Advances in Training Restricted Boltzmann Machines

KyungHyun Cho, Tapani Raiko and Juha Karhunen
Aalto University School of Science, Department of Information and Computer Science
firstname.lastname@aalto.fi

Recently *deep learning* has become a field of attention in machine learning (see [1, 2] for a review). Deep learning mainly focuses on applying a deep neural network to various machine learning applications where there are abundant amount of high-dimensional training samples either labeled or unlabeled. Especially, it was shown that deep models are able to extract efficient, hierarchical features. Many recent advances in deep learning have shown performances that are far superior to conventional approaches in many areas including image and speech recognition as well as multi-modal learning.

One of the most widely used models in deep learning is a deep belief network (DBN) which is built by stacking multiple layers of restricted Boltzmann machines (RBM). A restricted Boltzmann machine thus constitutes a basic building block with its simple bipartite structure. However, training even this simplified model has been difficult [6]. Training is sensitive to specific choice of learning hyper-parameters as well as the data representation.

We have recently proposed and are developing learning algorithms that could facilitate learning parameters of RBMs. First of all, we proposed the enhanced gradient update rules that are invariant to flipping of variables in RBMs [5]. As shown in Fig. 1 it effectively makes a gradient of weight parameters with respect to hidden units more orthogonal to each other, hence, leading to a reacher set of features learned by RBMs. Then, the adaptive learning rate adapts the learning rate on-the-fly [5], which otherwise, would require extensive cross-validation. Thirdly, we have shown that with a simple re-parameterization of the energy function training RBMs with Gaussian visible variables could become much easier [3]. Lastly, we have demonstrated that a better generative model could be learned if more advanced Markov-Chain Monte-Carlo (MCMC) method called parallel tempering is used [4].

We have released the code for MATLAB and Python implementing these recently announced algorithms for training RBMs on http://users.ics.aalto.fi/kcho.

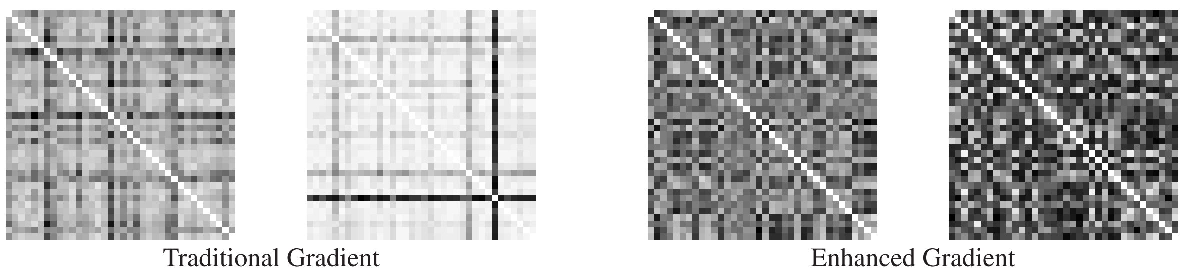

Figure 1: The angles between the update directions for the weights of an RBM with 36 hidden neurons. White pixels correspond to small angles, while black pixels correspond to orthogonal directions. From left to right: traditional gradient after 26 updates and after 364 updates, enhanced gradient after 26 updates and after 364 updates.

References

[1] Y. Bengio. Learning Deep Architectures for AI. *Foundations and Trends in Machine Learning*, 2:1–127, January 2009.

[2] Y. Bengio, A. C. Courville, and P. Vincent. Unsupervised feature learning and deep learning: A review and new perspectives. *CoRR*, abs/1206.5538, 2012.

[3] K. Cho, A. Ilin, and T. Raiko. Improved Learning of Gaussian-Bernoulli Restricted Boltzmann Machines. In *Proc. of the 20th Int. Conf. on Artificial Neural Networks*, ICANN 2011, 2011.

[4] K. Cho, T. Raiko, and A. Ilin. Parallel Tempering is Efficient for Learning Restricted Boltzmann Machines. In *Proc. of the Int. Joint Conf. on Neural Networks (IJCNN 2010)*, Barcelona, Spain, July 2010.

[5] K. Cho, T. Raiko, and A. Ilin. Enhanced Gradient and Adaptive Learning Rate for Training Restricted Boltzmann Machines. In *Proceedings of the Twenty-seventh International Conference on Machine Learning*, ICML 2011, 2011.

[6] A. Fischer and C. Igel. Empirical analysis of the divergence of Gibbs sampling based learning algorithms for restricted Boltzmann machines. In *Proc. of the 20th Int. Conf. on Artificial neural networks*, ICANN'10, pages 208–217, Berlin, Heidelberg, 2010. Springer-Verlag.