Probabilistic graphical models - Homework 2

December 11, 2020

The deadline for handing out this homework is January 15th (at midnight).

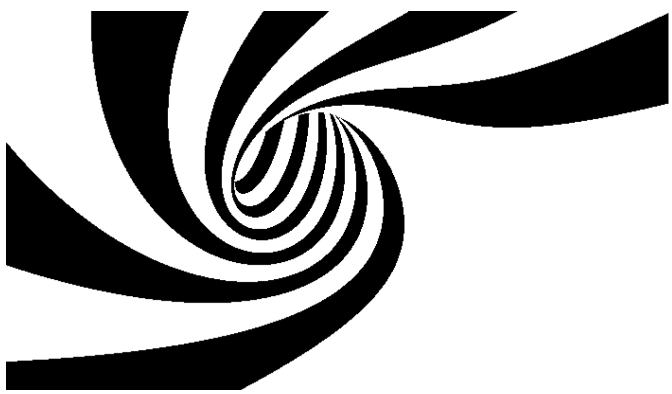
1 Noisy Ising

Consider a simple version of the Ising model:

$$p(x) = \frac{1}{Z_{\alpha,\beta}} \exp\left\{ \alpha \sum_{i=1}^{n} x_i + \beta \sum_{(i,j)\in E} \mathbf{1}(x_i = x_j) \right\}$$

where the x_i are either 0 or 1. Recall that $(i, j) \in E$ means that there is an edge between nodes i and j, and that $\mathbf{1}(x_i = x_j)$ is one if x_i and x_j have the same "colour", zero otherwise.

We do not observe x directly. Instead, we observe independent variables y_i which, conditional on $x_i = l$ are distributed according to a Gaussian distribution $N(\mu_l, 1)$. This type of model is often used for grayscale images, like this one:



(You can use another, more interesting gray-scale image, but don't forget to add some level of noise!)

We first treat the parameters α, β, μ_0, μ_1 as fixed.

1. Explain the trick we may use to adapt belief propagation to the problem of computing the distribution of a given x_i, conditional on all the y_i's. Is this algorithm exact? Explain. Implement the algorithm.

2. Explain how you may instead use MCMC to sample from the same conditional distribution. Compute any distribution you need to sample from, and explain why these distributions are easy to sample from. Implement the corresponding MCMC algorithm.

   We now wish to learn the parameters, based on data $y = (y_1, \ldots, y_n)$. For simplicity, we assume first α and β are fixed, and we want to learn μ_0 and μ_1.

3. Derive the update equation of the EM algorithm for this model. How could you use Question 2 to actually implement such a EM algorithm? Implement the corresponding EM algorithm.

4. Now assume that α and β are also unknown. What problem arises in this case, if we try to implement the EM algorithm?

5. Consider the following prior distribution for $(\alpha, \beta, \mu_0, \mu_1)$:

$$\pi(\alpha, \beta, \mu_0, \mu_1) \propto \varphi(\mu_0; m, s^2)\varphi(\mu_1; m, s^2) Z_{\alpha,\beta} \mathbf{1}_{[0,a]}(\alpha)\mathbf{1}_{[0,b]}(\beta)$$

   where $\varphi(\mu, m, s^2)$ is the density of a $N(m, s^2)$ distribution.

   Extend the MCMC algorithm of question 2 to sample from the joint distribution of $(\alpha, \beta, \mu_0, \mu_1, x)$, given the data, and implement it. Why do we choose this particular prior for (α, β)?

   Note: you may try different values for m, s^2, a, and b and see "what works".