Assignment 5 - Contrastive Divergence

Rhichard Koh (100842848)

Faculty of Science, Engineering & Information Technology, Durham College

Machine Learning – Clustering and Dimensionality Reduction (COSC-32000-01)

Dr. Uzair Ahmad

April 7, 2024

A learning algorithm called Contrastive Divergence is commonly employed to educate a particular class of stochastic neural system known as Restricted Boltzmann Machines, which themselves constitute a form of probabilistic model. Geoffrey Hinton's introduction of the algorithm proved highly influential within deep learning's domain, especially for unsupervised tasks seeking patterns within untagged data. CD is designed to efficiently approximate the gradient of the log-likelihood of the training data, which is generally difficult to compute directly in models like RBMs.

Importance in Training RBMs:

RBMs have a bipartite structure consisting of visible units that represent observed data and hidden units that represent latent features within the observed information. They are trained to learn a probability distribution over the input data. Computing the gradient of the log-likelihood with respect to the model parameters, which requires performing an intractable sum over all possible configurations of the hidden units, is where the traditional approach to train these models gets bogged down computationally due to the demands of such an intensive calculation.

Contrastive Divergence addresses this issue by providing a method to approximate the gradient without having to perform this exhaustive computation. This makes the training of RBMs feasible and efficient, enabling them to learn complex, high-dimensional data distributions effectively.

Overcoming Limitations of Traditional Gradient-Based Methods:

Traditional gradient-based methods for training RBMs require the computation of the exact gradient of the log-likelihood, involving terms that are computationally expensive to calculate. The partition function, one of these terms specifically known as such, proves particularly challenging in that it necessitates summing over all potential states of the network, an undertaking impractical for all but the smallest of networks.

Contrastive Divergence circumvents this by initiating a Markov Chain Monte Carlo (MCMC) sampling process from the observed data points and running it for a small number of steps (often just one). This approach significantly reduces the computational burden and still yields useful approximations to the true gradient.

Key Concepts:

- **Positive Phase**: This phase involves computing the expectation with respect to the probability distribution defined by the observed data and the current model parameters. It reflects the model's performance in reconstructing the input data and is associated with reinforcing the probabilities of the data points in the training set.

- **Negative Phase**: In this phase, the expectation is computed with respect to the distribution defined by the model itself, after being updated in the positive phase. This generally involves sampling from the model's distribution and serves to decrease the probability of configurations that the model generates on its own, which are not in the training set.
- **Contrastive Divergence Objective**: The objective of CD is to minimize the difference between the positive phase and the negative phase, effectively making the model's distribution more like the empirical distribution of the training data. By calibrating its configurable factors to heighten the plausibility of the witnessed information while reducing the plausibility of the examples spawned by itself, the model achieves this.

In summary, Contrastive Divergence offers a practical and efficient method for training RBMs by approximating the gradient of the likelihood. It balances reinforcing what the model should learn from the data (positive phase) against discouraging what it should not (negative phase), making it a powerful tool for unsupervised learning in deep neural networks.

Reference

Carreira-Perpinan, M. A., & Hinton, G. (2005, January). On contrastive divergence learning. In *International workshop on artificial intelligence and statistics* (pp. 33-40). PMLR.

Hinton, G. (2010, August 2). A Practical Guide to Training Restricted Boltzmann Machines. Toronto; University of Toronto. https://www.cs.toronto.edu/~hinton/absps/guideTR.pdf

Sutskever, I., & Tieleman, T. (2010, March). On the convergence properties of contrastive divergence. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics* (pp. 789-795). JMLR Workshop and Conference Proceedings.