Enhanced Gradient for Learning Boltzmann Machines

Tapani Raiko, KyungHyun Cho, and Alexander Ilin
Aalto University School of Science,
Department of Information and Computer Science
Firstname.Lastname@tkk.fi

Boltzmann machines are often used as building blocks in greedy learning of deep networks. However, training even a simplified model, known as restricted Boltzmann machine, can be extremely laborious: Traditional learning algorithms often converge only with the right choice of the learning rate scheduling and the scale of the initial weights. They are also sensitive to specific data representation: An equivalent Boltzmann machine can be obtained by flipping some bits and changing the weights and biases accordingly, but traditional learning rules are not invariant to such transformations. Without careful tuning of these training settings, traditional algorithms can easily get stuck at plateaus or even diverge.

We propose an enhanced gradient which is derived such that it is invariant to bit-flipping transformations. It is the interpolation between the updates of all the transformed models, where the weight update depends on the covariance rather than products of activations. The final enhanced update rules are, then, formulated as

$$\tilde{\nabla} w_{ij} = \mathrm{cov}_{\mathrm{d}}\left(x_i, x_j\right) - \mathrm{cov}_{\mathrm{m}}\left(x_i, x_j\right) \tag{1}$$

$$\tilde{\nabla} b_i = \left\langle x_i \right\rangle_{\mathrm{d}} - \left\langle x_i \right\rangle_{\mathrm{m}} - \sum_j \left\langle x_j \right\rangle_{\mathrm{dm}} \tilde{\nabla} w_{ij} \tag{2}$$

where x_i, $\mathrm{cov}_P\left(\cdot, \cdot\right)$, and $\left\langle \cdot \right\rangle_P$ indicate the state of the i-th neuron, the covariance between two variables and the expectation of a random variable computed under the distribution P, respectively.

Our experiments [1] confirm that the proposed improvements yield more stable training of RBMs with both contrastive divergence and parallel tempering. One experiment on Caltech 101 Silhouettes showed that the classification accuracy improved by more than 5% over the result reported in [2].

References

[1] K. Cho, T. Raiko, and A. Ilin. Enhanced Gradient and Adaptive Learning Rate for Training Restricted Boltzmann Machines. 2011. submitted to a conference.

[2] B. M. Marlin, K. Swersky, B. Chen, and N. de Freitas. Inductive Principles for Restricted Boltzmann Machine Learning. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pages 509–516, 2010.