Title: Supplementary material for computer practical class 2

Author: Georgios Karagiannis (georgios.karagiannis@durham.ac.uk) February 4, 2024

The goal's (roughly speaking...)

Consider a Bayesian statistical model

$$\begin{cases} z_i|w & \overset{\text{ind}}{\sim} f\left(z_{1:n}|w\right), \ i=1,...,n \\ w & \sim f\left(w\right) \end{cases} \tag{1}$$

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

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

the posterior distribution density

$$f\left(w|z_{1:n}\right) = \frac{L_n\left(w\right) f\left(w\right)}{\int L_n\left(w\right) f\left(w\right) \mathrm{d}w} \tag{2}$$

Goal 1

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

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

Goal 2

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

$$\begin{aligned} f_\tau\left(w|z_{1:n}\right) &\propto \exp\left(\frac{1}{\tau}\prod_{i=1}^{n} f\left(z_i|w\right) f\left(w\right)\right) &(3) \\ &\propto \exp\left(\frac{1}{\tau} L_n\left(w\right) f\left(w\right)\right) &(4) \end{aligned}$$

Gradient descent algorithm

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

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

   where $\epsilon_t \overset{\text{iid}}{\sim} \mathrm{N}(0, I)$.
3. Terminate if a termination criterion is satisfied; E.g., $t \leq T_{\max}$ for a prespecified $T_{\max} > 0$.