DD2434/FDD3434 Machine Learning, Advanced Course
Module 2 Exercise

November 2024

Contents

1 Directed Graphical Models (DGM) – Theory

1.1 D-separation rules

If two nodes are d-separated, they are independent. In Figure 1, S and R are independent from each other.

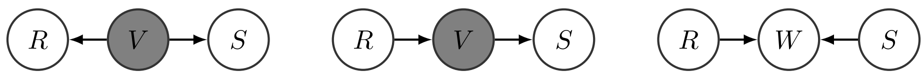


Figure 1: S and R are d-separated. Note that W is not observed and no descendants of W is observed either.

[Pearl 1988] example:
d-separated: dead battery → car won't start ← no gas
not d-separated: dead battery → **car won't start** ← no gas

A charged battery does not give information about the gas, but, a charge battery become informative (dependency) after observing that the car won't start (because then we know that gas was empty from knowing the battery was full and car didn't start)

A note regarding how to read conditional independence queries: If no conditioning is mentioned in the query of independence/dependence, then assume that all the shaded variables are given or conditioned on. But, if there is a specific conditioning in the query, only consider that, and, not all the shaded variables in the graphical model.

Two methods to determine if variables are independent or not:

1.1.1 Method 1 (D-separation)

This method is efficient when examining the dependence over the whole graph. We put a block symbol on the graph wherever we detect the d-separation using the rules in Figure 1. Then we can answer whether variables are d-separated or not by following the path between them i.e. if the path contains a block they are obviously d-separated. **Example** The blocks are illustrated as red point over the graph in Figure 2; the table shows the answers to d-separation which is gained by following the path between each pair of i and j.

i, j	d-separated given C
A, D	yes
A, B	no
D, G	yes
D, H	no
D, E	yes
D, F	no
B, H	yes

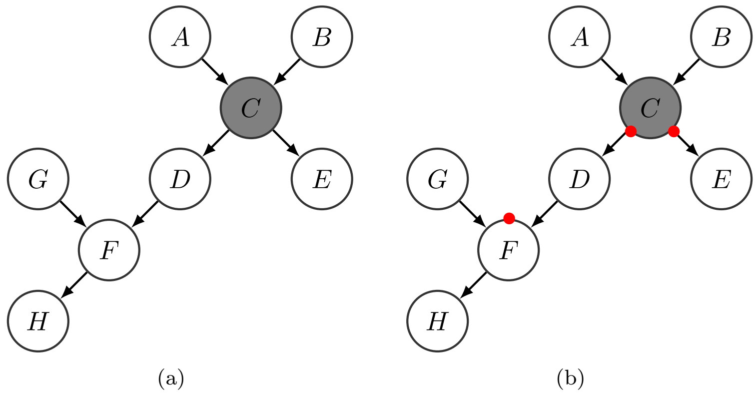


Figure 2: Blocks are shown in red on the graph in (b).

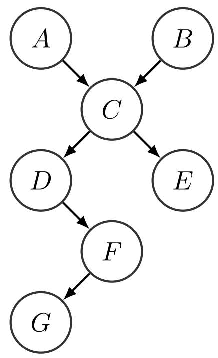
1.1.2 Method 2

This method is efficient when the dependencies are being examined among certain variables and not all the variables in the graph. We follow this procedure [1]:

- 1. Construct the "ancestral graph" of all variables mentioned in the probability expression. This is a reduced version of the original net, consisting only of the variables **mentioned** and all of their **ancestors** (parents, parents' parents, etc.)
- 2. "Moralize" the ancestral graph by "marrying" the parents. For each pair of variables with a common child, draw an undirected edge (line) between them. (If a variable has more than two parents, draw lines between every pair of parents.)
- 3. "Disorient" the graph by replacing the directed edges (arrows) with undirected edges (lines).
- 4. Delete the givens **(not the actual variables that we intend to judge their dependency - in some scenario the variables which are to be judged about their dependency, are also observed, and, them being observed does not mean that they will be deleted.)** and their edges. If the independence question had any given variables, erase those variables from the graph and erase all of their connections, too.
- 5. Read the answer off the graph.
  - If the variables are disconnected in this graph, they are guaranteed to be independent.
  - If the variables are connected in this graph, they are not guaranteed to be independent.* Note that "are connected" means "have a path between them," so if we have a path X-Y-Z, X and Z are considered to be connected, even if there's no edge between them.
  - If one or both of the variables are missing (because they were givens, and were therefore deleted), they are independent.

* We can say “the variables are dependent, as far as the Bayes net is concerned” or “the Bayes net does not require the variables to be independent,” but we cannot guarantee dependency using d-separation alone, because the variables can still be numerically independent (e.g. if $P(A|B)$ and $P(A)$ happen to be equal for all values of A and B).

Example Q: Are A and B conditionally independent, given D and F?



A: They are not required to be conditionally independent, using method 1.1.2 as below:

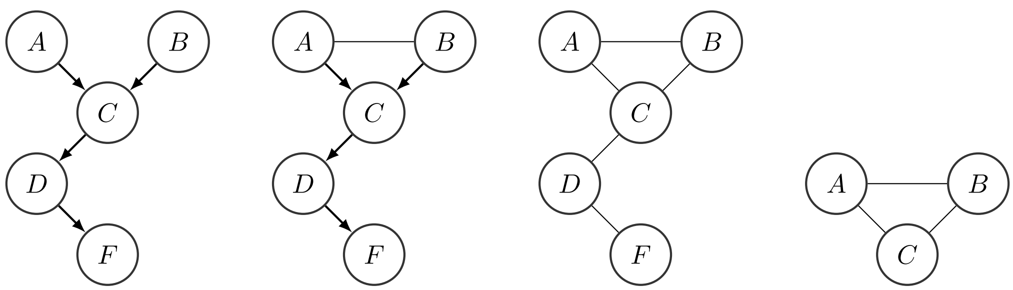
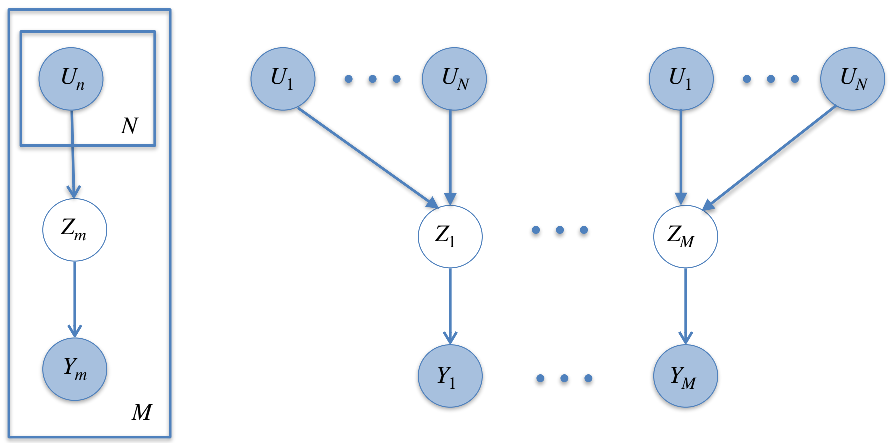


Figure 3: From left to right: 1. Draw ancestral graph 2. Moralize 3. Disorient 4. Delete givens

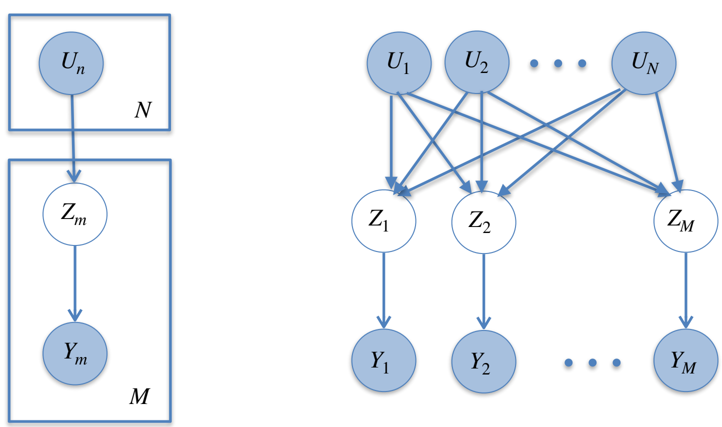
1.2 Plate notation

In plate notation, the node/graph is replicated N times (when N is written in the plate); moreover, any edge that crosses a plate boundary is replicated once for each subgraph repetition. The plate notation is used to illustrate a compact representation of the graphical model. Using such a simplified notation allows to have an overview of model parameters and variables. However, it is sometimes useful to unfold the plate, e.g., when evaluating the conditional dependencies. In Fig. 4, two examples of unfolding a plate graph are illustrated.



The plate graph

The corresponding unfolded graph



The plate graph

The corresponding unfolded graph

Figure 4: Two examples of unfolding a plate DGM.

2 Directed Graphical Models (DGM) – Exercises

2.1 Bayes Ball

Question: List all variables that are independent of A given evidence on the shaded node for each of the DGMs a), b) and c) below.

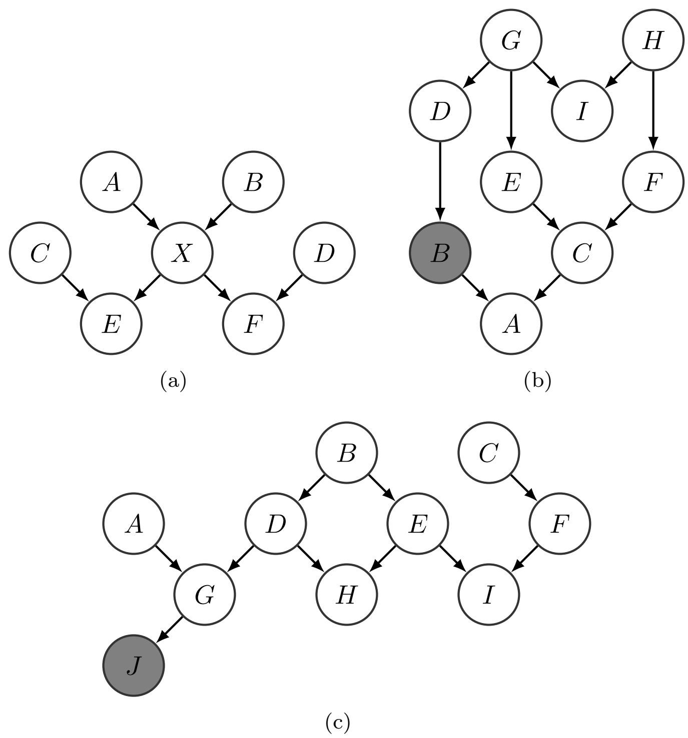


Figure 5: Some DGMs.

2.2 PyClone DGM

Consider the graphical model shown in Figure 6. Answer "yes" or "no" to each question:

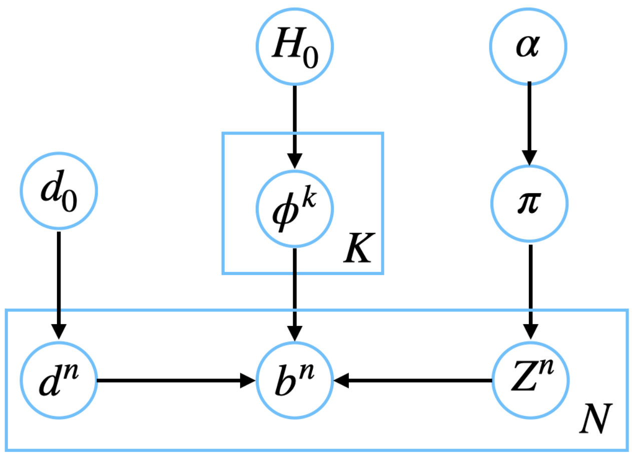


Figure 6: Graphical model of PyClone in plate notation

$$\pi \sim Dirichlet(\alpha) \tag{1}$$
$$Z^n \sim Categorical(\pi) \tag{2}$$
$$\phi^k \sim H_0 = Beta(a_0, a_1) \tag{3}$$
$$d^n \sim Poisson(d_0) \tag{4}$$
$$b^n \sim Bin(d^n, \phi^{Z^n}) \tag{5}$$

- $b^n \perp b^{n+1} \mid d^n, d^{n+1}$?
- $d^n \perp Z^n \mid \alpha, H_0$?
- $d^n \perp Z^n \mid b^n$?
- $\phi^k \perp d^n \mid d_0, \pi$?
- $b^{1:N} \perp \pi \mid Z^{1:N}$?

2.3 Bayes nets for a rainy day (Exercise 10.5 from Murphy [2])

Question: (Source: Nando de Freitas) In this question you must model a problem with 4 binary variables: G = "*gray*", V = "*Vancouver*", R = "*rain*" and S = "*sad*". Consider the directed graphical model describing the relationship between these variables shown in Figure 7 (and the probability tables shown in Table 1).

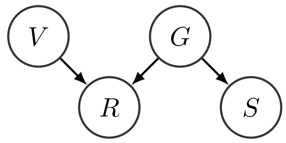


Figure 7: Bayesian net for a rainy day

Table 1: Probability tables of Bayes net for a rainy day

$V = 0$	$V = 1$
δ	$1 - \delta$

$G = 0$	$G = 1$
α	$1 - \alpha$

	$S = 0$	$S = 1$
$G = 0$	γ	$1 - \gamma$
$G = 1$	β	$1 - \beta$

	$R = 0$	$R = 1$
$VG = 00$	0.6	0.4
$VG = 01$	0.3	0.7
$VG = 10$	0.2	0.8
$VG = 11$	0.1	0.9

a. Write down an expression for $P(S = 1|V = 1)$ in terms of $\alpha, \beta, \gamma, \delta$.

b. Write down an expression for $P(S = 1|V = 0)$. Is this the same or different to $P(S = 1|V = 1)$? Explain why.

c. Find maximum likelihood estimates of α, β, γ using the following data set, where each row is a training case. (You may state your answers without proof.)

V	G	R	S
1	1	1	1
1	1	0	1
1	0	0	0

3 Hidden Markov Models (HMM) – Exercises

3.1 Forward-Backward Algorithm for Posteriors

Derive forward-backward algorithm for:

- **(a)** the marginal posterior distribution of one hidden variable, i.e, $p(z_n|x_{1:N})$
- **(b)** the joint posterior distribution of two hidden variables, i.e, $p(z_{n-1}, z_n|x_{1:N})$
- **(c)** the posterior predictive distribution, i.e, $p(x_{N+1}|x_{1:N})$