Algorithms for Graphical Models (AGM)

Bayesian nets

$Date: 2008/10/16 10:24:16 $

AGM-08

In this lecture

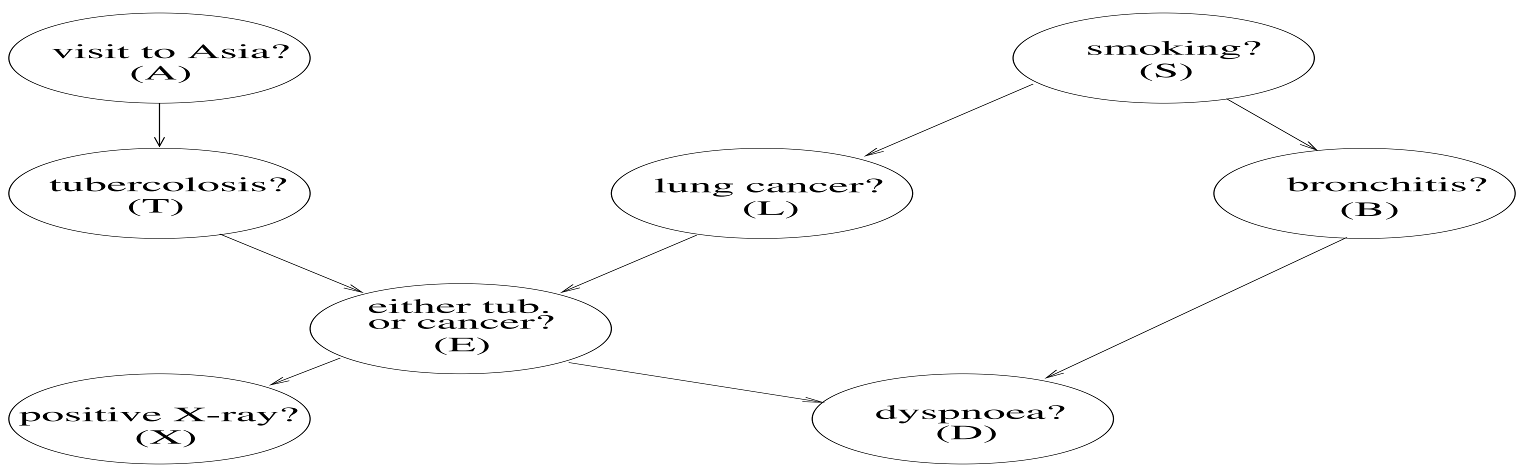
Bayesian nets:

- what they are
- as factored distributions
- conditional independence properties

Bayesian net structure

- The structure of a Bayesian net is given by a directed acyclic graph (DAG) whose nodes are the variables of a distribution.

- Here's one for the 'Asia' example:



Directed graph terminology

- If there's an arrow $A \to B$ in a directed graph, then A is a *parent* of B and B is a *child* of A.

- A directed graph (digraph) is *acyclic* if the graph contains no loops which follow the direction of the arrows.

- The DAG defines a partial order on nodes. A (total) ordering of nodes consistent with a given DAG is one consistent with this partial ordering.

Bayesian net parameters

- There is a *conditional probability table (CPT)* for each variable.

- The CPT for variable X defines a distribution over the values of X *for each joint instantiation of the parents of* X

- Here are the CPTs for Asia. Cue `asia_cpts` from `gPy.Examples`

The distribution represented by a Bayesian net

- Note that each CPT is a factor.

- The distribution represented by a Bayesian net is the product of all the CPTs.

- So BNs are factored representations of probability distributions and so everything we have said about such more general factored representations holds for BNs as well.

- But (unsurprisingly!) BNs have extra properties.

No normalisation needed

- A product of CPTs, one for each variable, defines a distribution directly, no matter what numbers are in the CPTs . . .

- . . . *as long as the corresponding digraph is acyclic*.

- There is no need to normalise: you will prove this as an exercise.

The moral graph

- If we ‘forget’ that the CPTs are CPTs and just treat them as factors, then the resulting interaction graph is called the *moral graph*.

- Unmarried parents get connected (since they are together in some CPT) and the arrows disappear (since the CPTs lose their CPTness).

- It’s not difficult to see that a BN *factorises according to its moral graph*.

Conditional independencies from a BN's moral graph

- Since a BN factorises according to its moral graph, we can read off some conditional independencies from its moral graph.

- For example, we have that $\{A\} \perp \{B, D, S\} | \{E, L, T\}$.

- However, there are further conditional independence relations which hold in a BN but cannot be deduced from the moral graph.

- To get these we need to exploit properties of CPTs . . .

The key property of CPTs

- For any configuration of its parents the values for the child must add up to one: this is what makes it a CPT.
- So if we marginalise away the child variable in a CPT we end up with a factor of ones.
- Factors with all data values = 1 can be deleted from a factored distribution without altering the distribution.
- So if a variable only appears in its own CPT then we can marginalise it away from the Bayesian net by simply removing this CPT!

Removing childless variables from Bayesian nets

- Let $P(A, T, E, L, S, B, D, X)$ be the joint distribution defined by the 'Asia' Bayesian net.

- The marginal distribution $P(A, T, E, L, S, B, X)$ [D marginalised away] is given by the BN with (i) the CPT for D deleted and thus (ii) the node D and all links to it deleted in the corresponding acyclic digraph.

- The BN for $P(A, T, E, L, S, X)$ [B marginalised away in addition] is produced by deleting B from the BN for $P(A, T, E, L, S, B, X)$.

- Clearly there is some sort of general principle here . . .

Ancestral sets

- In a given digraph, the *ancestors* of vertex X, denoted $an(X)$, is the set of nodes Y such that there is a directed path from Y to X.

- It's just the transitive closure of the parent relationship.

- A set of nodes $\mathbf{X}$ is an *ancestral set* iff for any node $X \in \mathbf{X}$ we have $an(X) \subseteq \mathbf{X}$.

- Let $An(\mathbf{Z})$ denote the smallest ancestral set containing a set of nodes $\mathbf{Z}$. (Just add ancestors, recursively.)

Ancestral sets by example

- Some ancestral sets in the 'Asia' digraph:
  $\emptyset, \{B, L, S\}, \{A, E, L, S, T\}, \{A, B, E, L, S, T\}, \{A, S\}$.

- And some sets which are not ancestral:
  $\{L\}, \{A, E, L, S\}$.

The key property of Bayesian nets

- (Notation: Let P be a joint distribution and $\mathbf{X}$ be some subset of its variables, then denote the distribution produced by marginalising P onto $\mathbf{X}$ as $P_{\mathbf{X}}$. If $\mathcal{G}$ is a graph, let $\mathcal{G}_{\mathbf{X}}$ be the graph formed by removing from $\mathcal{G}$ all nodes not in $\mathbf{X}$.)

- Let P be defined by a BN with acyclic digraph $\mathcal{G}$ and let $\mathbf{X}$ be an ancestral set in $\mathcal{G}$...

- ... then $P_{\mathbf{X}}$ is given by the BN with DAG $\mathcal{G}_{\mathbf{X}}$ (and the relevant CPTs deleted).

What this means for conditional independence

- If we want to know whether the BN structure implies $\mathbf{A} \perp \mathbf{B}|\mathbf{S}$ for disjoint sets of variables $\mathbf{A}, \mathbf{B}, \mathbf{S}$ ...

- ... first construct the DAG for the distribution $P_{An(\mathbf{A}\cup\mathbf{B}\cup\mathbf{S})}$.

- This is just $\mathcal{G}_{An(\mathbf{A}\cup\mathbf{B}\cup\mathbf{S})}$ where $\mathcal{G}$ is the original DAG.

- Then construct the moral graph for this (smaller) DAG and use this (smaller) moral graph to check for separation.

The directed global Markov property

This is from Lauritzen:

Let P factorise recursively according to $\mathcal{G}$. Then

$$\mathbf{A} \perp \mathbf{B} | \mathbf{S}$$

whenever $\mathbf{A}$ and $\mathbf{B}$ are separated by $\mathbf{S}$ in $(\mathcal{G}_{An(\mathbf{A} \cup \mathbf{B} \cup \mathbf{S})})^m$, the moral graph of the smallest ancestral set containing $\mathbf{A} \cup \mathbf{B} \cup \mathbf{S}$.

This property, which recursive factorisation implies, is the *directed global Markov property*.

Further conditional independence relations in ‘Asia’

- Using this technique from $\mathcal{G}_{\text{Asia}}$ we can deduce that $A \perp S|\emptyset$.
- And that $A \perp S|T$.
- But not that $A \perp S|E$. It could be that $A \perp S|E$ for a particular choice of CPTs, but this does not hold for all choices of CPTs.

The directed local Markov property

- In some DAG $\mathcal{G}\ldots$

- Let $nd(A)$ be the non-descendants of node A and $pa(A)$ be the parents of A.

- A distribution P obeys the *directed local Markov property* relative to $\mathcal{G}$, if each variable is independent of its non-descendants given its parents: $A \perp nd(A) | pa(A)$.

- It turns out that for any DAG $\mathcal{G}$, recursive factorisation, the global Markov property and the local Markov property are all equivalent.

Inference in Bayesian nets

- Since a BN is a factored distribution, one option is just to use variable elimination as previously described.

- There are also BN-specific options for inference, both exact and approximate.