ML accelerated Monte-Carlo

Discussions

Author

Sakib Matin

August 15, 2022

'Positive' and 'Negative' phase of gradient descent

Energy based models are trained using the Maximum Likelihood framework. Quite simply, we want to find parameters which maximize the likelihood of training data,

$$L = \prod_i p(x_i; \theta) \tag{1}$$

where $\{x_i\}$ is the set of training data, and θ represents all the parameters of the model.

The convention is to work with the log-likelihood.

$$\mathcal{L}_\theta = \langle \log p_\theta(x) \rangle_{\text{data}} \tag{2}$$

For an Energy based model, the log-likelihood is

$$\mathcal{L}_\theta = -\langle E(x; \theta) \rangle_{\text{data}} - \log Z_\theta \tag{3}$$

the Z_θ is the partition function which tells us about ensemble ('model' in CS nomenclature) averages, and does not depend on the training data. The partition function is typically intractable, and we have to approximate it using Markov chain methods such as Gibbs sampling. The terminology introduced by Hinton's group; the first term, which explicitly depends on the training data, contributes to the *positive phase* of the training and the second term which only depends on the model parameters, contributes to the *negative phase*.

Let's consider the case of an RBM, where the energy is given by

$$E(\boldsymbol{v}, \boldsymbol{h}) = -\sum_i a_i(v_i) - \sum_\mu b_\mu(h_\mu) - \sum_{i\mu} W_{i\mu} v_i h_\mu \tag{4}$$

To update the train an RBM, we can use the stochastic gradient to update the weights and biases (hidden and visible layers.) We will just take a look how to update the the visible layer biases, $a_i \to a_i - \eta \partial_{a_i} \mathcal{L}$. We have to compute the gradient of the log-likelihood function is

$$\begin{aligned}
-\partial_{a_i} \mathcal{L} &= \partial_{a_i} \left[\langle E(\boldsymbol{v}, \boldsymbol{h}) \rangle_{\text{data}} + \log Z \right] &(5)\\
&= \langle \partial_{a_i} E(\boldsymbol{v}, \boldsymbol{h}) \rangle_{\text{data}} + \partial_{a_i} \log Z &(6)\\
&= -\langle v_i \rangle_{\text{data}} + \partial_{a_i} \log Z &(7)\\
&= -\langle v_i \rangle_{\text{data}} + \langle v_i \rangle_{\text{ensemble}} &(8)
\end{aligned}$$

In the final step, we have used the fact that the derivative of the free energy with respect to a parameter is simply the ensemble expectation of the conjugate observable. Here $\langle v_i \rangle_{\text{data}}$ is the *positive phase* and $\langle v_i \rangle_{\text{ensemble}}$ is the *negative phase* of the gradient descent.

We can follow the same approach for the hidden biases and the Weights. This is left as an exercise for the reader.

[1] Pankaj Mehta, Marin Bukov, Ching-Hao Wang, Alexandre GR Day, Clint Richardson, Charles K Fisher, and David J Schwab. A high-bias, low-variance introduction to machine learning for physicists. *Physics reports*, 810:1–124, 2019.

[2] Robert H Swendsen and Jian-Sheng Wang. Nonuniversal critical dynamics in monte carlo simulations. *Physical review letters*, 58(2):86, 1987.

[3] Ulli Wolff. Collective monte carlo updating for spin systems. *Physical Review Letters*, 62(4):361, 1989.

[4] A Coniglio and W Klein. Clusters and ising critical droplets: a renormalisation group approach. *Journal of Physics A: Mathematical and General*, 13(8):2775, 1980.

[5] Cornelius Marius Fortuin and Piet W Kasteleyn. On the random-cluster model: I. introduction and relation to other models. *Physica*, 57(4):536–564, 1972.

[6] Daniel Kandel and Eytan Domany. General cluster monte carlo dynamics. *Physical Review B*, 43(10):8539, 1991.

[7] Olivier Breuleux, Yoshua Bengio, and Pascal Vincent. Quickly generating representative samples from an rbm-derived process. *Neural computation*, 23(8):2058–2073, 2011.

[8] Tijmen Tieleman and Geoffrey Hinton. Using fast weights to improve persistent contrastive divergence. In *Proceedings of the 26th annual international conference on machine learning*, pages 1033–1040, 2009.

[9] Radford M Neal. Mcmc using hamiltonian dynamics. arxiv e-prints, page. *arXiv preprint arXiv:1206.1901*, 2012.

[10] Nobuyuki Yoshioka, Yutaka Akagi, and Hosho Katsura. Transforming generalized ising models into boltzmann machines. *Physical Review E*, 99(3):032113, 2019.