

Title: Supplementary material for computer practical class 2

Author: Georgios Karagiannis (georgios.karagiannis@durham.ac.uk) February 6, 2024

The goal's (roughly speaking...)

Consider a Bayesian statistical model

$$\begin{cases} z_i|w & \stackrel{\text{ind}}{\sim} f(z_{1:n}|w), \quad i = 1, \dots, n \\ w & \sim f(w) \end{cases} \quad (1)$$

Here: observables are $\{z_i \in \mathcal{Z}\}_{i=1}^n$ and the unknown parameter is $w \in \Theta$.

The likelihood of the observables $\{z_i \in \mathcal{Z}\}_{i=1}^n$ given the parameter $w \in \Theta$ as

$$L_n(w) := f(z_{1:n}|w) = \prod_{i=1}^n f(z_i|w)$$

the posterior distribution density

$$f(w|z_{1:n}) = \frac{L_n(w) f(w)}{\int L_n(w) f(w) dw} \quad (2)$$

Goal 1

Find mode of $f(w|z_{1:n})$, that is the Maximum a posteriori estimator (MAP) of w i.e.

$$\begin{aligned} w^* &= \arg \max_w \{\log(f(w|z_{1:n}))\} \\ &= \arg \max_{w \in \Theta} \left(\underbrace{\log(L_n(w))}_{\text{(I)}} + \underbrace{\log(f(w))}_{\text{(II)}} \right) \end{aligned}$$

Goal 2

Generate a random sample $\{w^{(t)}\}$ from (2), in particular from

$$f_\tau(w|z_{1:n}) \propto \exp \left(\frac{1}{\tau} \prod_{i=1}^n f(z_i|w) f(w) \right) \quad (3)$$

$$\propto \exp \left(\frac{1}{\tau} L_n(w) f(w) \right) \quad (4)$$

Stochastic Gradient Langevin Dynamics (SGLD)

Provide:

- A seed (initial value) $w^{(0)}$, m , τ
- Learning rate $\eta_t : \mathbb{N} \rightarrow \mathbb{R}_+$; a positive non-increasing function such as $\eta_t \rightarrow 0$. E.g.
 - $\eta_t = t_0/t$, with $t_0 > 0$ user specified
 - η_t = something relatively tiny
- Temperature $\tau > 0$

Stochastic Gradient Langevin Dynamics (SGLD) with learning rate $\eta_t > 0$, batch size m , and temperature $\tau > 0$ is

1. Generate a random set $\mathcal{J}^{(t)} \subseteq \{1, \dots, n\}^m$ of m indices from 1 to n with or without replacement, and set a $\tilde{\mathcal{S}}_m = \{z_i; i \in \mathcal{J}^{(t)}\}$.
2. Compute

$$w^{(t+1)} = w^{(t)} + \eta_t \left(\frac{n}{m} \sum_{i \in \mathcal{J}^{(t)}} \nabla \log f(z_i | w^{(t)}) + \nabla \log f(w^{(t)}) \right) + \sqrt{2\eta_t \tau} \epsilon_t \quad (5)$$

where $\epsilon_t \stackrel{\text{iid}}{\sim} \mathcal{N}(0, I)$.

3. Terminate if a termination criterion is satisfied; E.g., $t \geq T_{\max}$ for a prespecified $T_{\max} > 0$.