Minimum Probability Flow Learning

Jascha Sohl-Dickstein, Peter Battaglino, Michael R. DeWeese
Redwood Center for Theoretical Neuroscience
University of California, Berkeley
jascha@berkeley.edu, pbb@berkeley.edu, deweese@berkeley.edu
http://redwood.berkeley.edu

Fitting probabilistic models to data is often difficult, due to the general intractability of models' partition functions. We propose a new parameter estimation technique that bypasses this difficulty. It works by establishing dynamics that would transform an observed data distribution into a model distribution, then minimizing the KL divergence between the data and the distribution produced by running the dynamics for an infinitesimal time (rather than the equilibrium distribution as is traditional). This technique extends the ideas behind minimum velocity learning [1] to arbitrary state spaces and a far broader class of dynamics. Its update rule resembles that for contrastive divergence [2], but it additionally provides an objective function and allows non-stochastic updates. We demonstrate application of this method to several cases, including the Ising model for which it outperforms current techniques by two orders of magnitude in learning time.

References

[1] J R Movellan. A minimum velocity approach to learning. unpublished draft, Jan 2008.
[2] Hinton. Training products of experts by minimizing contrastive divergence. Neural Comput (2002) vol. 14 (8) pp. 1771-1800

Topic: learning theory
Preference: poster