Bayesian Theory and Computation, Problem Set 1

Mar 2, 2021

Yutao He 1700012189

Code Document

Problem 2.

HMC:

$$
\begin{aligned}
&U(\theta, r) = \sum_{i=1}^{n} \frac{(y_i - \theta_1 - \theta_2^2)^2}{2\sigma_y^2} + \frac{\theta_1^2 + \theta_2^2}{2\sigma_\theta^2} \\
&K(r) = \frac{\theta_1^2 + \theta_2^2}{2} \\
&\frac{\partial U}{\partial \theta_1} = \sum_{i=1}^{n} \frac{\theta_1 + \theta_2^2 - y_i}{\sigma_y^2} + \frac{\theta_1}{\sigma_\theta^2} \\
&\frac{\partial U}{\partial \theta_2} = 2\theta_2 \sum_{i=1}^{n} \frac{\theta_1 + \theta_2^2 - y_i}{\sigma_y^2} + \frac{\theta_2}{\sigma_\theta^2} \\
&\frac{\partial K}{\partial r_1} = \theta_1, \ \frac{\partial K}{\partial r_2} = \theta_2
\end{aligned}
$$

The leap-frog function is exactly the same as the one in lecture note.

SGLD:

The step size $\epsilon_t = \frac{10^{-4}}{10^{-4}\sqrt{1+t}}$

The gradient calculations are similar to HMC, function $batch_grad_theta$ is to calculate the batch gradient of $p(x|\theta)$, and p_grad_theta is to calculate the gradient of prior distribution $p(\theta)$

SGHMC:

$M = I, \ C = 20, \epsilon_t = \frac{10^{-3}}{0.05\sqrt{1+t}}$, so that

$$
\begin{aligned}
\theta_i &= \theta_{i-1} + \epsilon_t M^{-1} r_{i-1} \\
r_i &= r_{i-1} + \epsilon_t g(\theta_i) - \epsilon_t C M^{-1} r_{i-1} + \mathcal{N}(0, 2C\epsilon_t)
\end{aligned}
$$

SGNHT:

Step size $\epsilon_t = \frac{10^{-3}}{0.05\sqrt{1+t}}$, $\xi_0 = A = 100$, SGNHT didn't have as desired performance despite its fast speed, probably resulting from the bad choice of parameters.

Problem 3.

E and M step are described in detail in hw3.pdf.