Inference and Representation: Bayesian Networks

Rahul G. Krishnan

New York University

Lab 2, September 14, 2015

Outline

Today's Recitation

- Bayesian Networks
- Conditional Independencies
- A small Bayesian Network
- A challenge appears......

Quick Recap

- You learned about random variables last week, in this weeks lecture you learned about how to combine them to capture relationships between them.
- Bayesian Networks are Directed Acyclic Graphs (DAGs)
- The nodes are random variables, the edges represent conditional probability distributions

Designing your own networks

As you think about interesting problems for your projects, you might wonder, what should I know when I design my Bayesian Network?

- How do you come up with the structure? Use prior knowledge.
- How do you model each of the random variables? Ask yourself (1) is the random variable positive? (2) how would X behave if i know the value of Y

Posterior Distribution

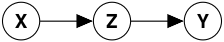
- Recall: In last weeks recitation, we talked about probablistic inference as predicting the parameters of some (unknown) distribution
- Posterior distribution: The distribution of unobserved (or latent) random variables conditioned on the observed data
- Often, this is the distribution we will be interested in approximating.

Bayesian Network

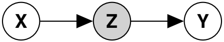
- Three important rules for conditional independance:
  - Cascade (or chain)
  - Common Parent (or common cause)
  - Common Child (or v-structure)

Cascade

- Cascade/Chain: The first structure to be aware of is a cascade or a chain
- **Found in:** Hidden Markov Models
- **Useful for:** Modeling hierarchical generative processes

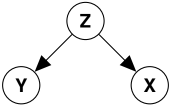


- $X \perp\!\!\!\perp Y | Z$



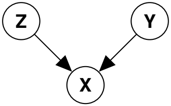
Common Parent

- **Found in:** Latent Dirichlet Allocation (a model you will study later), Naive Bayes (a model you already know)
- $X \perp\!\!\!\perp Y | Z$



V-structures/Common Child

- V-structures: Models the idea of combining multiple hypothesis (Y and Z) to create X
- **Found in:** Models of Semi-Supervised Learning (Kingma et. al)[1]

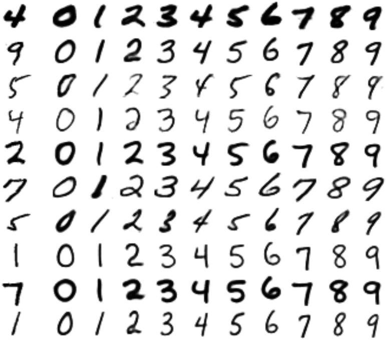


- $Z \perp\!\!\!\perp Y$, Z and Y are marginally independant
- However, conditioned on X $Z$and Y are dependant

[1] https://arxiv.org/pdf/1406.5298v2.pdf

An example in research

You will not be tested on this! :)



(b) MNIST analogies

Figure: Figure from Kingma et. al

Bayes Ball

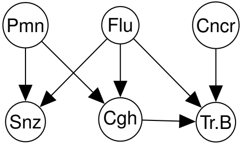
- The Bayes Ball algorithm is an answer to the question is X d-separated (directed-separated) from Y in a graph G
- Relies on local level rules to figure out whether a “ball” maybe pass between any node as it moves from X to Y or vice versa

Building a graphical model

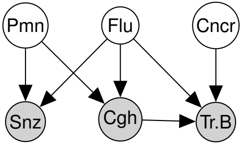
- Building a graphical model with the following random variables:
- Pmn: Pneumonia, Snz: Sneezing, Cou: Cough, Tr.B: Trouble Breathing
- Cncr: Cancer, Flu
- Observed vs Latent?

Graphical Model

Pmn: Pneumonia, Snz: Sneezing, Cou: Cough, Tr.B: Trouble Breathing



Bayes Ball

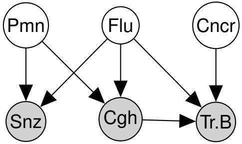


Is “pneumonia” conditionally independant of “cancer”?

Posterior Inference

Given the symptoms that you do observe, posterior inference in this graphical model would comprise estimating:

- $p(pmn = 1|snz, cgh, trb)$
- $p(pmn = 1, flu = 1|snz, cgh, trb)$
- $p(can = 0|snz, cgh, trb)$



A Challenge Appears!

- Look around, gather in groups of 4-5
- Make sure you have a worksheet.
- Important!: Put the names of all team members on it!
- You will not be marked on what you write.

Guidelines

"Everything should be made as simple as possible, but not simpler." - Einstein echoing William of Occam

- Minimize the number of edges you use! A fully connected network can represent any distribution but is difficult to perform inference in...
- *Use* conditional independencies!
- Scoring system: $\frac{N(N-1)}{2}$ - E where E is the number of edges you use in your graph and N is the number of nodes in your graph