Chapter 24

stevenjin8

July 18, 2021

Exercises

Exercise 1

$$p(x_1|x_2) = \mathcal{N}\left(x_1 \middle| \frac{3}{2} - \frac{1}{2}x_2, \frac{3}{4}\right)$$

The formula for $p(x_2|x_1)$ is the same but with the indices switched.

Exercise 2

Given the structure of this model, we know that the posterior of z_i is conditionally independent of $\mathbf{z}_{-i}$ given the cluster parameters. Thus, we can first sample z_i with

$$p(z_i = k|x_i, \boldsymbol{\theta}) \propto \mathcal{N}(x_i|\mu_k m, \sigma_k^2)\text{Cat}(z_i = k|\boldsymbol{\pi}).$$

Next, we sample the cluster parameters using the semi-conjugate posterior $p(\mu_k, \sigma_k^2) = p(\mu_k)p(\sigma_k^2) = \mathcal{N}(\mu_k|m_0, v_0^2)\text{IG}(\sigma_k^2|a_0, b_0)$. It doesn't matter if we sample μ_k or σ_k^2 first. We sample μ_k from

$$\mu_k \sim \mathcal{N}(m_{N_k}, v_{N_k}^2)$$

where $m_{N_k}, v_{N_k}^2$ is as defined in section 4.6.1, but only using the data points where $z_i = k$. Its the same process for σ_k^2, but with

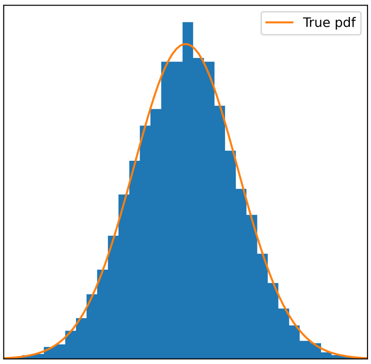
$$\sigma_k^2 \sim \text{IG}(a_{N_k}, b_{N_k})$$

where a_{N_k} and b_{N_k} are given in section 4.6.2.2.

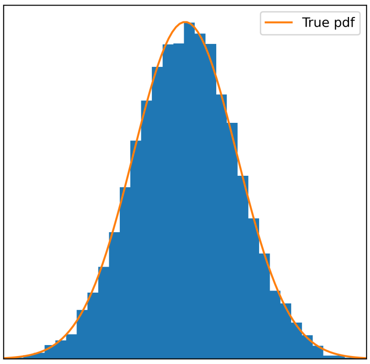
Exercise 3

I don't know any Matlab so I'm not going to try to modify the code. Gibbs sampling a Potts model is quite straightforward. Assuming that $w_{st} = J$ (see equation 19.22), we have

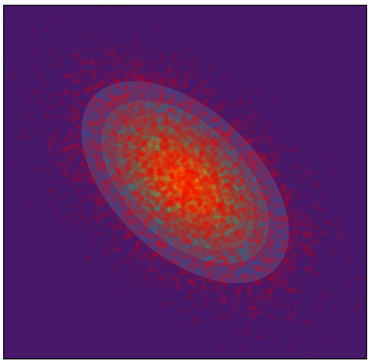
$$p(x_s = k|\mathbf{x}_{-s}) \propto e^{JN_{sk}} \tag{1}$$



(a) Empirical vs true marginal for x_1.



(b) Empirical vs true marginal for x_2.



(c) Empirical vs true marginal for x_1, x_2.

Figure 1: Gibbs sampling a 2-D Gaussian.

where N_{sk} is the number of neighbours of x_s with value k. To perform annealed sampling we first draw k' from a proposed distribution (such as equation 1 or a uniform distribution), and then we set α to

$$\alpha = \exp\left(\frac{JN_{sk} - JN_{sk'}}{T}\right)$$

Where k is the current value of x_s. We then set $x_s = k'$ with a probability of $\min(\alpha, 1)$. We can rewrite the equation for α as

$$\alpha = \exp\left(\frac{J}{T}(N_{sk} - N_{sk'})\right).$$

Thus we can see that increasing the temperature is equivalent to decreasing J, the coupling strength.

Exercise 5

Recall that a Student distribution can be seen as an infinite mixture of Gaussians $\mathcal{N}(x|\mu, \sigma^2/z)$ where z is unknown but follows a Gamma distribution $\mathrm{Ga}(\frac{v}{2}, \frac{v}{2})$. We can first sample z_i with equation 11.69 but using

$$\delta_i = \left(\frac{y_i - \mathbf{w}^T\phi(\mathbf{x}_i)}{\sigma}\right)^2$$

Now we have reduce the problem to weighted linear regression, here the weight for the ith data point is $1/z_i$. We can apply the results of section 7.6.1, but replacing σ^{-2} with $\sigma^{-2}\mathrm{diag}(\mathbf{z})$.

For σ^2, we can use the results of section 4.6.2.2, replacing equation 4.189 with

$$b_N = b_0 + \frac{1}{2}\sum_i^N (y_i - \mathbf{w}^T\phi(\mathbf{x}_i))^2 z_i.$$

I was on the fence about doing this question, but it turned out to be pretty cool as the result used results from four chapters (4, 7, 11, and 24) and gave me a lot of confidence on my knowledge retention. Of course, I had to look up all the results from previous chapters, but at least I know where to look.

Exercise 6

Recall that Probit regression states that the difference in utilities between two options is Gaussian with $\sigma^2 = 1$ and $\mu_i = \mathbf{w}^T\mathbf{x}_i$. We can sample z_i by drawing values $\mathcal{N}(\mathbf{w}^T\mathbf{x}_i, 1)$ until we draw a value with the same sign as y_i (there are probably better ways of doing this, but I want to go to bed soon). Sampling $\mathbf{w}$ is just applying the results of 7.6.1, but replacing y_i with z_i.