HOMEWORK 5 471

HW5 is due by Wednesday November 18th at 11:59 pm CT pm. Please upload your solutions to Canvas.

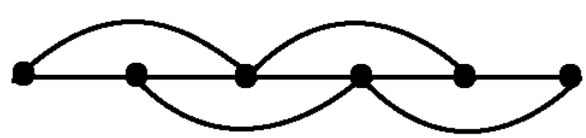
Attach the codes/command windows that you use to answer the questions.

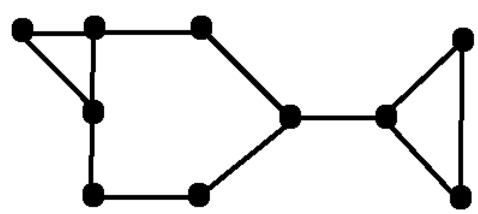
(1) Suppose that the set of variables $\{X_v\}_{v\in V}$ respect the graph $G = (V, E)$. That is, we assume that

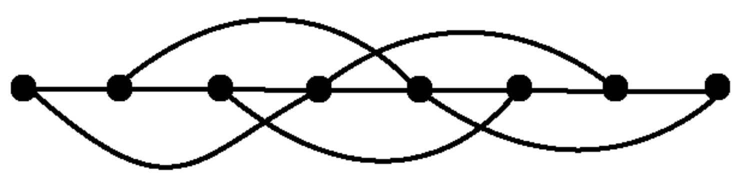
$$p_V(x_V) = \frac{1}{Z} \prod_{C\in\mathcal{C}(G)} \phi_C(x_C).$$

Assume also that each of the variables X_v may take values in $\{1, \dots, 10\}$.

- For each of the following graphs, find the best ordering of the variables, so that the computational complexity of obtaining the most likely configuration for p, is as low as possible (according to the theory).







- Would your answers to the previous questions change if we wanted to compute the normalization constant Z?

(2) (Most likely configuration for a hidden Markov model HMM) For this problem we will consider a random vector $(X_1, Y_1, X_2, Y_2, \dots, X_n, Y_n)$. We use the convention $X = (X_1, \dots, X_n)$, $x = (x_1, \dots, x_n)$, $Y = (Y_1, \dots, Y_n)$ and $y = (y_1, \dots, y_n)$. We assume that each X_i can only take one of the values 0 or 1 and that each Y_i can take values in $\mathbb{R}$ (the Y_i are continuous random variables).

Suppose that the joint distribution of X, Y is given by:

$$p(x,y) := p_1(x_1) \prod_{k=1}^{n-1} a(x_{k+1}, x_k) \prod_{l=1}^{n} f_l(y_l|x_l),$$

where

$$p_1(x_1) = 1/2, \quad \forall x_1 \in \{0,1\},$$
$$f_l(y_l|x_l) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-(x_l - y_l)^2/2\sigma^2), \quad \forall l \in \{1, \ldots, n\},$$
$$a(x_{k+1}, x_k) = q\mathbb{1}_{x_k = x_{k+1}} + (1-q)\mathbb{1}_{x_k \neq x_{k+1}}.$$

Download the dataset with $y_1, \ldots, y_{1000}$, which is in the file `yvalues` (saved using different formats for your convenience) and posted along with the HW assignment. It contains an observation of the y variable from the above model with $n = 1000$, $q = 0.95$ and $\sigma = 1$.

- What is the right interpretation for the function $p(x, y)$? Notice that it is not a p.m.f. because the Y_i variables are continuous!
- What graph is respected by the variables (X, Y)?
- For each y_i, compute the most likely X_i, that is,

  $$x_i^* := argmax_{x_i \in \{0,1\}} \mathbb{P}(X_i = x_i | Y_i = y_i)$$

  This does not require dynamic programming! Plot x_i^* versus i and y_i versus i on the same graph. Make sure the two are clearly distinguishable.
- Now compute the most probable sequence X given the sequence y, that is,

  $$x^{**} := argmax_{x \in \{0,1\}^n} \mathbb{P}(X = x | Y = y)$$

  This does require dynamic programming. Plot x_i^{**} versus i and y_i versus i on the same graph. Make sure the two are clearly distinuishable.

  Note: You may find it useful to maximize $\log p(x, y)$ instead of $p(x, y)$, otherwise you will run into numerical underflow/overflow problems. In the dynamic programming algorithm discussed in class, simply replace products with sums of logarithms. Take a look at the notes DP in a nutshel on Canvas for more details.

(3) (Gibbs sampling for HMMs.) Consider the HMM described in problem 2.

- Find a simple expression for the conditional distribution of X_i given $Y = y$ and $X_{V \setminus \{i\}} = x_{V \setminus \{i\}}$. You may need to handle the cases $i = 1$ and $i = n$ separately.
- Sketch a Gibbs sampling algorithm for sampling from the conditional distribution of X given $Y = y$. Update a single X_i at each step, and choose uniformly among the n different X_i. (This means we are using a randomized version of the Gibbs sampler.)
- Implement your algorithm from the previous step and generate a random sample from the conditional distribution of X given that $Y = y^*$ (where y^* is the observed data vector from problem 2). Discuss how long you let the algorithm run and why. Generate a few different samples (by continuing to let the algorithm run).
- Use a long Gibbs sampling run to approximate

  $$a_{y^*} := \mathbb{E}\Big[\exp\Big((\textstyle\sum_i X_i)/10 - 25\Big)\Big| Y = y^*\Big]$$

and compare your answer to $\exp\big((\sum_i x_i^{**})/10 - 25\big)$ using x^{**} from problem 2. Interpret your results.

Make a plot of your estimate of the conditional expectation versus how long the Gibbs sampler has run. Repeat this with a few different methods of initialization (you do not need to show the plots) and comment on whether Gibbs sampling in this case appears to be quickly forgetting the initialization.

(4) (Ising model) Consider the Ising mdoel with 50×50 electrons arranged on a square grid (as discussed in class and as suggested in the lecture notes in section 2.5). In class we discussed that the statistical equilibrium for a canonical ensemble for such system when the temperature is fixed to be equal to $T > 0$, is the distribution with pmf

$$p_T(x) = \frac{1}{Z_T} \exp\left(-\frac{1}{T} E(x)\right), \quad x \in \{0,1\}^{50\times 50},$$

where the energy $E : \{0,1\}^{50\times 50} \to \mathbb{R}$ is defined as

$$E(x) := -\sum_{u \sim v} x_u x_v;$$

In the above the sum ranges over all edges of the lattice (i.e. over all pairs of neighbours in the lattice).

- (MCMC) Use the Gibbs sampler to simulate samples from p_T.
  Note: The question is open ended, and the idea is that you conduct a serious numerical experiment for this problem. By this I mean that you should try different values of temperature, explore different initializations for the sampler, experiment how long the Gibbs samplers have to run for, etc etc. Discuss your findings and in particular describe the qualitative differences between samples for different values of temperature T (for example, for what values of T do you observe pattern formation?).
- (Exact computing Vs MCMC) Explain why DP is an unfeasible approach to simulate samples from p_T.