Graphical Models

Lecture 10, DD2431 Machine Learning

J. Sullivan, A. Maki

October 2013

Bayesian Networks

Synonyms:

- directed graphical models
- probability network
- belief network
- causal network
- knowledge map

Bayesian Networks

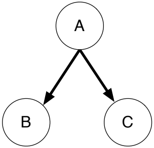
A **Bayesian Network** is a graphical representation of joint probability $P(\mathbf{X})$ where it is easy to see the dependencies which exist between the variables $X_1, \ldots, X_K$.

In general for a K-dimensional random variable $\mathbf{X}$

$$P(\mathbf{X}) \neq P(X_1)P(X_2) \cdots P(X_K)$$

Bayesian Networks

A **Bayesian Network** is a directed acyclic (no loops) graph.

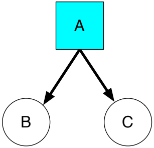


Directed graph is a set of *nodes* connected by *directed edges*.

For the **Bayesian Network**:

- Nodes represent the random variables.
- Directed edges indicate causal relationships between random variables.

Bayesian Networks

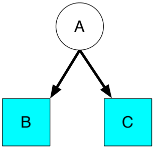


Knowing A provides *causal* support for values of B, C.

Example:

- A = presence of clouds
- B = will rain this afternoon
- C = current temperature

Bayesian Networks



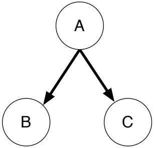
Knowing B, C provides *evidential* support for values of A.

Example:

- A = Type of disease
- B = Presence of fever
- C = Presence of nausea

Bayesian Networks

How does this graph represent the $P(A, B, C)$?



For each node X

- there is a conditional probability distribution,
- this conditional distribution is conditioned only on the node's parents

The joint probability distribution is the product of all these conditional distributions. Therefore for the above example

$$P(A, B, C) = P(A)\, P(B|A)\, P(C|A)$$

Bayesian Networks

Each node has a conditional probability distribution $P(X_k) = P(X_k|\,\text{parents}\,(X_k))$ that quantifies the effect of the parents on the node.

For a graph with K nodes corresponding to $X_1, \ldots, X_K$ the joint distribution $P(X_1, \ldots, X_K)$ is given by

$$P(\mathbf{X}) = P(X_1, \ldots, X_K) = \prod_{k=1}^{K} P(X_k|\,\text{parents}\,(X_k))$$

Example: Is anyone at home?

The first family member home in the evening has to bring in the firewood from the shed. To avoid wasting time checking the house or the shed, can I use other clues to decide if I am the first person home?

I know the following information:

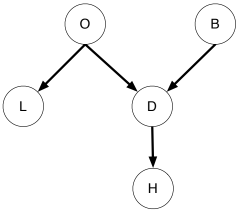
1. When nobody is home, the outside light is sometimes on.
2. When nobody is home, the dog is often left outside.
3. If the dog has stomach-troubles, it is also often left outside.
4. If the dog is outside, I will probably hear it barking (though it might not bark, or I might hear a different dog barking and think it's my dog).

Example: Is anyone at home?

The variables: (all binary)

1. *O - No one is home*
2. *L - The light is on*
3. *D - The dog is outside*
4. *B - The dog has stomach problems*
5. *H - I can hear the dog barking*

The network:



Structure

Probabilities: (learned or set by hand)

$P(O = 1) = .6;$

$P(B = 1) = .3;$

$P(H = 1|D = 1) = .3, \ P(H = 1|D = 0) = .8;$

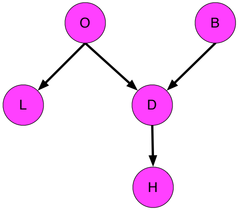
$P(L = 1|O = 1) = .3, \ P(L = 1|O = 0) = .6;$

$P(D = 1|O = 1, B = 1) = .05, P(D = 1|O = 1, B = 0) = .1,$
$P(D = 1|O = 0, B = 1) = .1, P(D = 1|O = 0, B = 0) = .2$

Inference in Bayes Net

- Given a network we can make predictions about certain variables given others
- Can compute joint probabilities
- Can also compute conditional probabilities:
  - Causal/predictive/top-down inference
  - Evidential/diagnostic/bottom-up inference

Computing Joint Probabilities



$P(O=1) = .6;$

$P(B=1) = .3;$

$P(H=1|D=1) = .3,\ P(H=1|D=0) = .8;$

$P(L=1|O=1) = .3,\ P(L=1|O=0) = .6;$

$P(D=1|O=1, B=1) = .05, P(D=1|O=1, B=0) = .1,$
$P(D=1|O=0, B=1) = .1, P(D=1|O=0, B=0) = .2$

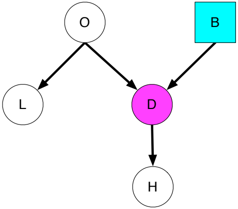
The joint probability from the network is:

$$P(O, L, D, B, H) = P(O)\, P(B)\, P(L|O)\, P(D|O, B)\, P(H|D)$$

Thus, for instance

$$\begin{aligned}\mathrm{P(O{=}0,\ L{=}0,\ D{=}1,\ B{=}1,\ H{=}1)} &= P(O=0)\, P(B=1)\, P(L=0|O=0)\times \\ &\quad P(D=1|O=0, B=1)\, P(H=1|D=1) \\ &= .4 \times .3 \times .4 \times .1 \times .3 \\ &= .0014\end{aligned}$$

Causal Inference



$P(O = 1) = .6;$

$P(B = 1) = .3;$

$P(H = 1|D = 1) = .3, \; P(H = 1|D = 0) = .8;$

$P(L = 1|O = 1) = .3, \; P(L = 1|O = 0) = .6;$

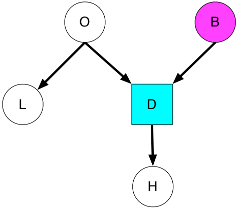
$P(D = 1|O = 1, B = 1) = .05, P(D = 1|O = 1, B = 0) = .1,$
$P(D = 1|O = 0, B = 1) = .1, P(D = 1|O = 0, B = 0) = .2$

Probability dog is outside if it has stomach problems.

Observe B want to infer D.

$$
\begin{aligned}
P(D = 1|B = 1) &= \sum_{o=0}^{1} P(D = 1, O = o|B = 1), \quad \leftarrow \text{marginalization} \\
&= \sum_{o=0}^{1} P(D = 1|O = o, B = 1)P(O = o|B = 1), \quad \leftarrow \text{chain rule} \\
&= \sum_{o=0}^{1} P(D = 1|O = o, B = 1)P(O = o) \quad \leftarrow \text{conditional independence} \\
&= .1 \times .4 + .05 \times .6 = .07
\end{aligned}
$$

Evidential Inference



$P(O = 1) = .6;$

$P(B = 1) = .3;$

$P(H = 1|D = 1) = .3,\ P(H = 1|D = 0) = .8;$

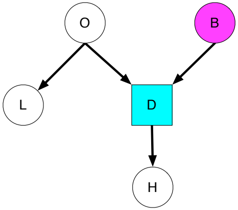
$P(L = 1|O = 1) = .3,\ P(L = 1|O = 0) = .6;$

$P(D = 1|O = 1, B = 1) = .05, P(D = 1|O = 1, B = 0) = .1,$
$P(D = 1|O = 0, B = 1) = .1, P(D = 1|O = 0, B = 0) = .2$

Probability dog does not have stomachache given it is inside.
Observe D want to infer B.

$$P(B = 0|D = 0) = \frac{P(D = 0|B = 0)P(B = 0)}{P(D = 0)} \quad \leftarrow \text{Bayes' rule}$$

Evidential Inference



$P(O = 1) = .6;$

$P(B = 1) = .3;$

$P(H = 1|D = 1) = .3, \; P(H = 1|D = 0) = .8;$

$P(L = 1|O = 1) = .3, \; P(L = 1|O = 0) = .6;$

$P(D = 1|O = 1, B = 1) = .05, P(D = 1|O = 1, B = 0) = .1,$
$P(D = 1|O = 0, B = 1) = .1, P(D = 1|O = 0, B = 0) = .2$

Probability dog does not have stomachache given it is inside.

Observe D want to infer B.

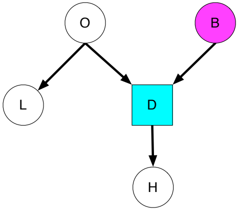
$$P(B = 0|D = 0) = \frac{\boxed{P(D = 0|B = 0)}\,P(B = 0)}{P(D = 0)} \quad \leftarrow \text{Bayes' rule}$$

Likelihood:

$$P(D = 0|B = 0) = \sum_{o=0}^{1} P(D = 0|O = o, B = 0)P(O = o) \quad \leftarrow \text{as in previous slide}$$

$$= .9 \times .6 + .8 \times .4 = .86$$

Evidential Inference



$P(O=1) = .6;$

$P(B=1) = .3;$

$P(H=1|D=1) = .3, \; P(H=1|D=0) = .8;$

$P(L=1|O=1) = .3, \; P(L=1|O=0) = .6;$

$P(D=1|O=1,B=1) = .05, P(D=1|O=1,B=0) = .1,$
$P(D=1|O=0,B=1) = .1, P(D=1|O=0,B=0) = .2$

Probability dog does not have stomachache given it is inside.

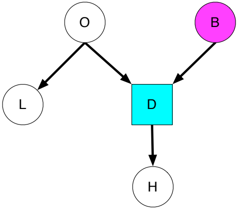
Observe D want to infer B.

$$P(B=0|D=0) = \frac{P(D=0|B=0)P(B=0)}{\boxed{P(D=0)}} \quad \leftarrow \text{Bayes' rule}$$

Evidence:

$$\begin{aligned} P(D=0) &= \sum_{o=0}^{1}\sum_{b=0}^{1} P(D=0|O=o,B=b)P(B=b,O=o) \\ &= \sum_{o=0}^{1}\sum_{b=0}^{1} P(D=0|O=o,B=b)\,P(B=b)\,P(O=o) = .881 \end{aligned}$$

Evidential Inference



$P(O=1)=.6;$

$P(B=1)=.3;$

$P(H=1|D=1)=.3,\ P(H=1|D=0)=.8;$

$P(L=1|O=1)=.3,\ P(L=1|O=0)=.6;$

$P(D=1|O=1,B=1)=.05, P(D=1|O=1,B=0)=.1,$
$P(D=1|O=0,B=1)=.1, P(D=1|O=0,B=0)=.2$

Probability dog does not have stomachache given it is inside.

Observe D want to infer B.

$$P(B=0|D=0)=\frac{P(D=0|B=0)P(B=0)}{P(D=0)} \quad \leftarrow \text{Bayes' rule}$$

Putting everything together:

$$P(B=0|D=0)=\frac{.86\times .7}{.881}=.683$$

Learning of Bayes Net

- **Trivial Case:**
  - Network structure known **and**
  - all variables observable (present in data vector).

  Can learn all conditional probabilities directly from data.
- **More difficult:**
  - Network structure known **and**
  - only some variables observable.

  Use the Expectation Maximization (EM) algorithm.
- **Most difficult:**
  - Network structure not known - subject of ongoing research (structural learning)

  Use some heuristic method to search through possible data structures.

  Learn dependencies from data.

Learning of Bayes Net

- **Trivial Case:**
  - Network structure known **and**
  - all variables observable (present in data vector).

  Can learn all conditional probabilities directly from data.
- **More difficult:**
  - Network structure known **and**
  - only some variables observable.

  Use the Expectation Maximization (EM) algorithm.
- **Most difficult:**
  - Network structure not known - subject of ongoing research (structural learning)

  Use some heuristic method to search through possible data structures.

  Learn dependencies from data.

Learning of Bayes Net

- **Trivial Case:**
  - Network structure known **and**
  - all variables observable (present in data vector).

  Can learn all conditional probabilities directly from data.
- **More difficult:**
  - Network structure known **and**
  - only some variables observable.

  Use the Expectation Maximization (EM) algorithm.
- **Most difficult:**
  - Network structure not known - subject of ongoing research (structural learning)

  Use some heuristic method to search through possible data structures.

  Learn dependencies from data.