Exercises on Bayesian Networks and Undirected Graphical Models

Raquel Urtasun and Tamir Hazan

This exercise session has 4 questions, each of which is 25 points, as well as an additional question which if you solve give you and additional 25 points. It is due on **April 20, at 1:30 pm**. Please bring the exercises to the class.

1 Question 1

Given a direct acyclic graph (DAG) over n vertices, we define $parents(i) \subset \{1, ..., n\}$ to be the set of parents the vertex i has in the graph. We also define $non-descendants(i) \subset \{1, ..., n\}$ to be the set of non-descendants of i in the graph. Let

$$p(x_1, ..., x_n) \stackrel{def}{=} \prod_{i=1}^{n} p(x_i | x_{parents(i)}).$$

Show that for this joint probability distribution it holds that

$$X_i \perp X_{non-descendants(i)} | X_{parents(i)}$$

with X_I the joint distribution of the variables X_i for $i \in I$.

2 Question 2

Consider the network shown in Figure 1, where we assume that all variables are binary, and that the the conditional probability of F_i, which is typically called noisy-or[1], is given by:

$$P(F_i = 0 | D_{parents(i)} = d_{parents(i)}) = (1 - \lambda_{i,0}) \prod_{D_j \in parents(i)} (1 - \lambda_{i,j})^{d_i}$$

where $\lambda_{i,j}$ is the noise parameter associated with parent D_j of variable F_i. This network architecture, called a $BN2O$ network is characteristic of several medical diagnosis applications, where the D_i variables represent diseases (e.g., flu, pneumonia), and the F_i variables represent medical findings (e.g., coughing, sneezing).

[1] The term noisy-or describes the behavior that or the or-function for which F_i is inactive, i.e. $F_i = 0$, if all its parents are inactive, i.e. $D_j = 0$ for every $j \in parents(i)$. The or-function is given by setting the noise parameters, $\lambda_{i,0} = 0, \lambda_{i,j} = 1$

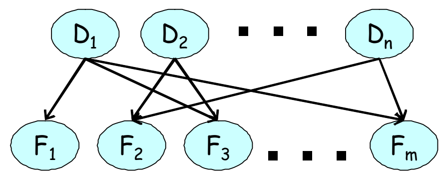
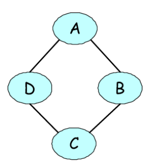


Figure 1: A two-layer noisy-or network.



$\epsilon_1[A, B]$

a^0	b^0	-3.4
a^0	b^1	-1.61
a^1	b^0	0
a^1	b^1	-2.3

$\epsilon_2[B, C]$

b^0	c^0	-4.61
b^0	c^1	0
b^1	c^0	0
b^1	c^1	-4.61

$\epsilon_3[C, D]$

c^0	d^0	0
c^0	d^1	-4.61
c^1	d^0	-4.61
c^1	d^1	0

$\epsilon_4[D, A]$

d^0	a^0	-4.61
d^0	a^1	0
d^1	a^0	0
d^1	a^1	-4.61

Figure 2: Graphical model and energy functions for an undirected graphical model $\mathcal{H}$.

Our general task is medical diagnosis: We obtain evidence concerning some of the findings, and we are interested in the resulting posterior probability over some subset of diseases. However, we are only interested in computing the probability of a particular subset of the diseases, so that we wish (for reasons of computational efficiency) to remove from the network those disease variables that are not of interest at the moment.

Considering a particular variable F_i, and assume (without loss of generality) that the parents of F_i are $D_1, ..., D_k$, and that we wish to maintain only the parents $D_1, ..., D_l$ for $l < k$. Show how we can construct a new noisy or CPD for F_i that preserves the correct joint distribution over $D_1, ...D_l, F_i$. Hint: Construct a $\lambda'_{i,0}$ accounting for the removed parents.

3 Question 3

Let $\mathcal{H}$ be the undirected graphical model in Fig. 2. Let the energy function be

$$\begin{aligned}
\epsilon'_1[a, b^i] &= \epsilon_1[a, b^i] + \lambda^i \\
\epsilon'_2[b^i, c] &= \epsilon_2[b^i, c] - \lambda^i
\end{aligned}$$

Show that the resulting energy function is equivalent for any constant λ^i.

4 Question 4

Let P satisfy $\mathcal{I}_l(\mathcal{H})$, and assume that X and Y are two nodes in $\mathcal{H}$ that are NOT connected directly by and edge. Prove that P satisfies $(X \perp Y | \mathcal{X} - \{X, Y\})$.

5 Bonus Question

Provide an example of a class of Markov networks $\mathcal{H}_n$ over n such that the size of the largest clique in $\mathcal{H}_n$ is constant, yet any Bayesian network I-map for $\mathcal{H}_n$ is exponentially large in n.