SI 151
Graphical Models and Bayes Nets

Xin Deng

April 5, 2020

Outline

- Introduction to Graphical models and Bayes Nets
- Simple inference and Simple learning

Graphical models

- Two types of graphical models:
  - Directed graphs (aka Bayes Networks)
  - Undirected graphs (aka Markov Random Fields)
- Graphical models allow combining:
  - Prior knowledge in form of dependencies/independencies
  - Prior knowledge in form of priors over parameters
  - Observed training data

Bayes Nets

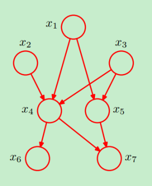
- A Bayes network includes $\begin{cases} \mathrm{DAG} \\ \mathrm{CPD's} \end{cases}$
  - Each node denotes a random variable
  - Edges denote dependencies
  - For each node X_i its CPD defines $P(X_i | Pa(X_i))$
  - The joint distribution over all variables is defined to be

$$P(X_1, \cdots, X_n) = \prod_i P(X_i | Pa(X - i))$$

- Each node is conditionally independent of its non-descendents, given only its immediate parents.

Bayes Nets

$\boxed{\text{Example:}}$ A directed acyclic graph describing the joint distribution over variables $x_1, \cdots, x_7$.



The corresponding decomposition of the joint distribution is given by

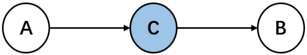
$$p(x_1, \cdots, x_7) = p(x_1)p(x_2)p(x_3)p(x_4|x_1, x_2, x_3)p(x_5|x_1, x_5)p(x_6|x_4)p(x_7|x_4, x_5)$$

Outline

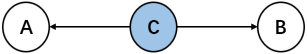
- Introduction to Graphical models and Bayes Nets
- Simple inference and Simple learning

Inference

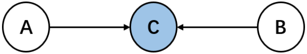
- Head-to-Tail



  - None of the variables are observed: A is not cond indep of B
  - Given C: A is cond indep of B
- Tail-to-Tail



  - None of the variables are observed: A is not cond indep of B
  - Given C: A is cond indep of B
- Head-to-Head



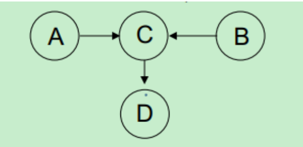
  - None of the variables are observed: A is cond indep of B
  - Given C: A is not cond indep of B

D-separation

Consider a general directed graph in which A, B, and C are arbitrary nonintersecting sets of nodes.

Any possible paths from any node in A to any node in B is said to be blocked if it includes a node such that either

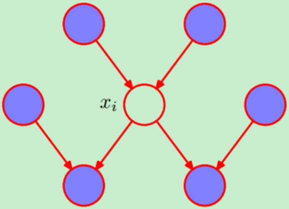
- the arrows on the path meet either head-to-tail or tail-to-tail at the node, and the node is in the set C, or
- the arrows meet head-to-head at the node, and neither the node, nor any of its descendants, is in the set C.



Inference

$\boxed{\text{Markov Blanket:}}$

The Markov blanket of a node $\mathbf{x}_i$ comprises the set of parents, children and co-parents of the node. It has the property that the conditional distribution of $\mathbf{x}_i$, conditioned on all the remaining variables in the graph, is dependent only on the variables in the Markov blanket.



Inference

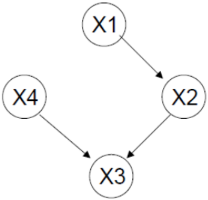
- In general, intractable (NP-hard)
- For certain cases, tractable
  - Assigning probability to fully observed set of variables
  - Or if just one variable unobserved
- Sometimes use Monte Carlo methods
  - Generate many samples according to the Bayes Net distribution, then count up the result

Inference

Example:

Conditional independence and marginal independence:

- $X_1 \perp\!\!\!\perp X_4 \mid X_3$?
- $X_1 \perp\!\!\!\perp X_4 \mid \{X_2, X_3\}$?
- $X_1 \perp\!\!\!\perp X_4 \mid \varnothing$?



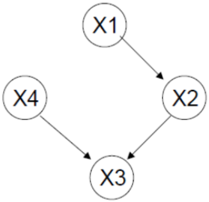
Inference

$\boxed{\text{Example:}}$

Conditional prob and marginal prob:

- $P(x_1 \mid x_2, x_3, x_4) = ?$

- $P(x_3) = ?$



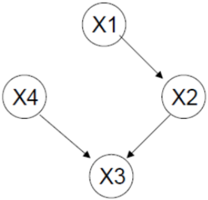
Inference

$\boxed{\text{Example:}}$

- $$P(x_1 \mid x_2, x_3, x_4) = \frac{P(x_1, x_2, x_3, x_4)}{P(x_2, x_3, x_4)}$$
  $$= \frac{P(x_1)P(x_4)P(x_2 \mid x_1)P(x_3 \mid x_2, x_4)}{P(x_3 \mid x_2, x_4)}$$
- $P(x_3) = \sum_{x_1, x_2, x_4} P(x_1, x_2, x_3, x_4)$

Generating a sample from joint distribution

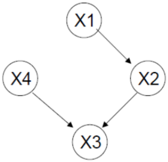
- For marginal prob:
  Take X_1 as an example and $P(X_1) = \theta_{X_1=1}$
  - Draw a value of r_1 uniformly from $[0, 1]$
  - $X_1 = \begin{cases} 1, & \text{if } r_1 < \theta_{X_1=1} \\ 0, & \text{else} \end{cases}$



Generating a sample from joint distribution

- For cond prob:
  - Draw a value r_1 for X_1
  - Draw value for X_4, for $X_2 \mid X_1$, for $X_3 \mid X_2, X_4$
- Generate a sample by $P(X_1, X_2, X_3, X_4) = P(x_1)P(x_4)P(x_2 \mid x_1)P(x_3 \mid x_2, x_4)$

Learning



- Learn the parameter

$$\theta_{\{X_3=1|X_2=1,X_4=1\}} = P(X_3 = 1 \mid X_2 = 1, X_4 = 1)$$

- Maximum likelihood estimate

$$\theta = \arg\min_{\theta} \log P(\mathcal{D} \mid \theta)$$

Learning

$$
\begin{aligned}
\log P(\mathcal{D} \mid \theta) &= \log \prod_i P(x_{1i}, x_{2i}, x_{3i}, x_{4i} \mid \theta) \\
&= \log \prod_i P(x_{1i} \mid \theta) P(x_{4i} \mid \theta) P(x_{2i} \mid x_{1i}, \theta) P(x_{3i} \mid x_{2i}, x_{4i}, \theta) \\
&= \sum_i \log P(x_{1i} \mid \theta) + \log P(x_{4i} \mid \theta) + \log P(x_{2i} \mid x_{1i}, \theta) + \log P(x_{3i} \mid x_{2i}, x_{4i}, \theta)
\end{aligned}
$$

$$
\frac{\partial \log P(\mathcal{D} \mid \theta)}{\partial \theta_{\{X_3=1|X_2=1,X_4=1\}}} = \sum_i \frac{\partial \log P(x_{3i} \mid x_{2i}, x_{4i}, \theta)}{\partial \theta_{\{X_3=1|X_2=1,X_4=1\}}}
$$

$$
\implies \theta_{\{X_3=1|X_2=1,X_4=1\}} = \frac{\#(X_3=1,X_2=1,X_4=1)}{\#(X_2=1,X_4=1)}
$$

Thanks !