

Assignment 3: Graphical Models and Bayesian Treatment of Probabilistic Models

Probabilistic and Unsupervised Learning

Maneesh Sahani

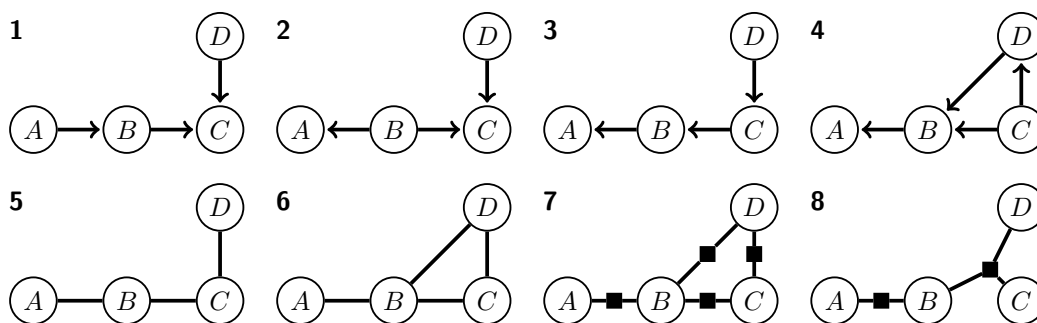
Due: Mon Nov 19, 2018

Note: Assignments are due at 11:00 AM (the start of lecture) on the day specified. The usual College late assignments policy will apply.

You must make a credible attempt at the main questions to receive credit for the bonus ones.

1. [25 marks] Conditional independencies and expressiveness of graphical models.

Consider the following graphical models:



- For graphs 2, 4, 6 and 8, write down **all** the conditional independence relationships that hold for variable C of the form $C \perp\!\!\!\perp \mathcal{X} | \mathcal{V}$, where \mathcal{X} and \mathcal{V} are sets of other variables. [15 marks]
- Two graphs are **equivalent** if they express *all* the same marginal and conditional independence relationships between their variables. A graph G is **subsumed** by graph H if all conditional independence relationships in H are exhibited in G . Divide the above 8 graphs into the smallest number of non-overlapping sets of equivalent graphs, and state which of these sets of equivalent graphs are subsumed by one of other sets. [10 marks]

2. [40 marks] **Constructing directed graphs and junction trees.** You are the doctor on the Starship Enterprise and you are attempting to use Bayesian methods to help your diagnosis abilities. You would like to represent your knowledge about the following seven binary random variables describing the state of your patients on any given visit

M = has the disease microsoftus
L = has the disease linuxitis
A = has the disease applois
V = is a vulcan (V=0 means "is a human")
H = has high temperature
P = likes pizza
B = has blue spots on face

You would like to build a directed graphical model which captures the following background knowledge:

Microsoftus is a rare disease.
Linuxitis and applois are very rare diseases.
There are about four times as many humans as vulcans on the ship.
Vulcans have higher probability of getting microsoftus than humans.
Most vulcans like pizza, some humans like pizza.
Microsoftus usually causes high temperature and blue spots on the face.
Linuxitis always causes high temperature.
Applois sometimes causes blue spots on the face.
A recent study suggests that excess pizza consumption increases risk of linuxitis.

- (a) Draw a directed graphical model representing the relationships between the above variables. Your graph should retain as much independence between the variables as possible, given the stated background knowledge. [5 marks]
- (b) For each variable in your graph, define a conditional probability table for that variable given the settings of its parents. Use the above background knowledge and convert those statements into probability tables which you think reasonably represent them. You will have to translate terms like "rare", "most", and "usually" into numerical values. [10 marks]
- (c) Construct an efficient junction tree (i.e., with minimally sized cliques) for your directed graph, drawing out the intermediate factor graph, undirected graph and chordal graph. Use the minimum deficiency search variable elimination order, and show the clique factors on the resulting junction tree in terms of the conditional probabilities. [10 marks]

You are told that a patient has blue spots on the face and a high temperature. As you rush to the ward you wonder whether you are about to encounter a career-defining case of applois.

- (d) Explain how to use Shafer-Shenoy propagation on the junction tree to compute the probability
 $P(\text{patient has Applois} \mid \text{patient has blue spots on face and high temperature})$.
Write out the expression to compute each of the messages that you need. [5 marks]
- (e) Do you think the probability is higher or lower than each of

$P(\text{patient has Apploisis})$

and

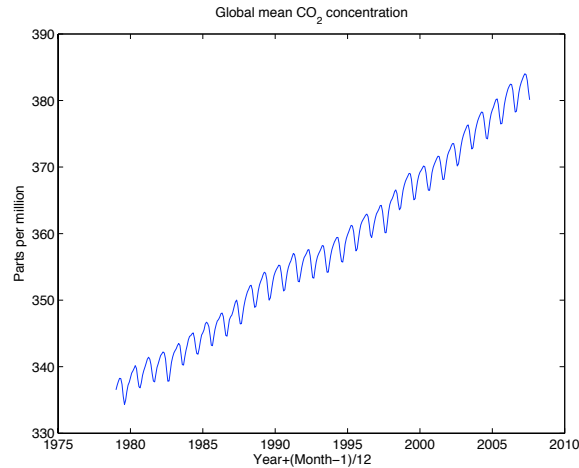
$P(\text{patient has Apploisis} \mid \text{patient has blue spots on face})?$

Explain. [5 marks]

You have your doubts about the recent report of pizza consumption causing linuxitis. You think it more likely that linuxitis induces a craving for pizza.

- (f) Draw the graphical model that would apply if your supposition were correct. Comparing this graph to one you drew previously, give at least two (conditional) independence relationships which would differ under the two hypotheses. [5 marks]
- (g) [**Bonus**] You gain access to the database used in the study. For each subject this records: the species, pizza consumption, and whether they had a high temperature and/or blue spots on their face. Explain how you would use these data and Bayesian model selection to compare your hypothesis to the one advanced by the original study. In particular, write down all the integrals you would need to compute in as much detail as you can. Do you think these data would be adequate to distinguish between the hypotheses? As always, explain. [10 bonus marks]

3. [35 marks] **Bayesian linear and Gaussian process regression.** The following time series of monthly mean global CO₂ concentrations can be obtained from the file [co2.txt](#) (original data obtained from <http://www.esrl.noaa.gov/gmd/ccgg/trends/>):



We will apply Bayesian linear and Gaussian process regression to predict the CO₂ concentration $f(t)$ as a function of time t , where $t = \text{Year} + (\text{Month} - 1)/12$.

- (a) First we model the function using linear regression, that is, using the functional form

$$f(t) = at + b + \epsilon(t),$$

with i.i.d. noise residual $\epsilon(t) \sim \mathcal{N}(0, 1)$ and prior $a \sim \mathcal{N}(0, 10^2)$, $b \sim \mathcal{N}(360, 100^2)$. Compute (using MATLAB or another package) the posterior mean and covariance over a and b given the CO₂ data. [10 marks]

- (b) Let $a_{\text{MAP}}, b_{\text{MAP}}$ be the MAP estimate in the question above. The residual is the difference between the observed function values and the predicted mean function values

$$g_{\text{obs}}(t) = f_{\text{obs}}(t) - (a_{\text{MAP}}t + b_{\text{MAP}}),$$

where $f_{\text{obs}}(t)$ is the observed value of the CO₂ concentration at time t .

Plot $g_{\text{obs}}(t)$. Do you think these residuals conform to our prior over $\epsilon(t)$? State, with justifications, which characteristics of the residual you think do or do not conform to our prior belief. [5 marks]

- (c) Write a MATLAB (or other language) function to generate samples drawn from a GP. Specifically, given a covariance kernel function $k(\cdot, \cdot)$ and a vector of input points \mathbf{x} , return a function $f(\mathbf{x})$ evaluated on the input points \mathbf{x} drawn randomly from a GP with the given covariance kernel and with zero mean. [10 marks]
- (d) Test your function by plotting sample functions drawn from the following kernel, for various settings of the hyperparameters

$$k(s, t) = \theta^2 \left(\exp \left(-\frac{2 \sin^2(\pi(s-t)/\tau)}{\sigma^2} \right) + \phi^2 \exp \left(-\frac{(s-t)^2}{2\eta^2} \right) \right) + \zeta^2 \delta_{s=t} \quad (1)$$

Describe the characteristics of the drawn functions, and how the characteristics of the functions depend on the parameters. [5 marks]

- (e) Suppose we were to consider modelling the residual function $g(t)$ using a zero mean GP with the covariance kernel above. Based on the plot of $g(t)$ and your explorations in the preceding part, what do you think will be suitable values for the hyperparameters of k ? [5 marks]
- (f) **[Bonus]** Extrapolate the CO₂ concentration levels to 2020 using the GP with covariance kernel k of eqn 1, and your chosen parameter values. Specifically, compute the predictive mean and variance of the residual $g(t)$ for every month between September 2007 and December 2020 given the observed residuals $g_{\text{obs}}(t)$. Plot the means and one standard deviation error bars of the extrapolated CO₂ concentration levels

$$f(t) = a_{\text{MAP}}t + b_{\text{MAP}} + g(t)$$

along with the observed CO₂ levels. Does the behaviour of the extrapolation conform to your expectations? How sensitive are your conclusions to settings of the kernel hyperparameters? [15 bonus marks]

- (g) **[Bonus]** Why is the above procedure not fully Bayesian? How would we go about modelling $f(t)$ in a Bayesian framework? [5 bonus marks]