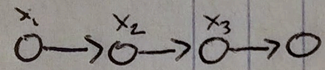
Directed Acyclic Graphical (DAG) Models

Markov Chains



$p(x_1,\dots,x_d) = p(x_1)\prod_{t=2}^{d} p(x_t \mid x_{t-1})$

$x_i \perp x_t \mid x_k \iff k$ is between i, t

Marginals: CK equations if $p()$ are gaussian or discrete

- Monte Carlo methods in general

Sampling: Ancestral Sampling

Decoding: Viterbi Algorithm

Message Passing: Forward-Backward Algorithm

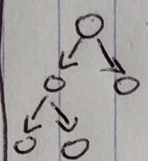
Estimators: - for discrete x_t, MLE is counts

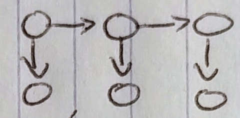
e.g. $p(x_t \mid x_{t-1}) \approx \hat{\theta}_t = \dfrac{\text{numbers of transitions from } x_{t-1} \text{ to } x_t}{\text{number of times we saw } x_{t-1}}$

- Monte Carlo Methods in general

Belief Networks / Bayesian Networks

Treewidth = 1





Hidden Markov Model

$p(x_1,\dots,x_d) = \prod_t p(x_t \mid pa(x_t))$

$x_i \perp x_t \mid x_k$ if x_i and x_t are d-separated

CK equations if $p()$ are gaussian or discrete

- Monte Carlo methods in general
- Ancestral Sampling
- Max-product Algorithm
- Belief Propagation
- for discrete x_t, MLE is counts
- with data, can do supervised learning

e.g. find $w = \arg\max_w p(x_t \mid pa(x_t), w)$

treat $p(x_t \mid pa(x_t))$ as $p(y \mid x)$

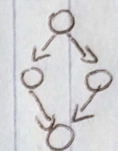
if $p(y \mid x) \sim N(w^T x, \frac{1}{\lambda})$

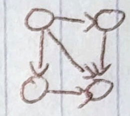
it's called a Gaussian Belief Network

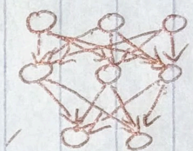
if $p(y \mid x) \sim \mathrm{Sigm}(w^T x)$

it's called Sigmoid Belief Network

Treewidth > 1







Deep Belief Net

$p(x_1,\dots,x_d) = \prod_t p(x_t \mid pa(x_t))$

$x_i \perp x_t \mid x_k$ if x_i and x_t are d-separated

- Monte Carlo methods
- Ancestral Sampling
- Monte Carlo methods
- Belief Propagation (approximate)
- for discrete x_t, MLE is counts
- can do supervised learning (Gaussian Belief Networks, Sigmoid Belief Networks)
- Monte Carlo Methods