CS181 Assignment 4: Probabilistic Graphical Models and Inference

Out Monday 1 April
Due Friday 12 April

Submit by **noon** via iSites dropbox.

Problem 1

[14 Points] You are building a directed graphical model (Bayesian network) to help you decide whether or not there is a virus on your computer. This virus, if present, would print silly messages to your screen. You have decided to use the following attributes, all of which are binary:

- **RD**: recent download – True if there has been a recent download from an untrusted site.
- **SMOS**: silly messages on screen – True if "you've been hacked" or some such message appears on screen.
- **VRAS**: virus reported by antivirus software – True if your antivirus software reports the presence of a virus.
- **MAI**: McAfee antivirus software installed – True if McAfee software is installed on the computer.
- **NAI**: Norton antivirus software installed – True if Norton software is installed on the computer.
- **VP**: virus present – True if a virus is actually present on the computer.

a. **[5 Points]** Draw the directed graph structure of the network. Imagine that the causal effects of these dependencies are sometimes noisy, e.g., sometimes virus software fails to detect the presence of a virus. Describe a parameterization of these noisy effects and write the resulting conditional probability tables for each node in the graph.

b. **[3 Points]** Either discuss an edge that could plausibly have been added to your network (an independence assumption that you made), or discuss an edge that could plausibly be removed from your network (an independence assumption you chose not to make). Explain briefly why you made your decision.

c. **[3 Points]** Add new variables

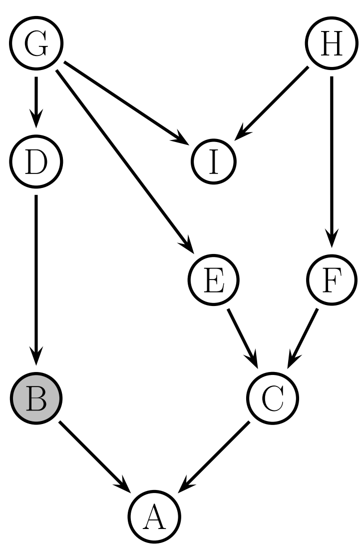
- **VRM**: virus reported by McAfee – True if a virus is reported by McAfeee antivirus software.
- **VRN**: virus reported by Norton – True if a virus is reported by Norton antivirus software.

in addition to the attributes above. Draw the graph structure for the Bayesian network with these new variables. Write conditional probability tables for VRM and VRN, and for any nodes whose parents are different from part (a).

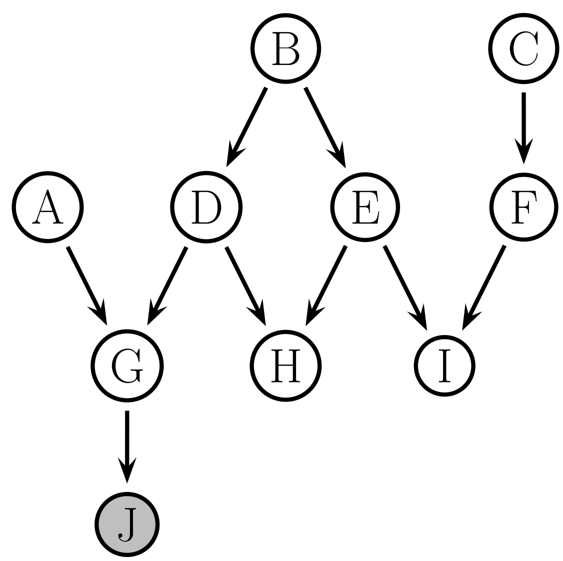
d. **[3 Points]** Briefly discuss whether including nodes VRM and VRN helps or hinders the modeling process.

Problem 2

[24 Points] Consider the following two directed graphical models:



(a)



(b)

a. **[6 Points]** Conditional independence

- **[3 Points]** For graph (a), list all variables which are independent of A, given knowledge of B.
- **[3 Points]** For graph (b), list all variables which are independent of A, given knowledge of J.

b. **[6 Points]** Factored distributions

- **[3 Points]** For graph (a), write the joint distribution $p(A, B, C, D, E, F, G, H, I)$ as a product of factors implied by the directed graph.
- **[3 Points]** For graph (b), write the joint distribution $p(A, B, C, D, E, F, G, H, I, J)$ as a product of factors implied by the directed graph.

c. **[6 Points]** Factor graphs

- **[3 Points]** Using the factorization from the previous problem, draw a factor graph representation of (a).
- **[3 Points]** Using the factorization from the previous problem, draw a factor graph representation of (b).

d. **[6 Points]** Undirected graphs

- **[3 Points]** Using the factorization from the previous problem, draw an undirected graphical model representation of (a).
- **[3 Points]** Using the factorization from the previous problem, draw an undirected graphical model representation of (b).

Problem 3

[25 Points] Consider the mixture of Gaussian distributions given by the following probability density function:

$$f(x) = 0.2 \cdot \mathcal{N}(x \mid \mu = 1, \sigma^2 = 25) + 0.3 \cdot \mathcal{N}(x \mid \mu = -2, \sigma^2 = 1) + 0.5 \cdot \mathcal{N}(x \mid \mu = 3, \sigma^2 = 4)$$

a. **[2 Points]** Plot this density function.

b. **[3 Points]** Since we can easily generate data from discrete distributions and from Gaussian distributions, we can generate data directly from this distribution. Describe how you would do this and produce a histogram of 500 samples drawn in this way.

c. **[10 Points]** Pretend that we don't know how to perform the "direct" sampling above. Implement a rejection sampler for this density. What was your upper bounding function? Plot it against the density function. Draw 500 samples and make a histogram. How many rejections did you get before you had 500 acceptances?

d. **[10 Points]** Continue to pretend that we don't know how to perform "direct" sampling. Instead of a rejection sampler, implement Metropolis–Hastings using a simple Gaussian proposal. Draw 500 samples and make a histogram. What was your acceptance rate? What was the variance of your proposal distribution? Try several different variances and describe how the results varied.