Parameter Estimation with Deformed Bregman Divergence

Takashi Takenouchi
Faculty of Systems Information Science, Department of Complex and Intelligent Systems
Future University Hakodate
116-2 Kamedanakano, Hakodate, Hokkaido 041-8655, Japan
e-mail: ttakashi@fun.ac.jp

Probabilistic models on discrete space are useful and parameter estimation of probabilistic models on discrete space is a popular and important issue. For example, the restricted Boltzmann machine (RBM) attracts increasing attention in the context of Deep learning [1]. The Maximum Likelihood Estimation (MLE) is popular method for parameter estimation, but constructions for probabilistic models on the discrete space are often difficult because of the normalization constant which sometimes requires exponential order computation. To avoid the problem, various kinds of methods have been proposed. The contrastive divergence [2] avoids the exponential order calculation using the Markov Chain Monte Carlo (MCMC) sampling. The score matching method [3] and the proper local scoring rules [4] utilize information of "neighbor" and estimate parameter without calculation of the normalization constant. [5] avoids the calculation of normalization constant by employing homogeneous divergence and a technique of empirical localization for unnormalized model.

In this paper, we focus on a deformed Bregman divergence [6] to estimate parameters of probabilistic models on discrete space. By combining the deformed Bregman divergence and the technique of the empirical localization, we propose an estimator, which can be constructed without calculation of the normalization constant and is asymptotically efficient as the MLE. Some experiments show that the proposed estimator attains comparable performance to the MLE with drastically lower computational cost.

Keywords: Fisher efficiency, Bregman divergence, Normalization constant

References

[1] G. E Hinton. A practical guide to training restricted boltzmann machines. *Momentum*, 9(1):926, 2010.

[2] G. E. Hinton. Training Products of Experts by Minimizing Contrastive Divergence. *Neural Computation*, 14(8):1771–1800, 2002.

[3] A. Hyvärinen. Some extensions of score matching. *Computational statistics & data analysis*, 51(5):2499–2512, 2007.

[4] A. P. Dawid, S. Lauritzen, and M. Parry. Proper local scoring rules on discrete sample spaces. *The Annals of Statistics*, 40(1):593–608, 2012.

[5] T. Takenouchi and T. Kanamori. Empirical localization of homogeneous divergences on discrete sample spaces. In *Advances in Neural Information Processing Systems*, pages 820–828, 2015.

[6] N. Murata, T. Takenouchi, T. Kanamori, and S. Eguchi. Information geometry of U-boost and bregman divergence. *Neural Computation*, 16(7):1437–1481, 2004.