IE 598 Inference in Graphical Models

Homework 1

Covers lecture slides

1. Overview
2. Graphical models
3. Markov property (partially)

Problem 1.1 (Exercise 2.5 in Koller/Friedman)

Let X, Y, Z be three disjoint subsets of random variables. We say X and Y are conditionally independent given Z if and only if

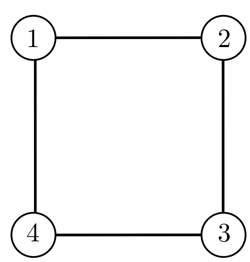
$$\mathbb{P}_{X,Y|Z}(x, y|z) \quad = \quad \mathbb{P}_{X|Z}(x|z)\, \mathbb{P}_{Y|Z}(y|z) \; .$$

Show that X and Y are conditionally independent given Z if and only if the joint distribution for the three subsets of random variables factors in the following form:

$$\mathbb{P}_{X,Y,Z}(x, y, z) \quad = \quad h(x, z)\, g(y, z) \; .$$

Problem 1.2 (Exercise 4.1 in Koller/Friedman)

In this problem, we will show by example that the distribution of a graphical model need not have a factorization of the form in the Hammersley-Clifford Theorem if the distribution is not strictly positive. In particular, we will consider a distribution on the following simple 4-cycle where each node is a binary random



variable, X_i, for $i \in \{1, 2, 3, 4\}$. Consider a probability distribution that assigns a probability 1/8 uniformly to each of the following set of values (X_1, X_2, X_3, X_4):

$$\begin{array}{cccc} (0,0,0,0) & (1,0,0,0) & (1,1,0,0) & (1,1,1,0) \\ (0,0,0,1) & (0,0,1,1) & (0,1,1,1) & (1,1,1,1) \end{array}$$

and assigns zero to all other configurations of (X_1, X_2, X_3, X_4).

(*a*) We first need to show that this distribution is Markov on our graph. To do this, it should not be difficult to see that what we need to show are the following conditions:

– The pair of variables X_1 and X_3 are conditionally independent given (X_2, X_4).
– The pair of variables X_2 and X_4 are conditionally independent given (X_1, X_3).

First, show that if we interchange X_1 and X_4 and interchange X_2 and X_3, we obtain the same distribution, i.e.., $\mathbb{P}(x_1, x_2, x_3, x_4) = \mathbb{P}(x_4, x_3, x_2, x_1)$. This implies that if we can show the first condition, then the other is also true.

(*b*) Show that whatever pair of values you choose for (X_2, X_4), we then know either X_1 or X_3 with certainty. For example, $(X_2 = 0, X_4 = 0)$ implies that $X_3 = 0$. Since we know either X_1 or X_3 with certainty, then conditioning on the other one of these obviously provides no additional information, trivially proving conditional independence.

(*c*) What we now need to show is that the distribution cannot be factorized in the way stated in the Hammersley-Clifford Theorem. We will do this by contradiction. Noting that the maximal cliques in our graph are just the edges and absorbing the normalization $1/Z$ into any of the pairwise compatibility functions, we know that if our distribution has the factorization implied by the Hammersley-Clifford Theorem, we can write it in the following form:

$$\mathbb{P}(x_1, x_2, x_3, x_4) \quad = \quad \psi_{12}(x_1, x_2)\, \psi_{23}(x_2, x_3)\, \psi_{34}(x_3, x_4)\, \psi_{41}(x_4, x_1)\ .$$

Show that assuming that our distribution has such a factorization leads to a contradiction by examining the values of $\mathbb{P}(0, 0, 0, 0)$, $\mathbb{P}(0, 0, 1, 0)$, $\mathbb{P}(0, 0, 1, 1)$, and $\mathbb{P}(1, 1, 1, 0)$.

Problem 1.3

Given a graph $G = (V, E)$, an *independent set* of G is a subset $S \subseteq V$ of the vertices, such that no two vertices in S is connected by an edge in E. Precisely, if $i, j \in S$ then $(i, j) \notin E$. We let $\mathrm{IS}(G)$ denote the set of all independent sets of G, and let $Z(G) = |\mathrm{IS}(G)|$ denote its size, i.e. the total number of independent sets in G. The number of independent sets $Z(G)$ is at least $1 + |V|$, since the empty set and all subsets with single vertex are always independent sets. We are interested in the uniform probability measure over S:

$$\mathbb{P}_{\mathrm{IS}(G)}(S) = \frac{1}{Z(G)} \mathbb{I}(S \in \mathrm{IS}(G))\ ,$$

where $\mathbb{I}(A)$ is an indicator function which is one if event A is true and zero if false.

(*a*) The set S can be naturally encoded by a binary vector $x \in \{0, 1\}^{|V|}$ by letting $x_i = 1$ if and only if $i \in S$. Denote by $\mathbb{P}_G(x)$ the probability distribution induced on this vector x according to $\mathbb{P}_{\mathrm{IS}(G)}(S)$. Show that $\mathbb{P}_G(x)$ is a pairwise graphical model on G.
[Hint: A pairwise graphical model on a graph $G = (V, E)$ is defined by a factorization of the form $\mathbb{P}_G(x) = (1/Z) \prod_{(i,j)\in E} \psi_{i,j}(x_i, x_j)$.]

(*b*) Let L_n be the line graph with n vertices, i.e. the graph with vertex set $V(L_n) = \{1, 2, 3, \dots, n\}$ and edge set $E(L_n) = \{(1, 2), (2, 3), \dots, (n-1, n)\}$. Derive a formula for $Z(L_n)$.
[Hint: Write a recursion over n, and solve it using a matrix representation.]

(*c*) With the above definitions, derive a formula for $\mathbb{P}_{L_n}(x_i = 1)$, for each $i \in \{1, \dots, n\}$. Plot $\mathbb{P}_{L_n}(x_i)$ versus i for $n = 11$. Describe the main features of this plot. Can you give an intuitive explanation?
[Hint: Use the recursion from previous subproblem.]

(*d*) The same probability distribution $\mathbb{P}_{L_n}(x)$ can be also represented with a Bayesian network. For example, $\mathbb{P}_G(x) = \mathbb{P}_{X_1}(x_1)\mathbb{P}_{X_2|X_1}(x_2|x_1)\mathbb{P}_{X_3|X_1,X_2}(x_3|x_1, x_2) \cdots \mathbb{P}_{X_{11}|X_1\cdots X_{10}}(x_{11}|x_1 \cdots x_{10})$. Using the recursions used in (*b*) and (*c*), write the conditional probability distributions for this Bayesian network.

Problem 1.4

We again consider the independent set explained in the previous problem. Now let $G = T_{k,\ell}$ denote the rooted tree with branching factor k and ℓ generations, that is the root has k descendants and each other node has one ancestor and k descendants except for the leaves. The total number of vertices is $(k^{\ell+1} - 1)/(k-1)$, and $T_{k,\ell=0}$ is the graph consisting only of the root. We let ϕ denote the root of $T_{k,\ell}$.

(a) Let $Z_\ell = Z(T_{k,\ell})$ denote the total number of independent sets of $G = T_{k,\ell}$. Let $Z_\ell(0)$ be the number of independent sets in $T_{k,\ell}$ such that the root is $x_\phi = 0$, and $Z_\ell(1)$ be the number of independent sets such that $x_\phi = 1$. It is immediate that $Z_0(0) = Z_0(1) = 1$. Derive a recursion expressing $(Z_{\ell+1}(0), Z_{\ell+1}(1))$ as a function of $(Z_\ell(0), Z_\ell(1))$.

(b) Using the above recursion, derive a recursion for the probability that the root belongs to a uniformly random independent set. Explicitly, derive a recursion for

$$p_\ell \quad = \quad \mathbb{P}_{T_{k,\ell}}(\{x_\phi = 1\}) \, .$$

(c) Program this recursion and plot p_ℓ as a function of $\ell \in \{0, 1, \ldots, 50\}$ for four values of k, e.g. $k \in \{1, 2, 3, 10\}$. Comment on the qualitative behavior of these plots.

(d) Prove that, for $k \leq 3$, the recursion converges to a unique value using Banach's fixed point theorem.

Problem 1.5 (Intersection lemma) In proving that pairwise Markov property implies global Markov property for undirected graphical models, we used the intersection lemma which states that if μ is strictly positive and

$$A\text{–}(C \cup D)\text{–}B, \qquad A\text{–}(B \cup D)\text{–}C,$$

then

$$A\text{–}D\text{–}(B \cup C) \, .$$

Here A–B–C if and only if $\mu(x_A, x_C | x_B) = \mu(x_A | x_B) \, \mu(x_C | x_B)$. From previous homework, we know that A–$(C \cup D)$–B if and only if $\mu(x_A, x_B, x_C, x_D) = a(x_A, x_C, x_D) \, b(x_B, x_C, x_D)$ for some function $a(\cdot)$ and $b(\cdot)$. Similarly, we have $\mu(x_A, x_B, x_C, x_D) = f(x_A, x_B, x_D) \, g(x_B, x_C, x_D)$.

(a) Show $f(x_A, x_B, x_D) = a'(x_A, x_D) \, b'(x_B, x_D)$ for some $a'(\cdot)$ and $b'(\cdot)$ and find one such pair of functions a' and b' in terms of $a(), b(), g()$.

(b) Substitute $f(\cdot)$ and prove A–D–$(B \cup C)$.

(c) Find a counter example when μ is not strictly positive.