MACHINE LEARNING DEPARTMENT

10-301/601 Introduction to Machine Learning

Machine Learning Department
School of Computer Science
Carnegie Mellon University

Bayesian Networks
+
Reinforcement Learning:
Markov Decision Processes

Matt Gormley & Henry Chai
Lecture 21
Nov. 8, 2021

Reminders

- **Homework 7: HMMs**
  - Out: Wed, Nov. 03
  - Due: Fri, Nov. 12 at 11:59pm

Q&A

Q: Lecture: Would you be so kind as to end lecture on time?

A:

Q&A

Q: <u>Lecture</u>: The larger-than-life in-class demonstrations are absolutely amazing. Could you do more of them?

A: Honestly, as the core material becomes increasingly complex it will be quite difficult, but we sure can try!

Q: <u>Lecture</u>: The larger-than-life in-class demonstrations are really boring and take up a lot of time. Could you do less of them?

A: Honestly, that would make our lives a lot easier, so we sure can try!

Q&A

Q: Lectures: Could you upload the slides a day ahead of time?

A: Yes, we can do that.

(Just a heads up that the slides might change slightly after that first upload.)

Q: Homework: Some of the multiple choice homework questions are ambiguous or you end up changing the questions later

A: We are trying to improve our own testing to try to catch these sorts of bugs early. They tend to come up specifically in these heavily constrained multiple choice problems.

Q&A

Q: Recitation: Some of the TAs handwriting is even worse than yours (some is much better), could you all work on that?

A: Ah. We hadn't thought of that – sorry! We've just instituted some digital handwriting practice for those who haven't had much. (We used to use chalkboards, but don't have those this semester.)

Q: Recitation: It'd be great if recitations left more time for students to solve the problems.

A: Sorry about that. We've been trying to pack more and more in and rushing a bit through the interactive-problem-solving parts as a result.

GRAPHICAL MODELS: DETERMINING CONDITIONAL INDEPENDENCIES

What Independencies does a Bayes Net Model?

- In order for a Bayesian network to model a probability distribution, the following must be true:

  Each variable is conditionally independent of all its non-descendants in the graph given the value of all its parents.

- This follows from $P(X_1 \ldots X_n) = \prod_{i=1}^{n} P(X_i \mid parents(X_i))$

$$= \prod_{i=1}^{n} P(X_i \mid X_1 \ldots X_{i-1})$$

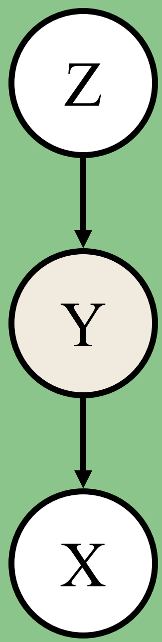
- But what else does it imply?

Slide from William Cohen

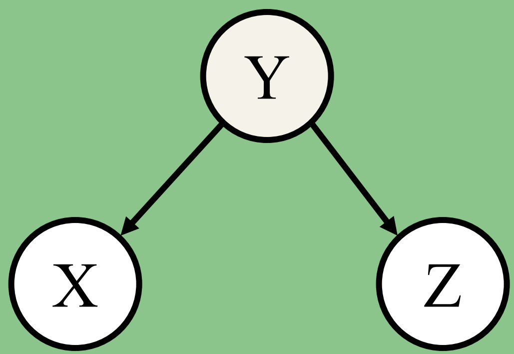
What Independencies does a Bayes Net Model?

Three cases of interest…

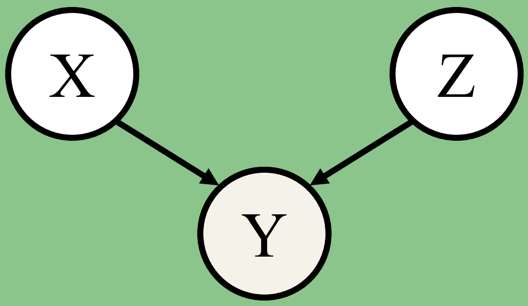
Cascade



Common Parent



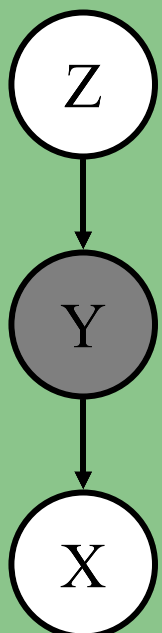
V-Structure



What Independencies does a Bayes Net Model?

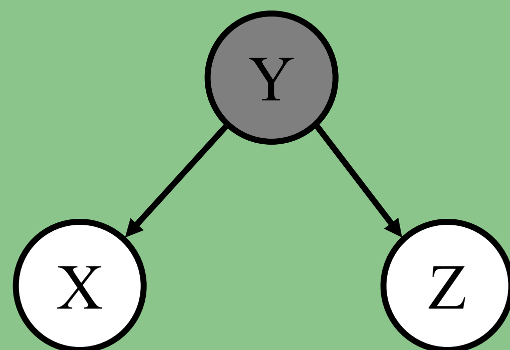
Three cases of interest…

Cascade



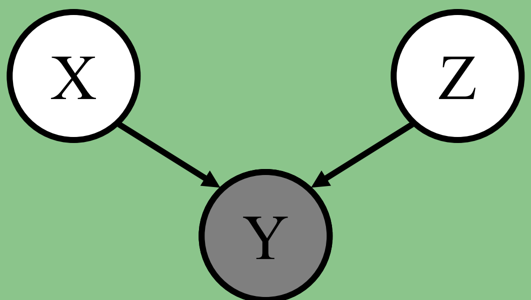
$$X \perp\!\!\!\perp Z \mid Y$$

Common Parent



$$X \perp\!\!\!\perp Z \mid Y$$

V-Structure



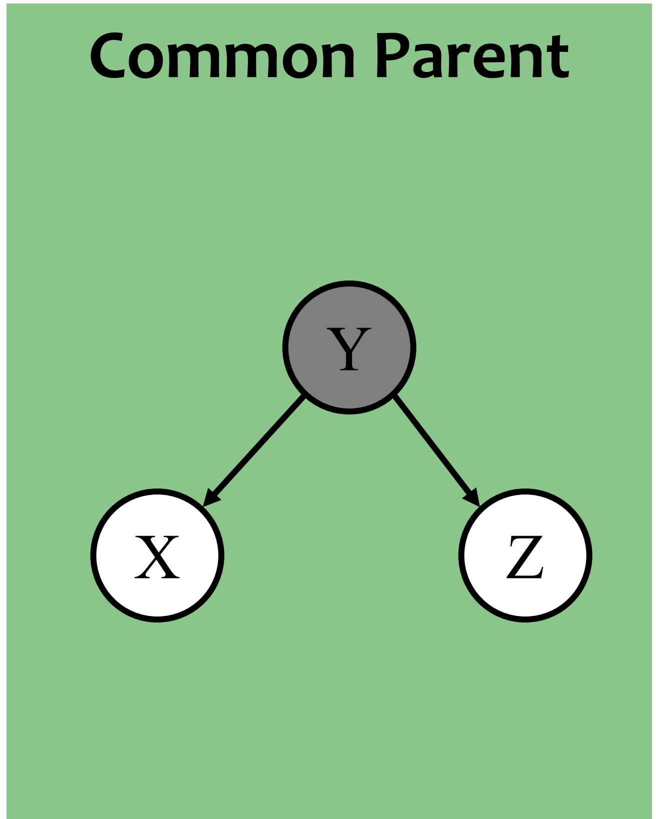
$$X \not\perp\!\!\!\perp Z \mid Y$$

Knowing Y **decouples** X and Z

Knowing Y **couples** X and Z

Whiteboard

Proof of conditional independence

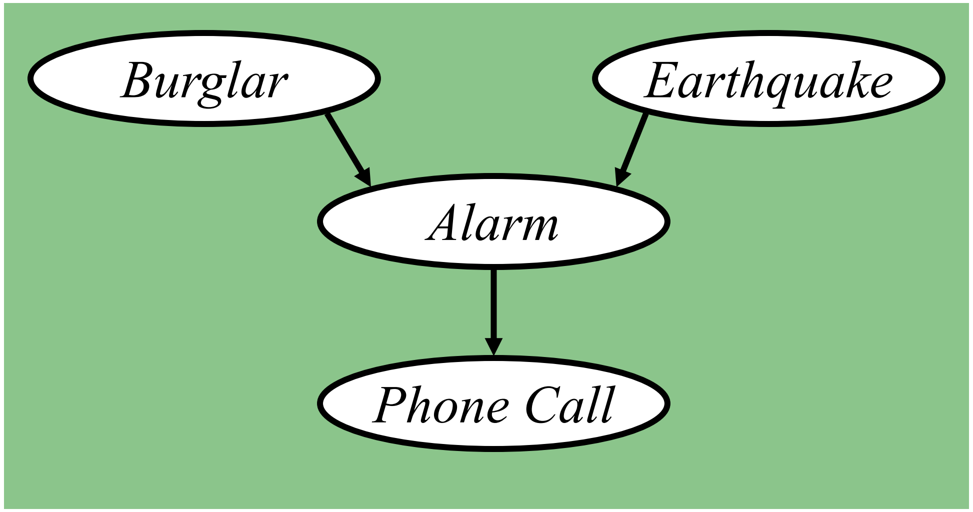


$$X \perp\!\!\!\perp Z \mid Y$$

(The other two cases can be shown just as easily.)

The "Burglar Alarm" example

- Your house has a twitchy burglar alarm that is also sometimes triggered by earthquakes.
- Earth arguably doesn't care whether your house is currently being burgled
- While you are on vacation, one of your neighbors calls and tells you your home's burglar alarm is ringing. Uh oh!



Quiz: True or False?

$$Burglar \perp\!\!\!\perp Earthquake \mid PhoneCall$$

Slide from William Cohen

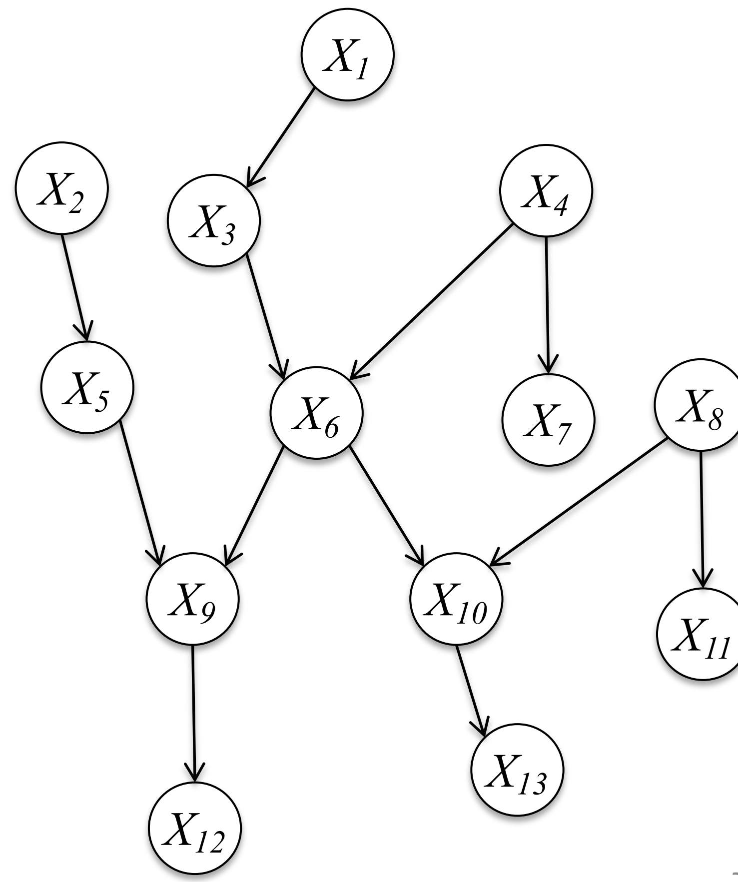
Question 1

A
B
C

Markov Blanket

Def: the **co-parents** of a node are the parents of its children

Def: the **Markov Blanket** of a node is the set containing the node's parents, children, and co-parents.

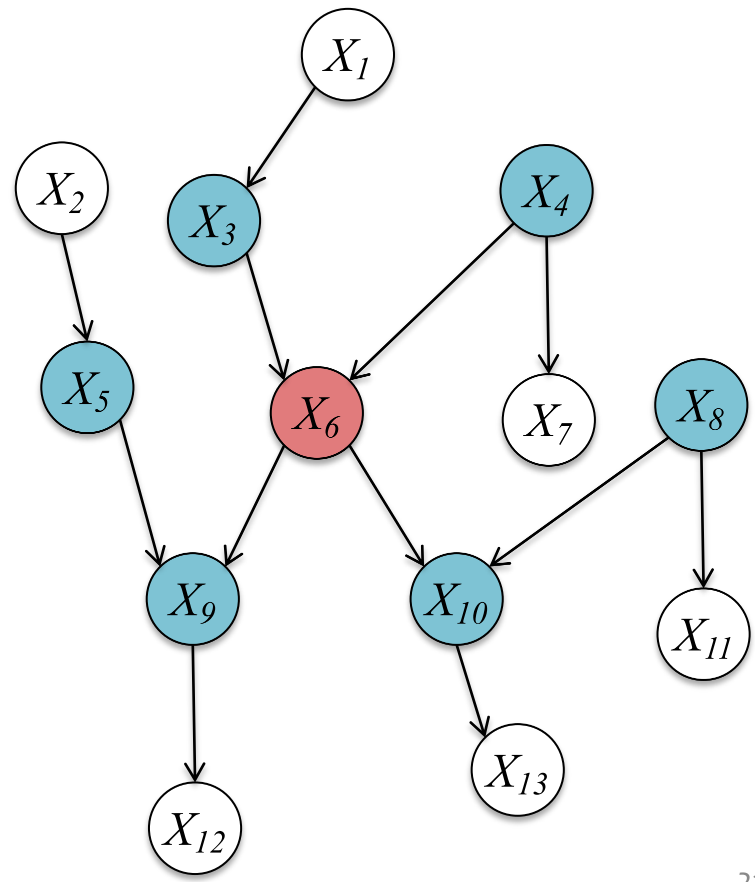


Markov Blanket

Def: the **co-parents** of a node are the parents of its children

Def: the **Markov Blanket** of a node is the set containing the node's parents, children, and co-parents.

Example: The Markov Blanket of X_6 is $\{X_3, X_4, X_5, X_8, X_9, X_{10}\}$



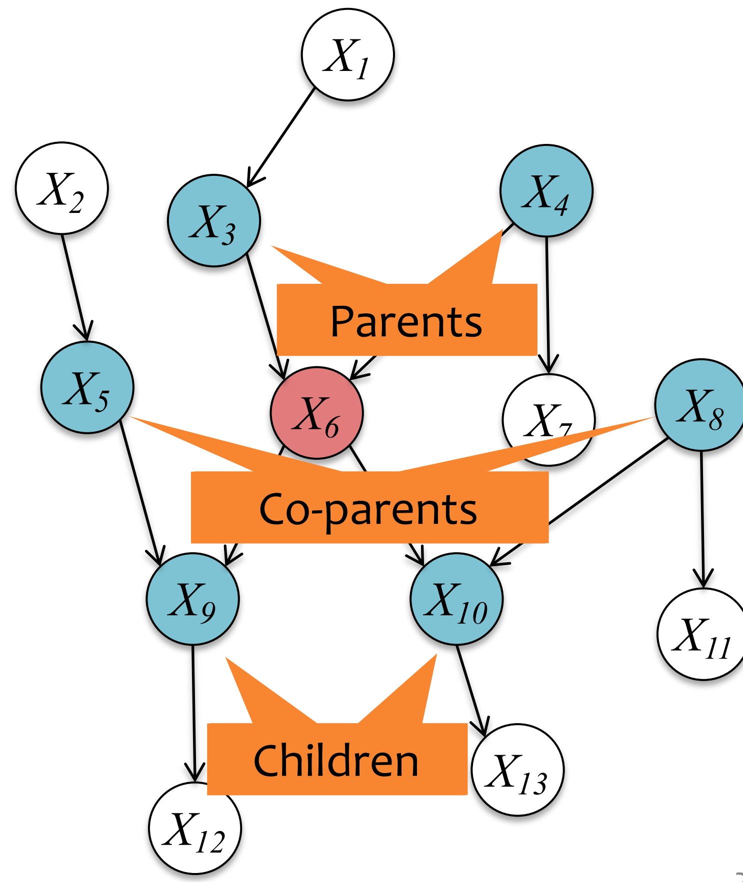
Markov Blanket

Def: the **co-parents** of a node are the parents of its children

Def: the **Markov Blanket** of a node is the set containing the node's parents, children, and co-parents.

Theorem: a node is **conditionally independent** of every other node in the graph given its **Markov blanket**

Example: The Markov Blanket of X_6 is $\{X_3, X_4, X_5, X_8, X_9, X_{10}\}$

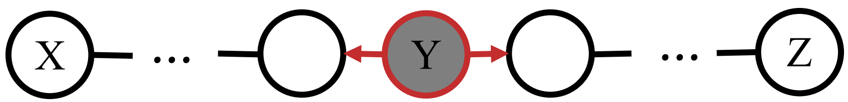


D-Separation

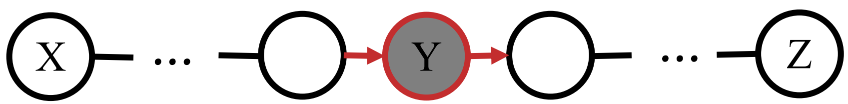
Definition #1:
Variables X and Z are **d-separated** given a **set** of evidence variables E (variables that are observed) iff every path from X to Z is "blocked".

A path is "blocked" whenever:

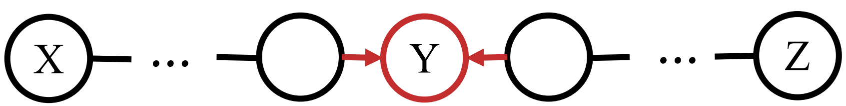
1. ∃Y on path s.t. Y ∈ E and Y is a "common parent"



2. ∃Y on path s.t. Y ∈ E and Y is in a "cascade"



3. ∃Y on path s.t. {Y, descendants(Y)} ∉ E and Y is in a "v-structure"



If variables X and Z are **d-separated** given a **set** of variables E
Then X and Z are **conditionally independent** given the **set** E

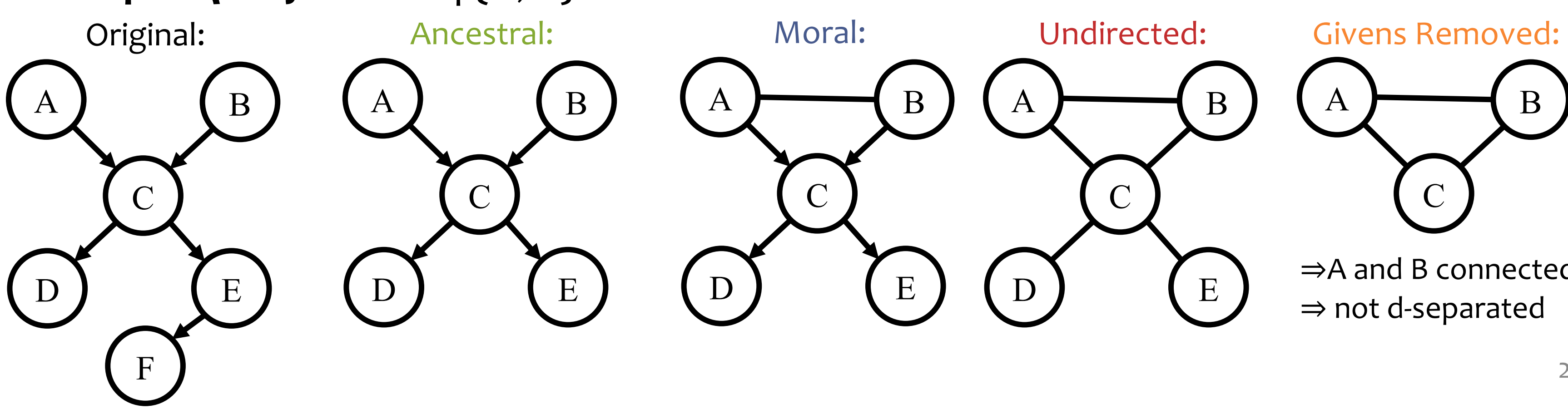
D-Separation

If variables X and Z are **d-separated** given a **set** of variables E
Then X and Z are **conditionally independent** given the **set** E

Definition #2:
Variables X and Z are **d-separated** given a **set** of evidence variables E iff there does **not** exist a path between X and Z in the **undirected ancestral moral** graph **with E removed**.

1. **Ancestral graph**: keep only X, Z, E and their ancestors
2. **Moral graph**: add undirected edge between all pairs of each node's parents
3. **Undirected graph**: convert all directed edges to undirected
4. Givens Removed: delete any nodes in E

Example Query: $A \perp\!\!\!\perp B \mid \{D, E\}$



SUPERVISED LEARNING FOR BAYES NETS

Recipe for Closed-form MLE

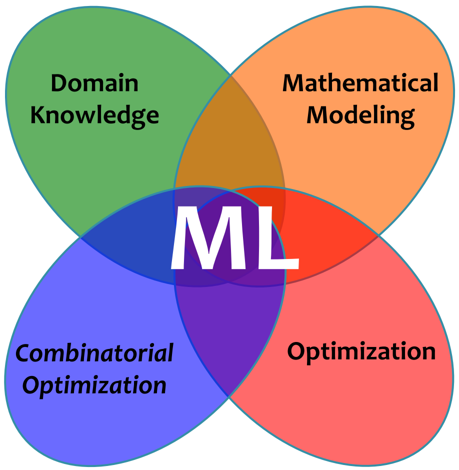
1. Assume data was generated i.i.d. from some model (i.e. write the generative story)
   $x^{(i)} \sim p(x|\boldsymbol{\theta})$
2. Write log-likelihood
   $\ell(\boldsymbol{\theta}) = \log p(x^{(1)}|\boldsymbol{\theta}) + \ldots + \log p(x^{(N)}|\boldsymbol{\theta})$
3. Compute partial derivatives (i.e. gradient)
   $\partial\ell(\boldsymbol{\theta})/\partial\theta_1 = \ldots$
   $\partial\ell(\boldsymbol{\theta})/\partial\theta_2 = \ldots$
   $\ldots$
   $\partial\ell(\boldsymbol{\theta})/\partial\theta_M = \ldots$
4. Set derivatives to zero and solve for $\boldsymbol{\theta}$
   $\partial\ell(\boldsymbol{\theta})/\partial\theta_m = 0$ for all $m \in \{1, \ldots, M\}$
   $\boldsymbol{\theta}^{MLE}$ = solution to system of M equations and M variables
5. Compute the second derivative and check that $\ell(\boldsymbol{\theta})$ is concave down at $\boldsymbol{\theta}^{MLE}$

Machine Learning

The **data** inspires the structures we want to predict

Our **model** defines a score for each structure

It also tells us what to optimize



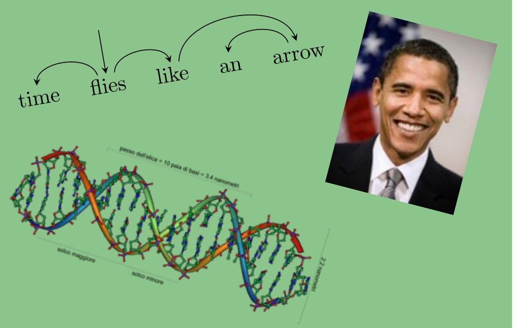
Learning tunes the parameters of the model

Inference finds {best structure, marginals, partition function} for a new observation

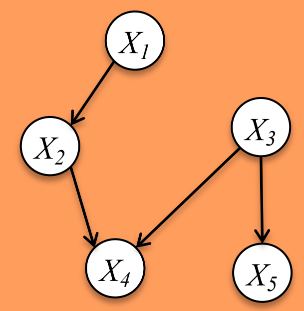
(**Inference** is usually called as a subroutine in learning)

Machine Learning

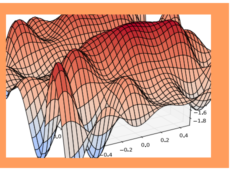
Data



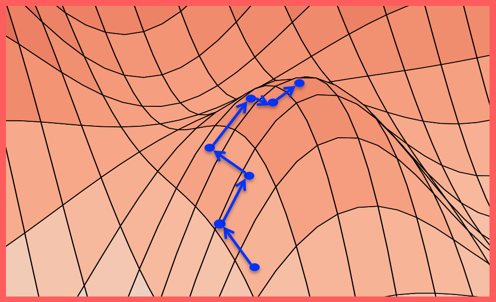
Model



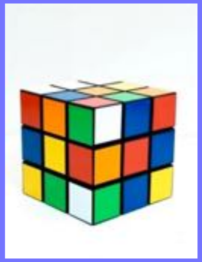
Objective



Learning

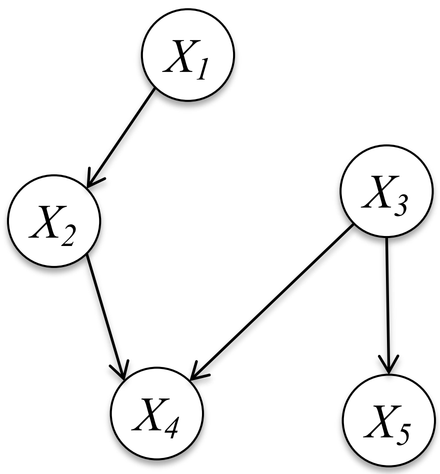


Inference



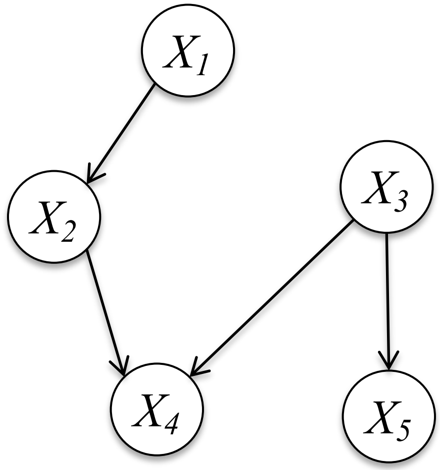
(**Inference** is usually called as a subroutine in learning)

Learning Fully Observed BNs



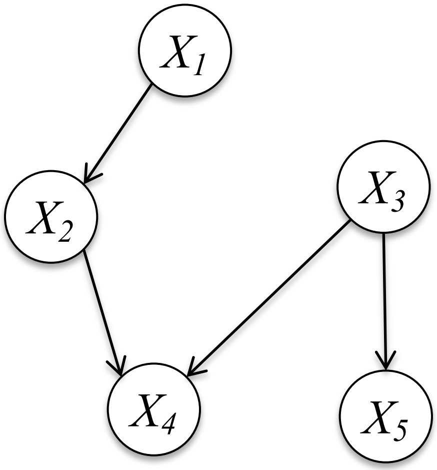
$$p(X_1, X_2, X_3, X_4, X_5) =$$
$$p(X_5|X_3)p(X_4|X_2, X_3)$$
$$p(X_3)p(X_2|X_1)p(X_1)$$

Learning Fully Observed BNs



$$
\begin{aligned}
p(X_1, X_2, X_3, X_4, X_5) = \\
p(X_5|X_3)p(X_4|X_2, X_3) \\
p(X_3)p(X_2|X_1)p(X_1)
\end{aligned}
$$

Learning Fully Observed BNs



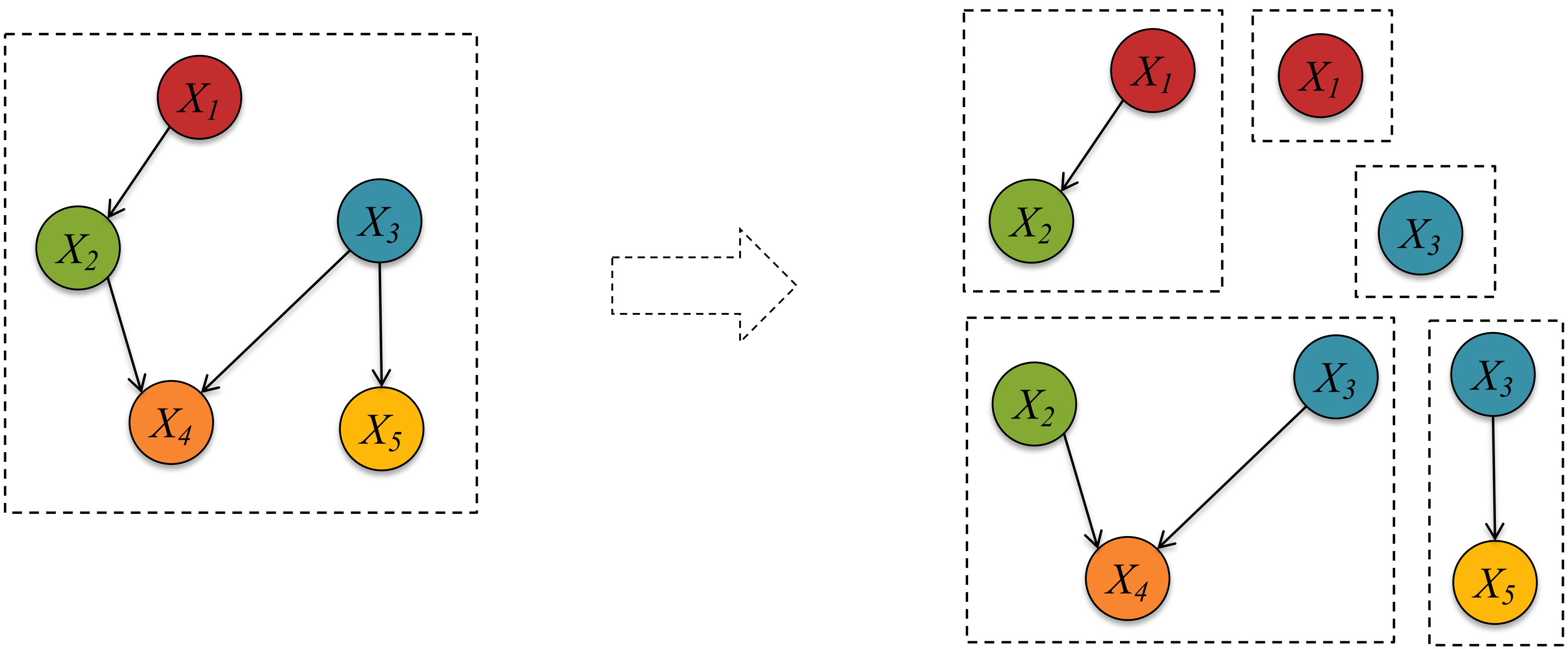
$$p(X_1, X_2, X_3, X_4, X_5) =$$
$$p(X_5|X_3)p(X_4|X_2, X_3)$$
$$p(X_3)p(X_2|X_1)p(X_1)$$

How do we learn these conditional and marginal distributions for a Bayes Net?

Learning Fully Observed BNs

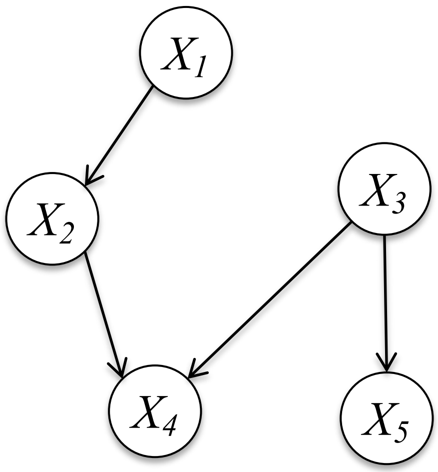
Learning this fully observed Bayesian Network is **equivalent** to learning five (small / simple) independent networks from the same data

$$p(X_1, X_2, X_3, X_4, X_5) = p(X_5|X_3)p(X_4|X_2, X_3) p(X_3)p(X_2|X_1)p(X_1)$$



Learning Fully Observed BNs

How do we **learn** these conditional and marginal distributions for a Bayes Net?



$$\boldsymbol{\theta}^* = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \log p(X_1, X_2, X_3, X_4, X_5)$$

$$= \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \log p(X_5|X_3, \theta_5) + \log p(X_4|X_2, X_3, \theta_4)$$

$$+ \log p(X_3|\theta_3) + \log p(X_2|X_1, \theta_2)$$

$$+ \log p(X_1|\theta_1)$$

$$\theta_1^* = \underset{\theta_1}{\operatorname{argmax}} \log p(X_1|\theta_1)$$

$$\theta_2^* = \underset{\theta_2}{\operatorname{argmax}} \log p(X_2|X_1, \theta_2)$$

$$\theta_3^* = \underset{\theta_3}{\operatorname{argmax}} \log p(X_3|\theta_3)$$

$$\theta_4^* = \underset{\theta_4}{\operatorname{argmax}} \log p(X_4|X_2, X_3, \theta_4)$$

$$\theta_5^* = \underset{\theta_5}{\operatorname{argmax}} \log p(X_5|X_3, \theta_5)$$

Example: Tornado Alarms

1. Imagine that you work at the 911 call center in Dallas
2. You receive six calls informing you that the Emergency Weather Sirens are going off
3. What do you conclude?

Example: Tornado Alarms

Hacking Attack Woke Up Dallas With Emergency Sirens, Officials Say

By ELI ROSENBERG and MAYA SALAM APRIL 8, 2017

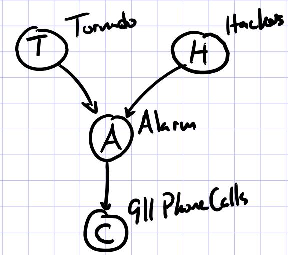
Warning sirens in Dallas, meant to alert the public to emergencies like severe weather, started sounding around 11:40 p.m. Friday, and were not shut off until 1:20 a.m. Rex C. Curry for The New York Times

1. Imagine that you work at the 911 call center in Dallas
2. You receive six calls informing you that the Emergency Weather Sirens are going off
3. What do you conclude?

Figure from https://www.nytimes.com/2017/04/08/us/dallas-emergency-sirens-hacking.html

Learning Fully Observed BNs

Ex: Tornado Alarms



$H \sim \text{Bernoulli}(\eta)$
$T \sim \text{Bernoulli}(\tau)$
$A \sim \text{Bernoulli}(\alpha_{H,T})$
$C \sim \text{Uniform}(\{1,\ldots,6\}) + A * \text{Uniform}(\{1,\ldots,6\})$

parameters

no parameters

integer

Dataset

i	T	H	A	C
1	0	0	0	2
2	0	0	0	6
3	0	0	0	4
⋮	1	0	0	3
	1	0	0	1
	1	0	1	10
	1	0	1	7
	0	1	0	2
	0	1	1	12
	0	1	0	5
⋮	1	1	1	10
12	1	0	0	2

MLE's in Closed Form

$$\ell(\eta,\tau,\alpha) = \log \prod_{i=1}^{12} p(t^{(i)}, h^{(i)}, a^{(i)}, c^{(i)} \mid \eta, \tau, \alpha)$$

$$= \sum_{i=1}^{12} \log p(t^{(i)} \mid \tau) + \log p(h^{(i)} \mid \eta) + \log p(a^{(i)} \mid t^{(i)}, h^{(i)}, \alpha) + \log p(c^{(i)} \mid a^{(i)})$$

$$\hat{\eta}, \hat{\tau}, \hat{\alpha} = \underset{}{\text{argmax}}\ \ell(\eta,\tau,\alpha)$$

$$\hat{\eta} = \underset{\eta}{\text{argmax}} \sum_{i=1}^{12} \log p(h^{(i)} \mid \eta) = \#(T=1)/N$$

$$\hat{\tau} = \underset{\tau}{\text{argmax}} \sum_{i=1}^{12} \log p(t^{(i)} \mid \tau) = \#(H=1)/N$$

$$\hat{\alpha} = \underset{\alpha}{\text{argmax}} \sum_{i=1}^{12} \log p(a^{(i)} \mid t^{(i)}, h^{(i)}, \alpha)$$

$$\hat{\alpha}_{t,h} = \frac{\#(A=1, T=t, H=h)}{\#(T=t, H=h)}$$

What are the MLEs?

$\hat{\eta} = 1/3$

$\hat{\tau} = 1/2$

$\hat{\alpha} =$

	H=0	H=1
T=0	0	1/3
T=1	2/3	1

INFERENCE FOR BAYESIAN NETWORKS

A Few Problems for Bayes Nets

Suppose we already have the parameters of a Bayesian Network…

1. How do we compute the probability of a specific assignment to the variables?
   P(T=t, H=h, A=a, C=c)

2. How do we draw a sample from the joint distribution?
   t,h,a,c ~ P(T, H, A, C)

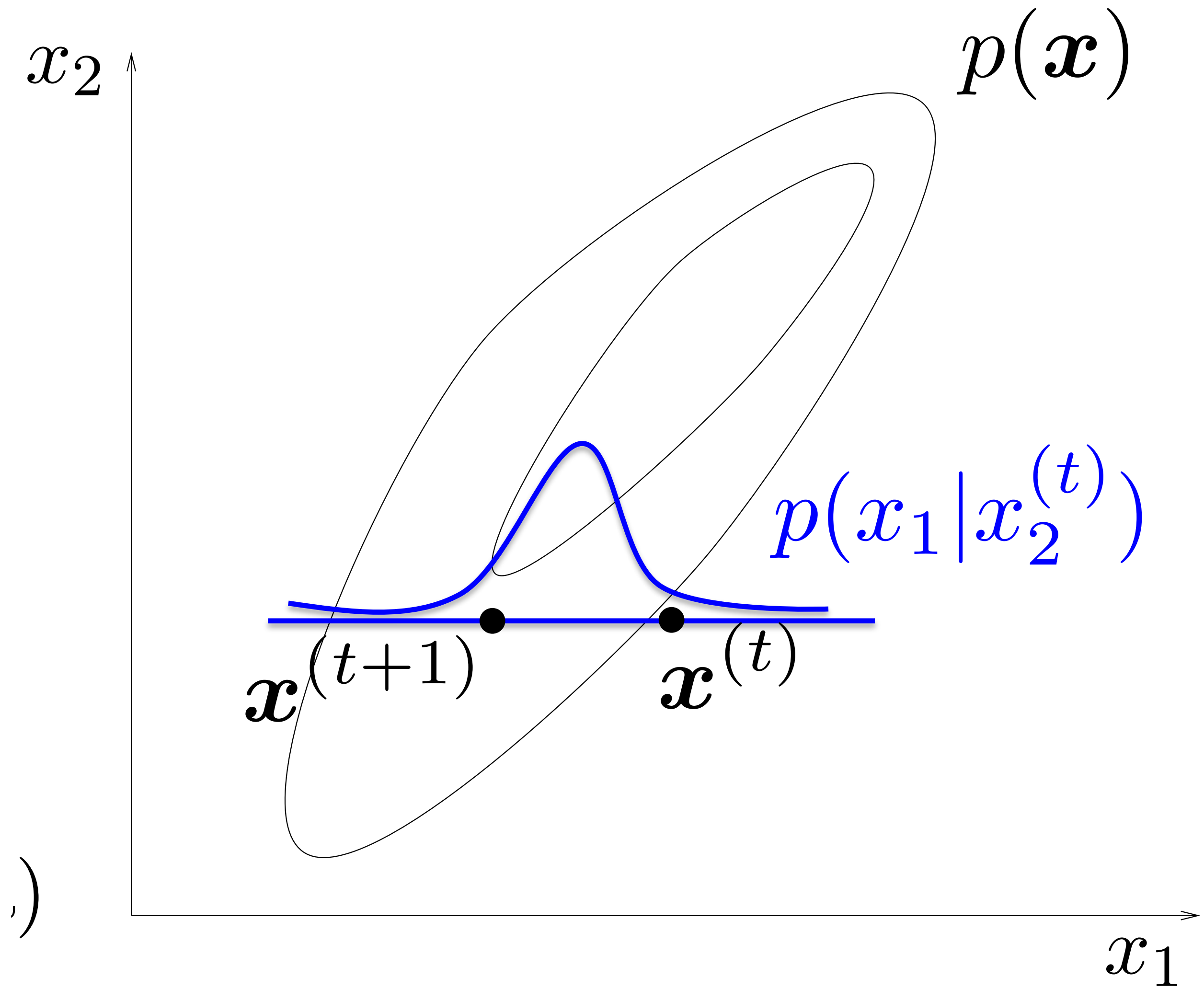
3. How do we compute marginal probabilities?
   P(A) = …

4. How do we draw samples from a conditional distribution?
   t,h,a ~ P(T, H, A | C = c)

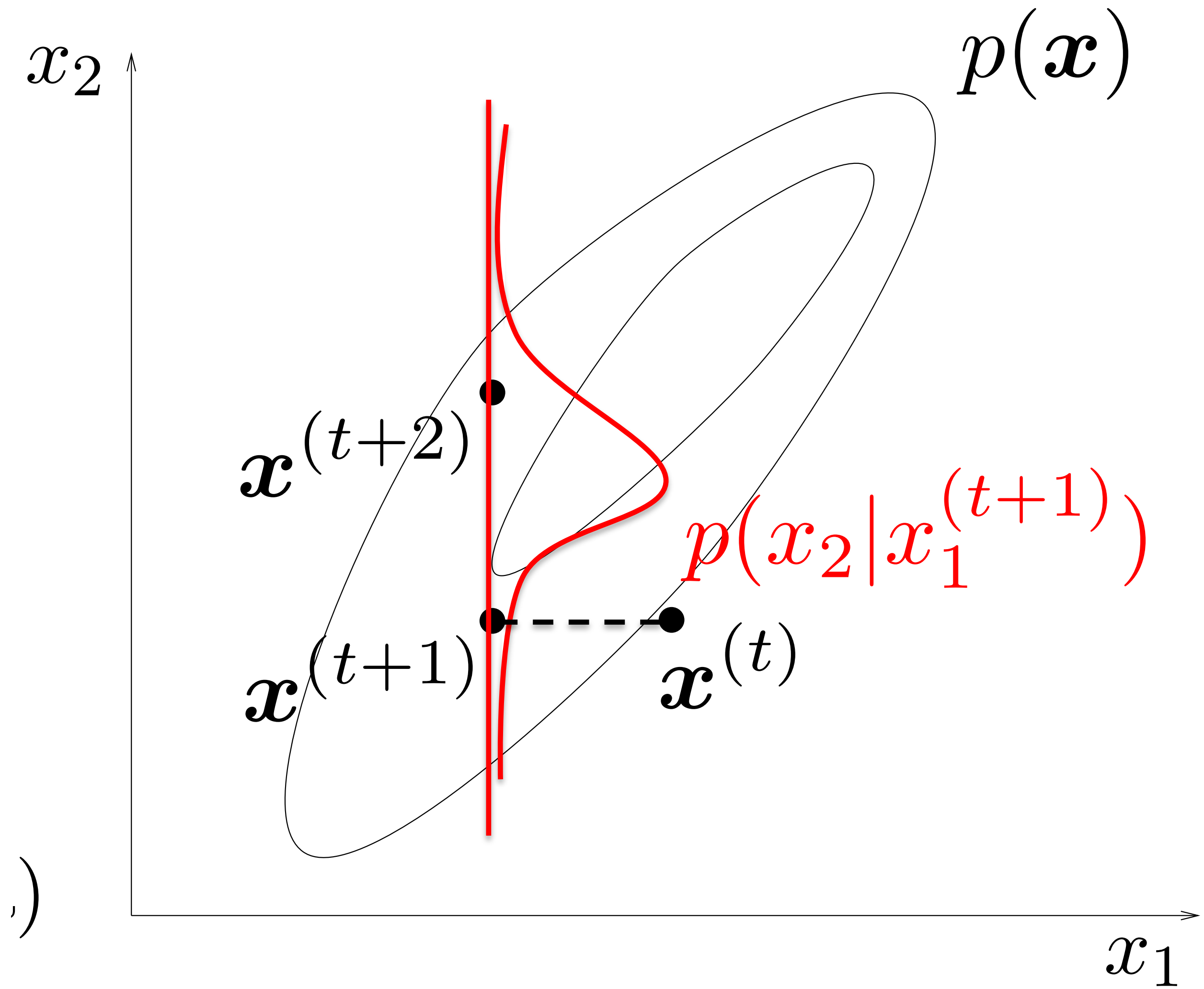
5. How do we compute conditional marginal probabilities?
   P(H | C = c) = …

Can we use samples?

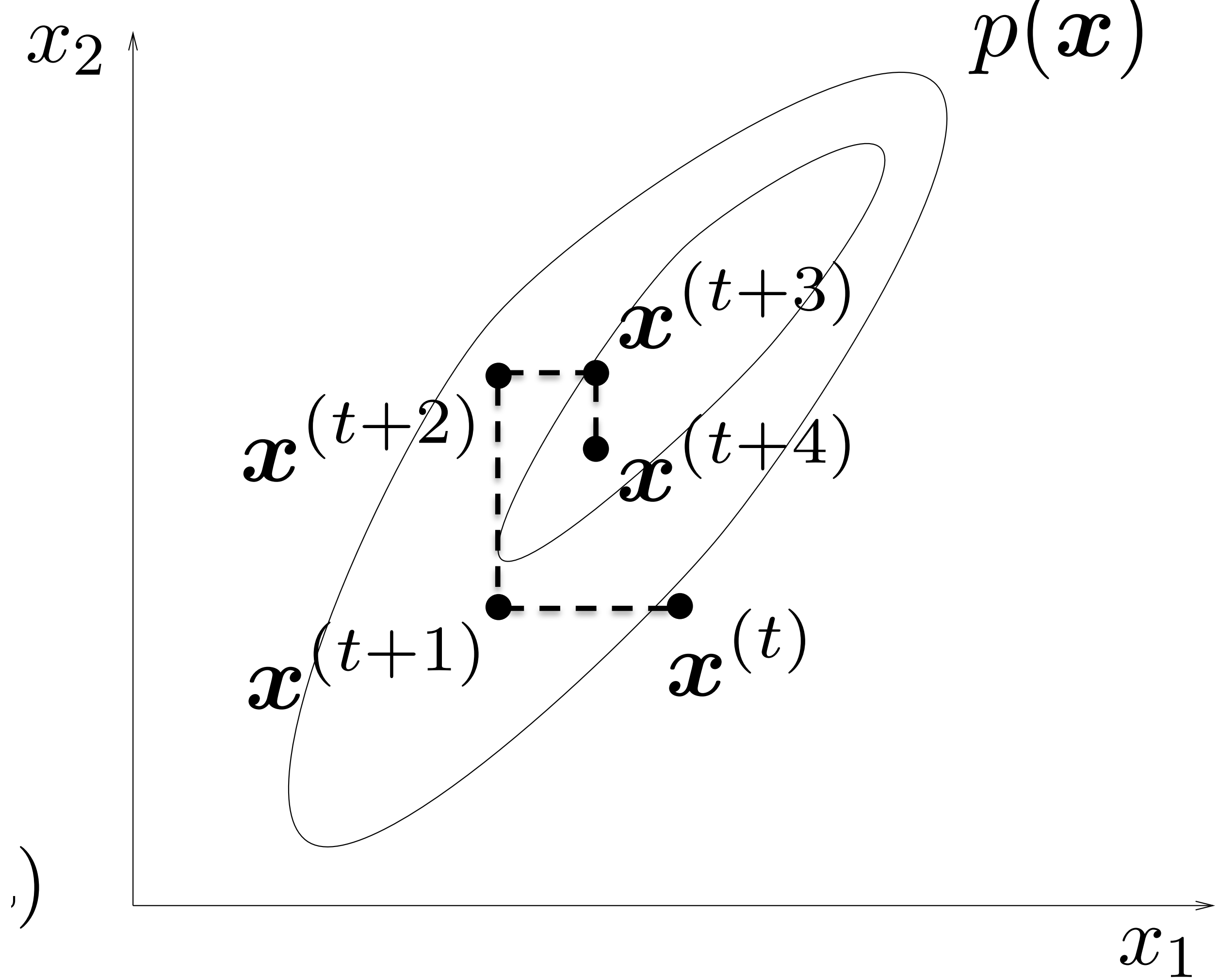
Gibbs Sampling



Gibbs Sampling



Gibbs Sampling



Gibbs Sampling

Question:
How do we draw samples from a conditional distribution?
$y_1, y_2, \ldots, y_J \sim p(y_1, y_2, \ldots, y_J \mid x_1, x_2, \ldots, x_J)$

(Approximate) Solution:

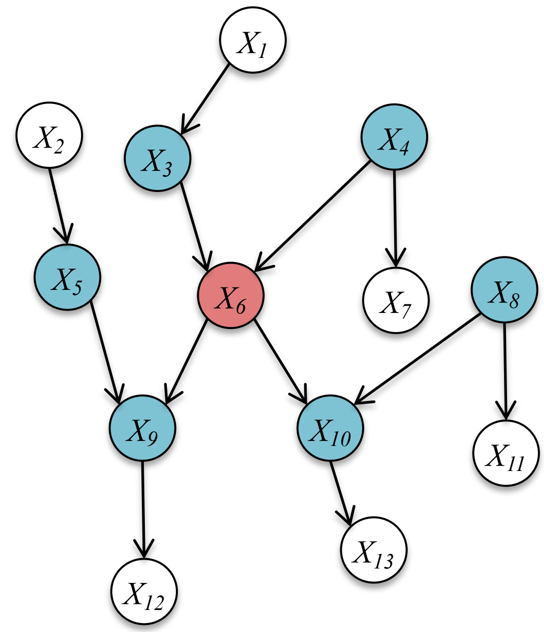
- Initialize $y_1^{(0)}, y_2^{(0)}, \ldots, y_J^{(0)}$ to arbitrary values
- For t = 1, 2, … :
  - $y_1^{(t+1)} \sim p(y_1 \mid y_2^{(t)}, \ldots, y_J^{(t)}, x_1, x_2, \ldots, x_J)$
  - $y_2^{(t+1)} \sim p(y_2 \mid y_1^{(t+1)}, y_3^{(t)}, \ldots, y_J^{(t)}, x_1, x_2, \ldots, x_J)$
  - $y_3^{(t+1)} \sim p(y_3 \mid y_1^{(t+1)}, y_2^{(t+1)}, y_4^{(t)}, \ldots, y_J^{(t)}, x_1, x_2, \ldots, x_J)$
  - …
  - $y_J^{(t+1)} \sim p(y_J \mid y_1^{(t+1)}, y_2^{(t+1)}, \ldots, y_{J-1}^{(t+1)}, x_1, x_2, \ldots, x_J)$

Properties:

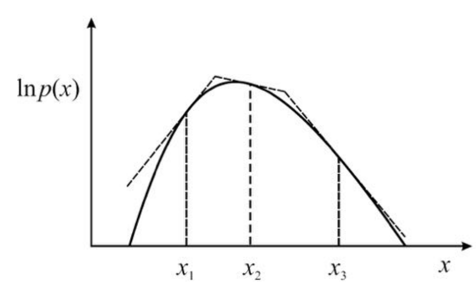
- This will eventually yield samples from $p(y_1, y_2, \ldots, y_J \mid x_1, x_2, \ldots, x_J)$
- But it might take a long time -- just like other Markov Chain Monte Carlo methods

Gibbs Sampling

Full conditionals only need to condition on the **Markov Blanket**



- Must be "easy" to sample from conditionals
- Many conditionals are log-concave and are amenable to adaptive rejection sampling



Learning Objectives

Bayesian Networks

You should be able to…

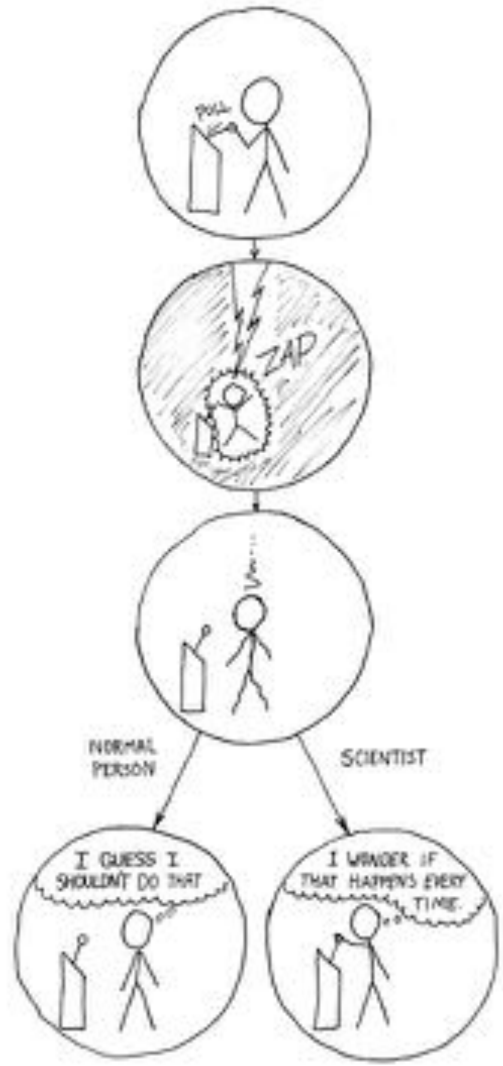
1. Identify the conditional independence assumptions given by a generative story or a specification of a joint distribution
2. Draw a Bayesian network given a set of conditional independence assumptions
3. Define the joint distribution specified by a Bayesian network
4. User domain knowledge to construct a (simple) Bayesian network for a real-world modeling problem
5. Depict familiar models as Bayesian networks
6. Use d-separation to prove the existence of conditional indenpendencies in a Bayesian network
7. Employ a Markov blanket to identify conditional independence assumptions of a graphical model
8. Develop a supervised learning algorithm for a Bayesian network
9. Use samples from a joint distribution to compute marginal probabilities
10. Sample from the joint distribution specified by a generative story
11. Implement a Gibbs sampler for a Bayesian network

Reinforcement Learning

Learning Paradigms

- Supervised Learning
  - Training data is (input, output)
  - Variants: active learning and online learning
- Unsupervised Learning
  - Training data is (input)
- Reinforcement Learning
  - Training data is (input, action, reward)

Reinforcement Learning (RL)



Source: https://www.xkcd.com/242/

Source: https://techobserver.net/2019/06/argo-ai-self-driving-car-research-center/

Source: https://www.wired.com/2012/02/high-speed-trading/

RL: Examples

Source: https://www.cnet.com/news/boston-dynamics-robot-dog-spot-finally-goes-on-sale-for-74500/

Source: https://twitter.com/alphagomovie

AlphaGo

Source: https://www.youtube.com/watch?v=WXuK6gekU1Y&ab_channel=DeepMind

RL: Challenges

- The algorithm has to gather its own training data
- The outcome of taking some action is often stochastic or unknown until after the fact
- Decisions can have a delayed effect on future outcomes (exploration-exploitation tradeoff)

RL: Outline

- Problem formulation
  - Time discounted cumulative reward
  - Markov decision processes (MDPs)
- Algorithms:
  - Value iteration and policy iteration (dynamic programming)
  - (Deep) Q-learning (temporal difference learning)

RL: Components

- State space, $\mathcal{S}$
- Action space, $\mathcal{A}$
- Reward function, $R : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$
- Transition probabilities, $p(s' \mid s, a)$
  - Deterministic transitions:

$$p(s' \mid s, a) = \begin{cases} 1 \text{ if } \delta(s, a) = s' \\ 0 \text{ otherwise} \end{cases}$$

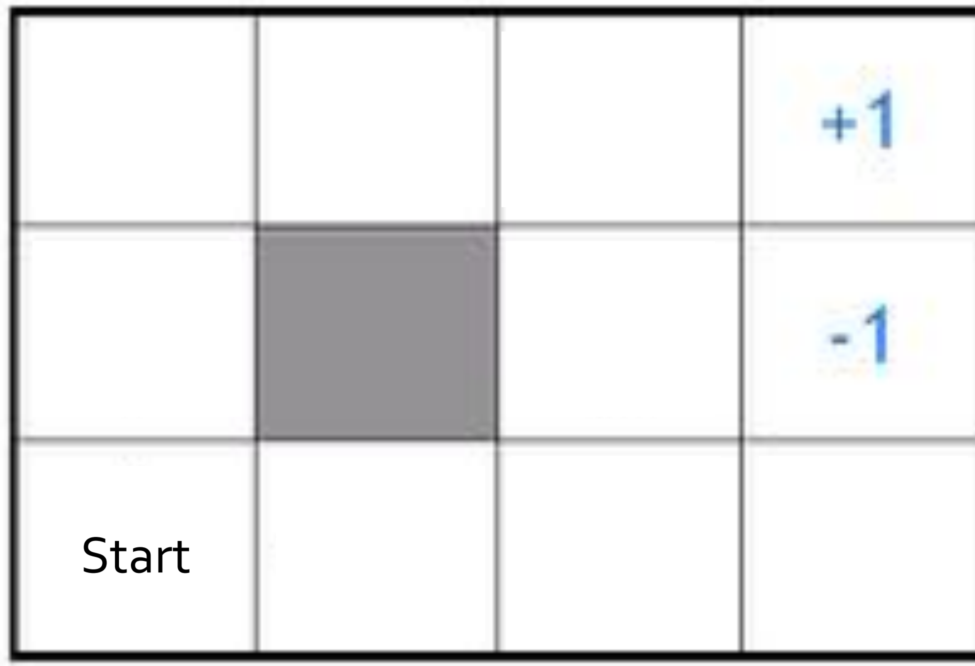
where $\delta(s, a)$ is a transition function

- Policy, $\pi : \mathcal{S} \rightarrow \mathcal{A}$
- Value function, $V^{\pi} : \mathcal{S} \rightarrow \mathbb{R}$
  - Measures the expected total payoff of starting in some state s and *executing* policy π

RL: Toy Example

$\mathcal{S}$ = all empty squares in the grid

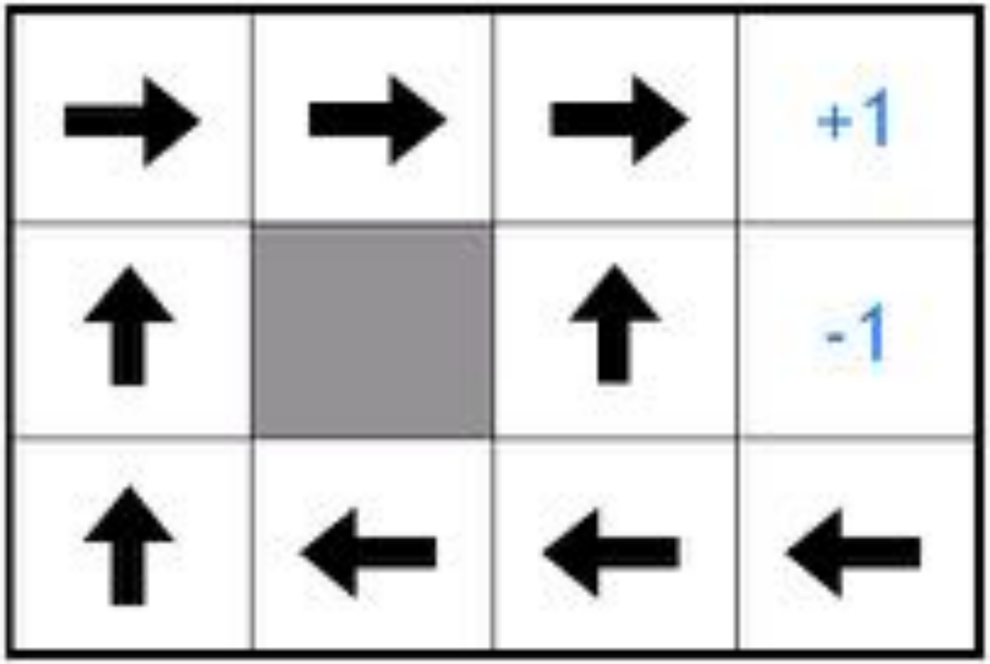
$\mathcal{A}$ = { up, down, left, right }



Example courtesy of Eric Xing, CMU MLD

RL: Poll Q2

Is this policy optimal?



Example courtesy of Eric Xing, CMU MLD

Question 2

A

B

C

Justify your answer to the previous question

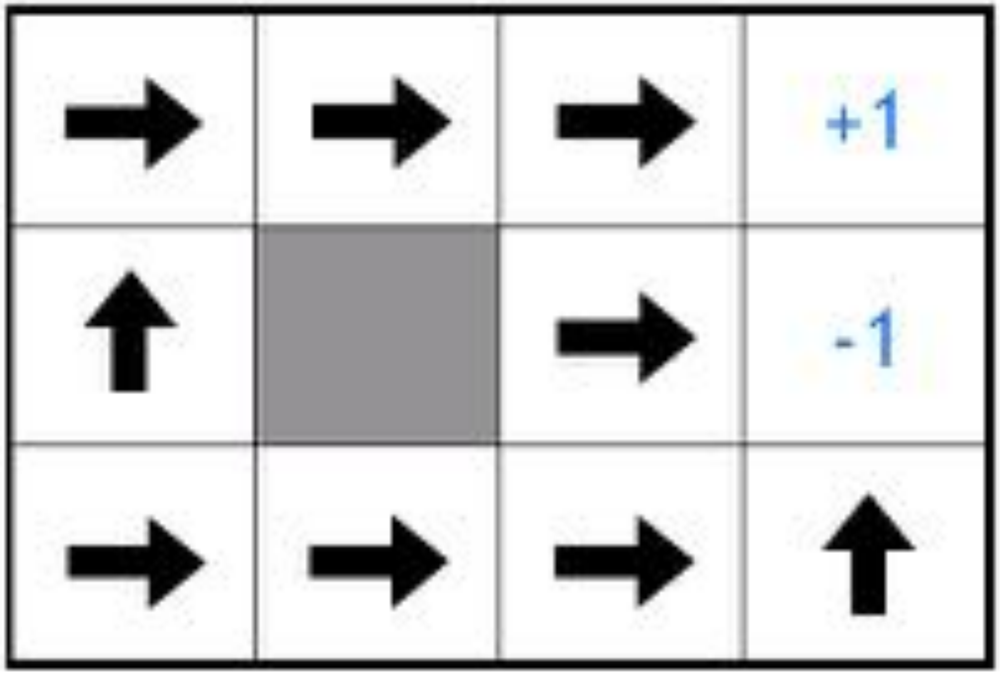
Join by Web

1. Go to **PollEv.com**
2. Enter **10301601POLLS**
3. Respond to activity

Instructions not active. **Log in** to activate

RL: Toy Example

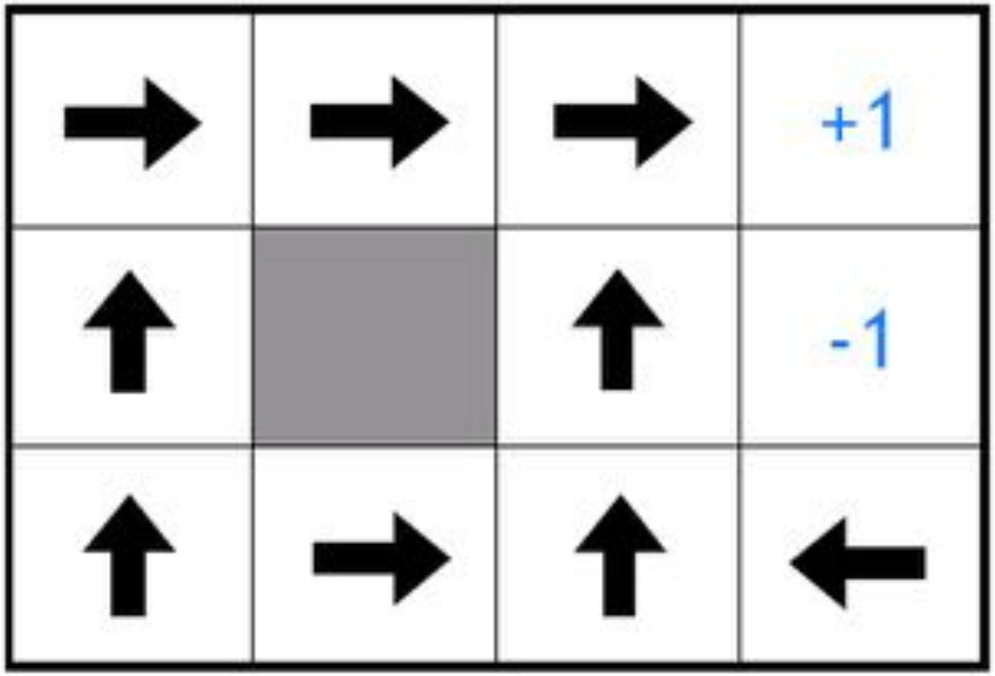
Optimal policy given reward of -2 for each step



Example courtesy of Eric Xing, CMU MLD

RL: Toy Example

Optimal policy given reward of -0.1 for each step



Example courtesy of Eric Xing, CMU MLD

RL: Objective Function

- Find a policy $\pi^* = \text{argmax}_\pi V^\pi(s) \; \forall \, s \in \mathcal{S}$

- $V^\pi(s) = \mathbb{E}[$discounted total reward of starting in state s and *executing* policy π forever$]$

$$= \mathbb{E}_{p(s' \mid s,\, a)}[R\big(s_0 = s, \pi(s_0)\big)$$

$$+ \gamma R\big(s_1, \pi(s_1)\big) + \gamma^2 R\big(s_2, \pi(s_2)\big) + \cdots]$$

$$= \sum_{t=0}^{\infty} \gamma^t \mathbb{E}_{p(s' \mid s,\, a)}\left[R\big(s_t, \pi(s_t)\big)\right]$$

where $0 < \gamma < 1$ is some discount factor for future rewards

Markov Decision Processes (MDP)

- In RL, the model for our data is an MDP:

1. Start in some initial state s_0
2. For time step t:
   1. Agent observes state s_t
   2. Agent takes action $a_t = \pi(s_t)$
   3. Agent receives reward $r_t = R(s_t, a_t)$
   4. Agent transitions to state $s_{t+1} \sim p(s' \mid s_t, a_t)$
3. Total reward is $\sum_{t=0}^{\infty} \gamma^t r_t$

- Makes the same Markov assumption we used for HMMs! The next state only depends on the current state and action.