Stochastic gradient Langevin dynamics

Apostolos Psaros

1 SGD-like updates

The Langevin MC method is an HMC method with $T = 1$; i.e., with a single leapfrog iteration. In fact, in the uncorrected version of Langevin MC (all candidates accepted) there is no need of the two-step procedure of HMC and no need for explicitly representing the momentum at all; one stochastic differential equation (Langevin equation) is used for picking new values of $\boldsymbol{w}$. For the t-th state, the Langevin equation is given as (see Eq. (5.28) in Neal, 1993)

$$\boldsymbol{w}^{(t+1)} - \boldsymbol{w}^{(t)} = -\frac{\lambda^2}{2}\nabla_{\boldsymbol{w}}\mathcal{L}(\boldsymbol{w}^{(t)}) + \lambda\boldsymbol{n} \tag{1}$$

where $\boldsymbol{n} = [n_1, \dots, n_k]^T$ and $n_j \sim \mathcal{N}(0,1)$ for all j. Also, recall that $\mathcal{L}(\boldsymbol{w}^{(t)}) = -\log p(\mathcal{D}|\boldsymbol{w}^{(t)}, \mathcal{H}) - \log p(\boldsymbol{w}^{(t)}|\mathcal{H})$. By writing $\lambda = \sqrt{\epsilon}$, Eq. (1) becomes

$$\boldsymbol{w}^{(t+1)} - \boldsymbol{w}^{(t)} = \frac{\epsilon}{2}\left[\nabla_{\boldsymbol{w}}\sum_{i=1}^{N}\log p(y_i|\boldsymbol{w}^{(t)}, x_i, \mathcal{H}) + \nabla_{\boldsymbol{w}}\log p(\boldsymbol{w}^{(t)}|\mathcal{H})\right] + \boldsymbol{\eta} \tag{2}$$

where $\boldsymbol{\eta} = [\eta_1, \dots, \eta_k]^T$ and $\eta_j \sim \mathcal{N}(0, \epsilon)$ for all j.

If at each iteration only a subset (mini-batch) of the available data is used, Eq. (2) becomes (Welling and Teh, 2011)

$$\boldsymbol{w}^{(t+1)} - \boldsymbol{w}^{(t)} = \frac{\epsilon}{2}\left[\nabla_{\boldsymbol{w}}\sum_{i\in S}\frac{N}{|S|}\log p(y_i|\boldsymbol{w}^{(t)}, x_i, \mathcal{H}) + \nabla_{\boldsymbol{w}}\log p(\boldsymbol{w}^{(t)}|\mathcal{H})\right] + \boldsymbol{\eta} \tag{3}$$

This is the same as the SGD update step with added Gaussian noise.

References

Neal, R. M. (1993). *Probabilistic Inference Using Markov Chain Monte Carlo Methods.* Department of Computer Science, University of Toronto Toronto, Ontario, Canada.

Welling, M. and Teh, Y. W. (2011). "Bayesian Learning via Stochastic Gradient Langevin Dynamics". *Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, pp. 681–688.