APPROXIMATE INFERENCE IN BAYESIAN NETWORKS

Today

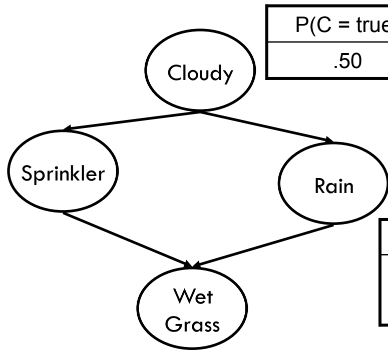
- Reading
  - AIMA 14.4 – 14.5
  - (AIMA Chapter 15.1-15.2, 15.5)
- Goals
  - Direct Sampling
  - Rejection Sampling
  - Likelihood Weighting
  - (Introduce Dynamic Bayesian Networks)

Approximate Inference

- Analogous to uninformed/informed search algorithms that use an incremental formulation
  - Direct sampling
  - Rejection sampling
  - Likelihood weighting
- Analogous to local search algorithms that use a complete-state formulation and make local modifications
  - Gibbs sampling (special case of MCMC methods)

Lecture proceeds on whiteboard!

Wet Grass Example



P(C = true)
.50

C	P(S=true|C)
T	.10
F	.50

C	P(R=true|C)
T	.80
F	.20

S	R	P(W=true|S,R)
T	T	.99
T	F	.90
F	T	.90
F	F	.01

Gibbs Sampling

- Analogous to a local search algorithm where we make local modifications to our current state
  - Initial state = random assignment of non-evidence variables
  - States = complete assignment of values to variables
  - Transition = sample a new value for each variable in turn

Draw state space for WetGrass example on board

Gibbs Sampling

- Analogous to a local search algorithm where we make local modifications to our current state
  - Initial state = random assignment of non-evidence variables
  - States = complete assignment of values to variables
  - Transition = sample a new value for each variable in turn
- Each step is recorded as a sample
- *In the limit, the probability of being in a state is proportional to that state's posterior probability*

Gibbs Sampling

- Gibbs sampling is an instance of a more general class of algorithms known as Markov Chain Monte Carlo (MCMC) algorithms
  - Note the use of the phrase "Markov chain" which we saw an example of earlier
- Other methods you might hear mentioned
  - Metropolis-Hastings (a generalization of Gibbs sampling)
  - Variational method
  - Belief propagation

Dynamic Bayesian Networks

Modeling uncertainty over time

- Sometimes, we want to model a *dynamic* process: the value of the random variables change over time
  - Price of a stock
  - Patient stats, e.g. blood pressure, heart rate, blood sugar levels
  - Traffic on California highways
  - Pollution, humidity, temperature, rain fall, storms
  - Sensor tracking and detection

Modeling uncertainty over time

- Tracy got a new job working at the Coop. She works the late shift and doesn't get off until 2am. When she works the late shift, I often observe her eyes are red the next day. But sometimes she stays up late doing homework, and her eyes are red anyways.
- What are questions we might be interested in asking?
- How can we model this domain as a Bayesian network?

Modeling uncertainty over time

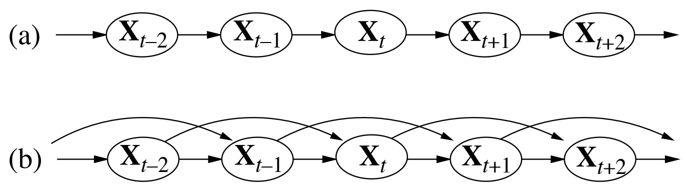
- Suppose we also know that if Tracy works the late shift one night she is less likely to work the late shift the next night.
- How does this change the model?

States and Evidence

- Model a dynamic process as a series of time slices
- Each time slice contains a set of random variables
  - We observe the value of some random variables called the evidence. Often denoted as E_t
  - We don't observe the value of some random variables called the state. Often denoted as X_t

Transition Model

- We're often interested in reasoning about the state variables X_t given the history $X_{0:t-1}$
- Markov Assumption: the state variable X_t depends on a *bounded* subset of $X_{0:t-1}$
  - First order Markov Process: $P(X_t | X_{0:t-1}) = P(X_t | X_{t-1})$
  - Second order Markov Process: $P(X_t | X_{0:t-1}) = P(X_t | X_{t-1}, X_{t-2})$



Transition Model

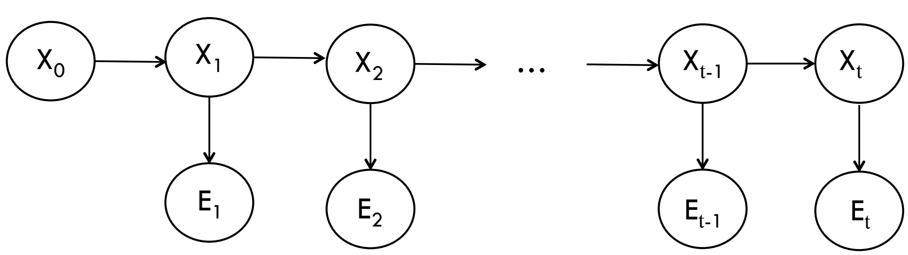
- We're often interested in reasoning about the state variables X_t given the history $X_{0:t-1}$
- Stationarity Assumption: the conditional distribution $P(X_t | X_{t-1})$ is the same for all t
  - Need to specify only one conditional distribution

Sensor (emission) model

- The state variables are responsible for generating (emitting) the evidence variables
- Sensor Markov Assumption: the evidence at time t is independent of every other random variable given the state at time t
  - As a result, your state should encompass all relevant information for specifying the evidence

Hidden Markov Model

- Hidden Markov Models involve three things:
  - Transition model: $P(X_t | X_{t-1})$
  - Emission (evidence) model: $P(E_t | X_t)$
  - Prior probability: $P(X_0)$



Inference Tasks

- Filtering: $P(X_t | e_{1:t})$
  - Decision making in the here and now
- Prediction: $P(X_{t+k} | e_{1:t})$
  - Trying to plan the future
- Smoothing: $P(X_k | e_{1:t})$ for $0 \leq k < t$
  - Gives a better (smoother) estimate than filtering by taking into account future evidence
- Most Likely Explanation (MLE): $\underset{x_{1:t}}{\text{argmax}}\ P(x_{1:t} | e_{1:t})$
  - e.g., speech recognition, sketch recognition