C19 : UNSUPERVISED MACHINE LEARNING : HILARY 2013-2014

FRANK WOOD

Questions :

(1) For the Bayesian linear regression model on page 13 of the lecture notes on graphical models (Lecture 1) show

(a) how to perform block MH within Gibbs on $\mathbf{w}$ and $\hat{t}$.
(b) how to perform Gibbs on $\mathbf{w}$ and $\hat{t}$.
(c) the analytic form of the posterior predictive.

(2) Compute the probability that it is cloudy given that we observe that the grass is wet using the Bayes net in Figure 1.

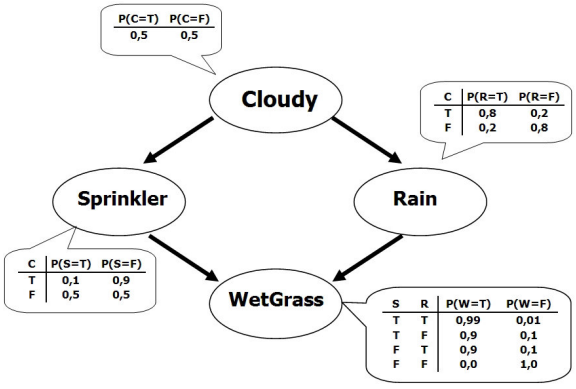

FIGURE 1. Bayes Net

(a) By enumeration and conditioning
(b) Using ancestral sampling and rejection
(c) Using Gibbs sampling

(3) In the preceding question, what is the most efficient way of computing the quantity of interest? *hint : it was not covered in the lectures*

(4) Show that the Gamma distribution is conjugate to the Poisson distribution.

(5) Show that the Gibbs transition operator satisfies the detailed balance equations and can be interpreted as an MH transition operator that always accepts.

(6) Implement a sampler for the LDA model on a corpus consisting of abstracts from NIPS papers over the last several years. You may wish to start from the support code provided on `http://www.robots.ox.ac.uk/~fwood/teaching/C19_hilary_2013_2014/`. Use the output to answer questions about the NIPS abstract corpus. The data for this model are in bagofwords_nips.mat, words_nips.mat, and title_nips.mat (also available on the website). When you load these files in Matlab the variables DS, WS, WO, and $titles$ will appear in your workspace. The variable WS is the entire corpus vectorized into a long row of words encoded as a row vector of integers. Each integer represents a word, WO is the dictionary; to look up the word corresponding to an integer use $WO\{i\}$. The variable DS is a row vector of the same length as WS that indicates from which document each word comes. If the document number is j then the title of that document can be found using $titles\{j\}$. Train LDA with 20 topics using an MCMC sampler in the collapsed representation (i.e. with all β_k's and θ_d's integrated out analytically) and use the single sample with the highest joint log likelihood to answer the following questions :

(a) What are the top ten most probable words for each of the 20 topics?
(b) One can consider the distribution over topics as a low dimensional representation of a document. We can use the dot product between topic distributions for two documents as a similarity metric. What are the ten most similar documents to the first document, "Connectivity Versus Entropy"?