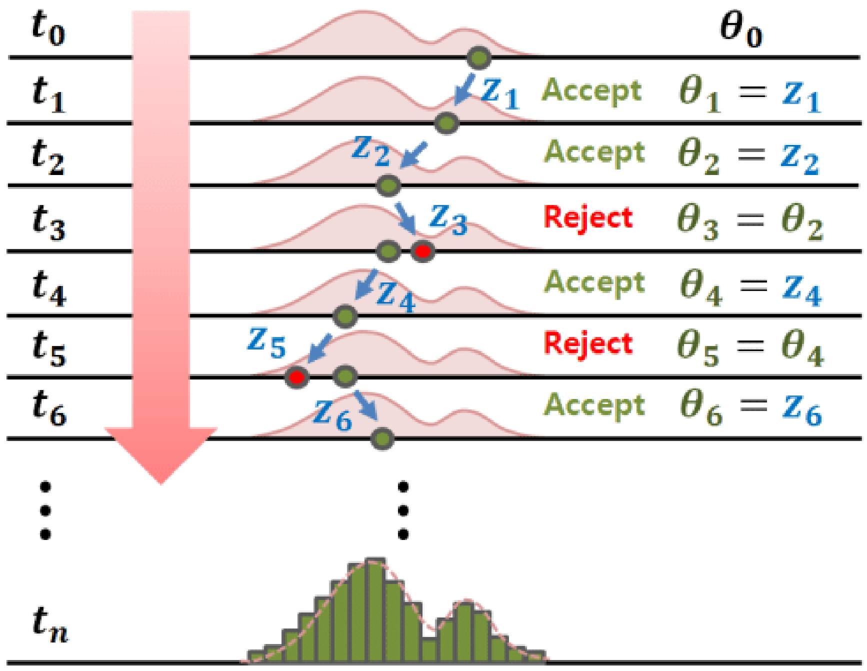
Recasts statistical inference as the solution to an optimization problem

We will learn how to conduct inference via,

- *Mean field and variational Bayes*
- *Stochastic variational*
- *Bethe free energy methods (Belief Propagation, Expectation Propagation)*

Monte Carlo Methods

Sample-based methods that simulate realizations from the model to perform inference



We will learn how to perform sample-based inference using:

- Rejection sampling
- Importance sampling
- Sequential importance sampling (particle filter)
- Markov chain Monte Carlo (MCMC) : Metropolis-Hastings
- MCMC : Gibbs Sampling

[Image: Jin, Seung-Seop et al. (2019)]

Implicit Models

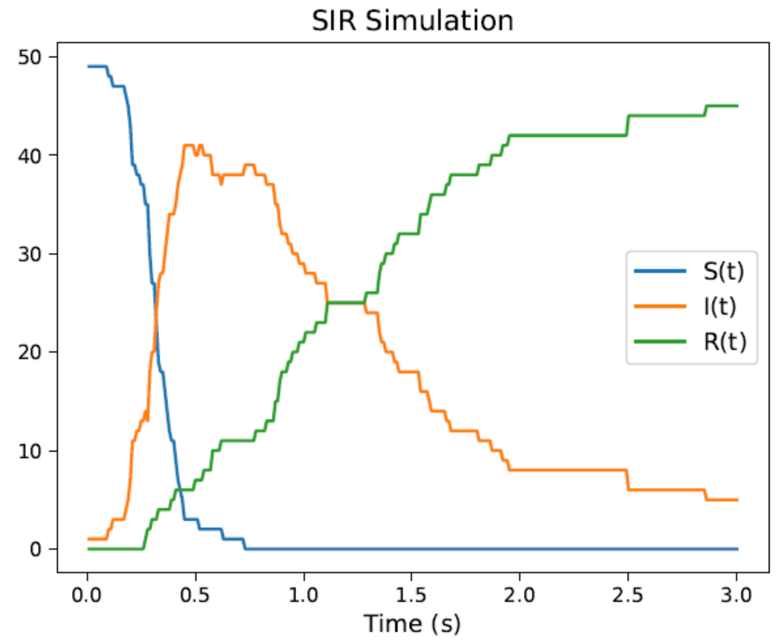
Some observation models are naturally defined via simulation processes…

SIR Model Example : Epidemiological model of disease among **S**usceptible, **I**nfected, **R**ecovered,

$$S(t+\Delta_t) = S(t) - \Delta I(t)$$
$$I(t+\Delta_t) = I(t) + \Delta I(t) - \Delta R(t)$$
$$R(t+\Delta_t) = R(t) + \Delta R(t)$$

With random additive noise,

$$\Delta I(t) \sim \text{Binomial}\left(S(t), \frac{\beta I(t)}{N}\right), \quad \Delta R(t) \sim \text{Binomial}(I(t), \gamma)$$

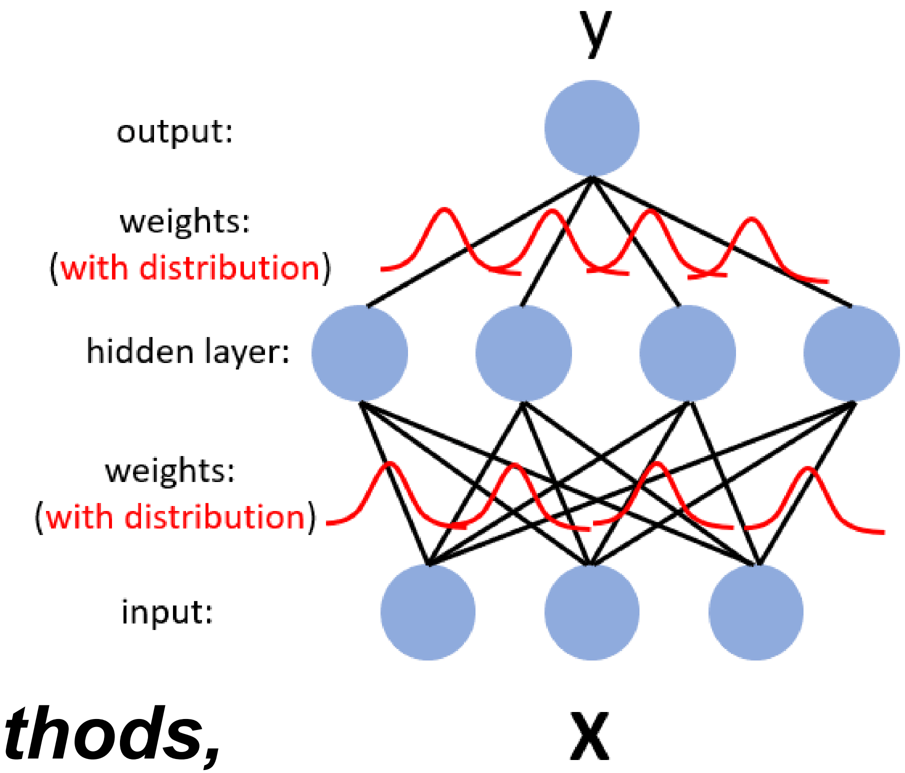


For some time interval Δ_t we can simulate S/I/R from initial conditions, but have no closed-form likelihood,

$$p(S, I, R \mid \beta, \gamma) = ???$$

Bayesian Deep Learning

Combine probabilistic reasoning with Deep Learning models



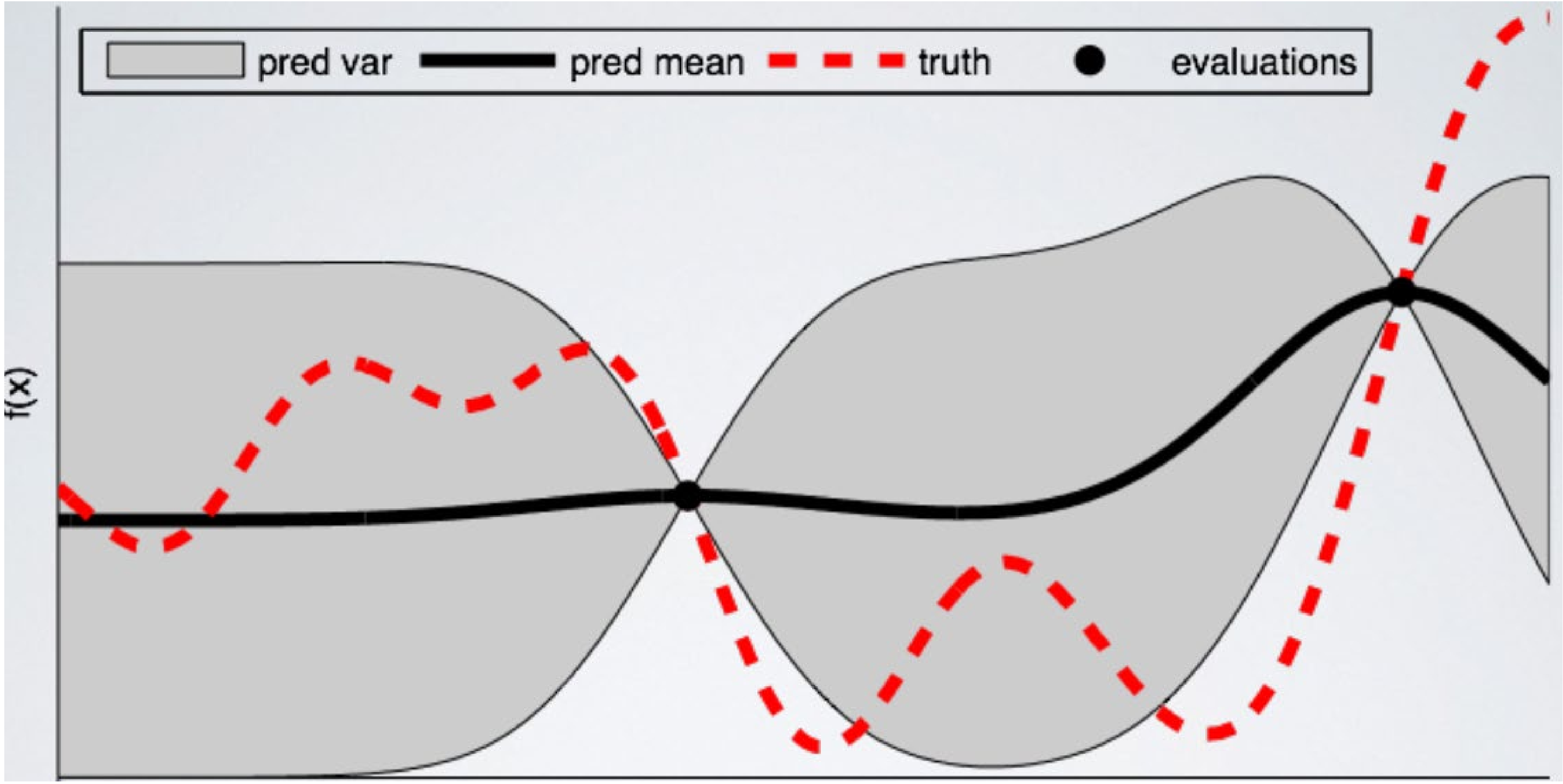
We will learn the following models and methods,

- *Variational autoencoder*
- *Bayesian Neural Network*
- *Structured Variational Autoencoders*
- *Dropout predictions*

[Image: Lin, Xi et al. (2018)]

Bayesian Optimization

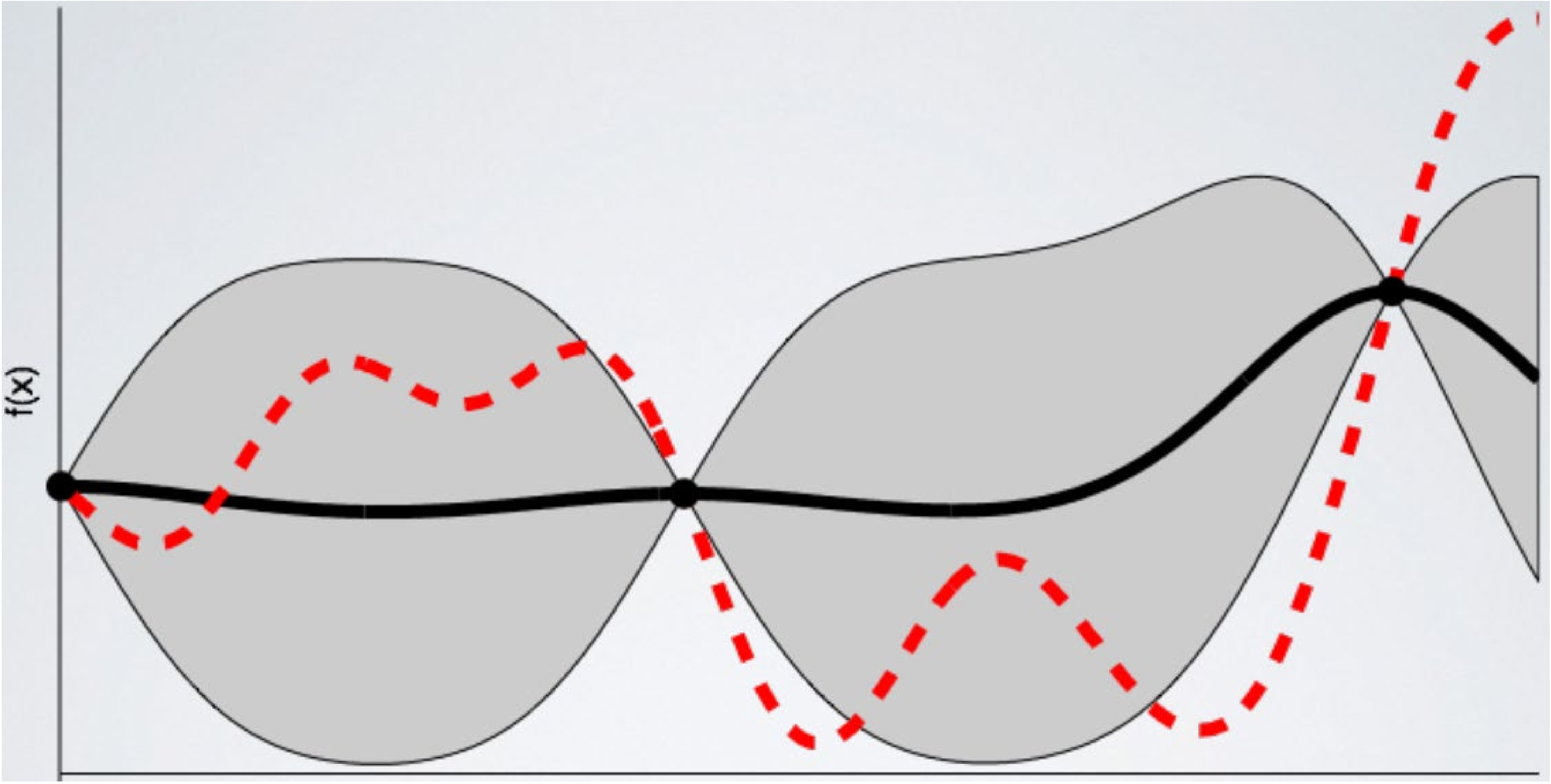
Global optimization of <u>random functions</u>: $\min_x f(x)$



[Source: Ryan Adams]

Bayesian Optimization

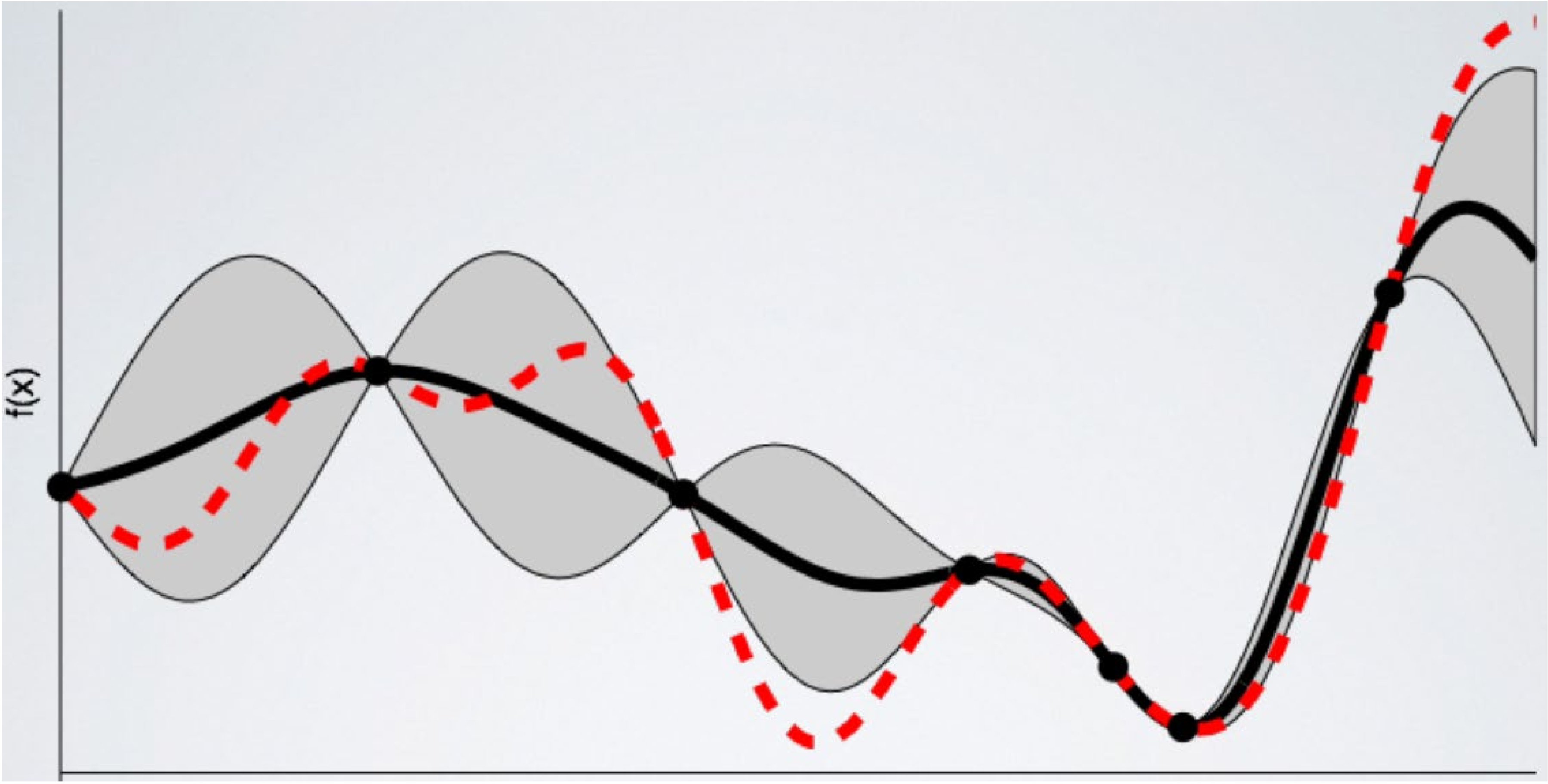
Iteratively updates distribution over function value (regression)



[Source: Ryan Adams]

Bayesian Optimization

The function is well-approximated around the minimizer



[Source: Ryan Adams]

Online Resources

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

http://pachecoj.com/courses/csc696h_fall22/

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

https://d2l.arizona.edu/d2l/home/1205997

We will use **Piazza** for discussion:

https://piazza.com/arizona/fall2022/csc696h1

Grading Breakdown

- 10% - Attendance / Participation
- 15% - Paper presentations
- 10% - Critical reading summaries
- 25% - Term project proposal
- 40% - Term project (presentation and writeup)

Note: This breakdown differs slightly from the syllabus. I will be updating the syllabus to reflect.

Grading Questions

- I will announce in class and/or Piazza when grading of each item is complete
- Officially, you have **1 week** to raise any grading concerns (from the completion of grading)
- If you don't receive a grade, but should have, you must tell me **within 1 week**

Necessary Coding and Math Background

Basic Probability and Statistics
- Marginal / Conditional Probability
- Bayes' Rule

Familiarity with Probabilistic Graphical Models
- Bayes Nets / Markov Random Fields / Factor Graphs
- Conditional Independence

Basic understanding of optimization
- Nonlinear vs. linear programming
- Gradient ascent
- Dynamic programming (we will cover the basics)

Basic Linear Algebra
- Basic vector / Matrix algebra
- Matrix inversion / rank / condition

Basic coding and data structures (for semester project)
- Ideally familiarity with Python / Numpy / Scipy
- Should be able to write / manage code on the order of 1,000 lines

Critical Reading Summaries

- Submit on D2L by the night **before** lecture
- A brief paragraph with at least one sentence on each:
  - What is the strength of this paper?
  - What is the primary weakness of this paper?
  - What would you have done differently to improve the paper?
- With each of the above:
  - **Be specific:** Reference figures / equations when possible
  - **Be original:** Consider aspects that are not obvious on a first reading
  - **Be concise:** Make your points in 3 to 6 sentences

Paper Presentations

- Each student will choose N assigned papers (N=?)
- Choose early:
  - Typically the “easy” papers are selected early
  - I will assign one to you if you don’t choose
- Plan for 45min presentations to leave room for discussion
- Board presentations are acceptable if it is appropriate for the material (e.g. lots of math)
  - Don’t use a board presentation for papers that don’t require them
  - If you’re unsure then check with me ahead of time

Term Project

1) Select a project topic
2) Write a proposal (10% of grade)
3) Execute the project
4) Present during final exam period (40% of grade)

- Individual projects only unless you have a compelling reason
- A component of your graduate research can be used but must be proposed and executed as a standalone sub-problem
- I will be asking that you share Github repos for each project

Late Policy

Please complete reading summaries by the **night before lecture** on that topic

- I will usually grade these at the end of the week, if yours isn't submitted by then then it will not be graded

The **project proposal** and **project report** need to be submitted on time!

- If you are having issues then notify me ahead of time and I will deal with it on a case-by-case basis

Academic Integrity Continued

- All reading summaries **must** be done **independently**
- Project proposals and reports **must** be done **independently**
- You may discuss project details with others, but **do the work yourself**
- **Cite any and all resources** (that includes your own work)
- **Do not** submit your previous work for the term project—it must be novel work (I will look at Github commits)

Good Rule Cite any external resource you use that may be considered plagiarism without citation.

Lectures and Attendance

In-Person Attendance

- I ask that students attend in-person
- I have Zoom meetings scheduled on D2L, but this **should be used sparingly**
- If you might have COVID-19 symptoms then use Zoom
- Attendance and participation will be graded (10% of grade)

Lecture Recordings

- All lectures will be recorded and posted online after-the-fact
- Recordings are accessible via D2L
- Recorded lectures should supplement lecture, **they should not** be used in place of lecture

Office Hours

Use scheduled office hours for

- Specific homework questions
- Clarification on lecture / reading topics
- General course-related questions

Details

- 2 hours per week
- Office hours will be held on Zoom, but schedule is TBD
- Message me on Piazza if you have a conflict with hours and I will try to schedule something for you
- I still need to set a time for hours – ignore the current time in the Syllabus / course webpage

Piazza

- Use Piazza for **all course communication**
- If you email me directly I may not see it (I get a lot of email)
- You can ask / answer questions related to the course
- Also post course-related material (e.g. if you find something on the internet that is interesting / useful to the course)

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

Questions? Comments? Thoughts?