PGMs week 2 Representation: Bayes Nets

1 WATCH the Koller videos on Bayesian Network Fundamentals

- *PGM5 - Bayesian Networks: Semantics & Factorization*
- *PGM6 - Bayesian Networks: Reasoning Patterns*
- *PGM7 - Bayesian Networks: Flow of Probabilistic Influence*
- *PGM8 - Bayesian Networks: Independencies Preliminaries*
- *PGM9 - Bayesian Networks: Independencies Bayesian Networks*
- *PGM10 - Bayesian Networks: Naive Bayes*
- *PGM11 - Bayesian Networks: Application: Diagnosis*
- *PGM12 - Bayesian Networks: Knowledge Engineering*

2 Comments on the videos

- *PGM5 Semantics & Factorization (17:21)* explains how a joint distribution, expressed as a product of factors, can equivalently be represented by a Directed Acyclic Graph termed a Bayes Net (BN). Although not a requirement, we typically implement the direction of the edges to correspond with causal relationships. We do so because this makes the parameters more intuitive to estimate, and the graph easier to interpret.
- *PGM6 Reasoning Patterns (10:00) - PGM9 Independencies in Bayesian Networks (18:19)* explains how in a BN you can directly see which variables affect which.
  - The idea of conditional independence is important.
  - You will also see that in some cases, observing a variable (i.e. it appears on the right hand side of a conditional) can cause some other variables to become conditionally independent, while in other cases it can cause them to become conditionally dependent. You will see that 'V-shaped' nodes (Barber calls them 'colliders', some other sources call them 'head-to-head' nodes) behave differently from the others. All of this culminates in the concept of d-separation.
  - Note that all the independencies that the structure of the graph G implies, *must* hold in the distribution P for G to be an I-MAP of P. However, there may be extra independencies in P that is not visible in the structure of G.
- *PGM10 Naive Bayes (09:53)* discusses a set of very commonly used simplifying assumptions, known as the Naive Bayes assumptions, and how it is represented via a BN.
- *PGM11 Application - Medical Diagnosis (09:20)* discusses a historical progression of medical diagnosis systems which played an important role in the development of modern BN's. Note how the initial rule-based expert system approaches (typical of the AI thinking in those days) became more and more probabilistic in nature, ultimately developing into BN modelling.
- *PGM12 Knowledge Engineering Example - SAMIAM (14:14)* discusses a BN toolbox called SAMIAM that you can download from the web at `http://reasoning.cs.ucla.edu/samiam/index.php`. You can use this to double check your results in the exercise below. However, for the moment rather focus on the examples here that illustrate "flow-of-influence" of variables in different configurations.

3 READ Barber chapter 3: BN's, aka Belief Networks.

This chapter mainly covers the same work as in the above videos, with possibly different formulations.

- As we have already seen last week, a distribution can be represented as a product of factors.
- On a given path 'colliders' (aka 'v-structures', aka 'head-to-head' nodes) behaves differently from other types of nodes. Also note that a node can be a collider on one path while not being so on another path.
- From these we get to the concept of d-separation which enables us to easily determine conditional independencies in a graph.
- Look carefully at the section on Markov equivalence – discussing when can two graphs represent the same set of independencies. They have to share the same 'skeleton' and the same set of 'immoralities'. Note the relationship between 'collider' nodes and immoralities.
- Once again, note that while certain independencies follows necessarily from the structure of G, there may be additional independencies resulting from the values of the distribution itself.
- Section 3.4 (Causality), although interesting, is optional. Judea Pearl (from whose work much of graphical models arose), is advancing an alternative to probability theory that explicitly takes interventions into account. Read his book 'Causality' if you want to follow this up.

4 CODE IT: Flow of influence in the Monty-Hall problem

4.1 Background

For this one we are going to consider the well-known Monty-Hall problem. Lets first state the original problem: In a TV game-show a participant must choose between one of three doors. Behind one of the doors there is an expensive car, the other two conceals somewhat smelly goats. The participant indicates a certain door after which the game-show host (Monty Hall) reveals a goat behind another door. You can assume that the host will never at this stage open the initial choice of the participant or the door with the car. Except for this, he has no preference for any specific door. The participant now may reconsider his choice. Under the assumption that he actually wants to win the car, should he stick with his original choice or should he change to the other remaining door?

Model the above situation with a BN. Use three random variables:

1. I is the initial choice of the participant – it takes on values $0, 1$ or 2 depending on his door of choice. The corresponding factor is $p(I)$.
2. C is where the car is – once again it takes on values $0, 1$ or 2 depending on where the car actually is. The corresponding factor is $p(C)$.
3. M is the door that Monty opens – once again it takes on values $0, 1$ or 2. Its factor is $p(M|I, C)$. This is the tricky one to set up, carefully include all the considerations that Monty have to keep in mind when doing so.

4.2 Solve the Monty-Hall problem

With this in place you can now easily answer the basic Monty-Hall question. Get the full joint $p(I, C, M) = p(M|I, C)p(I)P(C)$ (i.e. the product of all three above factors). Now simply pick an initial choice (e.g. $I = 0$) and a (valid) door that Monty opened (i.e. $M = 1$) via the observe/reduce operation to get $P(C|, I, M)$. That should tell you that it is better for the player to switch to the other door. Trivial. With that out of the way, lets more specifically focus on the concept of flow of influence.

4.3 Flow of influence via collider nodes

- The path $I \rightarrow M \leftarrow C$ forms a collider node at M. It is now easy to investigate various scenarios about the probabilistic dependence between I and C given that M is either unobserved or observed.

  - Set I to one of its allowed values, while leaving the other variables unobserved. Does this change the marginal for C? What is your conclusion about the dependence (or not) of C on I when M is unobserved?
  - Set M to one of the values it can take on, while leaving the other variables unobserved. Find the marginal for C. Now, with M still on the value you chose, set I to one of the two remaining values it can legally take on. Does this change the marginal for C?
  - Do the above in the opposite direction by setting a value for C and observing what it tells us about the distribution of I, for the cases where M is observed vs unobserved.

- Let's make things slightly more complex. The game remains more or less as it was, but we are now being prevented from observing M directly. Instead we get to observe a variable R that (unreliably) reports which door Monty opened. Define the factor $p(R|M)$ such that we have a 50% chance that R will report the correct value of M, and 25% chances that it will (erroneously) report one of the other two values for M. Repeat the above, but now with R observed/unobserved instead of M (which now is not being observed directly). Conclusions?
- Can you specify other equivalent networks by reversing the direction of some of the arrows?
- What happens to C when you set the observed value of R to be the same as I? What is the sense behind this?

4.4 Flow of influence via non-collider nodes

The path from $I \rightarrow M \rightarrow R$ forms a head-to-tail node at M. It is now easy to investigate various scenarios about the probabilistic dependence between I and R given that M is either unobserved or observed.

- Set I to one of its allowed values, while leaving the other variables unobserved. Does this change the marginal for R? What is your conclusion about the dependence (or not) of R on I when M is unobserved?
- Set M to one of the values it can take on, while leaving the other variables unobserved. Find the marginal for R. Now, with M still on the value you chose, set I to one of the two remaining values it can legally take on. Does this change the marginal for R?
- Do the above in the opposite direction by setting a value for R and observing what it tells us about the distribution of I, for the cases where M is observed vs unobserved.

NOTE: We have not investigated tail-to-tail non-collider nodes. Their influence blocking behaviour is the same as that of head-to-tail nodes.