CSC 535 – Probabilistic Graphical Models

Assignment Four

Due: 11:59pm (*) Friday, October 05.

(*) There is grace until 8am the next morning, as the instructor will not grade assignment before then. However, once the instructor starts grading assignments, no more assignments will be accepted.

Weight about 7 points

This assignment should be done individually

The purpose of this assignment is think a little bit about estimating quantities from a probability density followed by becoming very familiar with directed graphical models (Bayes nets).

Deliverables

Deliverables are specified below in more detail. For the high level perspective, you are to provide a program to output a few numbers and to create figures. You also need to create a PDF document that tells the story of the assignment including output, plots, and images that are displayed when the program runs. Even if the question does not explicitly remind you to put the resulting image into the PDF, if it is flagged with **($)**, you should do so. The instructor should not need to run the program to verify that you attempted the question. See

http://kobus.ca/teaching/grad-assignment-instructions.pdf

for more details about preparing write-ups. While it takes work, it is well worth getting better (and more efficient) at this. A substantive part of each assignment grade is reserved for exposition.

---

1. Denote $p(\Theta|X)$ for the probability distribution function for either model parameters, based on the data, X.

   Further assume that we want to use the squared loss function, $L(\Theta,\hat{\Theta})=(\Theta-\hat{\Theta})^2$. Use calculus to show that the loss is minimized by setting $\hat{\Theta}=\int \Theta\, p(\Theta|X)d\Theta$ **($)**.

   Notice that this result also applies to the predictive distribution, $p(x|X)$.

2. Construct your own example of "explaining away." **($)**. Your answer should begin with the specification of 3 random variables. Argue both intuitively, and also with an example based on a reasonable made up table for the joint distribution as we did in class **($)**.

   If you are having trouble with this, perhaps study the car battery, fuel tank, and electric gauge example in Bishop Chapter 8, bottom of page 376, with probabilities specified on page 377. (The chapter is on D2L and also on-line). Notice how the normalization (equation 8.33) makes it so that the "explanation" must lower the probability of the other explanation. Since observing one explanation influences the probability of the other explanation, they are not independent (conditioned on G).

3. For each of the factorizations below, **first** reorder the factors so that all variables first appear without being conditioned on (**$**). In other words, they first appear to the left of the "|" (including the case where there is no "|") . This is one canonical way to write the chain rule, i.e., P(A,B,C)=P(A)P(B|A)P(C|A,B). (The answer is not necessarily unique). **Second**, identify any differences between this form of the factorization and the canonical form of the chain rule where you choose the same ordering of the variables to the left of the "|" as in the factorization (**$**). This identifies where the factorization restricts the distribution from the general case. **Third**, draw the graph representing the factorization (**$**).

   Comment. The main point of this question is to understand that key property of our models, as simplifications, is that some links that could be there, are not.

   (a) p(A,B,C,D,E)=p(A|D,E)p(B|D)p(C)p(D|C)p(E)

   (b) p(A,B,C,D,E)=p(A|C)p(E|B)p(C)p(B)p(D|A,E)

   (c) p(A,B,C,D,E)=p(C|A,B,D,E)p(A|B,D)p(D|B)p(B)p(E|A,B,D)

4. For **each** of your graphs in the previous question (in the order given) use d-separation to argue whether or not i) C and B must be conditionally independent give D (**$**), and ii) E and C must be conditionally independent given B (**$**). [Note that "must be" is asking about properties that must be true, regardless of the distribution. There may be some specific distributions which have extra conditional independencies not evident by d-separation, but there will be other distributions where this would not be the case].

   (Please do (i) and (ii) for distribution (a), followed by (i) and (ii) for distribution (b), and then (i) and (ii) for distribution (c)).

5. Suppose that your undergraduate grades (G) influence the letters (L) that professors write for you (higher grade, better letter). Another thing that influences those letters is your interactions (I) with those professors (better interaction, better letter). Now suppose that the rank (R) of the graduate program you get into is influenced by your GRE test (T), your grades, and your letter (better letter increases probability that rank will be higher). Finally, suppose that whether or not you get an academic job (J) depends on the graduate program rank (better rank increases your chances), and the state (S) where the job is located (snarky political comment goes here).

   (a) Draw a directed graphical model for this problem (**$**).

   (b) Write down an expression for the probability distribution that corresponds to your model (**$**).

   (c) Argue that getting the job is independent of your grades, *conditioned on the rank of your grad program* (**$**). (Yup, eventually they are forgotten about!).

   (d) Are GRE scores independent of grades in your model? Provide a formal argument based on rules learned in class (**$**).

   (e) Are GRE scores independent of the quality of the letter in your model? Provide an argument based on rules learned in class (**$**).

   (f) Are GRE scores independent of the quality of the letter *given the on the rank of the grad program*? Provide an argument based on rules learned in class (**$**).

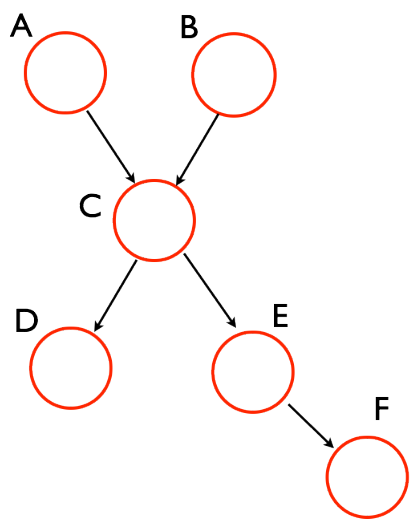
   (g) Now suppose that getting the job *also* depends on one's interaction with their professors (better interaction, higher chance). In other words, good interaction with the professor might be indicative of some qualities that may also help the candidate in a job interview. Is getting the job *still* independent of their grades *given the rank of their grad program*? Provide an argument based on rules learned in class (**$**).

   If you argue "no", then explain it in English as well (**$**). In particular address whether learning that a friend had really high grades increases or decreases the amount of money you are willing to bet that they get the job.

   (h) Using the extended model from (g), suppose you now have also come across your friend's letter (the professor forgot to shred the draft). Conditioned on your knowledge of both the *grad program* (and thus its rank) that your friend is in, and the *letter*, are their grades independent of them getting a job? (**$**). Provide an argument based on rules learned in class. (**$**).

   If you argue "no", then explain it in English as well (**$**). In particular address whether learning that your friend had really high grades increases or decreases the amount of money you are willing to bet that they get the job.

6. Consider this graph:



(d) Using the abstract definition of what a directed graphical model means in terms of local conditional independencies, write down all the local conditional independencies that it expresses ($). You can make use of the compact representation for X being conditionally independent of multiple variables by separating them with commas. (This is valid because conditioning on them considered as a group implies doing so for each one in particular). Also, you do not need to write down independencies that are symmetric; one of the two will suffice.

(e) Use this to derive an expression for the probability distribution expressed ($).

(f) Write down the probability distribution available by our first definition of a directed graphical model that specified a particular factorization of the distribution ($). Is it the same? ($). Do you expect it to be? (**$**).

Do one of the next two questions. There is no "right" answer to either of these problems until you state your assumptions and/or reasoning. So, I expect to get some variety of answers. A good answer will be clear about the assumptions, which will be neither two simplistic, nor everything depends on everything. It will also be consistent with the proposed network.

7. Provide a graphical model for detections of asteroids in an image taken at time T. Consider that there is a useful prior probability on the orbital parameters of asteroids, as most of them are in solar orbit in approximately the same plane as the other planets, and so you will find them near the "ecliptic" (projection of this plane onto the sky as seen from earth). Given an orbit, there is uniform probability of where along the orbit an asteroid is at time T. Together with the orbit, this tells you where an asteroid is in space in 3D coordinates. Given that, you can assume a function that tells you where in the image the asteroid is, including a special value for the case that the asteroid is not in the image and therefore can be ignored (this is the most common case by far). If the asteroids location is in the image, there is some chance that it will be detected. Further, the location of the moon influences detections. Ditto for the weather. Finally, the detector hallucinates detections (noise).

   Your mission is to provide a graphical model for detections of moving objects that might be asteroids. (**$**). Note that part of the problem is to figure out what the relevant random variables are. Be very clear in your writeup what the variables represent (**$**). Finally, ensure that you have a generative model by explaining how you get real data samples by ancestral sampling (**$**).

8. Provide a graphical model for the following problem. Consider a dancer with LED lights attached to appropriate parts of their body such as their center of mass, their forehead, and major joints (shoulders, knees, elbows, ankles)—you can specify these locations as you like. These 3D positions have structure that you can exploit. For example, the location of the foot is influenced by the location of the knee. Having developed a model for the 3D locations, consider a camera function with unknown parameters which have prior probability distributions. This camera model mediates how points in 3D become points in images. Notice that you may not see an LED due to occlusion. Provide a graphical model for this situation that includes the data we plan to observe (location of LED lights in the image) ($). Explain your notation and assumptions clearly **($)** — there are many possible answers, and you will want to make (and state!) some assumptions, but we would like a well explained model that makes sense and is consistent with your graphical model. Finally, ensure that you have a generative model by explaining how you get real data samples by ancestral sampling **($).**

What to Hand In

If you wrote code for this assignment, you should hand in a program hw4.<suffix> (e.g., hw4.m if you are working in Matlab) and the PDF file hw4.pdf with the story of your efforts into D2L.