

Assignment 3: Graphical Models and Bayesian Treatment of Probabilistic Models

Probabilistic and Unsupervised Learning

Maneesh Sahani

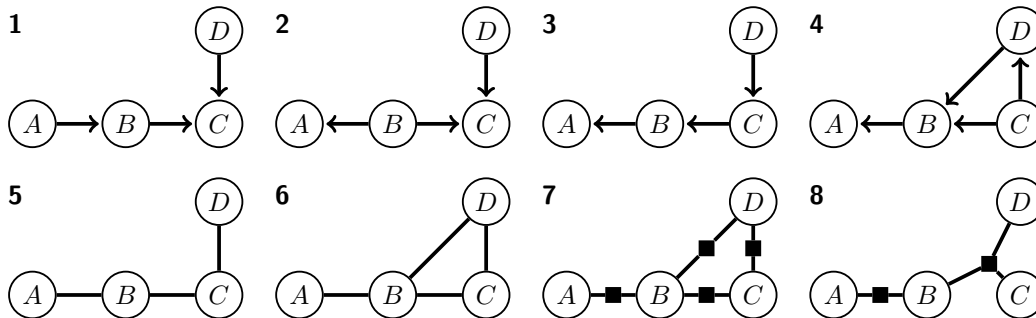
Due: Mon Nov 16, 2015

Note: all assignments for this course are to be handed in to the Gatsby Unit, **not** to the CS department. Please give all assignments to the lecturer or TAs at **the beginning of** lecture on or before the due date. Alternatively, you may bring coursework to Mr. Barry Fong at the Gatsby Unit reception.

Please attempt the first questions before 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 for variable C of the form $C \perp\!\!\!\perp \mathbf{X} | \mathbf{Y}$, where \mathbf{X} and \mathbf{Y} can be 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 Star Trek 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:

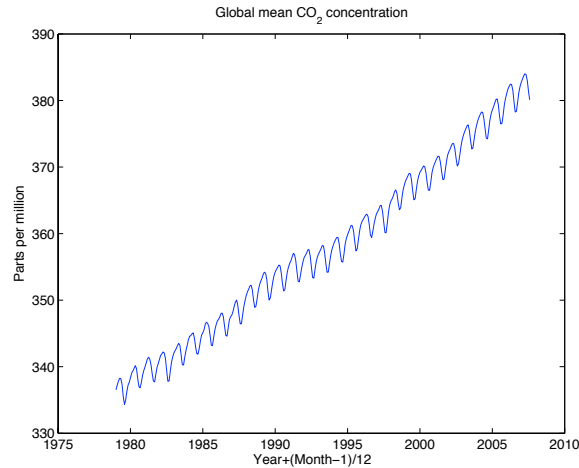
Microsoftus is a rare disease.
Linuxitis and applois are very rare diseases.
There are about four times as many humans as vulcans on the ship.
Vulcans have higher probability of getting microsoftus than humans.
Most vulcans like pizza, some humans like pizza.
Microsoftus usually causes high temperature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

- Draw a directed graphical model representing the relationships between the above variables. If you need to make any additional assumptions to draw your graph, state clearly what they are. [5 marks]
- 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 make up numbers for what terms like “rare”, “most”, and “usually” mean. [10 marks]
- Construct a junction tree 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. [10 marks]
- Using Shafer-Shenoy propagation on the junction tree, compute the probability

$P(\text{patient is a vulcan} \mid \text{patient has blue spots and high temperature})$

Show each message computed. You may compute the messages by hand, or using MATLAB or other package as you wish. Does this probability match your intuitions? [15 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 Bayesian? How would we go about modelling $f(t)$ in a Bayesian framework? [5 bonus marks]