Frontier of Statistical Physics and Information Processing 2013

Fitting Intractable Probabilistic Models Using Nonequilibrium Dynamics

Jascha Sohl-DicksteinI[1,2], Peter Battaglino[3], and Michael R. DeWeese[3]
1. Department of Applied Physics, Stanford University
2. Khan Academy
3. Department of Physics, University of California, Berkeley

Fitting probabilistic models to data is often extremely difficult, due to the general intractability of the partition function. Using concepts from nonequilibrium statistical physics, we introduce a technique for model fitting called Minimum Probability Flow learning (MPF) [1]. We demonstrate parameter estimation using MPF for several probabilistic models, including a product-of-experts model of natural images, and an Ising spin glass where it outperforms current techniques by an order of magnitude in convergence time with lower error in the recovered coupling parameters.

MPF functions by adjusting model parameters θ so as to minimize the Kullback-Leibler divergence between a data distribution $\mathbf{p}^{(0)}$ and a distribution $\mathbf{p}^{(\epsilon)}(\theta)$ which has been allowed to evolve for an infinitesimal time ϵ away from the data distribution towards an equilibrium model distribution $\mathbf{p}^{(\infty)}(\theta)$. The superscript indicates time under system dynamics initialized at the data distribution. This objective is equivalent to minimizing the initial flow of probability away from the data distribution $\mathbf{p}^{(0)}$. A cartoon depiction of this learning progression is provided in Figure 1.

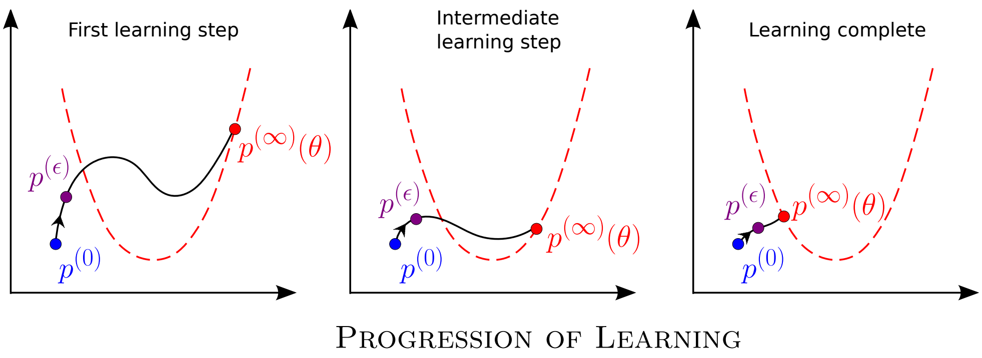

Figure 1: The three successive panels illustrate the sequence of parameter updates that occur during parameter estimation using MPF. Each set of axes represents the simplex of probability distributions. The dashed red curves indicate the family of model distributions $\mathbf{p}^{(\infty)}(\theta)$ parametrized by θ. The black curves indicate deterministic dynamics, governed by the master equation, that transform the data distribution $\mathbf{p}^{(0)}$ into the equilibrium model distribution $\mathbf{p}^{(\infty)}(\theta)$. The dynamics are a function of θ. MPF estimates parameters by adjusting θ so as to minimize the KL divergence between the data distribution $\mathbf{p}^{(0)}$ and the distribution $\mathbf{p}^{(\epsilon)}(\theta)$ resulting from running the dynamics for an infinitesimal time ϵ.

[1] Sohl-Dickstein, Jascha, Peter B. Battaglino, and Michael R. DeWeese. "New method for parameter estimation in probabilistic models: minimum probability flow." Physical review letters 107.22 (2011): 220601.