CS688: Graphical Models - Spring 2020

Assignment 4

Assigned: Wed, Apr 8th. Due: Mon, Apr 20th, 11:59am.

General Instructions: Submit a report with the answers to each question at the start of class on the date the assignment is due. You are encouraged to *typeset your solutions*. For your assignment to be considered "on time", you must upload a zip file containing all of your code to Gradescope by the due date. Make sure the code is sufficiently well documented that it's easy to tell what it's doing. You may use any programming language you like. For this assignment, you may **not** use existing code libraries for inference with CRFs or MRFs, or any code library for sampling. If you think you've found a bug with the data or an error in any of the assignment materials, please post a question to the Piazza discussion forum. Make sure to list in your report any outside references you consulted (books, articles, web pages, etc.) and any students you collaborated with.

Academic Honesty Statement: Copying solutions from external sources (books, web pages, etc.) or other students is considered cheating. Sharing your solutions with other students is also considered cheating. Any detected cheating will result in a grade of -100% on the assignment for all students involved, and potentially a grade of F in the course.

Introduction: In this part, you will experiment with Monte Carlo Inference. More specifically, you will experiment with Monte Carlo image denoising using grid-structured conditional random field models.

Binary Model: Consider a probability distribution over a vector $y \in \{-1, +1\}^d$ model of the form

$$p(y|b, w) = \frac{1}{Z} \prod_{i=1}^{d} \exp(b_i y_i) \prod_{(i,j)\in\text{pairs}} \exp(w_{ij} y_i y_j).$$

Here, each component y_i of y is in $\{-1, +1\}$, we can think of the "bias" b_i as "how much y_i wants to be $+1$", and we can think of the weights w_{ij} as "how much y_i and y_j want to have the same value".

Question 1. (*5 points*) **Derivation of conditional distribution** Mathematically derive a formula for $p(y_i = 1|y_{-i}, b, w)$, the conditional probability that $y_i = 1$ given the state of all other variables. Write your solution using the function $\sigma(s) := 1/(1 + \exp(-s))$. You will need to reference the set of all neighbors of i, $\text{nb}(i)$.

Question 2. (*5 points*) **Pseudo-Code for Gibbs Sampling** Write pseudo-code for the Gibbs sampling algorithm, where in each iteration, you update for $i = 1, i = 2, ..., i = d$. (This is slightly different from in class where each iteration used a single randomly chosen i.) You should take as input (1) the total number of iterations, and (2) the model parameters $\{b, w\}$. You should return a list of the samples of y that you get after each iteration.

Question 3. (*5 points*) **Samples** Consider an 30×30 4-connected grid structured model where we have a zero bias of $b_i = 0$, and a constant interaction strength of $w_{ij} = \bar{w}$. Now, implement the Gibbs sampling algorithm. For each value of $\bar{w} \in \{0, .1, .2, .3, .4, .5\}$, run a single chain Markov chain for 100 iterations. (Where, again, one iteration is a full pass over the entire grid.) Plot the final samples for each $\bar{w}$ value. Show each sample as an image with black pixels for variables with $y_i = -1$ and white pixels for variables with for $y_i = +1$. Make sure to label each image with the value of $\bar{w}$ it corresponds to.

Question 4. (*5 points*) **Discussion** Explain briefly (no more than a single paragraph of 4 sentances) why the images from the previous question look like they do, and why they change for various values of $\bar{w}$.

Question 5. (*5 points*) **Mixing Times** For the model in the previous problem, it is easy to see by symmetry that the true mean value of y is $\mathbb{E}[y] = 0$. To get an idea of average performance, for each value of $\bar{w} \in \{0, .1, .2, .3, .4, .5\}$ run *100 independent* Markov chains each for 100 iterations. (Where, again, one iteration is a full pass over the entire grid). For each iteration, compute the mean value of y (averaged over the 100 independent repetitions). Plot the 6 curves together on one graph, with the number of iterations on the x-axis and the mean value of y on the y-axis. Make sure to label the axes and the value of $\bar{w}$ for each curve.

Question 6. (*5 points*) **Discussion** Do your samples give the correct value of $\mathbb{E}[y]$? When do the samples look better or worse? Explain in a single paragraph of no more than four sentences why the curves in the previous problem look as they do.

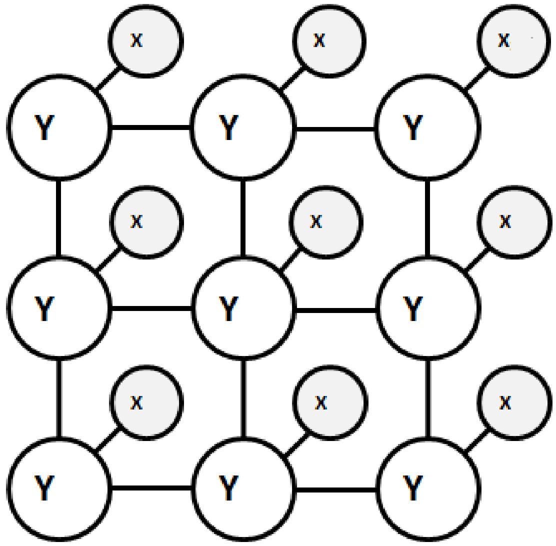
Data Set: The data for the remaining questions consists of a single 200×154 pixel binary image. The original image is `im_clean.png`, while the copy with noise is `im_noisy.png`. The two images are shown below. We will refer to the images on the left as the "true" image and the image with noise as the "observed" or "noisy" image.

(a) im_clean

(b) im_noisy

Image Denoising Model: Intuitively, given a "noisy image" x, we should think of this defining a probabilistic model over the "clean image" y, with the structure shown in the figure on the next page.
We can think of the model as being defined exactly as in the previous problem, except that $b(x, \theta)$ and $w(x, \theta)$ are now functions of the input x as well as some parameters θ

$$p(y|x, \theta) = \frac{1}{Z(x)} \prod_{i=1}^{d} \exp(b_i(x, \theta) y_i) \prod_{(i,j) \in \text{pairs}} \exp(w_{ij}(x, \theta) y_i y_j).$$



Question 7. (*10 points*) **Fixed Parameter Denoising** Make sure you scale the input x so that each value is either -1 or +1. Then, determine the parameters by the simple mapping of $b_i = .5x_i$ and $w_{ij} = 0.3$. Run 100 iterations of Gibbs sampling. Show an image of μ, the mean value of y you obtain over those samples, and report the mean per-pixel absolute difference of μ from the true output. (Make sure to scale your image appropriately so that all values between -1 and $+1$ are shown, and make sure that both μ and y are in the original space between -1 and $+1$ before computing a difference.)

Question 8. (*10 points*) **Varying Parameters** Consider setting $b_i = \theta_1 x_i$ and $w_{ij} = \theta_2$ where θ_1 and θ_2 are 2 unknown parameters. (In the previous problem we essentially used $\theta_1 = .5$ and $\theta_2 = 0.3$) Can you find a pair of θ_1 and θ_2 to give you lower error than the values from the previous problem? Limit the total number of iterations you do for any parameter setting to 100 iterations. Describe (a) how you tried to find better parameters (there is no single correct answer here) (b) show an image of your final denoised image and (c) report the final error.

Extra Credit Finally, here are some extra-credit problems. These are *far* more difficult than the above problems and have very small point values. *These problems are a terrible "value" in terms of the points you will earn for spending your time*. To maximize your score with limited time, you should make sure the above problems are done thoroughly and ignore these. We will be very stingy in giving credit for these problems- do them only for the glory, and at your own risk! These problems are more open-ended. As a result, you will need to carefully describe what you did for each problem.

Question 9. (*5 points*) **MCMC: Varying w_{ij} parameters** Come up with a scheme to have w_{ij} vary as a function of x, rather than being constant. (1) Describe the informal motivation for the scheme (2) describe formally what you have done (3) Give an image of the final results (4) list your final error, and how much this improves on the previous problem.

Question 10. (*5 points*) **MCMC: Faster mixing** Above, we used a fixed update order of $s = 1, 2, ..., d$. Can you come up with a different update order that leads to faster mixing? (Meaning that you get closer to the true mean value of 0 during the 100 iterations. (1) Describe what your improved update order is and (2) make another version of the same graph from Question 5, showing your improvement. (3) Discuss why you think this mixes faster.